

Theoretical and computational chemistry editor's pick 2024

Edited by

Sam P. De Visser and Thomas S. Hofer

Published in

Frontiers in Chemistry



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5769-3
DOI 10.3389/978-2-8325-5769-3

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Theoretical and computational chemistry editor's pick 2024

Topic editors

Sam P. De Visser — The University of Manchester, United Kingdom

Thomas S. Hofer — University of Innsbruck, Austria

Citation

De Visser, S. P., Hofer, T. S., eds. (2024). *Theoretical and computational chemistry editor's pick 2024*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5769-3

Table of contents

- 05 Expected value of first Zagreb connection index in random cyclooctatetraene chain, random polyphenyls chain, and random chain network
Zahid Raza, Shehnaz Akhter and Yilun Shang
- 14 Assessment of host–guest molecular encapsulation of eugenol using β -cyclodextrin
Camila Auad Beltrão de Freitas, Clauber Henrique Souza Costa, Kauê Santana da Costa, Simone Patrícia Aranha da Paz, José Rogério A. Silva, Cláudio Nahum Alves and Jerônimo Lameira
- 26 Study on the molecular mechanism of anti-liver cancer effect of *Evodiae fructus* by network pharmacology and QSAR model
Peng-Yu Chen and Lin-Tao Han
- 44 Thiophene and diaminobenzo- (1,2,5-thiadiazol)- based DAD-type near-infrared fluorescent probe for nitric oxide: A theoretical research
X. Y. Lin, S. H. Sun, Y. T. Liu, Q. Q. Shi, J. J. Lv and Y. J. Peng
- 51 Computational investigation on the antioxidant activities and on the M^{pro} SARS-CoV-2 non-covalent inhibition of isorhamnetin
Maciej Spiegel, Giada Ciardullo, Tiziana Marino and Nino Russo
- 64 Structural and electronic properties of H_2 , CO , CH_4 , NO , and NH_3 adsorbed onto $Al_{12}Si_{12}$ nanocages using density functional theory
Liu-Kun Li, Yan-Qiu Ma, Kang-Ning Li, Wen-Li Xie and Bin Huang
- 73 The location of the chemical bond. Application of long covalent bond theory to the structure of silica
Stephen A. Miller
- 91 PVD growth of spiral pyramid-shaped WS_2 on SiO_2/Si driven by screw dislocations
Yassine Madoune, DingBang Yang, Yameen Ahmed, Mansour M. Al-Makeen and Han Huang
- 99 Computational pharmacology and computational chemistry of 4-hydroxyisoleucine: Physicochemical, pharmacokinetic, and DFT-based approaches
Imad Ahmad, Aleksey E. Kuznetsov, Abdul Saboor Pirzada, Khalaf F. Alsharif, Maria Daglia and Haroon Khan
- 114 Analysis of aptamer-target binding and molecular mechanisms by thermofluorimetric analysis and molecular dynamics simulation
Hong-Li Zhang, Cong Lv, Zi-Hua Li, Song Jiang, Dan Cai, Shao-Song Liu, Ting Wang and Kun-He Zhang

- 126 **When do tripdouplet states fluoresce? A theoretical study of copper(II) porphyrin**
Xingwen Wang, Chenyu Wu, Zikuan Wang and Wenjian Liu
- 141 **Molecular properties of a triazole–Ce(III) complex with antioxidant activity: structure, spectroscopy, and relationships with related derivatives. Influence of the ligands in the complex**
M. Alcolea Palafox, Nataliya P. Belskaya, Lozan Todorov, Nadya Hristova-Avakumova and Irena P. Kostova



OPEN ACCESS

EDITED BY
Anakuthil Anoop,
Indian Institute of Technology
Kharagpur, India

REVIEWED BY
Parvez Ali,
Qassim University, Saudi Arabia
Rogers Mathew,
Indian Institute of Technology
Hyderabad, India

*CORRESPONDENCE
Zahid Raza,
✉ zraza@sharjah.ac.ae
Yilun Shang,
✉ yilun.shang@northumbria.ac.uk

SPECIALTY SECTION
This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 12 October 2022
ACCEPTED 07 December 2022
PUBLISHED 04 January 2023

CITATION
Raza Z, Akhter S and Shang Y (2023),
Expected value of first Zagreb
connection index in random
cyclooctatetraene chain, random
polyphenyls chain, and random
chain network.
Front. Chem. 10:1067874.
doi: 10.3389/fchem.2022.1067874

COPYRIGHT
© 2023 Raza, Akhter and Shang. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Expected value of first Zagreb connection index in random cyclooctatetraene chain, random polyphenyls chain, and random chain network

Zahid Raza^{1*}, Shehnaz Akhter² and Yilun Shang^{3*}

¹Department of Mathematics, College of Sciences, University of Sharjah, Sharjah, United Arab Emirates,
²School of Natural Sciences, National University of Sciences and Technology, Islamabad, Pakistan,
³Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom

The Zagreb connection indices are the known topological descriptors of the graphs that are constructed from the connection cardinality (degree of given nodes lying at a distance 2) presented in 1972 to determine the total electron energy of the alternate hydrocarbons. For a long time, these connection indices did not receive much research attention. Ali and Trinajstić [Mol. Inform. 37, Art. No. 1800008, 2018] examined the Zagreb connection indices and found that they compared to basic Zagreb indices and that they provide a finer value for the correlation coefficient for the 13 physico-chemical characteristics of the octane isomers. This article acquires the formulae of expected values of the first Zagreb connection index of a random cyclooctatetraene chain, a random polyphenyls chain, and a random chain network with l number of octagons, hexagons, and pentagons, respectively. The article presents extreme and average values of all the above random chains concerning a set of special chains, including the meta-chain, the ortho-chain, and the para-chain.

KEYWORDS

average value, expected value, random cyclooctatetraene chain, random polyphenyls chain, zagreb connection indices

1 Introduction

Graph theory is vital to various disciplines, including the chemical and biological sciences. One of the objectives of chemical graph theory is its primary and significant role in studying physico-chemical reactions and biological activities and pointing out the structural properties of molecular graphs, *etc.*, Topological descriptors have played a significant role in achieving the desired properties of molecular graphs. Topological descriptors are molecular structural invariants that theoretically and mathematically explain the connectivity characteristics of nano-materials and chemical compounds. Therefore, topological indices produce sharper approaches to measuring their behavior and characteristics.

For 20 years, hydrocarbons and their derivatives have received attention from researchers because these compounds only have two members, carbon and hydrogen. We can acquire various types of hydrocarbon derivatives by replacing their molecular hydrogen atoms with different types of other atomic groups. A large number of valuable hydrocarbons are available in plants and some valuable characteristics of hydrocarbons are important to chemical raw materials and fuel.

Throughout this article, the vertex and edge sets of a graph \mathcal{H} are represented as $V(\mathcal{H})$ and $E(\mathcal{H})$, respectively. We denote the degree of a vertex $v \in V(\mathcal{H})$ by $d_{\mathcal{H}}(v)$, which is defined as the cardinality of edges joined with v . Let the order and size of \mathcal{H} be n and m , respectively. The l -degree of a given vertex $v \in V(\mathcal{H})$, presented by $d_l(v)$, is the cardinality of set of vertices of $V(\mathcal{H})$ whose distance from v is l , where $d_1(v) = d_{\mathcal{H}}(v)$ and $d_2(v) = \tau_v$ [this is known as the connection number of v (Todeschini and Consonni, 2000)].

Suppose that \mathbb{Z} is a collection of all connected simple graphs. There is a function $\mathcal{P}: \mathbb{Z} \rightarrow \mathbb{R}^+$ that describes a topological invariant if for any two isomorphic members \mathcal{M}_1 and \mathcal{M}_2 of \mathbb{Z} , we have $\mathcal{P}(\mathcal{M}_1) = \mathcal{P}(\mathcal{M}_2)$. Thousands of degree and distance-related topological invariants have been proposed, but some are better known because of their high predictive power for many characteristics like density, boiling point, molecular weight, refractive index, etc., Topological invariants have so many implementations in numerous areas of sciences such as drug discovery, physico-chemical research, toxicology, biology, and chemistry. To date, topological indices are the most notable field of graphical research. For more discussion on numerous invariants, we refer readers to studies by (Gutman, 2013; Akhter et al., 2016; Akhter and Imran, 2016; Akhter et al., 2017; Akhter et al., 2018; Akhter et al., 2020).

The Zagreb indices are the most notable invariants, and they have many valuable applications in chemistry. In 1972 Gutman and Trinajstić (Gutman and Trinajstić, 1972) established the first vertex degree dependent Zagreb index of a graph \mathcal{H} . Two renowned Zagreb indices of a graph \mathcal{H} can be described in the following manner:

$$\begin{aligned}\mathcal{M}_1(\mathcal{H}) &= \sum_{x \in V(\mathcal{H})} d_{\mathcal{H}}^2(x), \\ \mathcal{M}_2(\mathcal{H}) &= \sum_{xy \in E(\mathcal{H})} d_{\mathcal{H}}(x)d_{\mathcal{H}}(y).\end{aligned}$$

Motivated by how influential they have become and the many important applications of primary Zagreb indices, Naji et al. (Soner and Naji, 2016; Gutman et al., 2017) presented the concept of Zagreb connection indices (leap Zagreb indices), constructed from the second degrees of the vertices of a graph \mathcal{H} . The first, second, and modified Zagreb connection indices of \mathcal{H} can be defined as:

$$\begin{aligned}ZC_1(\mathcal{H}) &= \sum_{y \in V(\mathcal{H})} \tau_y^2, \\ ZC_2(\mathcal{H}) &= \sum_{xy \in E(\mathcal{H})} \tau_x \tau_y, \\ ZC_1^*(\mathcal{H}) &= \sum_{x \in V(\mathcal{H})} d_{\mathcal{H}}(x)\tau_x.\end{aligned}$$

The chemical applications of ZC_1 were presented in (8), indicating that the given index has a wide co-relation with the physical characteristics of chemical compounds, for instance, boiling point, enthalpy of evaporation, entropy, acentric factor, and standard enthalpy of vaporization. Let f_l present the cardinality of the subset of vertices of \mathcal{H} with connection number l . The next formula for the first Zagreb connection index is equal to the above definition.

$$ZC_1(\mathcal{H}) = \sum_{0 \leq l \leq n-2} f_l(G)l^2. \quad (1.1)$$

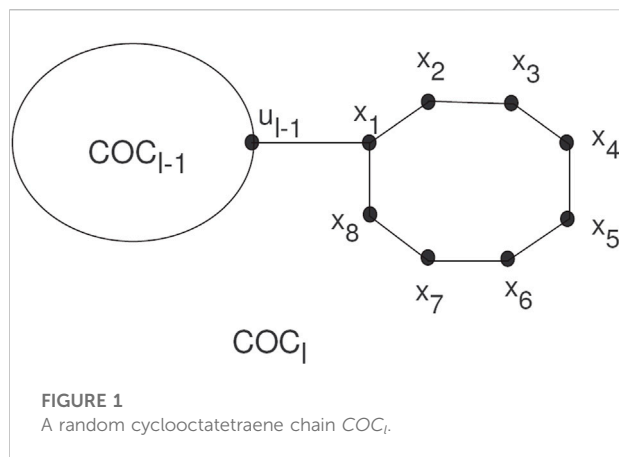
Naji and Soner (2018), (Gutman et al., 2017) determined the leap Zagreb descriptors of some graph operations and families. Leap Zagreb indices are presented in a recently published survey (Gutman et al., 2020). In (39), the authors establish sharp bounds for the leap Zagreb indices of trees and unicyclic graphs and also determined the corresponding extremal graphs. For more studies on Zagreb connection indices, we refer the readers to (Ducoffe et al., 2018a; Ali and Trinajstić, 2018; Shao et al., 2018a; Basavanagoud and Chitra, 2018; Ducoffe et al., 2018b; Khalid et al., 2018; Manzoor et al., 2018; Du et al., 2019; Fatima et al., 2019; Tang et al., 2019; Ye et al., 2019; Raza, 2020a; Bao et al., 2020; Raza, 2020b; Cao et al., 2020; Naji et al., 2020; Raza, 2022).

Huang et al. (2014) determined the expected values for Kirchhoff indices of random polyphenyl and spiro chains. Ma et al. (2017), Yang and Zhang. (2012), and Qi et al. (2022) independently acquired the expected value of Wiener indices of random polyphenyl chain and random spiro chain. Zhang et al. (2020) have provided expected values of the Schultz, Gutman, multiplicative degree-Kirchhoff, and additive degree-Kirchhoff indices of random polyphenylene chains. Raza and Imran. (2021) obtained expected values of modified second Zagreb, symmetric difference, inverse symmetric, and augmented Zagreb indices in random cyclooctane chains. Zhang et al. (2021) established the formulae for expected values of Sombor indices of a general random chain. Recently, many studies have explored the expected values of different topological indices. For further information, we refer readers to the following studies (Raza, 2020b; Fang et al., 2021; Raza, 2021; Jahanbanni, 2022; Raza et al., 2022).

Motivated by the above research, the present study determined the explicit formulae for expected values of the first Zagreb connection index of the random cyclooctatetraene chain, random polyphenyls chain, and random chain network with l octagons, hexagons, and pentagons, respectively. Moreover, we examined the average and extreme values of the Zagreb connection index among all the above-mentioned random chains corresponding to their set.

2 The first Zagreb connection index of random cyclooctatetraene chain

Cyclooctatetraene, having chemical formula C_8H_8 , is an organic compound whose full name is '1, 3, 5, 7 - cyclooctene.



Its structure is a cyclic polyolefin-like benzene, but it is not aromatic, see (Willis et al., 1952; Mathews and Lipscomb, 1959; Traetteberg et al., 1970). It has the same chemical characteristics as unsaturated hydrocarbons and is easy to construct explosive organic peroxides, (Milas and NolanPetrus, 1958; Donald and Whitehead, 1969; Garavelli et al., 2002; Schwamm et al., 2019).

Spiro compounds are valuable types of cycloalkanes in organic chemistry. A spiro union is a join of two rings that have a common atom between both rings, and a join of a direct union among the rings is known as a free spiro union in spiro compounds. In a cyclooctatylene chain, octagons are

joined by cut vertices or cut edges. A random cyclooctatetraene chain COC_l , has l octagons, and can be constructed by a cyclooctatetraene chain COC_{l-1} with $l-1$ octagons attached to a new octagon G_l by a bridge (see Figure 1).

The COC_l is a cyclooctatetraene chain with $l \geq 2$ having G_1, G_2, \dots, G_l octagons. The new octagon can be joined by four different schemes, which give the local orderings. We use these as $COC_l^1, COC_l^2, COC_l^3, COC_l^4$ (see Figure 2).

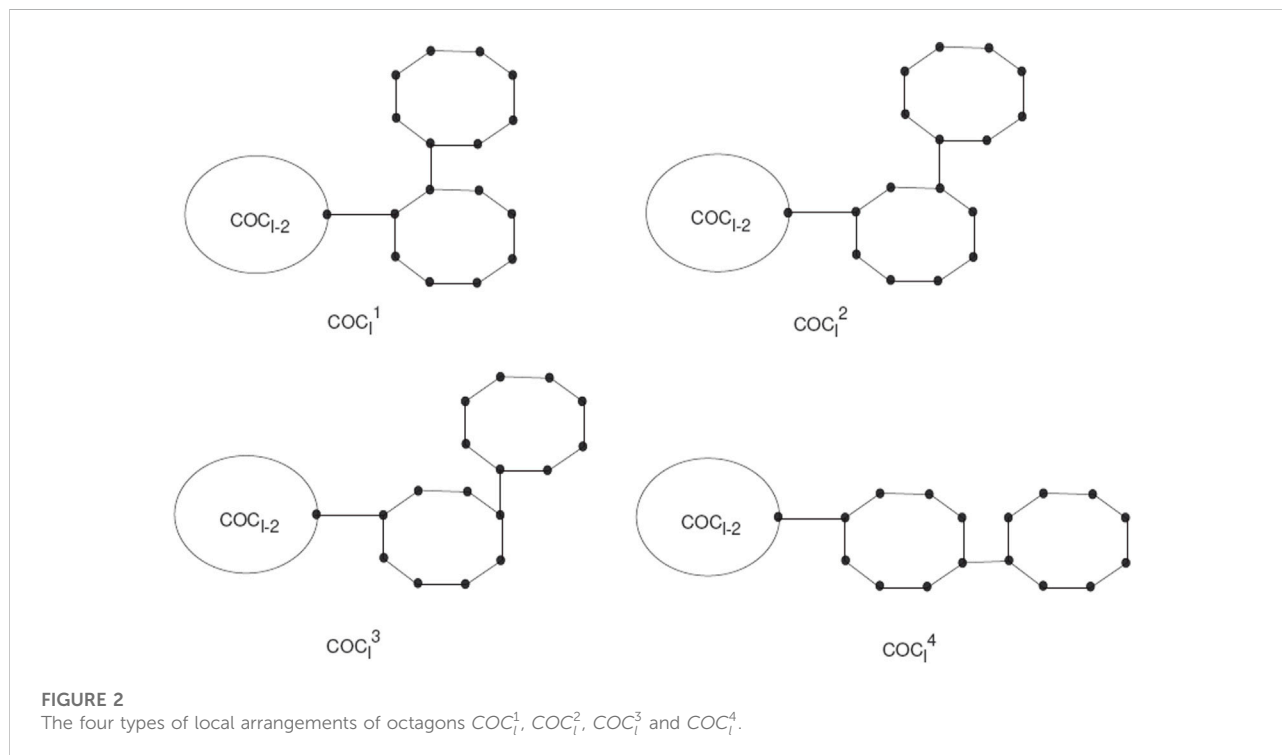
A random cyclooctatetraene chain $COC_l(k_1, k_2, k_3)$ is a cyclooctatetraene chain constructed by step-by-step attachment of new octagons. At every step $p = 2, 3, \dots, l$ a random choice is constructed from one of the four possible chains:

- 1 $COC_{p-1} \rightarrow COC_p^1$ with probability k_1 ,
- 2 $COC_{p-1} \rightarrow COC_p^2$ with probability k_2 ,
- 3 $COC_{p-1} \rightarrow COC_p^3$ with probability k_3 ,
- 4 $COC_{p-1} \rightarrow COC_p^4$ with probability $k_4 = 1 - k_1 - k_2 - k_3$,

Where all the given probabilities are constant. In this section, we will discuss the expected value for the first Zagreb connection index among random cyclooctatetraene chains with l octagons.

Theorem 2.1. For $l \geq 2$, the expected value for the first Zagreb connection index of random cyclooctatetraene chain COC_l is

$$E(ZC_1(COC_l)) = (8k_1 + 2k_2 + 76)l - 16k_1 - 4k_2 - 44.$$



Proof. **Case-I:** When $l = 2$, we get the result by direct calculations as:

$$E(ZC_1(COC_l)) = 10 \times (2)^2 + 4 \times (3)^2 + 2 \times (4)^2 = 108.$$

Case-II: When $l \geq 3$, it is obvious that $f_2(COC_l)$, $f_3(COC_l)$, $f_4(COC_l)$ and $f_5(COC_l)$ depends on the four possible cases as following:

1 If $COC_{l-1} \rightarrow COC_l^1$ with probability k_1 , we acquire

$$\begin{aligned} f_2(COC_l^1) &= f_2(COC_{l-1}) + 4, \\ f_3(COC_l^1) &= f_3(COC_{l-1}) + 2, \\ f_4(COC_l^1) &= f_4(COC_{l-1}) = 2, \\ f_5(COC_l^1) &= f_5(COC_{l-1}) + 2. \end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned} ZC_1(COC_l^1) &= ZC_1(COC_{l-1}) + 4 \times 2^2 \\ &\quad + 2 \times 3^2 + 2 \times 5^2 \\ &= ZC_1(COC_{l-1}) + 84. \end{aligned}$$

2 If $COC_{l-1} \rightarrow COC_l^2$ with probability k_2 , we acquire

$$\begin{aligned} f_2(COC_l^2) &= f_2(COC_{l-1}) + 3, \\ f_3(COC_l^2) &= f_3(COC_{l-1}) + 2, \\ f_4(COC_l^2) &= f_4(COC_{l-1}) + 3, \end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned} ZC_1(COC_l^2) &= ZC_1(COC_{l-1}) + 3 \times 2^2 + 2 \times 3^2 \\ &\quad + 3 \times 4^2 \\ &= ZC_1(COC_{l-1}) + 78. \end{aligned}$$

3 If $COC_{l-1} \rightarrow COC_l^3$ with probability k_3 , we acquire

$$\begin{aligned} f_2(COC_l^3) &= f_2(COC_{l-1}) + 2, \\ f_3(COC_l^3) &= f_3(COC_{l-1}) + 4, \\ f_4(COC_l^3) &= f_4(COC_{l-1}) + 2, \end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned} ZC_1(COC_l^3) &= ZC_1(COC_{l-1}) + 2 \times 2^2 + 4 \times 3^2 \\ &\quad + 2 \times 4^2 \\ &= ZC_1(COC_{l-1}) + 76. \end{aligned}$$

4 If $COC_{l-1} \rightarrow COC_l^4$ with probability $1 - k_1 - k_2 - k_3$, we acquire

$$\begin{aligned} f_2(COC_l^4) &= f_2(COC_{l-1}) + 2, \\ f_3(COC_l^4) &= f_3(COC_{l-1}) + 4, \\ f_4(COC_l^4) &= f_4(COC_{l-1}) + 2, \end{aligned}$$

By using above the values in Eq. 1.1, we get

$$\begin{aligned} ZC_1(COC_l^4) &= ZC_1(COC_{l-1}) + 2 \times 2^2 + 4 \times 3^2 \\ &\quad + 2 \times 4^2 \\ &= ZC_1(COC_{l-1}) + 76. \end{aligned}$$

Now

$$\begin{aligned} E_l^i &= E(ZC_1(COC_l)) = k_1 ZC_1(COC_l^1) + k_2 ZC_1(COC_l^2) + k_3 ZC_1(COC_l^3) \\ &\quad + (1 - k_1 - k_2 - k_3) ZC_1(COC_l^4) \\ &= ZC_1(COC_{l-1}) + 8k_1 + 2k_2 + 76. \end{aligned} \quad (2.1)$$

Note that $E[E_l^i] = E_l^i$. By applying the expression operator to Eq. 2.1 and also $l \geq 3$, we get

$$E_l^i = E_{l-1}^i + 8k_1 + 2k_2 + 76. \quad (2.2)$$

The Eq. 2.2 is a first-order non-homogeneous linear difference result with constant coefficients. We easily see that the general solution of the homogeneous equation of Eq. 2.2 is $E^i = C$. Suppose $E^i = bl$ is a particular solution of Eq. 2.2, using E^i into Eq. 2.2, we acquire

$$b = 8k_1 + 2k_2 + 76.$$

Finally the general solution of Eq. 2.2 is

$$\begin{aligned} E_l^i &= E^i + E^{i'} \\ &= E(ZC_1(COC_l)) = (8k_1 + 2k_2 + 76)l + C. \end{aligned}$$

Applying the initial condition $l = 3$, we get the following

$$C = -16k_1 - 4k_2 - 44.$$

Therefore

$$\begin{aligned} E_l^i &= E(ZC_1(COC_l)) \\ &= (8k_1 + 2k_2 + 76)l - 16k_1 - 4k_2 - 44. \end{aligned}$$

If $k_1 = 1$ (respectively, $k_2 = 1$) and $k_2 = k_3 = k_4 = 0$ (respectively, $k_1 = k_3 = k_4 = 0$), then $COC_l = M_l$ (respectively, $COC_l = O_l^1$). Similarly, if $k_3 = 1$ (respectively, $k_4 = 1$) and $k_1 = k_2 = k_4 = 0$ (respectively, $k_1 = k_2 = k_3 = 0$), then $COC_l = Q_l^2$ (respectively $COC_l = L_l$). By Theorem 2.1, we can acquire the first Zagreb connection index of the cyclooctatetraene meta-chain M_l , ortho-chains O_l^1 , O_l^2 and para-chain L_l as:

$$\begin{aligned} ZC_1(M_l) &= 84l - 60, ZC_1(O_l^1) = 78l - 48, \\ ZC_1(O_l^2) &= 76l - 44, ZC_1(L_l) = 76l - 44. \end{aligned}$$

Corollary 2.2. For a random cyclooctatetraene chain $COC_l(l \geq 3)$, the para-chain L_l and ortho chain O_l^1 , and the meta-chain M_l achieves the minimum and the maximum of $E(ZC_1(COC_l))$, respectively.

Proof. Using Theorem 2.1, we acquire

$$E_l^i = E(ZC_1(COC_l)) = (8l - 16)k_1 + (2l - 4)k_2 + 76l - 44.$$

By taking partial derivatives, we acquire $\frac{\partial E}{\partial k_1} = 8l - 16 > 0$, $\frac{\partial E}{\partial k_2} = 2l - 4 > 0$. When $k_1 = k_2 = k_3 = 0$ (i.e. $k_4 = 1$), the para-chain L_l and ortho chain O_l^1 achieve the minimum of $E(ZC_1(COC_l))$, that is $COC_l \cong L_l$ or $COC_l \cong O_l^1$. If $k_3 = 1 - k_1 - k_2$ ($0 \leq k_1, k_2 \leq 1$), we have

$$E_l^i = E(ZC_1(COC_l)) \\ = (8l - 16)k_1 + (2l - 4)k_2 + 76l - 44.$$

But $k_1 = k_2 = 0$ (when $k_3 = 1$), $E(ZC_1(COC_l))$ can not attain the maximum value. If $k_1 = 1 - k_2$ ($0 \leq k_2 \leq 1$), we acquire

$$E_l^i = E(ZC_1(COC_l)) \\ = (8l - 16)(1 - k_2) + (2l - 4)k_2 + 76l - 44.$$

Therefore $\frac{\partial E}{\partial k_2} = -6l + 12 < 0$. Thus $E(ZC_1(COC_l))$ achieves the maximum value, if $k_2 = 0$ ($k_1 = 1$), that is $COC_l \cong M_l$.

3 The first Zagreb connection index of a random polyphenyl chain

Polyphenyls showed a molecular graph corresponding to a type of macrocyclic aromatic hydrocarbons, and these molecular graphs of polyphenyls construct a polyphenyl structure. Polyphenyls and their derivatives have applications in drug synthesis, organic synthesis, heat exchangers, etc., and have received attention from chemists. A random polyphenyl chain PPC_l with l hexagons can be constructed by a polyphenyl chain PPC_{l-1} using $l-1$ hexagons attached to a new hexagon G_l by a bridge (see Figure 3).

The PPC_l will be a polyphenyl chain with $l \geq 2$ having G_1, G_2, \dots, G_l hexagons. PPC_l is the meta-chain M_l , the ortho-chain O_l^1 and the para-chain L_l . The new hexagon can be

joined in three arrangements, which construct the local orderings. We use these as $PPC_l^1, PPC_l^2, PPC_l^3$ (see Figure 4).

A random polyphenyl chain $PPC_l(k_1, k_2)$ is a polyphenyl chain constructed by step-by-step attachment of new hexagons. At every step $p = 2, 3, \dots, l$, a random choice construct one of the three possible chains:

- 1 $PPC_{p-1} \rightarrow PPC_p^1$ with probability k_1 ,
- 2 $PPC_{p-1} \rightarrow PPC_p^2$ with probability k_2 ,
- 3 $PPC_{p-1} \rightarrow PPC_p^3$ with probability $k_3 = 1 - k_1 - k_2$,

Where all the given probabilities are constant. In this section, we discuss the expected value for the first Zagreb connection index of the random polyphenyl chain with l hexagons.

Theorem 3.1. For $l \geq 2$, the expected value for the first Zagreb connection index of the random polyphenyl chain PPC_l is

$$E(ZC_1(PPC_l)) = (8k_1 + 2k_2 + 68)l - 16k_1 - 4k_2 - 44.$$

Proof. Case-I: When $l = 2$, one can get

$$E(ZC_1(PPC_l)) = 6 \times (2)^2 + 4 \times (3)^2 + 2 \times (4)^2 = 92.$$

Case-II: When $l \geq 3$, it is obvious that $f_2(PPC_l), f_3(PPC_l), f_4(PPC_l)$ and $f_5(PPC_l)$ depends on the four possible cases, as follows:

- 1 If $PPC_{l-1} \rightarrow PPC_l^1$ having probability k_1 , we acquire

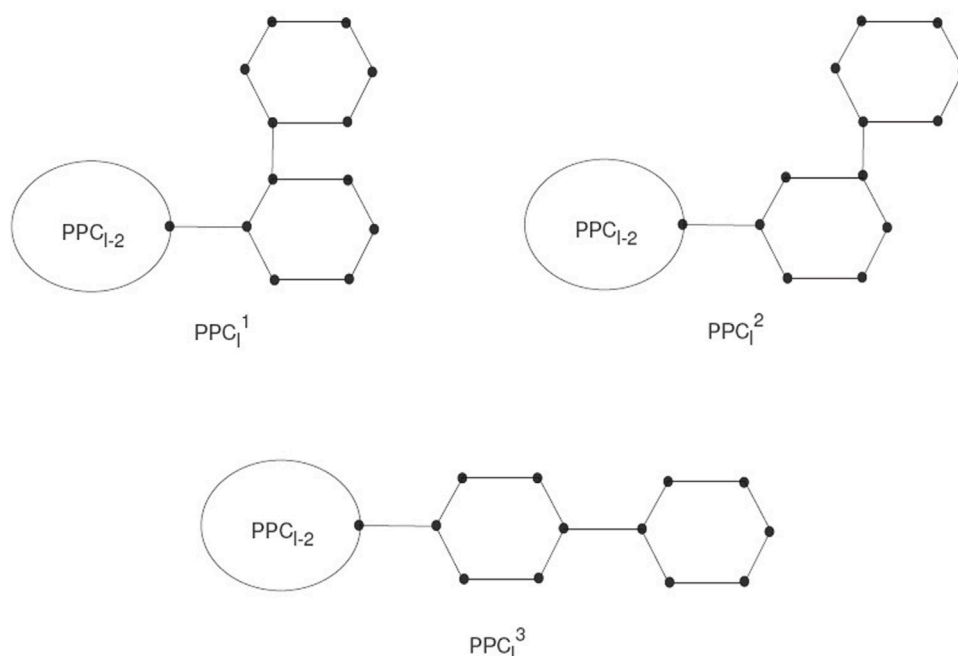
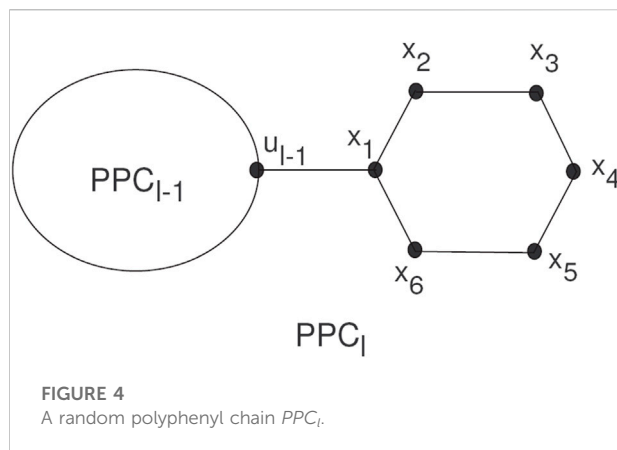


FIGURE 3
The three types of local arrangements of hexagons PPC_l^1, PPC_l^2 , and PPC_l^3 .



$$\begin{aligned}f_2(PPC_l^1) &= f_2(PPC_{l-1}) + 2, \\f_3(PPC_l^1) &= f_3(PPC_{l-1}) + 2, \\f_4(PPC_l^1) &= f_4(PPC_{l-1}) = 2, \\f_5(PPC_l^1) &= f_5(PPC_{l-1}) + 2.\end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned}ZC_1(PPC_l^1) &= ZC_1(PPC_{l-1}) + 2 \times 2^2 + 2 \times 3^2 \\&\quad + 2 \times 5^2 \\&= ZC_1(PPC_{l-1}) + 76.\end{aligned}$$

2 If $PPC_{l-1} \rightarrow PPC_l^2$ having probability k_2 , we acquire

$$\begin{aligned}f_2(PPC_l^2) &= f_2(PPC_{l-1}) + 1, \\f_3(PPC_l^2) &= f_3(PPC_{l-1}) + 2, \\f_4(PPC_l^2) &= f_4(PPC_{l-1}) + 3,\end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned}ZC_1(PPC_l^2) &= ZC_1(PPC_{l-1}) + 1 \times 2^2 + 2 \times 3^2 \\&\quad + 3 \times 4^2 \\&= ZC_1(PPC_{l-1}) + 70.\end{aligned}$$

3 If $PPC_{l-1} \rightarrow PPC_l^3$ having probability k_3 , we acquire

$$\begin{aligned}f_2(PPC_l^3) &= f_2(PPC_{l-1}) = 6, \\f_3(PPC_l^3) &= f_3(PPC_{l-1}) + 4, \\f_4(PPC_l^3) &= f_4(PPC_{l-1}) + 2,\end{aligned}$$

By using the above values in Eq. 1.1, we get

$$\begin{aligned}ZC_1(PPC_l^3) &= ZC_1(PPC_{l-1}) + 4 \times 3^2 + 2 \times 4^2 \\&= ZC_1(PPC_{l-1}) + 68.\end{aligned}$$

Now

$$\begin{aligned}E_l^i &= E(ZC_1(PPC_l)) \\&= k_1 ZC_1(PPC_l^1) + k_2 ZC_1(PPC_l^2) \\&\quad + (1 - k_1 - k_2) ZC_1(PPC_l^3) \\&= ZC_1(PPC_{l-1}) + 8k_1 + 2k_2 + 68.\end{aligned} \quad (3.1)$$

Note that $E[E_l^i] = E_l^i$. By applying the expression operator to Eq. 3.1 and also $l \geq 3$, we get

$$E_l^i = E_{l-1}^i + 8k_1 + 2k_2 + 68. \quad (3.2)$$

The result Eq. 3.2 is a first-order non-homogeneous linear difference equation with constant coefficients. The general solution of the homogeneous side is Eq. 3.2 is $E^i = C$. Suppose $E^{it} = bl$ is a particular result of Eq. 3.2, using E^{it} into Eq. 3.2, we acquire

$$b = 8k_1 + 2k_2 + 68.$$

Finally the general solution of Eq. 3.2 is given by

$$\begin{aligned}E_l^i &= E^i + E^{it} \\&= E(ZC_1(PPC_l)) = (8k_1 + 2k_2 + 68)l + C.\end{aligned}$$

Applying the initial condition $l = 3$, we get following

$$C = -16k_1 - 4k_2 - 44.$$

Therefore

$$E_l^i = E(ZC_1(PPC_l)) = (8k_1 + 2k_2 + 68)l - 16k_1 - 4k_2 - 44.$$

If $k_1 = 1$ (respectively, $k_2 = 1$) and $k_2 = k_3 = 0$ (respectively, $k_1 = k_3 = 0$), then $PPC_l = M_l$ (respectively, $PPC_l = O_l$). Similarly, if $k_3 = 1$ and $k_1 = k_2 = 0$, then $PPC_l = L_l$. By Theorem 3.1, we can acquire the first Zagreb connection index of polyphenyl chains like meta M_b , ortho O_b , and para L_b as

$$\begin{aligned}ZC_1(M_l) &= 76l - 60, & ZC_1(O_l) &= 70l - 48, \\ZC_1(L_l) &= 68l - 44.\end{aligned}$$

Corollary 3.2. For a random polyphenyl chain $PPC_l (l \geq 3)$, the para-chain L_l and the meta-chain M_l achieves the minimum and the maximum $E(ZC_1(PPC_l))$, respectively.

Proof. From Theorem 3.1, we obtain

$$\begin{aligned}E_l^i &= E(ZC_1(PPC_l)) \\&= (8l - 16)k_1 + (2l - 4)k_2 + 68l - 44.\end{aligned}$$

By taking partial derivatives, we acquire $\frac{\partial E}{\partial k_1} = 8l - 16 > 0$, $\frac{\partial E}{\partial k_2} = 2l - 4 > 0$. When $k_1 = k_2 = 0$ (i.e. $k_3 = 1$), the para-chain L_l has the minimum of $E(ZC_1(COC_l))$, that is $PPC_l \cong L_l$. If $k_1 = 1 - k_2$ ($0 \leq k_2 \leq 1$), we acquire

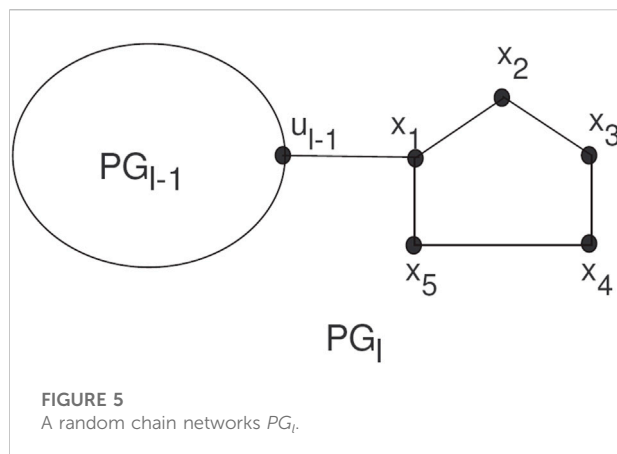
$$\begin{aligned}E_l^i &= E(ZC_1(PPC_l)) \\&= (8l - 16)(1 - k_2) + (2l - 4)k_2 + 68l - 44.\end{aligned}$$

Therefore $\frac{\partial E}{\partial k_2} = -6l + 12 < 0$. Thus $E(ZC_1(PPC_l))$ achieves the maximum value, if $k_2 = 0$ ($k_1 = 1$), that is $PPC_l \cong M_l$.

4 The first Zagreb connection index of random chain network PG_l

The random chain networks PG_l with l pentagons can be constructed by PG_{l-1} having $l-1$ pentagons attached to a new pentagon H_l by a bridge (see Figure 5).

The PG_l will be a random chain network with $l \geq 2$, and H_1, H_2, \dots, H_l pentagons. For $l \geq 3$, there are two ways to attach pentagons at the end and get PG_l^1 and PG_l^2 , (see Figure 6). For



such a random chain network, any step for $q = 2, 3, 4, \dots, l$ can be constructed by two possible chains with given probabilities k_1 and k_2 , respectively:

- 1 $PG_{q-1} \rightarrow PG_q^1$ with probability k_1 ,
- 2 $PG_{q-1} \rightarrow PG_q^2$ with probability $k_2 = 1 - k_1$,

Where all the given probabilities are constant.

This section discusses the expected value for the first Zagreb connection index of the random chain network with l pentagons. The proof of [Theorem 4.1](#) is the same as the proofs of [Theorem 2.1](#) and [Theorem 3.1](#); therefore, we omit it here.

Theorem 4.1. For $l \geq 2$, the expected value for the first Zagreb connection index of random chain network PG_l is $E(ZC_1(PG_l)) = (6k_1 + 66)L - 12k_1 - 48$.

If $k_1 = 1$ (respectively, $k_2 = 1$) and $k_2 = 0$ (respectively, $k_1 = 0$), then $PG_l = PG_l^1$ (respectively, $PG_l = PG_l^2$). By [Theorem 4.1](#), we can acquire the first Zagreb connection index of the meta-chain PG_l^1 and para-chain PG_l^2 , as

$$ZC_1(PG_l^1) = 72l - 60, \quad ZC_1(PG_l^2) = 66l - 48.$$

Corollary 4.2. For a random chain network $PG_l (l \geq 3)$, the para-chain PG_l^2 and the meta-chain PG_l^1 achieves the minimum and the maximum of $E(ZC_1(PG_l))$, respectively.

5 The average values for the first Zagreb connection index

This section finds the average values for the first Zagreb connection index concerning the sets of all cyclooctatetraene chains with l octagons, polyphenyl chains with l hexagons, and chain networks with l pentagons. Let \mathbb{G}_l , \mathbb{R}_l and \mathbb{Q}_l be the sets of all cyclooctatetraene chains, polyphenyl chains, and random chain network, respectively. The average values for the first Zagreb connection index for the sets \mathbb{G}_l , \mathbb{R}_l and \mathbb{Q}_l are given below:

$$\begin{aligned} ZC_1^{avg}(\mathbb{G}_l) &= \frac{1}{|\mathbb{G}_l|} \sum_{\mathcal{H} \in \mathbb{G}_l} ZC_1(\mathcal{H}), \\ ZC_1^{avg}(\mathbb{R}_l) &= \frac{1}{|\mathbb{R}_l|} \sum_{\mathcal{H} \in \mathbb{R}_l} ZC_1(\mathcal{H}), \\ ZC_1^{avg}(\mathbb{Q}_l) &= \frac{1}{|\mathbb{Q}_l|} \sum_{\mathcal{H} \in \mathbb{Q}_l} ZC_1(\mathcal{H}). \end{aligned}$$

The average value concerning sets \mathbb{G}_l , \mathbb{R}_l , and \mathbb{Q}_l are expected values for the first Zagreb connection index of the random chains. From [Theorem 2.1](#), [Theorem 3.1](#) and [Theorem 4.1](#), we have.

Theorem 5.1. The average value for the first Zagreb connection index concerning the set \mathbb{G}_l is given as:

$$ZC_1^{avg}(\mathbb{G}_l) = \frac{157}{2}l - 49.$$

After calculation, we acquire

$$ZC_1^{avg}(\mathbb{G}_l) = \frac{1}{4} (ZC_1(M_l) + ZC_1(O_l^1) + ZC_1(O_l^2) + ZC_1(L_l)).$$

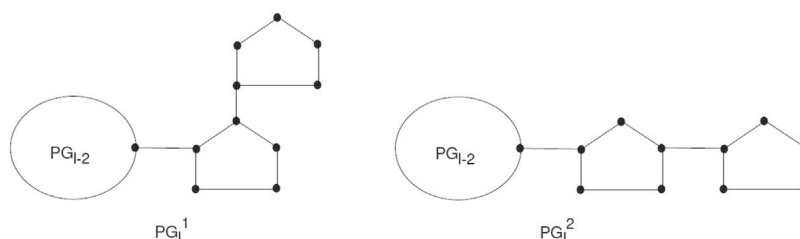


FIGURE 6
The two types of local arrangements of pentagons PG_l^1 and PG_l^2 .

Theorem 5.2. The average value for the first Zagreb connection index concerning \mathbb{R}_l is

$$ZC_1^{avg}(\mathbb{R}_l) = \frac{214}{3}l - \frac{152}{3}.$$

After calculation, we acquire

$$ZC_1^{avg}(\mathbb{R}_l) = \frac{1}{3} (ZC_1(M_l) + ZC_1(O_l) + ZC_1(L_l)).$$

Theorem 5.3. The average value for the first Zagreb connection index concerning \mathbb{Q}_l is $ZC_1^{avg}(\mathbb{Q}_l) = 69l - 54$. It is also:

$$ZC_1^{avg}(\mathbb{Q}_l) = \frac{1}{2} (ZC_1(PG_l^1) + ZC_1(PG_l^2)).$$

6 Conclusion

This study computed the expected values of the first Zagreb connection index in a random cyclooctatetraene chain, random polyphenyls chain, and random chain network with l , octagons, hexagons, and pentagons, respectively. It has discussed the maximum chain and the minimum chain of the COC_b , PPC_b , and PG_b , respectively, concerning the expected values of these chains. The average values discussed in all of the above are considered random chains for unique chains.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

References

- Akhter, S., Imran, M., Farahan, M. R., and Javaid, I. (2017). On topological properties of hexagonal and silicate networks. *Hac. J. Math. Stat.* 48 (3), 1–13. doi:10.15672/hjms.2017.541
- Akhter, S., Imran, M., Gao, W., and Farahani, M. R. (2018). On topological indices of honeycomb networks and graphene networks. *Hac. J. Math. Stat.* 47 (1), 19–35. doi:10.15672/hjms.2017.464
- Akhter, S., Imran, M., and Iqbal, Z. (2020). Mostar indices of SiO_2 nanostructures and melem chain nanostructures. *Int. J. Quantum Chem.* 121 (5). doi:10.1002/qua.26520
- Akhter, S., and Imran, M. (2016). On molecular topological properties of benzenoid structures. *Can. J. Chem.* 94 (8), 687–698. doi:10.1139/cjc-2016-0032
- Akhter, S., Imran, M., and Raza, Z. (2016). On the general sum-connectivity index and general Randić index of cacti. *J. Inequal. Appl.* 2016 (1), 300–309. doi:10.1186/s13660-016-1250-6
- Ali, A., and Trinajstić, N. (2018). A novel/old modification of the first Zagreb index. *Mol. Inf.* 37 (6–7), 1800008. doi:10.1002/minf.201800008
- Bao, L. J., Raza, Z., and Javaid, M. (2020). Zagreb connection numbers for cellular neural networks. *Discrete Dyn. Nat. Soc.* 2020, 1–8. doi:10.1155/2020/8038304
- Basavanagoud, B., and Chitra, E. (2018). On the leap Zagreb indices of generalized xyz-point-line transformation graphs $T_{xyz}(G)$, when $z = 1$. *Int. J. Math. Comb.* 2 (44–66).
- Basavanagoud, B., and Jakkannavar, P. (2018). Computing the first leap Zagreb index of some nanostructures. *Inter. J. Appl. Math.* 6 (2–B), 141–150.
- Cao, J., Ali, U., Javaid, M., and Huang, C. (2020). Zagreb connection indices of molecular graphs based on operations. *Complexity* 2020, 1–15. doi:10.1155/2020/7385682
- Donald, H. L., and Whitehead, M. A. (1969). Molecular geometry and bond energy. III. cyclooctatetraene and related compounds. *J. Am. Chem. Soc.* 91 (2), 238–242. doi:10.1021/ja01030a003
- Du, Z., Ali, A., and Trinajstić, N. (2019). Alkanes with the first three maximal/minimal modified first Zagreb connection indices. *Mol. Inf.* 38 (4), 1800116. doi:10.1002/minf.201800116
- Ducoffe, G., Marinescu-Ghemeci, R., and Obreja, C. (2018). International symposium on symbolic and numeric algorithms for scientific computing (SYNASC). *Timisoara, Rom.* 141–148.
- Ducoffe, G., Marinescu-Ghemeci, R., and Obreja, C. (2018). Proceedings of the 16th cologne-twente workshop on graphs and combinatorial optimization. *Paris, Fr. CNAM* 65–68.
- Fang, X., You, L., and Liu, H. (2021). The expected values of Sombor indices in random hexagonal chains, phenylene chains, and Sombor indices of some chemical graphs. *Int. J. Quantum. Chem.* 121 (17). doi:10.1002/qua.26740

Author contributions

Investigation: ZR, SA, and YS; Writing: ZR, SA, and YS; Review: ZR, SA, and YS.

Funding

This research was funded by the University of Sharjah.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.1067874/full#supplementary-material>

- Fatima, N., Bhatti, A. A., Ali, A., and Gao, W. (2019). Zagreb connection indices of two dendrimer nanostars. *Acta Chem. Iasi* 27 (1), 1–14. doi:10.2478/achi-2019-0001
- Garavelli, M., Bernardi, F., Cembran, A., Castaño, O., Frutos, L. M., Merchán, M., et al. (2002). Cyclooctatetraene computational photo- and thermal chemistry: A reactivity model for conjugated hydrocarbons. *J. Am. Chem. Soc.* 124 (46), 13770–13789. doi:10.1021/ja020741v
- Gutman, I. (2013). Degree-based topological indices. *Croat. Chem. Acta* 86, 351–361. doi:10.5562/cca2294
- Gutman, I., Milovanović, E., and Milovanović, I. (2020). Beyond the Zagreb indices. *AKCE Int. J. Graphs Comb.* 17 (1), 74–85. doi:10.1016/j.akcej.2018.05.002
- Gutman, I., Naji, A. M., and Sonner, N. D. (2017). The first leap Zagreb index of some graph operations. *Commun. Comb. Optim.* 2 (2), 99–117.
- Gutman, I., and Trinajstić, N. (1972). Graph theory and molecular orbitals, Total π electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* 17 (4), 535–538. doi:10.1016/0009-2614(72)85099-1
- Huang, G. H., Kuang, M. J., and Deng, H. Y. (2014). The expected values of Kirchhoff indices in the random polyphenyl and spiro chains. *ARS Math. Contem.* 9 (2), 197–207. doi:10.26493/1855-3974.458.7b0
- Jahanbanni, A. (2022). The expected values of the first Zagreb and Randić indices in random polyphenyl chains. *Polycycl. Aromat. Compd.* 42 (4), 1851–1860. doi:10.1080/10406638.2020.1809472
- Khalid, S., Kok, J., and Ali, A. (2018). Zagreb connection indices of TiO_2 nanotubes. *Chem. Bulg. J. Sci. Edu.* 27 (1), 86–92.
- Ma, L., Bian, H., Liu, B. J., and Yu, H. Z. (2017). The expected values of the Wiener index in the random phenylene and spiro chains. *ARS Comb.* 130, 267–274.
- Manzoor, S., Fatima, N., Bhatti, A. A., and Ali, A. (2018). Zagreb connection indices of some nanostructures. *Acta Chem. Iasi* 26 (2), 169–180. doi:10.2478/achi-2018-0011
- Mathews, F. S., and Lipscomb, W. N. (1959). The structure of silver cyclooctatetraene nitrate. *J. Phys. Chem.* 63 (6), 845–850. doi:10.1021/j150576a017
- Milas, N., and NolanPetrus, J., Jr. (1958). Notes-ozonization of cyclooctatetraene. *J. Org. Chem.* 23 (4), 624–625. doi:10.1021/jo01098a611
- Naji, A. M., Davvaz, B., and Mahde, S. S. (2020). A study on some properties of leap graphs. *Commun. Comb. Optim.* 5 (1), 9–17.
- Naji, A. M., and Sonner, N. D. (2018). The first leap Zagreb index of some graph operations. *Int. J. Appl. Graph Theor.* 2 (1), 7–18.
- Qi, J., Fang, M., and Geng, X. (2022). The expected value for the Wiener index in the random spiro chains. *Polycycl. Aromat. Compd.* 1–11.
- Raza, Z., and Imran, M. (2021). Expected values of some molecular descriptors in random cyclooctane chains. *Symmetry* 13 (11), 2197. doi:10.3390/sym13112197
- Raza, Z. (2020). Leap Zagreb connection numbers for some networks models. *J. Chem.* 20 (6), 1407–1413. doi:10.22146/ijc.53393
- Raza, Z., Naz, K., and Ahmad, S. (2022). Expected values of molecular descriptors in random polyphenyl chains. *Emerg. Sci. J.* 6 (1), 151–165. doi:10.28991/esj-2022-06-01-012
- Raza, Z. (2020). The expected values of arithmetic bond connectivity and geometric indices in random phenylene chains. *Heliyon* 6 (7), e04479. doi:10.1016/j.heliyon.2020.e04479
- Raza, Z. (2021). The expected values of some indices in random phenylene chains. *Eur. Phys. J. Plus* 136 (11–15), 91. doi:10.1140/epjp/s13360-021-01082-y
- Raza, Z. (2022). Zagreb connection indices for some Benzenoid systems. *Polycycl. Aromat. Compd.* 42 (4), 1814–1827. doi:10.1080/10406638.2020.1809469
- Schwamm, R. J., Anker, M. D., Lein, M., and Coles, M. P. (2019). Reduction vs. Addition: The reaction of an aluminyl anion with 1, 3, 5, 7 – cyclooctatetraene. *Angew. Chem. Int. Ed.* 58 (5), 1489–1493. in English) 131. doi:10.1002/anie.201811675
- Shao, Z., Gutman, I., and Li, Z. (2018). Leap Zagreb indices of trees and unicyclic graphs. *Commun. Comb. Optim.* 3 (2), 179–194.
- Shao, Z., Gutman, I., Li, Z., Wang, S., and Wu, P. (2018). Leap Zagreb indices of trees and unicyclic graphs. *Commun. Comb. Optim.* 3 (2), 179–194.
- Soner, N. D., and Naji, A. M. (2016). The k -distance neighborhood polynomial of a graph. *Int. J. Math. Comput. Sci.* 3 (9), 2359–2364.
- Tang, J. H., Ali, U., Javaid, M., and Shabbir, K. (2019). Zagreb connection indices of subdivision and semi-total point operations on graphs. *J. Chem.*
- Todeschini, R., and Consonni, V. (2000). *Handbook of molecular descriptors*. Weinheim: Wiley VCH.
- Traetteberg, M., Hagen, G., and Cyvin, S. J. (1970). IV. 1, 3, 5, 7-Cyclooctatetraene. *Z. Für Naturforsch. B* 25 (2), 134–138. doi:10.1515/znb-1970-0201
- Willis, B. P., George, C. P., and Kenneth, S. P. (1952). The structure of cyclooctatetraene. *J. Am. Chem. Soc.* 74 (13), 3437–3438. doi:10.1021/ja01133a524
- Yang, W., and Zhang, F. (2012). Wiener index in random polyphenyl chains. *MATCH Commun. Math. Comp. Chem.* 68, 371–376.
- Ye, A., Qureshi, M. I., Fahad, A., Aslam, A., Jamil, M. K., Zafar, A., et al. (2019). Zagreb connection number index of nanotubes and regular hexagonal lattice. *Open Chem.* 17 (1), 75–80. doi:10.1515/chem-2019-0007
- Zhang, L., Li, Q., Li, S., and Zhang, M. (2020). The expected values for the Schultz index, Gutman index, multiplicative degree Kirchhoff index, and additive degree Kirchhoff index of a random polyphenylene chain. *Discrete Appl. Math.* 282, 243–256. doi:10.1016/j.dam.2019.11.007
- Zhang, W., You, L., Liu, H., and Huang, Y. (2021). The expected values and variances for Sombor indices in a general random chain. *App. Math. Comp.* 411, 126521. doi:10.1016/j.amc.2021.126521



OPEN ACCESS

EDITED BY

William Tiznado,
Andres Bello University, Chile

REVIEWED BY

Venkatesan Thimmakondur,
San Diego State University, United States
Israel V. M. V. Enoch,
Karunya Institute of Technology and
Sciences, India
Osvaldo Andres Yáñez Osses,
University of the Americas, Chile

*CORRESPONDENCE

Jerônimo Lameira,
✉ lameira@ufpa.br

SPECIALTY SECTION

This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 05 October 2022

ACCEPTED 06 December 2022

PUBLISHED 09 January 2023

CITATION

Freitas CABd, Costa CHS, da Costa KS,
da Paz SPA, Silva JRA, Alves CN and
Lameira J (2023), Assessment of
host–guest molecular encapsulation of
eugenol using β -cyclodextrin.
Front. Chem. 10:1061624.
doi: 10.3389/fchem.2022.1061624

COPYRIGHT

© 2023 Freitas, Costa, da Costa, da Paz,
Silva, Alves and Lameira. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Assessment of host–guest molecular encapsulation of eugenol using β -cyclodextrin

Camila Auad Beltrão de Freitas ¹,
Clauber Henrique Souza Costa ¹, Kauê Santana da Costa ²,
Simone Patrícia Aranha da Paz ³, José Rogério A. Silva ¹,
Cláudio Nahum Alves ¹ and Jerônimo Lameira ^{1*}

¹Laboratório de Planejamento e Desenvolvimento de Fármacos, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, Pará, Brazil, ²Laboratório de Simulação Computacional, Instituto de Biodiversidade, Universidade Federal do Oeste do Pará, Unidade Tapajós, Santarém, Pará, Brazil, ³Laboratório de Caracterização Mineral, Universidade Federal do Pará, Belém, Pará, Brazil

Eugenol is a natural compound with well-known repellent activity. However, its pharmaceutical and cosmetic applications are limited, since this compound is highly volatile and thermolabile. Nanoencapsulation provides protection, stability, conservation, and controlled release for several compounds. Here, eugenol was included in β -cyclodextrin, and the complex was characterized through X-ray diffraction analysis (XRD) and Fourier-transform infrared spectroscopy (FTIR). Additionally, we used molecular dynamics simulations to explore the eugenol– β -cyclodextrin complex stability with temperature increases. Our computational result demonstrates details of the molecular interactions and conformational changes of the eugenol– β -cyclodextrin complex and explains its stability between temperatures 27°C and 48°C, allowing its use in formulations that are subjected to varied temperatures.

KEYWORDS

repellents, nanoencapsulation, eugenol, molecular dynamics, molecular modeling

Introduction

Mosquitos are the main vectors of viral diseases that manifest predominantly in tropical and subtropical regions of the world, such as dengue, yellow fever, Zika, and Chikungunya (Paixão et al., 2018; Higuera and Ramírez, 2019; Barreto-Vieira et al., 2020). The chemical protection conferred by repellents against mosquitoes has been an effective alternative to prevent their contact with the human skin (Ray, 2015; Mapossa et al., 2021). Natural products remain an interesting source of new bioactive compounds with different applications (Rayan et al., 2017; Galúcio et al., 2019; Do Nascimento et al., 2020; Newman and Cragg, 2020; Santana et al., 2021), and these structures have been widely investigated as a repellent against mosquitoes (Tabanca et al., 2016; da Costa et al., 2019a). However, these compounds have been reported to have a short shelf life, in part, due to their volatile nature (Kayaci et al., 2013; Tan et al., 2019; Beltrán Sanahuja and Valdés García, 2021). Eugenol (4-allyl-2-methoxyphenol) is a volatile and lipophilic phenolic natural

compound belonging to the class of phenylpropanoid, and it is mainly found in the essential oils of plants. Eugenol can be also produced synthetically by the allylation of guaiacol with allyl chloride (Moyer et al., 2002; Kuskoski et al., 2003). Eugenol is responsible for clove aroma, and it is well known for its wide range of biological activities, such as antibacterial, antioxidant, anesthetic, and anti-inflammatory (Jaganathan et al., 2011; Kamatou et al., 2012; Roth-Walter et al., 2014; Xu et al., 2016; Mateen et al., 2019; Methods, 2021). The U S Food and Drug Administration also considers eugenol as a safe food additive for human use (Kamatou et al., 2012; El-Saber Batiha et al., 2020), and studies have demonstrated that eugenol has efficient repellent activity against different mosquitos species, such as *Aedes aegypti* (Miot et al., 2011; Afify and Potter, 2020), *Culex quinquefasciatus* (Afify et al., 2019), and *Anopheles gambiae* (Lupi et al., 2013; Thomas et al., 2017).

Several experimental studies have reported the formation of inclusion complexes of eugenol with encapsulating agents, such as β -cyclodextrin, to reduce the undesirable effects (localized irritation of the skin and allergic contact dermatitis), increase its aqueous solubility, and prolong its biological activity (Yang and Song, 2005; Abarca et al., 2016; Gong et al., 2016; Kfoury et al., 2018; de Freitas et al., 2021). β -cyclodextrin (cyclohepta-amylose) is a cyclic oligosaccharide formed by D-glucose monomers which are produced by the enzymatic degradation of starch (Szejtli, 1998; Wüpper et al., 2021), and it is particularly interesting for the encapsulation of volatile compounds, thus representing a viable and efficient strategy to retain and modulate the release of volatile and hydrophobic compounds (Abarca et al., 2016; Kfoury et al., 2018; Zheng et al., 2020). Cyclodextrin inclusion complexation is widely used in food, cosmetics, agrochemical, and pharmaceutical industries to increase the stability of several volatile organic compounds due to its hydrophobic cavities and hydrophilic exterior (Loftsson and Brewster, 1996; Szejtli, 1998; Muthu Vijayan Enoch and Swaminathan, 2004; Enoch and Swaminathan, 2005; Xu et al., 2021), which creates a physical barrier between the nucleus and the shell materials (Anaya-Castro et al., 2017).

The α -, β -, and γ -cyclodextrins are subclasses of cyclodextrins widely used for nanoencapsulation, and they could be differentiated by the presence of 6, 7, and 8 glucopyranose units, respectively, that determine the size of their internal cavity (Szejtli, 1998; Saha et al., 2016; Jansook et al., 2018). These cyclodextrins have a truncated cone-shaped molecular form, and their hydrophobic cavities have a remarkable ability to form non-covalent inclusion complexes with a variety of compounds (Lee et al., 2020; Pena et al., 2022). During the formation of an inclusion complex, water molecules are displaced to the outside of the lipophilic cavity, due to the presence of new lipophilic guest molecules that induce a new equilibrium (Anaya-Castro

et al., 2017). This water displacement and the formation of a stable complex depends on the binding forces present in the inclusion complex (e.g., hydrophobic interactions, van der Waals attractions, hydrogen bonds, and electrostatic interactions) and temperature (Alvira, 2018; Lee et al., 2020).

Temperature is an important variable to assess the stability of the inclusion complexes, and understanding its influences on the formation of intermolecular interactions and mass loss is crucial to the experimental tests that evaluate the repellent efficiency (Kayaci and Uyar, 2011; Abarca et al., 2016; Celebioglu et al., 2018). Several studies have performed computational analyses to investigate the formation of inclusion complexes formed between the oligosaccharides and the natural products (de Sousa et al., 2016; Mustafa et al., 2021; Rezaeisadat et al., 2021). Similarly, experimental studies have investigated the formation of these complexes between eugenol and its derivatives with β -cyclodextrin (Alvira, 2018; Joardar et al., 2020) and have identified a slower controlled release of eugenol at elevated temperatures, such as 50°C, 75°C, and 100°C (Kayaci et al., 2013; Celebioglu et al., 2018).

In the present study, we analyzed the eugenol- β -cyclodextrin inclusion complex through X-ray diffraction analysis (XRD) and Fourier-transform infrared spectroscopy (FTIR) and its chemical stability and binding affinity using molecular dynamics (MD) simulations and binding free energy calculations, respectively. The representative structures of the analyzed systems are shown in Supplementary Figure S1.

Materials and methods

Chemical reagents

The eugenol (CAS: 97-53-0, medium molecular weight: 164.2 g/mol, and purity: 99%) and the commercial β -cyclodextrin (CAS: 7585-39-9, medium molecular weight: 1,134.98 g/mol, and purity: 97%) were obtained from Sigma Aldrich Laboratory (São Paulo, Brazil).

Synthesis of inclusion complexes

The inclusion complex was formed through co-precipitation and solvent evaporation (Ayala-Zavala et al., 2008), in which the hydroalcoholic solution of β -cyclodextrin was incorporated into an alcoholic solution of eugenol to obtain molar ratios 1:1, 2:1, and 2:3 in duplicate, to compare and analyze the influence of concentration molar in the final product of the complexes. The physical mixture was obtained by maceration, in grade and pistil, until the formation of a paste and homogenized for 15 min; then, it was allowed to rest for 24 h in an isolated environment, dried at 50°C for 12 h, and stored as other inclusion complexes.

Characterization of the inclusion complexes

X-ray diffraction

The measurements were recorded in a divergent beam diffractometer (model: Empyrean from PANalytical) with a θ - θ goniometer, ceramic X-ray tube sealed with a cobalt anode, monochromatic radiation of Co-K α 1 ($\lambda = 1.789$ Å), long fine focus 1800W, and a Fe K β filter. The detector PIXel3D 2×2 area was used with an active length of 3.3473° (2θ - 2θ) and 255 active channels. The following instrumental conditions were applied in the analyses: voltage of 40 kV and current of 40 mA, solar encapsulation slits of 0.04° rad (in the incident and diffracted beams), 2° - 80° (2θ) sweep range, and 0.04° step size in 2θ with 1 s time/step in continuous scan mode. Phase identification was performed using PANalytical's HighScore Plus 4.8.0 software.

Powder X-ray diffractometry is a useful method to confirm the formation of complex powder or microcrystalline states; therefore, the XRD technique is only applied to materials (solid-state matter). In the present study, it was applied to analyze the materials: free β -cyclodextrin (β -CD), eugenol- β -cyclodextrin (EG- β -CD) complexes, and physical mixture (PM).

Fourier transform infrared spectrometry

The spectroscopic investigations were performed to identify the functional groups of the eugenol- β -cyclodextrin complex in the middle infrared spectral region—Middle-IR ($4,000$ – 400 cm^{-1})—using a Thermo Scientific Fourier transform infrared spectrometer (model: Nicolet iS50 FTIR), a KBr (potassium bromide) beam splitter, an IR source, and a KBr DTGS detector. The measurements were obtained by transmission with KBr pellets (0.15 g) + sample (0.002 g), with an average of 100 scans and a resolution of 8 cm^{-1} . Data were acquired using OMNIC software. As a pre-treatment, the samples were dried at 105°C for 24 h.

Molecular modeling studies

Molecular docking

To investigate the most stable conformation of eugenol in complex with β -cyclodextrin, we performed molecular docking using AutoDock Vina (Trott and Olson, 2009). This computational method allowed us to describe the molecular interactions of the eugenol with the internal cavity (lipophilic) and external surface (hydrophilic) of the β -cyclodextrin nanoparticle. Herein, we used the crystallographic structure of β -cyclodextrin (PDB code: 3EDJ) as the starting point to perform the simulations (Buedenbender and Schulz, 2009). We used the following spatial coordinates for the docking grid: $X = 69.62$, $Y =$

67.16 , and $Z = 40.77$, with dimensions of $X = 40$, $Y = 40$, and $Z = 40$ Å. The docking simulations were performed with 10 runs, and a total of 10 conformations per compound were set to perform the docking. The formation of the intermolecular interactions, such as H-bond, π -interactions, and hydrophobic interactions, were analyzed using BIOVIA Discovery Studio (BIOVIA, 2017).

Molecular dynamics simulations

To perform the MD simulations, we selected the lowest energy structure of the eugenol/ β -cyclodextrin complex obtained from the docking simulations. First, the atomic charges of eugenol were calculated using the restrained electrostatic potential atomic partial charges (RESP) protocol (Wang et al., 2000; Wang et al., 2004) using the Hartree-Fock method with the 6-31G* basis set (Bayly et al., 1993) available in the Gaussian09 program (Frisch et al., 2009). The carbohydrate force field Glycam06 (Kirschner et al., 2008) was used to treat β -cyclodextrin, and the general AMBER force field (GAFF) was used to treat the complex formed with the eugenol (Wang et al., 2004). The complex was solvated in a cubic water box using the TIP3P model (Jorgensen et al., 1983; Jorgensen et al., 1996), and the distance between the box wall and atoms of the system was set to 12.0 Å.

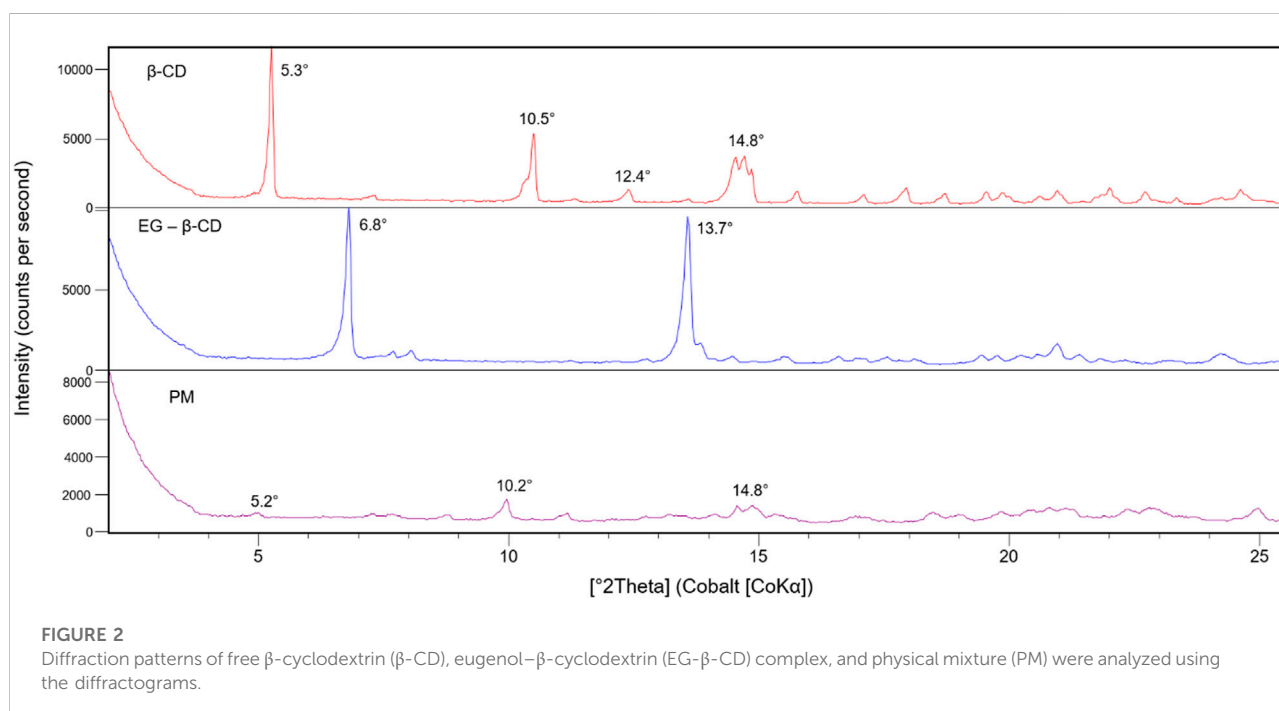
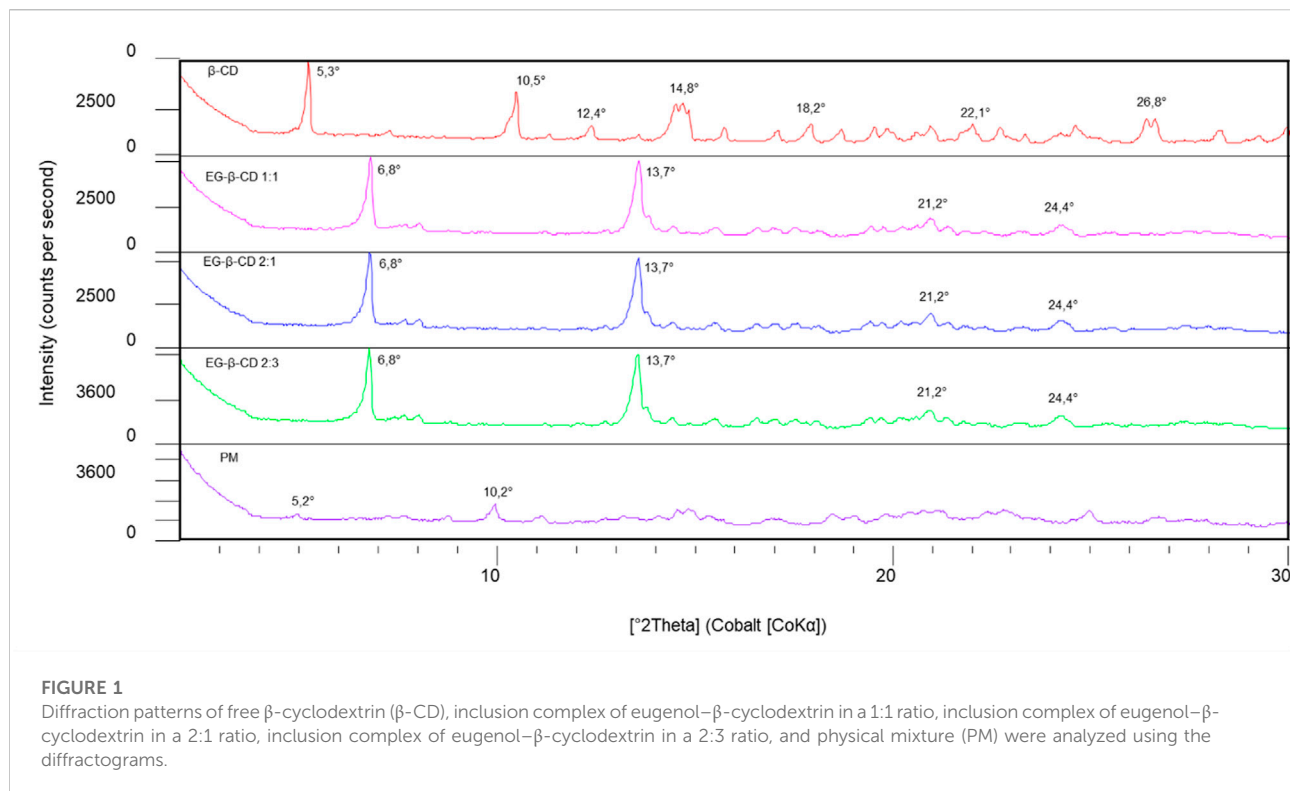
The geometry and the inter- and intra-atomic distances of hydrogen molecules, water molecules, and the eugenol- β -cyclodextrin complex were optimized in seven minimization steps using 100,000 cycles of steepest descent and the conjugate gradient method (Hestenes and Stiefel, 1952). The β -cyclodextrin-eugenol complex was investigated in four different temperatures: 27°C , 38°C , 48°C , and 58°C . In the MD simulations, the systems were heated to their final temperature (300 K) to equilibrate the density and maintain the constant pressure (1 atm). The SHAKE algorithm (AndersenRattle, 1983) was applied for all hydrogen molecules of the analyzed systems. A total time of 50 ns of MD simulation was performed using NPT ensemble.

Results and discussion

Experimental characterization of inclusion complexes

X-ray diffraction (XRD)

The diffraction pattern of the β -cyclodextrin-eugenol complexes in different stoichiometric proportions showed a similar trend, demonstrating only some different intensities between the 2:3 ratio and the others, in which the peaks are presented in greater intensity, in 3600 counts (Figure 1). However, all analyzed complexes show considerable differences when compared to the diffraction pattern of free β -cyclodextrin and the physical mixture (Figure 1). The free β -cyclodextrin diffractogram presents many Bragg reflections, highlighting the high-intensity



peaks in 2θ (CoK α): 5.3°, 10.5°, 12.4°, 14.8°, 18.2°, 22.1°, and 26.8°, as observed in studies carried out with β -cyclodextrin and other guests (Wang et al., 2011; Gong et al., 2016; Yang et al., 2016; Jiang et al., 2019), having as characteristic peak

the angle 2θ at 5.3° provided by the database of software used.

We found a significant difference between the diffractograms of free β -cyclodextrin and the β -cyclodextrin and eugenol

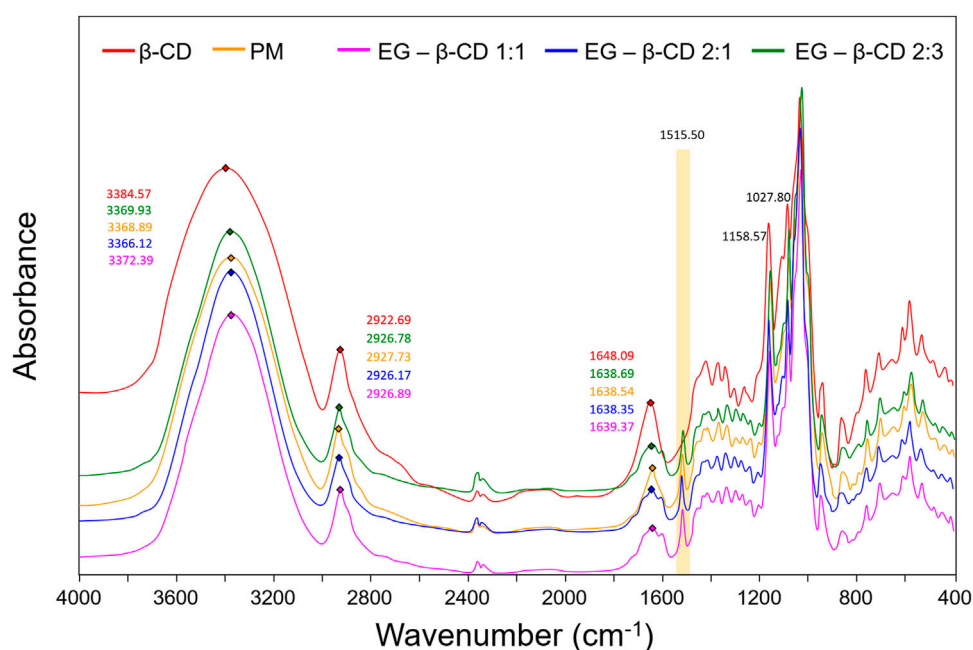


FIGURE 3

Comparison between the FTIR spectra for the β -cyclodextrin, physical mixture of eugenol and β -cyclodextrin, inclusion complex of eugenol- β -cyclodextrin in a 1:1 ratio, inclusion complex of eugenol- β -cyclodextrin in a 2:1 ratio, and inclusion complex of eugenol- β -cyclodextrin in a 2:3 ratio.

complexes, indicating the occurrence of encapsulation and the interaction between eugenol and β -cyclodextrin, given a reordering in the crystal structure, by the disappearance of peaks in 10.5° and $12.4^\circ 2\theta$ (CoK α) and contraction of the unit cell with decreasing dhkl, that is, by increasing the angle, as observed for 6.8 , 13.7 , 21.2° and $24.4^\circ 2\theta$ (CoK α), data that corroborate the results previously found (Yang and Song, 2005; Abarca et al., 2016; Dos Passos Menezes et al., 2017), which may be associated with changes in the molecular organization of β -cyclodextrin during the production of complexes.

The peak shifts indicate that the diffraction pattern of free β -cyclodextrin was altered when the eugenol was incorporated into the host molecule cavity. In studies that present the physical mixture diffractogram between eugenol and β -cyclodextrin, crystalline peaks of β -cyclodextrin were detected at a lower intensity, indicating that there was no marked difference in the crystalline form of β -cyclodextrin (Abarca et al., 2016; Gong et al., 2016), as seen in Figure 2. Furthermore, it is also important to note that the peak intensities in the eugenol- β -cyclodextrin complex were attenuated in relation to the same peaks in the free β -cyclodextrin spectrum, indicating greater structural disorder or loss in the degree of crystallinity for the complex (Figure 2). This fact is attributed to the rapid precipitation of the complex during preparation, which makes regular crystal growth insufficient (Yang and Song, 2005).

In this analysis, the diffractograms of the pure species were compared with the values obtained of the complex (Cao et al., 2005). The differences obtained from the analyses, such as the appearance or disappearance of peaks or changes in relative intensities, evidenced the formation of the inclusion complex since the principle of the complexation is associated with an increase in the degree of amorphization of the substances involved in the formation of the complex in the solid-state (Ribeiro et al., 2008; Gao et al., 2020).

Fourier transform infrared spectrometry (FTIR)

The FTIR technique is a very helpful tool to prove the interaction of both guest and host molecules in their inclusion complexes in a solid phase (Singh et al., 2010). Figure 3 shows the FTIR spectra for 1) β -cyclodextrin, 2) physical mixture of eugenol and β -cyclodextrin, 3) inclusion complex of eugenol- β -cyclodextrin in a 1:1 ratio, 4) inclusion complex of eugenol- β -cyclodextrin in a 2:1 ratio, and 5) inclusion complex of eugenol- β -cyclodextrin in a 2:3 ratio. The spectra bands which characterize the absorption regions of β -cyclodextrin are associated with the stretches, referring to the symmetric and asymmetric deformation of the hydroxyl group (OH) in the range of $3,600$ – $3,000\text{ cm}^{-1}$, that showed the characteristic bands of primary and secondary OH groups in 3384.57 cm^{-1} .

We also observed the CH bands of the β -cyclodextrin ring and methyl groups in the range of $2,940$ – $2,840\text{ cm}^{-1}$. These

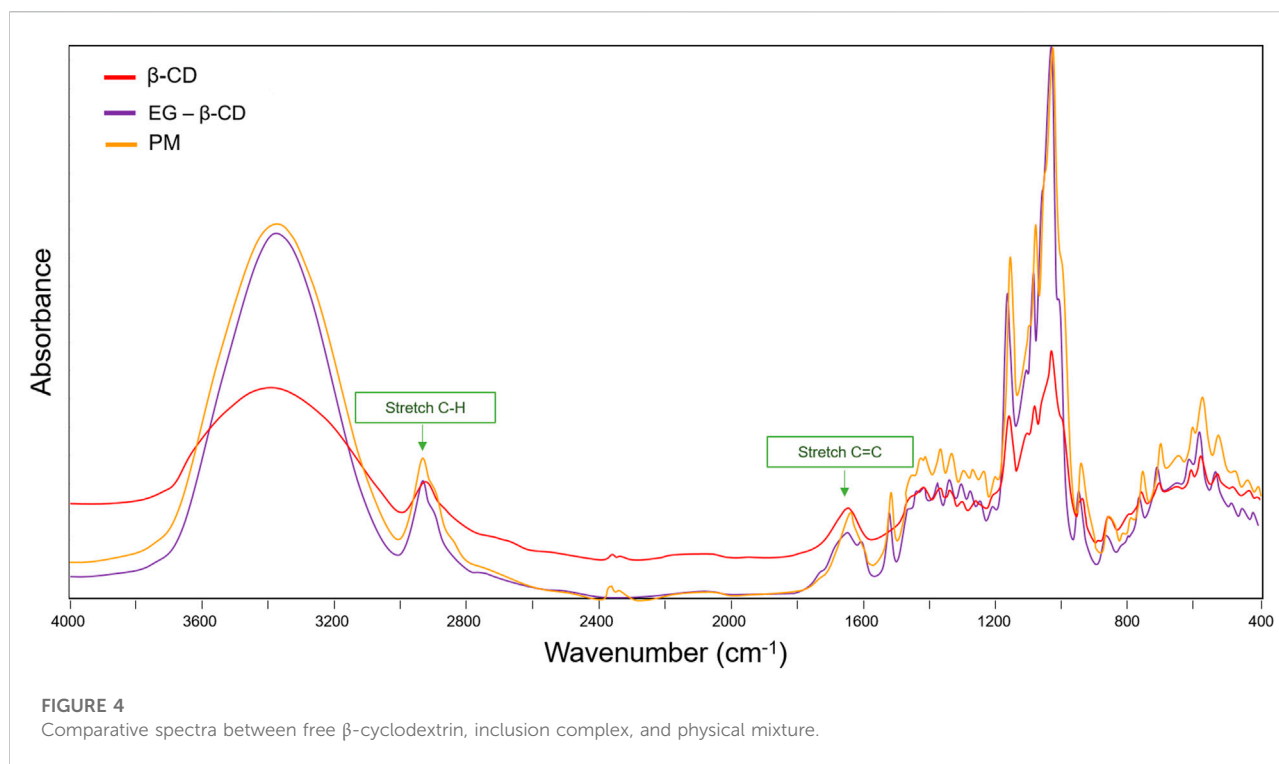


TABLE 1 Comparison between the intensity of free β -cyclodextrin, inclusion complexes, and physical mixture.

Functional group	Wavenumber (cm^{-1})				
	β -CD	Inclusion complex 1:1	Inclusion complex 2:1	Inclusion complex 2:3	PM
$\nu[\text{OH}]$ symmetric and antisymmetric	3,384.57	3,372.39	3,366.12	3,369.93	3,368.89
$\nu[\text{CH}]$	2,922.69	2,926.89	2,926.17	2,926.78	2,927.73
$\nu[\text{C-O-C}]$	1,158.57	1,156.85	1,156.64	1,156.92	1,157.10
$\nu[\text{O-H}]$ bending vibration	1,027.80	1,027.42	1,028.06	1,027.57	1,028.02

results were also described in previous studies (Celebioglu et al., 2018; Hadian et al., 2018), and the bands between $1,700\text{ cm}^{-1}$ and $1,600\text{ cm}^{-1}$ are associated with HOH bonds which are also abundant in the compound (Abarca et al., 2016). IR spectra are particularly sensitive to the presence of water; therefore, the spectral region of interest presents this contribution assigned to the HOH bending mode (Venuti et al., 2015). In addition, the spectra display the OH bending vibration in the range of $1,030\text{--}1,015\text{ cm}^{-1}$ and C-O-C stretch, between $1,159$ and $1,143\text{ cm}^{-1}$. These elongation vibrations were described previously in the literature (Celebioglu et al., 2018; Gao et al., 2020). A broad hydroxyl band of pure β -cyclodextrin spectrum (Figures 3, 4) shows the maximum absorption at $3,384.57\text{ cm}^{-1}$, found in the FTIR spectrum of the inclusion complexes which is a

good indication of their formation due to the stretching vibrations of the different β -cyclodextrin OH groups (Yang and Song, 2005; Kayaci et al., 2013; Celebioglu et al., 2018; Bríñez-Ortega et al., 2020).

Table 1 shows some increase and decrease in intensity changes due to the insertion of the part ring into the electron-rich cavity of β -cyclodextrin. Some bands showed little or no changes in the band upon complexation, implying that the inclusion of guest molecules in the CD cavity does not affect this vibrational mode. Although the spectrum of the inclusion complex appears almost similar to that of β -cyclodextrin alone, these results indicate the formation of the inclusion complex due to the weak interactions when partial inclusion of the ligand occurs (Stepniak et al., 2015; Vestland et al., 2015).

In relation to the presence of eugenol in the composition of the complex, it can be proven by the presence of absorption regions at 1515.50 cm^{-1} (Figure 3), related to the C=C bonds of the aromatic ring of the compound that usually appears in the region between 1650 and 1250 cm^{-1} corresponding to the vibration of the C=C groups and CH flexion of the alkene/aromatic groups of the eugenol (Yang and Song, 2005; Kayaci et al., 2013; Celebioglu et al., 2018). Additionally, frequencies between $1,200\text{ cm}^{-1}$ and $1,000\text{ cm}^{-1}$ are related to the presence of CO stretches (Yang and Song, 2005; Scremin et al., 2018). In addition, for eugenol, the characteristic bands are found in the ranges of $3,380$ – $3,360\text{ cm}^{-1}$ due to the axial stretching of the OH group (Rodríguez et al., 2021). Therefore, the results presented previously indicate the efficiency of the encapsulation process of the eugenol molecules by the β -cyclodextrin (Wang et al., 2011).

Generally, the comparative analysis of the physical mixture with the inclusion complex (Figure 4) presents the simple sum of the β -cyclodextrin and ligand bands (Zheng et al., 2020). A similar result was observed in the spectra. In the region of the absorption bands of the C–H stretches, we noted an overlap of the complex bands with that of β -cyclodextrin, without changes in the wavenumbers of the pure components. Similar results were noted in the region between $1,700\text{ cm}^{-1}$ and 1650 cm^{-1} , where stretch bands C=C with subtle deformation were observed. These results suggest that the simple mixture of the two components in the solid phase is not enough to prove the formation of the inclusion complex, once the shape, intensity, and position of the peaks vary. These observations, combined with the results obtained by XRD, can be considered strong evidence of the formation of the inclusion complex.

Molecular modeling analyses

Molecular interactions and formation of host–guest complex model

Molecular modeling analyses have been widely applied to assess the conformational, magnetic, and electronic properties of molecules (de Castro et al., 2014; de Sousa et al., 2016; de Souza Farias et al., 2021; Mustafa et al., 2021). Here, these computational analyses of the host–guest interactions were performed to better understand the formation of the eugenol– β -cyclodextrin complex and to provide additional insights into the complex model, especially when it is subjected to different temperatures in an aqueous solution.

The molecular analyses of the inclusion complex using molecular docking demonstrated that eugenol predominately had van der Waals interactions with the interior cavity of the β -cyclodextrin. Previous studies demonstrated that the geometric optimization of eugenol using density functional theory led eugenol to project its structure onto the external surface of the host (Mahboub, 2014; Chowdhry et al., 2015). Figure 5

shows the final conformation acquired by eugenol when complexed with β -cyclodextrin, obtained from molecular docking. The complex showed binding energy equal to $-4.0\text{ kcal mol}^{-1}$, and two hydrogen bond interactions were formed in the complex which contributed to the stability of eugenol in the inclusion complex. These molecular findings agree with the results reported previously (Joardar et al., 2020).

Molecular dynamics simulations of eugenol– β -cyclodextrin complex at different temperatures

Experimental thermal analyses, in general, reveal marked structural differences between isolated molecules and the inclusion complexes. In addition, these analyses exhibit a typical sharp melting endotherm at temperatures over 250°C , which is indicative of the anhydrous and crystalline state of the analyzed molecules; it also exhibits effects regarding their dehydration and degradation process (Seo et al., 2010; Piletti et al., 2017; Rodríguez et al., 2021). Since previous studies have identified a slower controlled release of eugenol at elevated temperatures, such as 50°C , 75°C , and 100°C (Kayaci et al., 2013; Piletti et al., 2017; Celebioglu et al., 2018), we analyzed the eugenol– β -cyclodextrin complex using a gradual temperature increase to evaluate the behavior and stability of the complex in aqueous solution, which is also compatible with its applications in repellent formulations (Mapossa et al., 2021). Computationally, we demonstrated that when the temperature is increased, it directly affects the interaction and stability of the complex (see RMSD plot, Figure 6A).

Between temperatures 27°C and 48°C , the complex showed stability, and eugenol maintained its interactions. However, we noted that for the temperature of 58°C , the RMSD plot showed high fluctuations (Figure 6), and eugenol was shown to be unstable in the complex, thus losing some of its interactions with the host and leaving the β -cyclodextrin cavity. Low stability of the inclusion complex can be attributed to poor interaction and orientation of eugenol inside the β -cyclodextrin cavity. These results demonstrate that the high temperatures impair the formation of the eugenol– β -cyclodextrin complex, reducing its interactions and leading to the departure of the eugenol from the complex. We conjecture that the complex will remain stable in solution when subjected to moderate temperatures due to the protection and stability provided by the β -cyclodextrin.

Since the binding affinity depends on the molecular interactions formed between eugenol and the β -cyclodextrin surface and directly affects the temperature increase (Supplementary Figures S2, S3, S4), we investigated the β -cyclodextrin–eugenol complex using different temperatures. The electrostatic (E_{elec}) and van der Waals (E_{vdw}) energies of the analyzed systems (Figure 7) were calculated using the LIE implemented in the Cpptraj program (<https://amberhub.chpc.utah.edu/lie/>) (Åqvist et al., 2002; Brandsdal et al., 2003; Roe and Cheatham, 2013). At 20 ns and 30 ns of MD trajectory, the complex showed an E_{elec} value of approximately $-100\text{ kcal mol}^{-1}$;

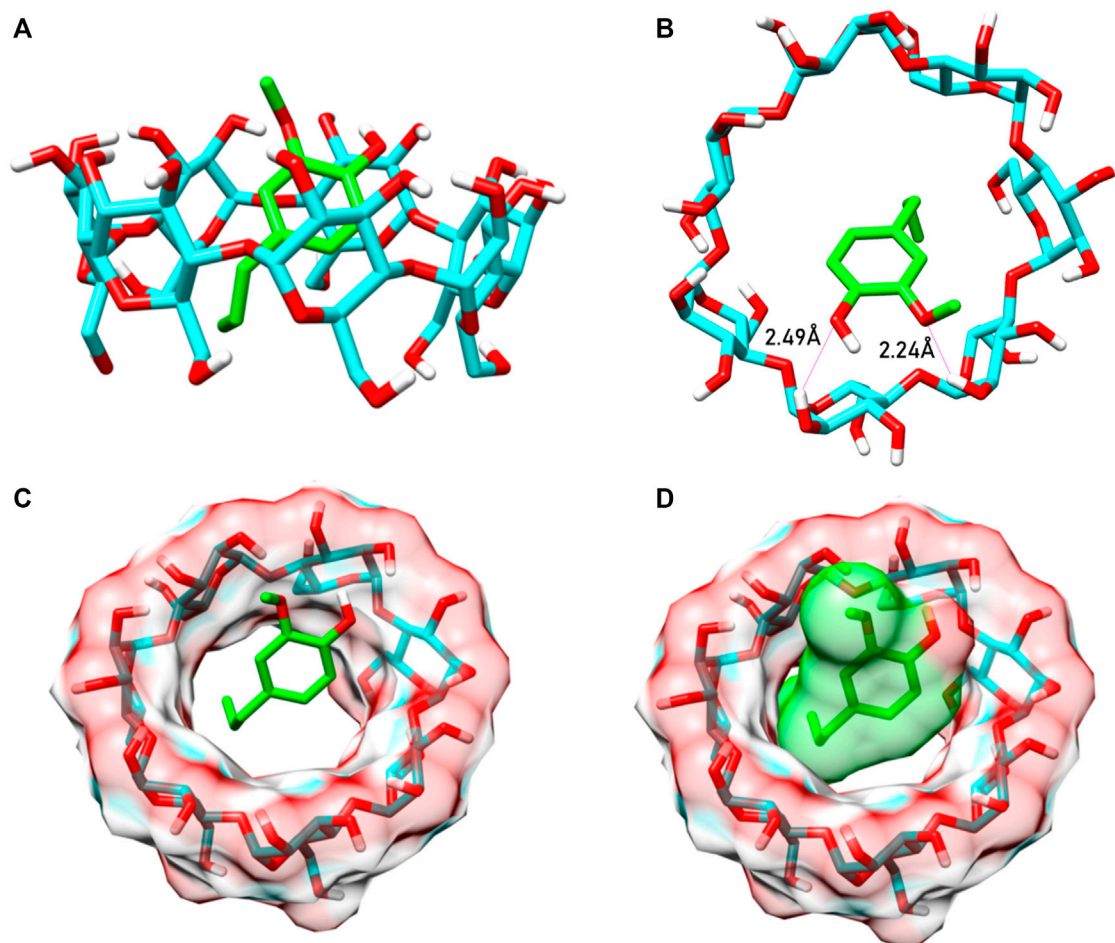


FIGURE 5

Inclusion complex formed between the eugenol (green) and the β -cyclodextrin (cyan). (A) 3D structure of BCD with eugenol forming the complex. (B) Exerting hydrogen bonds with the main electronegative group structure of EG. (C) EG (in green) inside the BCD cavity. (D) Eugenol with VDW surface, demonstrating binding in the BCD cavity.

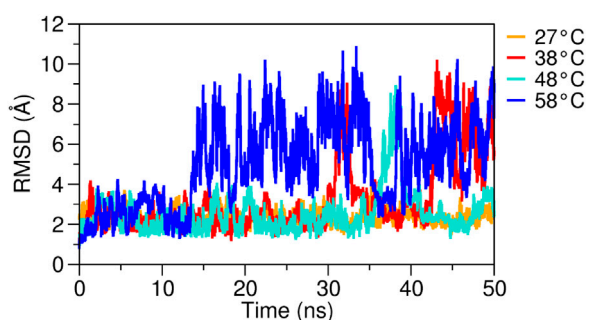


FIGURE 6

RMSD plots of the eugenol- β -cyclodextrin complex were evaluated at different temperatures. Complex analyzed over 50 ns of MD simulation at 27°C, 38°C, 48°C, and 58°C.

however, with temperature increase, the energies tend to move toward values near zero. In contrast, the E_{vdw} increases with increasing temperature, thus indicating the presence of repulsion forces caused by the instability of the complex. Recently, a study demonstrated that the binding affinity in host-guest systems including β -cyclodextrin may be estimated with a root mean square error $<1.5 \text{ kcal mol}^{-1}$ from the experimental results using the LIE method (Montalvo-Acosta et al., 2018), which is closely related to the error of $<1 \text{ kcal mol}^{-1}$ found in the experimental values obtained in the prediction of the relative binding affinity for a vast range of protein-ligand systems (Gapsys et al., 2020).

Additionally, the binding free energy was calculated using the MM/GBSA method for different temperatures, considering the total complexation time (Figure 8). The LIE and MM/GBSA are

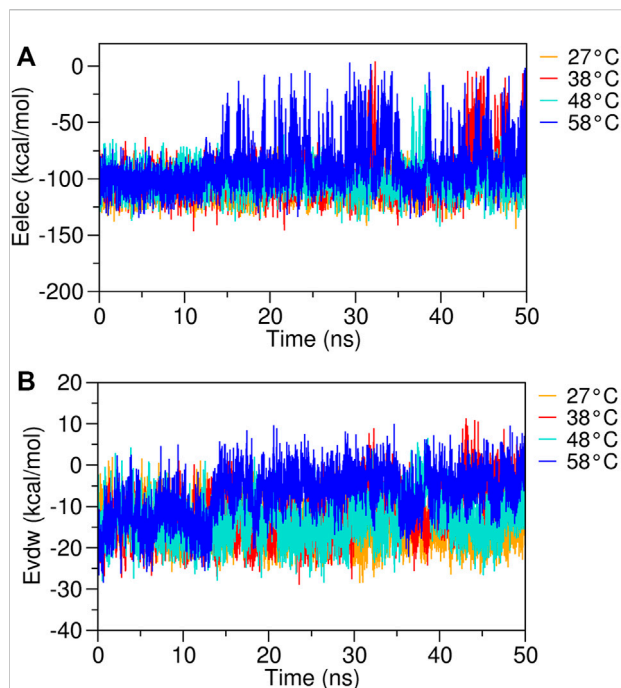


FIGURE 7
RMSD plots of the eugenol- β -cyclodextrin complex were obtained from different temperatures. **(A)** electrostatic energy (Eelec) and **(B)** van der Waals energy (Evdw) complex analyzed over 50 ns of MD simulation at 27°C, 38°C, 48°C, and 58°C.

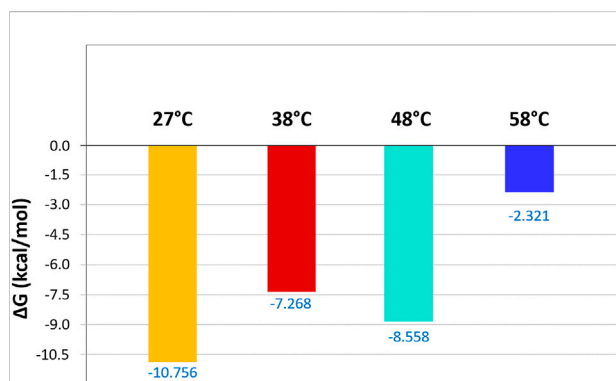


FIGURE 8
Binding free energies (kcal mol^{-1}) were calculated using the MM/GBSA method for the eugenol- β -cyclodextrin complex. The inclusion complex was analyzed at different temperatures (27°C, 38°C, 48°C, and 58°C).

both end-point methods widely applied to investigate the binding free energy of biomolecular complexes (Cardoso et al., 2021; Costa et al., 2021; da Costa et al., 2019b; de Oliveira et al., 2020; Fonseca et al., 2020), and their results tend to show a trend similar to the energy values obtained from the experimental methods (Hansson et al., 1998; Aqvist and Marelius, 2001; Zhu et al., 2017; Rifai et al., 2020). The

energy at room temperature (27°C) showed a highly stable complex, with eugenol demonstrating a high binding affinity with the β -cyclodextrin. However, with the increase of the temperature, the complex stability is affected and we have a decrease in the free energy of the analyzed systems, thus reducing the affinity of eugenol and causing its departure from the complex. This result indicates that the eugenol- β -cyclodextrin complex is stable at moderate temperatures and guarantees the permanence of the molecule inside the β -cyclodextrin with stronger binding.

Conclusion

Here, we obtained eugenol- β -cyclodextrin inclusion complexes through co-precipitation and solvent evaporation. Then, the inclusion complex was characterized using the X-Ray diffraction and Fourier transform infrared spectroscopy, confirming the formation of the host-guest inclusion complex. Additionally, our computational analyses demonstrated that the eugenol- β -cyclodextrin complex remains stable between temperatures 27°C and 48°C. In contrast, high temperatures impair the formation of the eugenol- β -cyclodextrin complex, reducing its interactions and leading to its premature departure from the complex which is consistent with controlled release of the repellent.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their financial support. The authors are also thankful for access to the computational resources of the Supercomputer Santos Dumont (SDumont), provided by the Laboratório de Computação Científica (LNCC). We also would like to thank to the financial support of the Pro-reitoria de Pesquisa da UFPA (PROPESP/UFPA).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.1061624/full#supplementary-material>

References

- Abarca, R. L., Rodríguez, F. J., Guarda, A., Galotto, M. J., and Bruna, J. E. (2016). Characterization of beta-cyclodextrin inclusion complexes containing an essential oil component. *Food Chem.* 196, 968–975. doi:10.1016/j.foodchem.2015.10.023
- Afify, A., Betz, J. F., Riabinina, O., Lahondère, C., and Potter, C. J. (2019). Commonly used insect repellents hide human odors from *Anopheles* mosquitoes. *Curr. Biol.* 29 (21), 3669–3680. e5. doi:10.1016/j.cub.2019.09.007
- Afify, A., and Potter, C. J. (2020). Insect repellents mediate species-specific olfactory behaviours in mosquitoes. *Malar. J.* 19 (1), 127–210. doi:10.1186/s12936-020-03206-8
- Alvira, E. (2018). Theoretical study of the β -cyclodextrin inclusion complex formation of eugenol in water. *Molecules* 23 (4), 928. doi:10.3390/molecules23040928
- Anaya-Castro, M. A., Ayala-Zavala, J. F., Muñoz-Castellanos, L., Hernández-Ochoa, L., Peydecastaing, J., and Durrieu, V. (2017). β -Cyclodextrin inclusion complexes containing clove (*eugenia caryophyllata*) and Mexican oregano (*lippia berlandieri*) essential oils: Preparation, physicochemical and antimicrobial characterization. *Food Packag Shelf Life* 14, 96–101. doi:10.1016/j.fpsl.2017.09.002
- AndersenRattle, H. C. (1983). Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52 (1), 24–34. doi:10.1016/0021-9991(83)90014-1
- Åqvist, J., Luzhkov, V. B., Brandsdal, B. O., Åqvist, J., Luzhkov, V. B., and Brandsdal, B. O. (2002). Ligand binding affinities from MD simulations. *Acc. Chem. Res.* 35 (6), 358–365. doi:10.1021/ar010014p
- Åqvist, J., and Marelus, J. (2001). The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High. Throughput Screen* 4 (8), 613–626. doi:10.2174/1386207013330661
- Ayala-Zavala, J. F., Soto-Valdez, H., González-León, A., Álvarez-Parrilla, E., Martín-Belloso, O., and González-Aguilar, G. A. (2008). Microencapsulation of cinnamon leaf (*cinnamomum zeylanicum*) and garlic (*allium sativum*) oils in β -cyclodextrin. *J. Incl. Phenom. Macrocycl. Chem.* 60 (3–4), 359–368. doi:10.1007/s10847-007-9385-1
- Barreto-Vieira, D. F., Couto-Lima, D., Jácome, F. C., Caldas, G. C., and Barth, O. M. (2020). Dengue, yellow fever, Zika and chikungunya epidemic arboviruses in Brazil: Ultrastructural aspects. *Mem. Inst. Oswaldo Cruz* 115, e200278. doi:10.1590/0074-02760200278
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97 (40), 10269–10280. doi:10.1021/j100142a004
- Beltrán Sanahuja, A., and Valdés García, A. (2021). New trends in the use of volatile compounds in food packaging. *Polym. (Basel)* 13 (7), 1053. doi:10.3390/polym13071053
- Biovia, D. S. (2017). *Discovery Studio modeling environment*. San Diego: Dassault Systèmes. Release.
- Brandsdal, B. O., Österberg, F., Almlöf, M., Feierberg, I., Luzhkov, V. B., and Åqvist, J. (2003). Free energy calculations and ligand binding. *Adv. Protein Chem.* 66, 123–158. doi:10.1016/S0065-3233(03)66004-3
- Bríñez-Ortega, E., De Almeida, V. L., Lopes, J. C. D., and Burgos, A. E. (2020). Partial inclusion of bis(1, 10-phenanthroline) silver(i) salicylate in β -cyclodextrin: Spectroscopic characterization, *in vitro* and *in silico* antimicrobial evaluation. *Acad Bras Cienc* 92 (3), e20181323–23. doi:10.1590/0001-3765202020181323
- Buedenbender, S., and Schulz, G. E. (2009). Structural base for enzymatic cyclodextrin hydrolysis. *J. Mol. Biol.* 385 (2), 606–617. doi:10.1016/j.jmb.2008.10.085
- Cao, F., Guo, J., and Ping, Q. (2005). The physicochemical characteristics of freeze-dried scutellarin-cyclodextrin tetra-component complexes. *Drug Dev. Ind. Pharm.* 31 (8), 747–756. doi:10.1080/03639040500216220
- Cardoso, R., Valente, R., Souza da Costa, C. H., da, S., Gonçalves Vianez, J. L., Santana da Costa, K., et al. (2021). Analysis of kojic acid derivatives as competitive inhibitors of tyrosinase: A molecular modeling approach. *Molecules* 26 (10), 2875. doi:10.3390/molecules26102875
- Celebioglu, A., Yildiz, Z. I., and Uyar, T. (2018). Fabrication of electrospun eugenol/cyclodextrin inclusion complex nanofibrous webs for enhanced antioxidant property, water solubility, and high temperature stability. *J. Agric. Food Chem.* 66 (2), 457–466. doi:10.1021/acs.jafc.7b04312
- Chowdhry, B. Z., Ryall, J. P., Dines, T. J., and Mendham, A. P. (2015). Infrared and Raman spectroscopy of eugenol, isoeugenol, and methyl eugenol: Conformational analysis and vibrational assignments from density functional theory calculations of the anharmonic fundamentals. *J. Phys. Chem. A* 119 (46), 11280–11292. doi:10.1021/acs.jpca.5b07607
- Costa, C. H. S., Santos, A. M., Alves, C. N., Martí, S., Moliner, V., Santana, K., et al. (2021). Assessment of the PETase conformational changes induced by poly(ethylene terephthalate) binding. *Proteins Struct. Funct. Bioinforma.* 89 (10), 1340–1352. doi:10.1002/prot.26155
- da Costa, K. S., Galúcio, J. M., Da Costa, C. H. S., Santana, A. R., Dos Santos Carvalho, V., Do Nascimento, L. D., et al. (2019a). Exploring the potentiality of natural products from essential oils as inhibitors of odorant-binding proteins: A structure- and ligand-based virtual screening approach to find novel mosquito repellents. *ACS Omega* 4 (27), 22475–22486. doi:10.1021/acsomega.9b03157
- da Costa, K. S., Galúcio, J. M., de Jesus, D. A., Gomes, G. C., Lima e Lima, A. H., Taube, P. S., et al. (2019b). Targeting peptidyl-prolyl cis-trans isomerase NIMA-interacting 1: A structure-based virtual screening approach to find novel inhibitors. *Curr. Comput. Aided Drug Des.* 15, 605–617. doi:10.2174/1573409915666191025114009
- de Castro, E. A. S., de Oliveira, D. A. B., Farias, S. A. S., Gargano, R., and Martins, J. B. L. (2014). Structure and electronic properties of azadirachtin. *J. Mol. Model* 20 (2), 2084. doi:10.1007/s00894-014-2084-0
- de Freitas, C. A. B., de Araújo, R. C. S., da Paz, S. P. A., de Araújo Silva, J. R., Alves, C. N., and Lameira, J. (2021). Obtenção E caracterização de complexo de inclusão de β -ciclodextrina E eugenol/ preparation and characterization of β -cyclodextrin inclusion complex of eugenol. *Braz. J. Dev.* 7 (3), 33056–33070. Obtenção E Caracterização De Complexo De Inclusão De B-Ciclodextrina E. doi:10.34117/bjdv7n3-838
- de Oliveira, M. D., Araújo, J. de O., Galúcio, J. M. P., Santana, K., and Lima, A. H. (2020). Targeting shikimate pathway: *In silico* analysis of phosphoenolpyruvate derivatives as inhibitors of EPSP synthase and DAHP synthase. *J. Mol. Graph Model* 101, 107735. doi:10.1016/j.jmgm.2020.107735
- de Sousa, S. M. R., Guimarães, L., Ferrari, J. L., De Almeida, W. B., and Nascimento, C. S. (2016). A DFT investigation on the host/guest inclusion process of pilocaine into β -cyclodextrin. *Chem. Phys. Lett.* 652, 123–129. doi:10.1016/j.cplett.2016.04.053

- de Souza Farias, S. A., da Costa, K. S., and Martins, J. B. L. (2021). Analysis of conformational, structural, magnetic, and electronic properties related to antioxidant activity: Revisiting flavan, anthocyanidin, flavanone, flavonol, isoflavone, flavone, and flavan-3-ol. *ACS Omega* 6 (13), 8908–8918. doi:10.1021/acsomega.0c06156
- Do Nascimento, L. D., de Moraes, A. A. B., da Costa, K. S., Galúcio, J. M. P., Taube, P. S., Costa, C. M. L., et al. (2020). Bioactive natural compounds and antioxidant activity of essential oils from spice plants: New findings and potential applications. *Biomolecules* 10 (7), 988–1037. doi:10.3390/biom10070988
- Dos Passos Menezes, P., Dos Santos, P. B. P., Dória, G. A. A., de Sousa, B. M. H., Serafini, M. R., Nunes, P. S., et al. (2017). Molecular modeling and physicochemical properties of supramolecular complexes of limonene with α - and β -cyclodextrins. *AAPS PharmSciTech* 18 (1), 49–57. doi:10.1208/s12249-016-0516-0
- El-Saber Batiha, G., Alkzami, L. M., Wasef, L. G., Beshbishy, A. M., Nadwa, E. H., and Rashwan, E. K. (2020). Syzygium aromaticum L. (Myrtaceae): Traditional uses, bioactive chemical constituents, pharmacological and toxicological activities. *Biomolecules* 10 (2), 202. doi:10.3390/biom10020202
- Enoch, I. V. M. V., and Swaminathan, M. (2005). Dual fluorescence and photoprototropic characteristics of 2-Aminodiphenylsulphone- β -Cyclodextrin inclusion complex. *J. Incl. Phenom. Macrocycl. Chem.* 53 (3–4), 149–154. doi:10.1007/s10847-005-2633-3
- Fonseca, E. C. M., da Costa, K. S., Lameira, J., Alves, C. N., and Lima, A. H. (2020). Investigation of the target-site resistance of EPSP synthase mutants P106T and T102I/P106S against glyphosate. *RSC Adv.* 10 (72), 44352–44360. doi:10.1039/D0RA09061A
- Frisch, M. J., Nathan, A. J., and Scobell, A., (2009). *Gaussian 09*. Wallingford CT: Gaussian, Inc., 2111–2113. Gaussian Inc.: Wallingford CT.
- Galúcio, J. M., Monteiro, E. F., de Jesus, D. A., Costa, C. H., Siqueira, R. C., Santos, G. B. D., et al. (2019). *In silico* identification of natural products with anticancer activity using a chemo-structural database of Brazilian biodiversity. *Comput. Biol. Chem.* 83, 107102. doi:10.1016/j.compbiolchem.2019.107102
- Gao, S., Jiang, J., Li, X., Liu, Y., Zhao, L., Fu, Y., et al. (2020). Enhanced physicochemical properties and herbicidal activity of an environment-friendly clathrate formed by β -cyclodextrin and herbicide cyanazine. *J. Mol. Liq.* 305, 112858. doi:10.1016/j.jmolliq.2020.112858
- Gapsys, V., Pérez-Benito, L., Aldeghi, M., Seeliger, D., van Vlijmen, H., Tresadern, G., et al. (2020). Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* 11 (4), 1140–1152. doi:10.1039/C9SC03754C
- Gong, L., Li, T., Chen, F., Duan, X., Yuan, Y., Zhang, D., et al. (2016). An inclusion complex of eugenol into β -cyclodextrin: Preparation, and physicochemical and antifungal characterization. *Food Chem.* 196, 324–330. doi:10.1016/j.foodchem.2015.09.052
- Hadian, Z., Maleki, M., Abdi, K., Atyabi, F., Mohammadi, A., and Khaksar, R. (2018). Preparation and characterization of nanoparticle β -cyclodextrin:geraniol inclusion complexes. *Iran. J. Pharm. Res.* 17 (1), 39–51.
- Hansson, T., Marelus, J., and Åqvist, J. (1998). Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* 12 (1), 27–35. doi:10.1023/A:1007930623000
- Hestenes, M. R., and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* (1934) 49 (6), 409. doi:10.6028/jres.049.044
- Higuera, A., and Ramirez, J. D. (2019). Molecular epidemiology of dengue, yellow fever, Zika and chikungunya arboviruses: An update. *Acta Trop.* 190, 99–111. doi:10.1016/j.actatropica.2018.11.010
- Jaganathan, S. K., Mazumdar, A., Mondhe, D., and Mandal, M. (2011). Apoptotic effect of eugenol in human colon cancer cell lines. *Cell Biol. Int.* 35 (6), 607–615. doi:10.1042/CBI20100118
- Jansook, P., Ogawa, N., and Loftsson, T. (2018). Cyclodextrins: Structure, physicochemical properties and pharmaceutical applications. *Int. J. Pharm.* 535 (1–2), 272–284. doi:10.1016/j.ijpharm.2017.11.018
- Jiang, L., Yang, J., Wang, Q., Ren, L., and Zhou, J. (2019). Physicochemical properties of catechin/ β -cyclodextrin inclusion complex obtained via Co-precipitation. *CyTA - J. Food* 17 (1), 544–551. doi:10.1080/19476337.2019.1612948
- Joardar, A., Meher, G., Bag, B. P., and Chakraborty, H. (2020). Host-guest complexation of eugenol in cyclodextrins for enhancing bioavailability. *J. Mol. Liq.* 319, 114336. doi:10.1016/j.jmolliq.2020.114336
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79 (2), 926–935. doi:10.1063/1.445869
- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118 (45), 11225–11236. doi:10.1021/ja9621760
- Kamatou, G. P., Vermaak, I., and Viljoen, A. M. (2012). Eugenol—from the remote Maluku islands to the international market place: A review of a remarkable and versatile molecule. *Molecules* 17 (6), 6953–6981. doi:10.3390/molecules17066953
- Kayaci, F., Ertas, Y., and Uyar, T. (2013). Enhanced thermal stability of eugenol by cyclodextrin inclusion complex encapsulated in electrospun polymeric nanofibers. *J. Agric. Food Chem.* 61 (34), 8156–8165. doi:10.1021/jf402923c
- Kayaci, F., and Uyar, T. (2011). Solid inclusion complexes of vanillin with cyclodextrins: Their formation, characterization, and high-temperature stability. *J. Agric. Food Chem.* 59 (21), 11772–11778. doi:10.1021/jf202915c
- Kfoury, M., Landy, D., and Fourmentin, S. (2018). Characterization of cyclodextrin/volatile inclusion complexes: A review. *Molecules* 23 (5), 1204. doi:10.3390/molecules23051204
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29 (4), 622–655. doi:10.1002/jcc.20820
- Kuskoski, E. M., Vega, J. M., Rios, J. J., Fett, R., Troncoso, A. M., and Asuero, A. G. (2003). Characterization of anthocyanins from the fruits of bagueçu (eugenia umbelliflora berg). *J. Agric. Food Chem.* 51 (18), 5450–5454. doi:10.1021/jf030014z
- Lee, J., Lee, S.-S., Lee, S., and Oh, H. B. (2020). Noncovalent complexes of cyclodextrin with small organic molecules: Applications and insights into host-guest interactions in the gas phase and condensed phase. *Molecules* 25 (18), 4048. doi:10.3390/molecules25184048
- Loftsson, T., and Brewster, M. E. (1996). Pharmaceutical applications of cyclodextrins. 1. Drug solubilization and stabilization. *J. Pharm. Sci.* 85 (10), 1017–1025. doi:10.1021/js950534b
- Lupi, E., Hatz, C., and Schlagenhauf, P. (2013). The efficacy of repellents against Aedes, Anopheles, Culex and ixodes spp. – a literature review. *Travel Med. Infect. Dis.* 11 (6), 374–411. doi:10.1016/j.tmaid.2013.10.005
- Mahboub, R. (2014). Structural conformational study of eugenol derivatives using semiempirical methods. *Adv. Chem.* 2014, 1–5. doi:10.1155/2014/490358
- Mapossa, A. B., Focke, W. W., Tewo, R. K., Androsch, R., and Kruger, T. (2021). Mosquito-repellent controlled-release formulations for fighting infectious diseases. *Malar. J.* 20 (1), 165. doi:10.1186/s12936-021-03681-7
- Mateen, S., Shahzad, S., Ahmad, S., Naeem, S. S., Khalid, S., Akhtar, K., et al. (2019). Cinnamaldehyde and eugenol attenuates collagen induced arthritis via reduction of free radicals and pro-inflammatory cytokines. *Phytomedicine* 53, 70–78. doi:10.1016/j.phymed.2018.09.004
- Methods, D. (2021). Venugopal jayapal mahatma gandhi medical college & research Institute. *Pillaiyar Kuppm - 607402, Puducherry, India* 12 (1), 330–335. doi:10.13040/IJPSR.0975-8232.12(1).330-35
- Miot, H. A., Lauterbach, G. d. P., Ribeiro, F. A. H., Júnior, F. E. L., Hercos, G. N., Madeira, N. G., et al. (2011). Comparison among homemade repellents made with cloves, picaridin, and andiroba, and soybean oil against Aedes aegypti bites. *Rev. Soc. Bras. Med. Trop.* 44 (6), 793–794. doi:10.1590/S0037-86822011000600030
- Montalvo-Acosta, J. J., Pacak, P., Gomes, D. E. B., and Cecchini, M. (2018). A linear interaction energy model for cavitand host-guest binding affinities. *J. Phys. Chem. B* 122 (26), 6810–6814. doi:10.1021/acs.jpcc.8b03245
- Moyer, R. A., Hummer, K. E., Finn, C. E., Frei, B., and Wrostad, R. E. (2002). Anthocyanins, phenolics, and antioxidant capacity in diverse small fruits: Vaccinium, rubus, and ribes. *J. Agric. Food Chem.* 50 (3), 519–525. doi:10.1021/jf011062r
- Mustafa, S. F. Z., Arsad, S. R., Mohamad, H., Abdallah, H. H., and Maarof, H. (2021). Host-guest molecular encapsulation of cucurbit[7]Uril with dillapiole congeners using docking simulation and density functional theory approaches. *Struct. Chem.* 32 (3), 1151–1161. doi:10.1007/s11224-020-01708-4
- Muthu Vijayan Enoch, I. V., and Swaminathan, M. (2004). Inclusion complexation of 2-amino-7-bromofluorene by β -cyclodextrin: Spectral characteristics and the effect of PH. *J. Fluoresc.* 14 (6), 751–756. doi:10.1023/B:JOFL.0000047226.60559.09
- Newman, D. J., and Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* 83 (3), 770–803. doi:10.1021/acs.jnatprod.9b01285
- Paixão, E. S., Teixeira, M. G., and Rodrigues, L. C. (2018). Zika, chikungunya and dengue: The causes and threats of new and reemerging arboviral diseases. *BMJ Glob. Health* 3, e000530–e000536. doi:10.1136/bmjgh-2017-000530
- Pena, G. A., da Costa Lopes, A. S., de Moraes, S. H. S., do Nascimento, L. D., dos Santos, F. R. R., da Costa, K. S., et al. (2022). Host-guest inclusion complexes of natural products and nanosystems: Applications in the development of repellents. *Molecules* 27 (8), 2519. doi:10.3390/molecules27082519

- Piletti, R., Bugiereck, A. M., Pereira, A. T., Gussati, E., Dal Magro, J., Mello, J. M. M., et al. (2017). Microencapsulation of eugenol molecules by β -cyclodextrine as a thermal protection method of antibacterial action. *Mater. Sci. Eng. C Mater. Biol. Appl.* 75, 259–271. doi:10.1016/j.msec.2017.02.075
- Ray, A. (2015). Reception of odors and repellents in mosquitoes. *Curr. Opin. Neurobiol.* 34, 158–164. doi:10.1016/j.conb.2015.06.014
- Rayan, A., Raiyn, J., and Falah, M. (2017). Nature is the best source of anticancer drugs: Indexing natural products for their anticancer bioactivity. *PLoS One* 12 (11), e0187925. doi:10.1371/journal.pone.0187925
- Rezaeisadat, M., Salehi, N., and Bordbar, A.-K. (2021). Inclusion of levodopa into β -cyclodextrin: A comprehensive computational study. *ACS Omega* 6 (37), 23814–23825. doi:10.1021/acsomega.1c02637
- Ribeiro, A., Figueiras, A., Santos, D., and Veiga, F. (2008). Preparation and solid-state characterization of inclusion complexes formed between miconazole and methyl- β -cyclodextrin. *AAPS PharmSciTech* 9 (4), 1102–1109. doi:10.1208/s12249-008-9143-8
- Rifai, E. A., van Dijk, M., and Geerke, D. P. (2020). Recent developments in linear interaction energy based binding free energy calculations. *Front. Mol. Biosci.* 7, 114. doi:10.3389/fmolb.2020.00114
- Rodríguez, F. J., Guarda, A., Galotto, M., and J. Galotto, M. (2021). Beta-Cyclodextrin: Eugenol inclusion complexes: Characterization and antifungal capacity. *ESPOCH Congr. Ecuadorian J. S.T.E.A.M.* 1 (1), 463–477. doi:10.18502/espoch.v1i1.9584
- Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9 (7), 3084–3095. doi:10.1021/ct400341p
- Roth-Walter, F., Moskovskich, A., Gomez-Casado, C., Diaz-Perales, A., Oida, K., Singer, J., et al. (2014). Immune suppressive effect of cinnamaldehyde due to inhibition of proliferation and induction of apoptosis in immune cells: Implications in cancer. *PLoS One* 9 (10), e108402. doi:10.1371/journal.pone.0108402
- Saha, S., Roy, A., Roy, K., and Roy, M. N. (2016). Study to explore the mechanism to form inclusion complexes of β -cyclodextrin with vitamin molecules. *Sci. Rep.* 6 (1), 35764. doi:10.1038/srep35764
- Santana, K., do Nascimento, L. D., Lima e Lima, A., Damasceno, V., Nahum, C., Braga, R. C., et al. (2021). Applications of virtual screening in bioprospecting: Facts, shifts, and perspectives to explore the chemo-structural diversity of natural products. *Front. Chem.* 9, 662688. doi:10.3389/fchem.2021.662688
- Scremin, F. R., Veiga, R. S., Silva-Buzanello, R. A., Becker-Algeri, T. A., Corso, M. P., Torquato, A. S., et al. (2018). Synthesis and characterization of protein microcapsules for eugenol storage. *J. Therm. Anal. Calorim.* 131 (1), 653–660. doi:10.1007/s10973-017-6302-8
- Seo, E.-J., Min, S.-G., and Choi, M.-J. (2010). Release characteristics of freeze-dried eugenol encapsulated with β -cyclodextrin by molecular inclusion method. *J. Microencapsul.* 27 (6), 496–505. doi:10.3109/02652041003681398
- Singh, R. B., Mahanta, S., and Guchhait, N. (2010). Spectral modulation of charge transfer fluorescence probe encapsulated inside aqueous and non-aqueous β -cyclodextrin nanocavities. *J. Mol. Struct.* 963 (1), 92–97. doi:10.1016/j.molstruc.2009.10.024
- Stepniak, A., Belica-Pacha, S., Rozalska, S., Długonski, J., Urbaniak, P., and Palecz, B. (2015). Study on a host–guest interaction of β -cyclodextrin with tebuconazole in water. *J. Mol. Liq.* 211, 288–293. doi:10.1016/j.molliq.2015.07.023
- Szejtli, J. (1998). Introduction and general overview of cyclodextrin Chemistry. *Chem. Rev.* 98 (5), 1743–1754. doi:10.1021/cr970022c
- Tabanca, N., Bernier, R. U., Agramonte, M. N., Tsikolia, M., and Bloomquist, R. J. (2016). Discovery of repellents from natural products. *Curr. Org. Chem.* 20 (25), 2690–2702. doi:10.2174/1385272820666160421151503
- Tan, K., Faierstein, G. B., Xu, P., Barbosa, R. M. R., Buss, G. K., and Leal, W. S. (2019). A popular Indian clove-based mosquito repellent is less effective against *Culex quinquefasciatus* and *Aedes aegypti* than DEET. *PLoS One* 14 (11), e0224810. doi:10.1371/journal.pone.0224810
- Thomas, A., Mazigo, H. D., Manjurano, A., Morona, D., and Kweka, E. J. (2017). Evaluation of active ingredients and larvicidal activity of clove and cinnamon essential oils against *Anopheles gambiae* (sensu lato). *Parasit. Vectors* 10 (1), 411. doi:10.1186/s13071-017-2355-6
- Trott, O., and Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2), 455–461. NA-NA. doi:10.1002/jcc.21334
- Venuti, V., Rossi, B., D'Amico, F., Mele, A., Castiglione, F., Punta, C., et al. (2015). Combining Raman and infrared spectroscopy as a powerful tool for the structural elucidation of cyclodextrin-based polymeric hydrogels. *Phys. Chem. Chem. Phys.* 17 (15), 10274–10282. doi:10.1039/C5CP00607D
- Vestland, T. L., Jacobsen, Ø., Sande, S. A., Myrset, A. H., and Klaveness, J. (2015). Compactible powders of omega-3 and β -cyclodextrin. *Food Chem.* 185, 151–158. doi:10.1016/j.foodchem.2015.03.132
- Wang, J., Cao, Y., Sun, B., and Wang, C. (2011). Physicochemical and release characterisation of garlic oil- β -cyclodextrin inclusion complexes. *Food Chem.* 127 (4), 1680–1685. doi:10.1016/j.foodchem.2011.02.036
- Wang, J., Cieplak, P., and Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21 (12), 1049–1074. doi:10.1002/1096-987x(200009)21:12<1049:aid-jcc3>3.0.co;2-f
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25 (9), 1157–1174. doi:10.1002/jcc.20035
- Wüpper, S., Lüersen, K., and Rimbach, G. (2021). Cyclodextrins, natural compounds, and plant bioactives—A nutritional perspective. *Biomolecules* 11 (3), 401–421. doi:10.3390/biom11030401
- Xu, J. G., Liu, T., Hu, Q. P., and Cao, X. M. (2016). Chemical composition, antibacterial properties and mechanism of action of essential oil from clove buds against *Staphylococcus aureus*. *Molecules* 21 (9), 1194–1213. doi:10.3390/molecules21091194
- Xu, W., Li, X., Wang, L., Li, S., Chu, S., Wang, J., et al. (2021). Design of cyclodextrin-based functional systems for biomedical applications. *Front. Chem.* 9, 635507. doi:10.3389/fchem.2021.635507
- Yang, Y., and Song, L. X. (2005). Study on the inclusion compounds of eugenol with α - β - γ - and heptakis (2, 6-di-O-Methyl)- β -Cyclodextrins. *J. Incl. Phenom. Macrocycl. Chem.* 53 (1–2), 27–33. doi:10.1007/s10847-005-0247-4
- Yang, Z., Yao, X., Xiao, Z., Chen, H., and Ji, H. (2016). Preparation and release behaviour of the inclusion complexes of phenylethanol with β -cyclodextrin. *Flavour Fragr. J.* 31 (3), 206–216. doi:10.1002/ffj.3302
- Zheng, D., Xia, L., Ji, H., Jin, Z., and Bai, Y. (2020). A cyclodextrin-based controlled release system in the simulation of *in vitro* small intestine. *Molecules* 25 (5), 1212. doi:10.3390/molecules25051212
- Zhu, J., Li, X., Zhang, S., Ye, H., Zhao, H., Jin, H., et al. (2017). Exploring stereochemical specificity of phosphotriesterase by MM-PBSA and MM-GBSA calculation and steered molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 35 (14), 3140–3151. doi:10.1080/07391102.2016.1244494



OPEN ACCESS

EDITED BY

Burak Tüzün,
Cumhuriyet University, Turkey

REVIEWED BY

Xiaoxiao Huang,
Shenyang Pharmaceutical University,
China
Neeraj Kumar,
Northwestern University, United States
Mahmoud A. Al-Sha'er,
Zarqa Private University, Jordan
Huagang Sheng,
Shandong University of Traditional
Chinese Medicine, China

*CORRESPONDENCE

Lin-Tao Han,
✉ hanlintao@hbtcm.edu.cn

SPECIALTY SECTION

This article was submitted to
Theoretical and Computational
Chemistry, a section of the journal
Frontiers in Chemistry

RECEIVED 03 October 2022

ACCEPTED 15 December 2022

PUBLISHED 09 January 2023

CITATION

Chen P-Y and Han L-T (2023), Study on
the molecular mechanism of anti-liver
cancer effect of Evodiae fructus by
network pharmacology and
QSAR model.
Front. Chem. 10:1060500.
doi: 10.3389/fchem.2022.1060500

COPYRIGHT

© 2023 Chen and Han. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Study on the molecular mechanism of anti-liver cancer effect of Evodiae fructus by network pharmacology and QSAR model

Peng-Yu Chen¹ and Lin-Tao Han^{1,2*}

¹Faculty of Pharmacy, Hubei University of Chinese Medicine, Wuhan, China, ²Key Laboratory of Traditional Chinese Medicine Resources and Prescription, Ministry of Education, Wuhan, China

Introduction: Evodiae Fructus (EF) is the dried, near ripe fruit of *Euodia rutaecarpa* (Juss.) Benth in *Rutaceae*. Numerous studies have demonstrated its anti-liver cancer properties. However, the molecular mechanism of Evodiae fructus against liver cancer and its structure-activity connection still require clarification.

Methods: We utilized network pharmacology and a QSAR (2- and 3-dimensional) model to study the anti-liver cancer effect of Evodiae fructus. First, by using network pharmacology to screen the active substances and targets of Evodiae fructus, we investigated the signaling pathways involved in the anti-liver cancer actions of Evodiae fructus. The 2D-QSAR pharmacophore model was then used to predict the pIC₅₀ values of compounds. The hiphop method was used to create an ideal 3D-QSAR pharmacophore model for the prediction of Evodiae fructus compounds. Finally, molecular docking was used to validate the rationality of the pharmacophore, and molecular dynamics was used to disclose the stability of the compounds by assessing the trajectories in 10 ns using RMSD, RMSF, Rg, and hydrogen bonding metrics.

Results: In total, 27 compounds were acquired from the TCMSP and TCM-ID databases, and 45 intersection targets were compiled using Venn diagrams. Network integration analysis was used in this study to identify SRC as a primary target. Key pathways were discovered by KEGG pathway analysis, including PD-L1 expression and PD-1 checkpoint pathway, EGFR tyrosine kinase inhibitor resistance, and ErbB signaling pathway. Using a 2D-QSAR pharmacophore model and the MLR approach to predict chemical activity, ten highly active compounds were found. Two hydrophobic features and one hydrogen bond acceptor feature in the 3D-QSAR pharmacophore model were validated by

Abbreviations: BP, biological processes; CC, cell components; C-T-P, compounds-targets-pathways; EF, Evodiae fructus; GFA, genetic function approximation; GO, gene ontology; HCC, hepatocellular carcinoma; ICCA, intrahepatic cholangiocarcinoma; KEGG, kyoto encyclopedia of genes and genomes; MF, molecular functions; MLR, multiple linear regression; PLS, partial least squares; PPI, protein-protein interaction; QSAR, quantitative structure activity relationship; TCM, traditional Chinese medicine.

training set chemicals. The results of molecular docking revealed that 10 active compounds had better docking scores with SRC and were linked to residues *via* hydrogen and hydrophobic bonds. Molecular dynamics was used to show the structural stability of obacunone, beta-sitosterol, and sitosterol.

Conclusion: Pharmacophore 01 has high selectivity and the ability to distinguish active and inactive compounds, which is the optimal model for this study. Obacunone has the optimal binding ability with SRC. The pharmacophore model proposed in this study provides theoretical support for further screening effective anti-cancer Chinese herbal compounds and optimizing the compound structure.

KEYWORDS

Euodiae fructus, liver cancer, network pharmacology, molecular docking, QSAR model, molecular dynamics simulation

1 Introduction

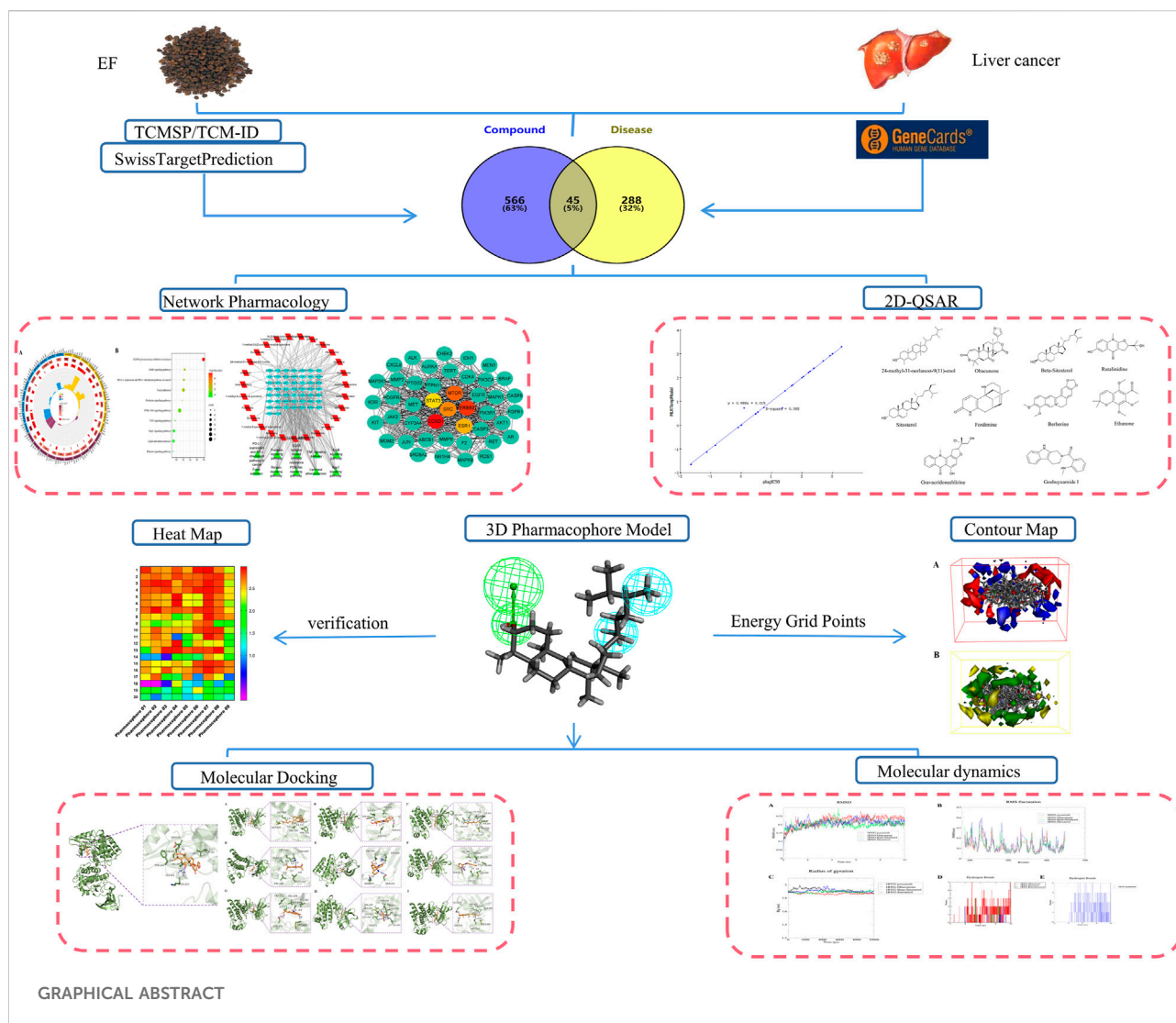
Liver cancer, including intrahepatic cholangiocarcinoma (ICC) and hepatocellular carcinoma (HCC), is the second-largest cause of cancer-related mortality worldwide and a significant public health issue (Kassebaum et al., 2014). Its incidence rate and mortality have increased annually, and over the past 20 years, it has become the cancer factor in the United States that has caused the biggest increase in mortality (Sia et al., 2017). In 2017, More than half of all liver cancer cases worldwide—5,70,000 new cases—occurred in China. The mortality rate was 26.26/1,00,000, and men had a greater mortality rate than women (37.55/1,00,000 vs. 14.45/1,00,000) (He et al., 2021). Cao et al. (2021) conducted a secondary analysis of cancer statistics from around the world. In 2020, cancer cases in China accounted for 24% of newly confirmed cases worldwide and 30% of cancer deaths worldwide. The death rate from liver cancer rose to second in China in 2020.

Currently, the clinical treatment of liver cancer includes surgery and drug chemotherapy. On the one hand, liver cancer has an extremely dismal prognosis; only 5%–15% of patients are candidates for surgical resection. Therefore, it is only appropriate for early-stage patients with some liver regeneration potential (Anwanwan et al., 2020). On the other hand, the most widely prescribed medication for individuals with advanced stages is the kinase inhibitor sorafenib. However, less than one-third of patients will be able to fully benefit from the course of their treatment. Sorafenib resistance is visible after use, and problems including toxicity and ineffectiveness can also result from prolonged exposure to chemotherapy medicines (El-Serag et al., 2008). Given the poor prognosis of liver cancer, scientists and physicians have been looking for new treatment options to improve patient survival.

Natural products provide distinct advantages in cancer treatment. Natural plant extracts and natural chemicals, as well as traditional Chinese medicines, have gained a lot of attention in recent years for their high-efficiency and low-toxicity anti-cancer

characteristics (Cho et al., 2015). Evodiae Fructus (EF), a traditional medicinal plant, is the dried and nearly ripe fruit of the *Rutaceae* *Euodia rutaecarpa* (Juss.) Benth. EF contains alkaloids, terpenes, flavonoids, phenolic acids, steroids, and phenylpropanoids (Li and Wang, 2020). Modern pharmacological researches have demonstrated its biological properties, including cardioprotective, antibacterial, anti-inflammatory, and anti-tumor effects (Yoon et al., 2013; Zhao et al., 2019; Shan et al., 2020). According to several research, evodiamine, and rutaecarpine in EF had hepatoprotective properties (Yoon et al., 2013). Meanwhile, evodiamine may induce apoptosis in liver cancer cells *via* the WWOX-dependent pathway, as well as the Akt pathway and others, and then exert antitumor effects (Hu et al., 2017; Yang et al., 2017). Furthermore, EF extracts can prevent the development of a variety of cancers, including colon cancer, cervical cancer, and others (Dong et al., 2010; Park et al., 2017). However, the molecular mechanism of EF against liver cancer is currently understudied.

By combining bioinformatics and network analysis, network pharmacology is particularly well suited for the analysis of complex pharmacological mechanisms of multi-compound (Xia and Tang, 2021). The quantitative structure activity relationship (QSAR) method is widely used to investigate the relationship between the physicochemical properties of chemicals and their biological activities in order to obtain a mathematical statistical model for predicting the activity of target chemicals, with differences in structural properties leading to different bioactivities of compounds as the basic principle (Verma et al., 2010). Recently, it has been fashionable to employ QSAR pharmacophore models to investigate the structure-activity connections of TCM and natural compounds in order to uncover their biological activities (Elekofehinti et al., 2021; Li et al., 2022). It is well known that small molecule drugs frequently bind to macromolecular receptors to perform specific biological functions. Molecular docking methods have been widely used in modern drug design to investigate the conformation of ligands within the



binding sites of macromolecular target proteins and to predict their binding mode and binding capacity (Ferreira et al., 2015).

Nowadays, the molecular docking method is an important technique in the field of computer-assisted drug research, and it has become an increasingly important tool for drug discovery. Molecular docking creates drugs based on receptor characteristics and the mode of interaction between receptors and drugs. It can simulate the interaction of molecules and proteins at the atomic level to elucidate the binding sites and binding characteristics of small molecules on target proteins, as well as the fundamental biochemical processes. Molecular docking can also predict ligand and receptor conformation and calculate parameters such as affinity to evaluate binding. This technique is accurate and low-cost, and it is currently used primarily for drug design and the elucidation of biochemical processes (Meng et al., 2011).

In this study, active compounds and key targets of EF were first screened using the network pharmacology method, followed

by GO and KEGG analysis to investigate the molecular mechanism. The pIC₅₀ of active compounds in EF were successfully predicted by the 2D-QSAR pharmacophore model, and the hiphop method was used to construct the 3D-QSAR pharmacophore model. Finally, molecular docking was used to confirm the binding modes of ligands to proteins. In the graphic abstract, the main concept under study is displayed.

2 Materials and methods

2.1 Network pharmacology

2.1.1 Chemical compounds and targets acquisition

We use the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP,

<http://tcmbspw.com/tcmssp.php>) and the Traditional Chinese Medicine Information Database (TCM-ID, <http://bidd.groupp/TCMID/index.html>) to obtain the chemical compounds in EF. The OB value (Oral bioavailability) and DL value (Drug-likeness) used as a reference for filter chemicals are 30% and .18%, respectively. Their structures were retrieved in SDF format from the PubChem database. Then we utilized the SwissTargetPrediction (<http://www.swisstargetprediction.ch/>) to find compound targets, entered “liver cancer” at GeneCards (<https://www.genecards.org/>) to find genes associated with liver cancer (score > 30), and then we analyzed the intersection targets with the aid of the venny2.1.0 platform to find the appropriate targets of EF against liver cancer.

2.1.2 GO/KEGG analysis

DAVID Bioinformatics Resources (<https://david.ncicrf.gov/home.jsp>) is a well-known bioinformatics resource system for functional annotation and enrichment analysis of gene lists (Sherman et al., 2022). We submitted the intersection targets of EF treatment for liver cancer to the DAVID database, chose *Homo sapiens* as the species, and obtained items from Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). It is used to annotate biological functions and analyze signal pathways of key targets.

2.1.3 C-T-P and PPI network construction

The “Compounds-Targets-Pathways” (C-T-P) network is frequently used to analyze the interactions between compounds, targets, and pathways (Chen et al., 2021). Furthermore, the analysis of Protein-protein Interaction (PPI) networks can contribute to a better understanding of disease molecular mechanisms by systematically analyzing and discovering important targets (O'Reilly et al., 2019). The PPI network was obtained using the STRING database (<https://string-db.org/>). The C-T-P network was built using the Cytoscape 3.9.1 software, and we ranked compounds by degree value to determine their significance. The key targets in the PPI network were examined using Cytoscape 3.9.1's cytohubba plugin analysis.

2.2 Construction of the QSAR model

We obtained 45 SRC inhibitors based on the literature and the Selleck website (<https://www.selleck.cn/>) (Martin et al., 2006; Bain et al., 2007; Schenone et al., 2007; Hiscox and Nicholson, 2008; Fathi et al., 2019; Ma et al., 2020). The 3D structural formulas were downloaded from PubChem, and energy was minimized through batch processing using Discovery Studio Software (Discovery Studio 2019; BIOVIA; San Diego, United States). Using the “Creat Training and Test Data” function module, all SRC inhibitor compounds were randomly divided into a training set (35 compounds) and a test set (ten

compounds) (Supplementary Figures S7, S8). The resulting activity values of the training and test set compounds ranged across four orders of magnitude, ensuring the model's accuracy. Quantitative structure-activity relationships (QSAR) are a powerful computational method for analyzing data based on chemical structure. The QSAR pharmacophore model was created by establishing a statistical mathematical link between calculated chemical descriptors of molecular structure and experimentally measured values of these molecules' biological activity, which can be used to predict biological activity with a variety of target chemical products (Muratov et al., 2020).

2.2.1 Construction of the 2D-QSAR pharmacophore

In this study, 2D-QSAR models were built by calculating the molecular properties of the training and test sets using Discovery Studio software. The forward selection by partial least squares (PLS) method and the stepwise multiple linear regression (MLR) method validated by the external test set prediction method were used to build the 2D-QSAR model (Hajalsiddig et al., 2020). Furthermore, the genetic function approximation method (GFA) is an intelligent regression algorithm that simulates biological evolution and natural selection in nature. Because GFA provided a better fit to the training set, it was also commonly used for QSAR modeling (Wang et al., 2012). The pIC_{50} value is typically used to describe the biological activity of a substance. The “calculate molecular properties” module was used to calculate a number of molecular descriptors included AlogP, molecular weight, the total number of bonds, the minimum energy of conformation (kcal/mol), volume, surface area, and other properties (Dwivedi et al., 2011; Imran et al., 2015). Some statistical parameters, such as the coefficient of determination (r^2), adjusted r^2 (r^2 adj), and prediction (PRESS) r^2 (r^2 pred), determine the accuracy of our constructed model.

2.2.2 Construction of the 3D-QSAR pharmacophore

Similarly, we built 3D-QSAR models with Discovery Studio software, typically using the HipHop method. The hip-hop method uses the three-dimensional structures of a series of known target inhibitors/activators to describe the common features of biological activities, develop pharmacophore models, and finally generate the best quantitative pharmacophore model for 3D querying (Wang et al., 2008). First, we screened 10 compounds with higher EF activities and imported their structures' SDF format (3D) into Discovery Studio software. The principal value and maximum feature value of each compound are then defined. Compounds were designated as active when their IC_{50} was less than 1 μ M, the Principal value was set to 2, and the Maxomitfeat value was set to 0. When their IC_{50} was greater than 1 μ M, both the Principal and Maxomitfeat values were set to 1 (Rampogu et al., 2018).

Feature mapping is used to identify compound feature elements and investigate molecules that contain those main feature elements. Furthermore, we performed pharmacophore feature element selection to select those included in the HypoGen module, which generally included five types of feature elements: hydrophobic, hydrogen bond donor (Donor), hydrogen bond acceptor (Acceptor), positively charged ion center (Ionizable Positive), and aromatic ring center (Ring Aromatic). Next, the Maximum Conformation was set to 255 and the Energy Threshold to 10 (Jiang et al., 2016). Finally, the compounds from the test set were used to validate the pharmacophore model.

The 3D structure of a set of ligands was used to calculate the potential energy in the discovery Studio software's 3D-QSAR method, and the potential energy was then used as a descriptor to build the model. Such a model correlates the molecular field and activity and links the three-dimensional structure and biological activity (Ahmed et al., 2017). The equation is as follows:

$$\text{Activity (Predicted)} = \sum_{i=1}^{NEP} CEP(i)VEP(i) + \sum_{i=1}^{NVDW} CVDW(i)VVDW(i)$$

where, N_{EP} represents the number of descriptors of electrostatic potential (EP), $CEP(i)$ is the model coefficient of EP descriptor, and $VEP(i)$ is the electrostatic potential value on the grid points. Furthermore, N_{VDW} is the number of descriptors of Van Der Waals (VDW) interactions, $CVDW(i)$ is the model coefficient of VDW descriptor I , and $VVDW$ is the VDW interaction potential energy on the grid points. In summary, the predicted activity of a compound is the linear sum of the model coefficients multiplied by their corresponding grid values.

2.3 Molecular docking

Molecular docking is currently a key method in drug design since it enables research of ligand binding modalities, stable ligand receptor complex intermolecular interactions, and binding energy predictions (Huang and Zou, 2010). The top ten active compounds in the pharmacophore model were chosen as ligand molecules, and the key target SRC was chosen as the target protein for AutoDock molecular docking tests. We obtained SRC crystal structure 1BYG from the PDB database. To improve reliability, we chose reference drugs such as Ponatinib and Dasatinib, which are SRC inhibitors that are already in clinical use (Ceppi et al., 2012; Roskoski, 2015). Furthermore, STAUROSPORINE (STU), the original SRC ligand, is used as a control (1BYG).

The SRC protein structure (1BYG) was obtained from the Protein Data Bank (<https://www.rcsb.org/>). First, the proteins were processed with PyMOL software, which removed water molecules and extracted ligands while recording the docking site

coordinates (Getbox plugin) of the chosen ligands. The small molecules were then preprocessed with AutoDockTools1.5.6, and the proteins were hydrogenated and charged. After that, one profile (including protein and ligand structure files as well as docking site information) was created for AutoDock Vina docking. Finally, information about the docking results was obtained from the PLIP website (<https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index>), and the docking results were visualized using PyMOL software.

2.4 Molecular dynamics simulation

GROMACS is used for molecular dynamics (MD) simulations of protein-ligand complexes. This experiment included compounds with high docking scores that formed complexes with proteins such as 1BYG-Obacunone, 1BYG-Beta-sitosterol, and 1BYG-Sitosterol. Meanwhile, 1BYG-Ponatinib was used as a control. We used the Acyppe portal and the AMBER force field to generate topology files for the ligands, with SPC as the water model. The relevant scripts were run to add water models, ions, and balance charges. The systems were stabilized by running energy minimization (100 ps) with the steepest descent algorithm. After that, we used the Verlet algorithm for NVT equilibration for 100 ps (2fs steps) and the Berenson algorithm for NPT equilibration for 100 ps (2fs steps) (Adelusi et al., 2021). To ensure that atoms did not jump out of the PBC, the "trjconv" module was used for system processing (periodic boundary conditions). The stability was assessed using root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and hydrogen bond analysis.

3 Results and discussion

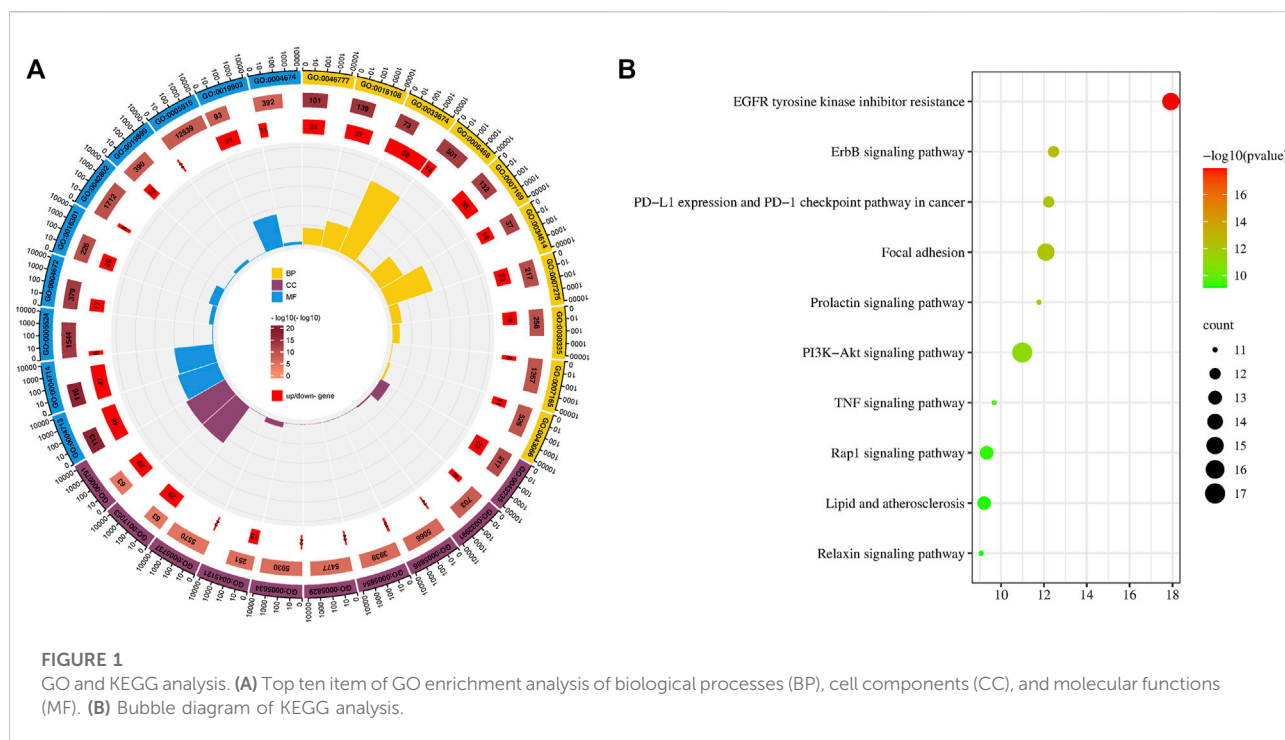
3.1 Network pharmacology analysis

3.1.1 Chemical compounds and targets acquisition

With OB 30% and DL .18 as thresholds, 27 compounds from EF (Supplementary Table S1) were screened from the TCMSP and TCM-ID databases. SwissTargetPrediction identified a total of 611 compound targets. Genecards identified 333 liver cancer targets, and Venn analysis revealed 45 intersecting targets that could be used by EF to play pharmacological roles in the anti-liver cancer process (Supplementary Figure S1; Supplementary Table S2).

3.1.2 GO/KEGG analysis results

We screened the top ten items in the GO and KEGG enrichment analyses in ascending order of p -value (Figure 1). The results of the Go analysis indicate that the biological



processes (BP) mainly involve the development of multicellular organisms, cellular response to reactive oxygen species, protein autophosphorylation, peptidyl-tyrosine phosphorylation, positive regulation of kinase activity, protein phosphorylation, and transmembrane receptor protein tyrosine kinase signaling pathway. Cell Components (CC) include the receptor complex, the macromolecular complex, the plasma membrane, the nucleoplasm, the cytosol, the nucleus, the membrane raft, the cytoplasm, and so on. Protein tyrosine kinase activity, transmembrane receptor protein tyrosine kinase activity, ATP binding, protein kinase activity, kinase activity, identical protein binding, enzyme binding, protein binding, and so on are examples of Molecular Functions (MF) (Supplementary Table S3).

A total of 138 pathways with significant meaning ($P < 0.05$) were enriched by KEGG analysis. The top ten key pathways include EGFR tyrosine kinase inhibitor resistance, ErbB signaling pathway, PD-L1 expression and PD-1 checkpoint pathway in cancer, Focal adhesion, Prolactin signaling pathway, PI3K-Akt signaling pathway, TNF signaling pathway, Rap1 signaling pathway, Lipid and atherosclerosis, and others. The EGFR and ErbB signaling pathways are the highest ranked in KEGG and are upstream in other pathways. EGFR and ErbB are members of a family of cell membrane protein receptors that can receive stimuli and send signals downstream, triggering a series of regulatory processes for both life activities and diseases. EGFR inhibition has been shown in studies to inhibit HCC cell survival, migration, and invasion (Chen et al., 2019; Jin et al., 2021). The

ErBB-PI3K-AKT pathway can promote the growth and spread of hepatocellular carcinoma (Ni et al., 2020). Overexpression of focal adhesion kinase (FAK) occurs frequently in human HCC tissues, and simultaneous overexpression of FAK increases AR expression, which leads to HCC formation in mice (Shang et al., 2019). Certainly, there is also evidence that the Prolactin signaling pathway, the PI3K-Akt signaling pathway, the TNF signaling pathway, and other pathways play important roles in the process of liver cancer (Dumaual et al., 2012; Meng et al., 2021; Miethe et al., 2021).

3.1.3 Network analysis

We can further investigate the relationship between EF compounds, targets, and pathways by building the C-T-P network with Cytoscape 3.9.1. The C-T-P network, as shown in Figure 2A, demonstrates interactions between compounds-targets-pathways. This network contains 314 edges that represent interactions between nodes. The network also included 88 nodes: 27 shared by red square nodes representing compounds, 51 shared by blue diamond nodes representing targets, and 10 shared by green triangle nodes representing signaling pathways. The higher the degree value of the node, the more critical this node is. This network reflects the characteristics of multi-compound and multi-target interactions in TCM. Furthermore, it demonstrated that these chemicals have a better potential for bioactivity against liver cancer. In the C-T-P network, higher degree values for compounds such as gravacridoneshlorine, gossypetin, quercetin, isorhamnetin,



their degree values (Supplementary Table S4). The results showed that SRC, which was thought to be the crucial target in this investigation, had the greatest degree value.

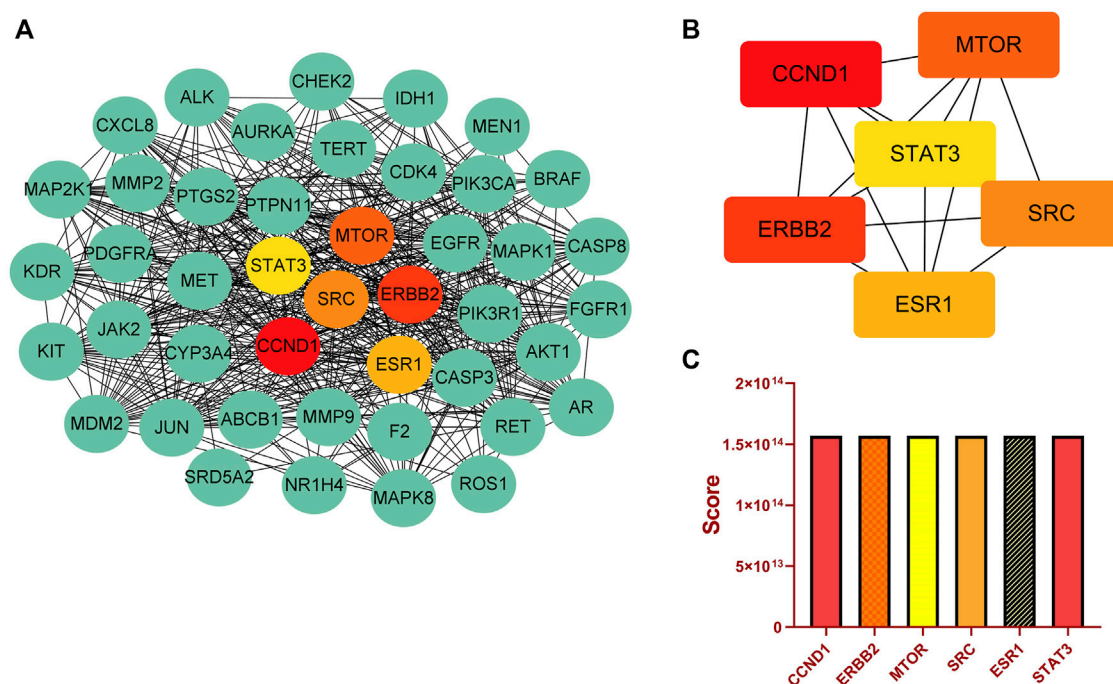


FIGURE 3

Key targets analysis. (A) PPI network. (B) Top six key targets derived from cytoHubba's MCC algorithm. (C) Scores for the top six key targets.

TABLE 1 Based on the 2D-QSAR test set compound experimental and predicted activity pIC₅₀ (μM).

Training no.	Experiment pIC ₅₀	Predicted GFA	Residual ^a	Predicted MLR	Residual ^b	Predicted PLS	Residual ^c
1	1.046	-.676	1.722	.603	.443	3.213	2.167
2	2.699	2.858	.159	1.958	.741	2.661	.038
3	-1.758	-1.285	.474	-1.329	.429	-.898	.861
4	2.301	1.271	1.030	2.049	.252	2.399	.098
5	1.695	.968	.727	1.498	.197	1.082	.613
6	3.000	2.858	.142	3.255	.255	2.187	.813
7	-.146	.662	.808	-.427	.281	.488	.634
8	2.046	1.783	.263	2.543	.497	2.148	.102
9	3.301	4.176	.875	3.365	.064	2.226	1.075
10	.208	.968	.760	.695	.488	.899	.691
r ²		.9117		.988		.800	
r ² (adj)		.8928		.981		.787	
r ² (pred)		.8564		.975			

^aGFA-Experiment pIC₅₀.

^bMLR-Experiment pIC₅₀.

^cPLS-Experiment pIC₅₀.

Moreover, we identified important target proteins in the PPI network using the Cytoscape cytohubba plugin. There are eleven computational methods available to Cytohubba, which

can examine the key nodes in bioinformatic networks. At the moment, MCC is regarded as being the best method (Chin et al., 2014). In addition, The PPI network showed high degree

TABLE 2 The activity pIC₅₀ (μM) of PD Based on the 2D-QSAR model.

Pubchem CID	Compound	Predicted (MLR) pIC ₅₀
5319810	1-methyl-2-[(Z)-undec-6-enyl]-4-quinolone	−1.059
13967189	1-methyl-2-nonyl-4-quinolone	−.674
5319811	1-methyl-2-undecyl-4-quinolone	−.880
162926950	24-methyl-31-norlanost-9(11)-enol	1.733
5281642	6-OH-Luteolin	−1.040
12457	Berberine	.318
222284	beta-Sitosterol	1.374
624052	Ethanone	.318
5317303	Evocarpine	−1.266
189454	Evodiamide	.023
442088	Evodiamine	−.056
58757248	Fordimine	.725
5317827	Goshuyamide_I	.142
5317828	GoshuyamideII	−.265
5280647	gossypetin	−1.035
163102888	Gravacridoneshlerine	.151
56967381	hydroxyevodiamine	.005
5281654	Isorhamnetin	−.944
5319506	N-(2-Methylaminobenzoyl)tryptamine	−.393
119041	Obacunone	1.491

values for CCND1, ERBB2, MTOR, SRC, ESR1, and STAT3 (Figures 3A, B). Histograms depict the score values of these targets (Figure 3C; Supplementary Table S5). We discovered that SRC not only plays an important role in the C-T-P network, but also ranks high in the PPI network. SRC has a high score in comparison to the other core proteins in the PPI network, and there are close interactions. Using KEGG analysis, we discovered that SRC is involved in a number of significant signaling pathways. EGFR/ErbB-2 can activate downstream SRC proteins, and SRC can then encourage liver cancer cell growth and metastasis through focal adhesion and the PI3K-Akt signaling pathway (Ren and Schaefer, 2002; You et al., 2018; Mo et al., 2020; Luo et al., 2021). Meanwhile, a large body of literature has reported on the role of SRC in liver cancer and has shown that SRC could be a therapeutic target in liver cancer (Tong et al., 2015; Zhao et al., 2020; Yang et al., 2021). Additionally, Liao et al. (2014) demonstrated through experiments that EF extract might reduce SRC expression to reduce hepatotoxicity. Therefore, we believe SRC was the most important key target in this study.

3.2 2D-QSAR model analysis

3.2.1 Construction of 2D-QSAR pharmacophore model

To build the 2D-QSAR model, we used the GFA, MLR, and PLS linear fitting algorithms in the Discovery Studio software. The linear regression equations (Supplementary Figures S3–S5) revealed that the correlation regression coefficients r^2 of the three algorithms (GFA, MLR, and PLS) were .912, .988, and .800, respectively, demonstrating their good predictive ability and prediction accuracy for pharmacophore models, particularly MLR. Then, using these three techniques, we created 2D-QSAR models to forecast the biological activities (pIC₅₀) of the compounds in the test set. The results showed that (Table 1) the predicted pIC₅₀ values of the compounds based on GFA and PLS algorithms showed relatively large deviations from the experimental pIC₅₀ values, indicating lower accuracy.

3.2.2 Prediction of EF activity from 2D-QSAR models

The activity of the EF compounds screened was predicted by our 2D-QSAR model using the MLR method (Table 2). The top highly

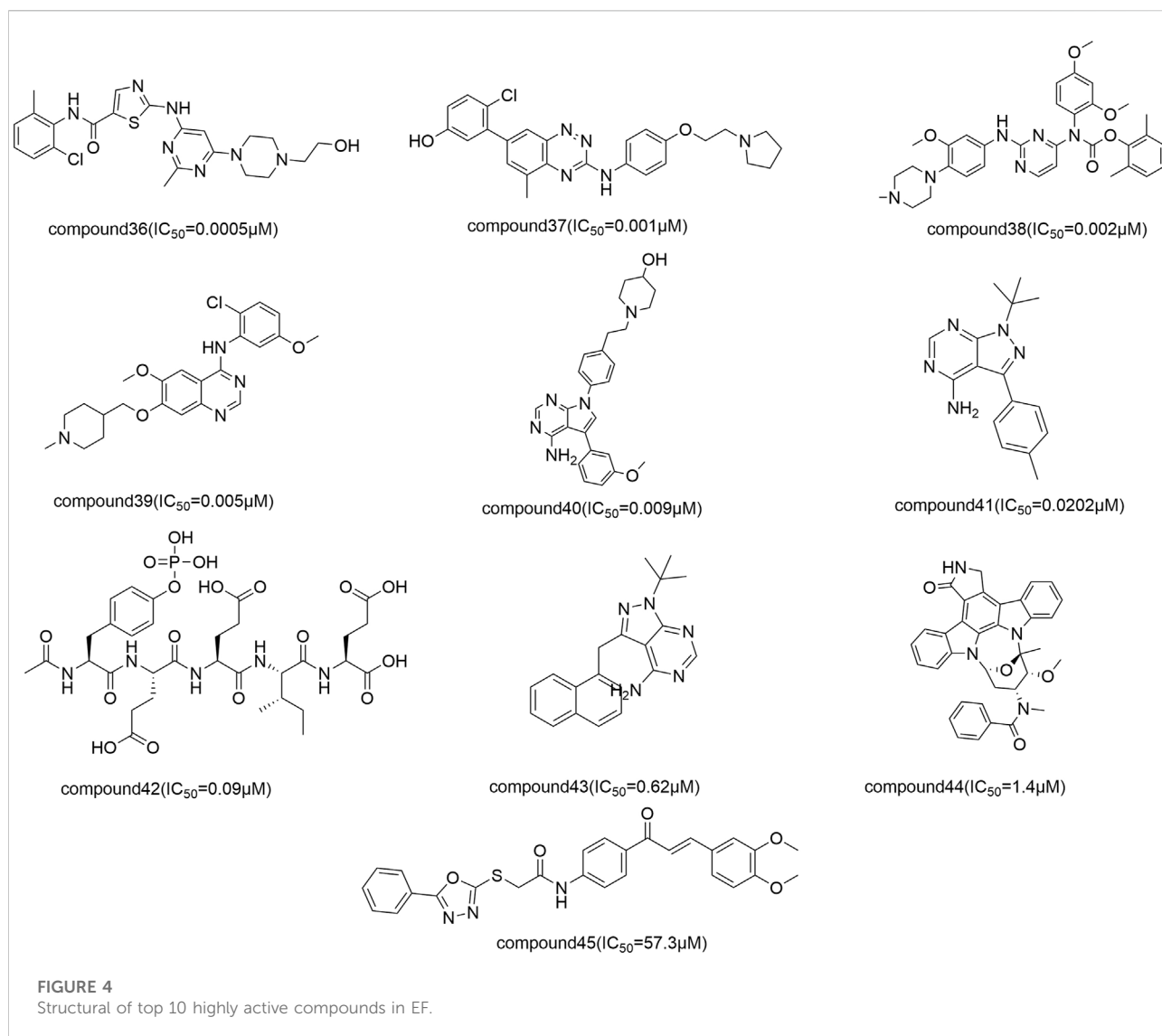
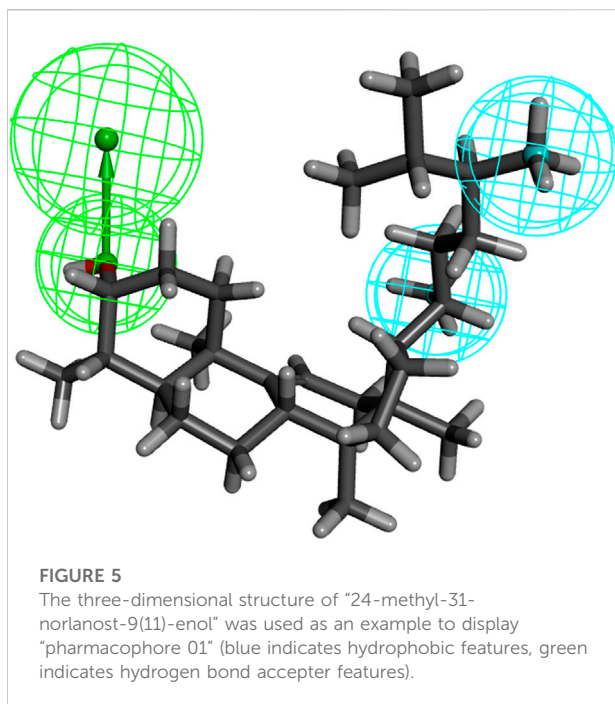


TABLE 3 Common characteristic parameters of active compounds displayed by pharmacophore models.

Pharmacophore	Feature	Rank	Direct hit	Partial hit	Max hit
01	HHA	57.220	1111111111	0000000000	3
02	HHA	57.042	1111111111	0000000000	3
03	HHA	55.151	1111111111	0000000000	3
04	HHA	52.597	1111111111	0000000000	3
05	HHA	49.170	1111111111	0000000000	3
06	HHA	48.528	1111111111	0000000000	3
07	HHA	42.982	1111111111	0000000000	3
08	HHA	41.428	1111111111	0000000000	3
09	HHA	32.624	1111111111	0000000000	3

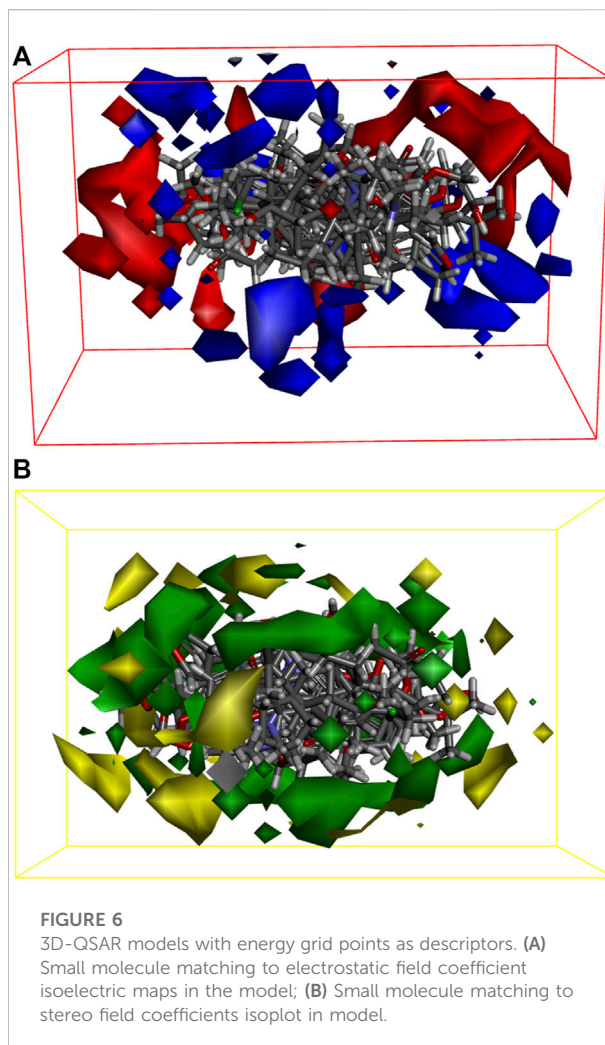


active compounds were 24-methyl-31-norlanost-9(11)-enol, Obacunone, Beta-Sitosterol, Berberine, Sitosterol, Rutalinidine, Fordimine, Ethanone, Gravacridoneshlinine, and Goshuyamide I, with pIC_{50} values of 1.733, 1.491, 1.374, 1.374, .725, .470, .318 (Figure 4). Several studies have confirmed the therapeutic effects of some of these compounds in the treatment of liver cancer. Obacunone in mandarin (*Citrus reticulata* Blanco) has been shown in studies to inhibit a variety of human cancer cell lines, including leukemia (HL-60), ovarian cancer (SKOV-3), cervical cancer (Hela), gastric cancer (NCI-SNU-1), liver cancer (HepG2), and breast cancer (MCF-7). Obacunone had an IC_{50} of $65.13 \pm 5.39 \mu\text{m}$ against HepG2 cells in an MTT assay (Tian et al., 2001). Mary (Ditty and Ezhilarasan, 2021) found that β -Sitosterol can increase cellular ROS levels, causing cell membrane damage and mitochondrial toxicity, as well as promoting HepG2 cell apoptosis. Berberine is an alkaloid that has been shown to inhibit the growth of various cancers. Ren et al. (2022) discovered that berberine inhibited hepatocarcinogenesis in mice by antagonizing the ATX-LPA-LPAR2-P38-leptin axis. Liu and Bai (2020) discovered that Sitosterol has effect on liver cancer with the help of analysis of network pharmacology. Ultimately, we screened out the top 10 compounds based on pIC_{50} , which may have better anti-liver cancer activity.

3.3 3D-QSAR model analysis

3.3.1 Construction of the 3D-QSAR model by the HipHop method

We investigated the common chemical properties of the compounds and built a 3D-QSAR model by superimposing



the three-dimensional molecular structure. The HipHop method of the Discovery Studio software was used to generate nine 3D-QSAR pharmacophore models. Table 3 displays the pharmacophore model's matching parameters with the ten active compounds. According to the results, the 01 pharmacophore model had the highest rank score value of 57.220. The “HHA” shown in its “Feature” item indicates that it has two hydrophobic features and one hydrogen bond acceptor feature. The pharmacophore features in this model were directly matched to ten small molecules, and the Partial Hit revealed no partial matches to the pharmacophore for the ten small molecules. Furthermore, Max Hit indicated that three pharmacodynamic features could all be matched. The spatial distribution of Pharmacophore 01 is depicted in Figure 5. (Supplementary Figure S6 showed the spatial distribution of pharmacophore 02–09). As a result, we chose Pharmacophore 01 for further analysis and validation as the best constructed in this study.

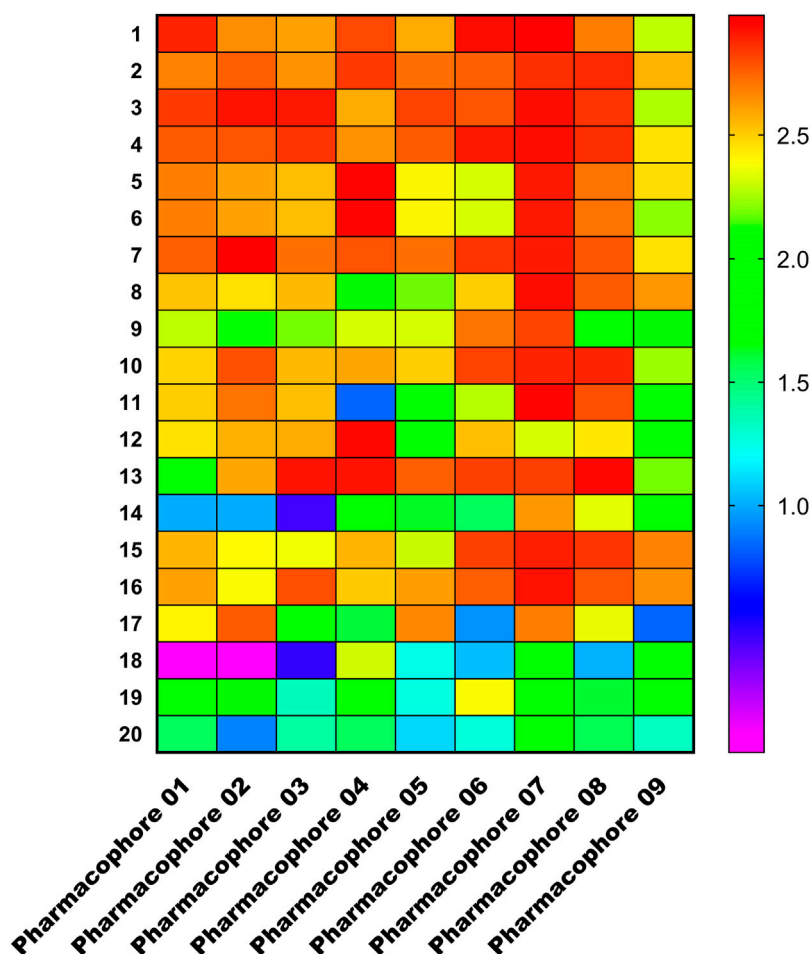


FIGURE 7
Heat map of the training set compounds predicted by the nine pharmacophore models.

3.3.2 Construction of the 3D-QSAR model based on steric and electrostatic fields of small molecules

We built 3D-QSAR models based on the steric and electrostatic fields of the small molecules to investigate the non-covalent interactions in the structure of EF active compounds. Our 3D-QSAR model incorporates and visualizes the common important structural characteristics of a series of active compounds. The contour plots of the electrostatic field coefficients of the small molecules that match the 3D-QSAR models are shown in Figure 6A. In this system, the more negatively charged the substituents in the red region and the more positively charged the substituents in the blue region, the higher the activity of the compound. The contour plots of the steric field coefficients of small molecules that match the 3D-QSAR model are shown in Figure 6B. In this system, the yellow region indicates that increasing the volume of substituents is

detrimental to compound activity, whereas the blue region indicates that increasing the volume of substituents is beneficial to compound activity. The results showed that the electrostatic effects of the substituents, as well as the spatial distribution of the functional groups of the active compounds, affected their biological activities and may have contributed to the binding of key target proteins for liver cancer. Based on the force field information, this part of the study can help us screen and optimize the active compounds in EF.

3.3.3 Verification of 3D-QSAR pharmacophore model

We chose the top 20 training set compounds based on activity to verify that pharmacophores can correctly distinguish between active and inactive molecules. Figure 7 depicts a matching heatmap of the pharmacophore with the training set compounds, where the ordinate represents the

TABLE 4 Matching results of nine pharmacophore models for compounds of the training set.

Compound no.	01	02	03	04	05	06	07	08	09
1	2.886	2.649	2.604	2.805	2.585	2.934	2.965	2.682	2.289
2	2.677	2.754	2.638	2.838	2.718	2.757	2.858	2.874	2.559
3	2.834	2.924	2.912	2.583	2.818	2.777	2.931	2.851	2.261
4	2.767	2.782	2.846	2.640	2.764	2.911	2.933	2.866	2.450
5	2.681	2.598	2.535	2.956	2.399	2.325	2.905	2.708	2.462
6	2.681	2.598	2.535	2.956	2.399	2.325	2.905	2.708	2.219
7	2.756	2.983	2.717	2.787	2.719	2.850	2.903	2.773	2.453
8	2.526	2.457	2.543	1.880	2.183	2.498	2.933	2.767	2.632
9	2.284	2.126	2.194	2.321	2.326	2.706	2.810	2.037	1.817
10	2.482	2.788	2.544	2.594	2.499	2.818	2.889	2.881	2.242
11	2.502	2.711	2.530	.836	2.035	2.271	2.949	2.790	1.660
12	2.444	2.568	2.582	2.942	1.874	2.529	2.328	2.435	1.996
13	2.034	2.593	2.919	2.919	2.755	2.824	2.827	2.938	2.192
14	1.000	1.000	.455	1.978	1.629	1.539	2.625	2.345	1.933
15	2.552	2.391	2.364	2.557	2.303	2.822	2.899	2.845	2.671
16	2.601	2.387	2.797	2.504	2.610	2.761	2.923	2.781	2.644
17	2.404	2.766	1.959	1.600	2.662	.935	2.684	2.358	.839
18	.000	.000	.485	2.307	1.239	1.046	1.929	1.008	2.048
19	1.806	1.810	1.338	1.971	1.254	2.380	1.868	1.605	2.003
20	1.533	.903	1.395	1.541	1.100	1.271	1.932	1.552	1.333

TABLE 5 The affinity of compounds with SRC(1BYG) (kcal.mol⁻¹).

Active compound	Protein(PDBID)	Affinity
24-methyl-31-norlanost-9(11)-enol	SRC(1BYG)	−8.6
Obacunone	SRC(1BYG)	−9.4
beta-Sitosterol	SRC(1BYG)	−9.0
Sitosterol	SRC(1BYG)	−9.3
Fordimine	SRC(1BYG)	−7.8
Rutalinidine	SRC(1BYG)	−8.3
Berberine	SRC(1BYG)	−9.1
Ethanone	SRC(1BYG)	−6.6
Gravacridoneshlorine	SRC(1BYG)	−8.9
Goshuyamide_I	SRC(1BYG)	−8.7
Ponatinib	SRC(1BYG)	−9.8
Dasatinib	SRC(1BYG)	−8.7
STU	SRC(1BYG)	−10.9

20 training ensemble scores ranked in order of activity, the abscissa represents the pharmacophore 01–09, and the Fitvalue ranges from low to high, as indicated by the gradual purple to red color change. The darker the red color of each rectangular block in the figure, the greater the Fitvalue, thus the greater the corresponding compound activity. In theory, the active compounds in the training set should be red and orange, while the inactive compounds should be blue and purple, indicating that this pharmacophore can distinguish between active and inactive compounds. We discovered that the Fitvalue of the more active compound in pharmacophore 01 was higher than that of the less active compound, and that its fitvalue had a certain trend, indicating that pharmacophore 01 had ability to discriminate between active and inactive compounds. Pharmacophore 01 had the highest rank value.

Additionally, we used receiver operating characteristic (ROC) curve analysis to validate the pharmacophore (Al-Sha'er et al., 2022; Taha et al., 2014). ROC plots are a quantitative measure of whether a test can distinguish between two populations (typically active/inactive

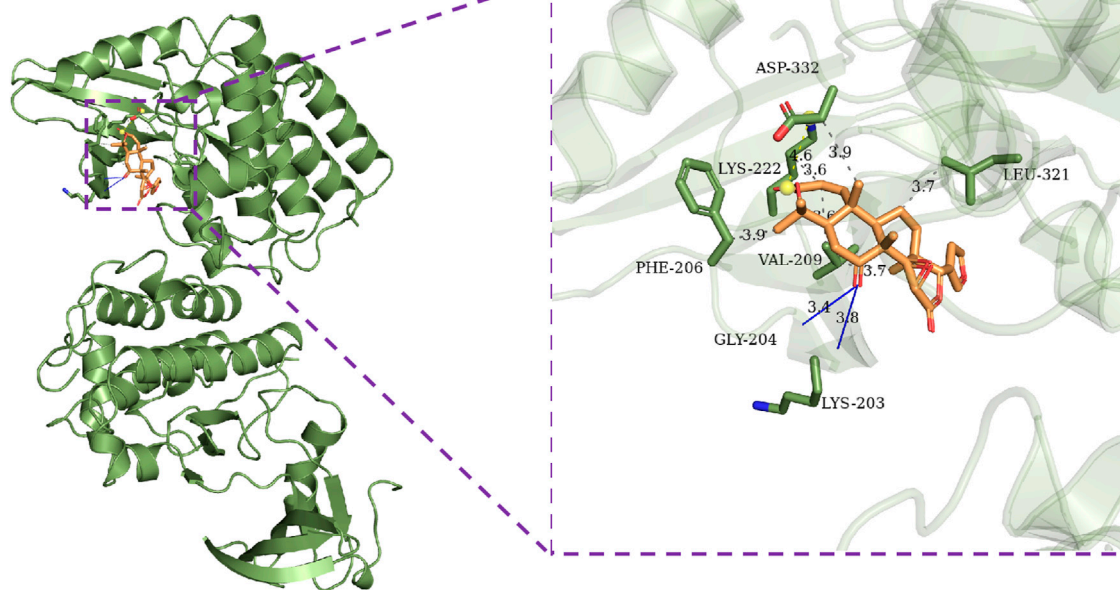


FIGURE 8

Molecular docking results of Obacunone with SRC (1BYG). Gray dashed lines represent hydrophobic interactions, blue lines represent hydrogen bonding interactions, and yellow dashed lines represent salt bridges.

compounds). It can be compared to other data sets. The area under the curve was used to evaluate test accuracy in ROC analysis (AUC). A test set of active and inactive compounds was used in our 3D-QSAR model to validate the pharmacophore model's selectivity. On the one hand, sensitivity is defined as the model's ability to identify positives, namely the extent to which an active molecule is distinguished from the inactive, or the proportion of the predicted active that is actually active:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

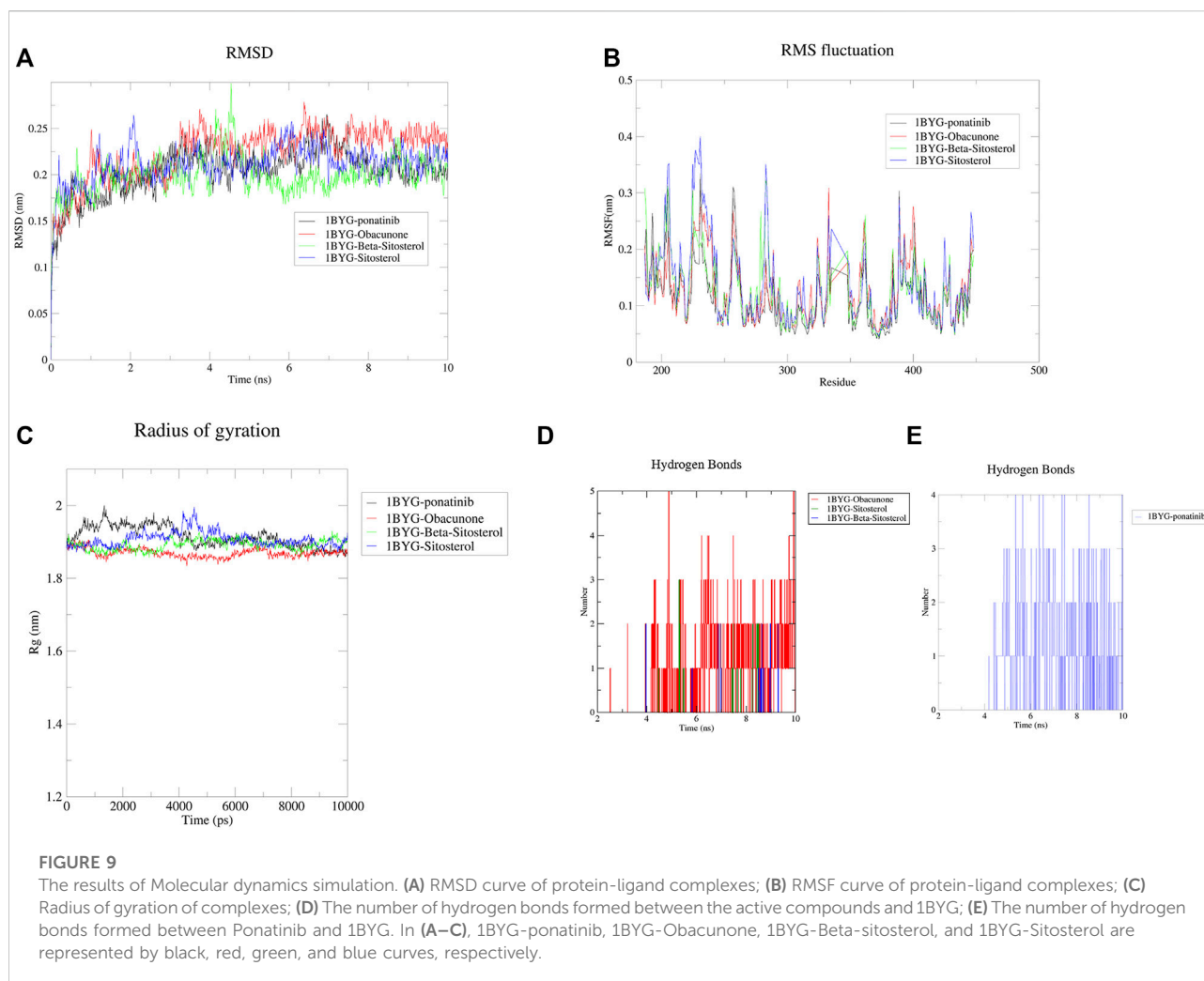
Specificity, on the other hand, is the ability of the model to determine negatives, or its discriminatory power for inactive compounds:

$$\text{Specific TN} / (\text{TN} + \text{FP})$$

ROC analysis results show that (Supplementary Figure S9; Supplementary Table S6), the ROC score of the pharmacophore 01 model is .768, and the sensitivity and specificity scores are both high. We believe pharmacophore 01 is the most effective pharmacophore hypothesis for distinguishing between active and inactive compounds. Therefore, we considered pharmacophore 01 was the best 3D-QSAR pharmacophore model. This means that pharmacophore model 01 may have anti-cancer effects on EF. The Fitvalue of the other pharmacophore models is shown in Table 4.

3.4 Molecular docking verification

This section of the study used molecular docking to validate the rationality of EF's 3D-QSAR pharmacophore model built from ten highly active compounds. The optimal conformation was chosen for each compound, and the RMSD of the conformations was less than 2. As shown in Table 5, the docking scores of the ten active compounds with SRC were all less than -7 kcal.mol^{-1} , with the exception of Ethanone. This implies that they play an important role in EF's anti-cancer process. Supplementary Figure S9 depicts their docking conditions. These compounds all bind to the amino acid residues of SRC *via* hydrophobic effects and hydrogen bonding interactions and have a similar docking mode. More specifically, Obacunone had the highest docking score with SRC, its antihepatoma activity *in vitro* has previously been reported, and its three-dimensional molecular docking state is shown in Figure 8. Obacunone primarily binds to SRC *via* hydrophobic and hydrogen bonding interactions. Obacunone could form hydrophobic interactions with amino acid residues PHE-206, VAL-209, THR-266, and PHE-333, with hydrophobic bond distances of 3.44, 2.61, 3.54, and 3.55 Å, respectively. Meanwhile, obacunone could form hydrogen bonding interactions with GLU-205 and PHE-206 at distances of 2.84 and 3.31 Å, respectively. Obacunone also has one salt bridge interaction with LYS-222. In addition, in order to improve the reliability of molecular docking, some reference drugs were used for docking, such as Ponatinib, Dasatinib, STU (Table 5). Their docking scores were -9.8 , -8.7 , and $-10.9 \text{ kcal.mol}^{-1}$,



respectively. Some of the screened compounds have affinity values similar to the reference drugs, such as Obacunone, Beta-Sitosterol, Sitosterol, and Berberine. In general, these reference drugs interact with SRC *via* hydrophobic and hydrogen bond interactions. Hydrophobic interactions are formed by amino acid residues ILE-201, VAL-209, TYR-268, and LEU-321, and hydrogen bonds are formed by THR-266 and MET-269. Meanwhile, Ponatinib has some salt bridge interactions (Supplementary Figure S10). Overall, the screened compounds have similar docking modes to the reference drugs, and these docking interactions are consistent with the common features of pharmacophore 01.

3.5 MD simulations

After a 10 ns molecular dynamics simulation, we could further examine how the screened compounds interacted with the protein SRC (1BYG). Obacunone, Beta-sitosterol, and sitosterol were chosen for MD simulation analysis in this work because they

had superior docking scores (less than -9 kcal.mol^{-1}) than other compounds. Figure 9A shows that throughout the entire procedure, the RMSD of 1BYG Obacunone, 1BYG Beta sitosterol, and 1BYG Sitosterol maintained between .1 and .3, and the fluctuation became stable after 10 ns. It indicates that the 10 ns trajectory conformation of compounds does not reveal substantial structural differences from ponatinib, implying that the ligand and complex structures are stable.

The root mean square fluctuation (RMSF) curve was used to investigate the local changes in protein chain residues. In the 10 ns trajectory files, the fluctuation profiles of the amino acid residues of the complexes (1BYG-Obacunone, 1BYG-Beta-sitosterol, and 1BYG-Sitosterol) were visualized in Figure 9B. RMSF value is lower than 0.4 nm, and a stable fluctuation is generated at about .15, which provides an appropriate basis for subsequent research. 1BYG-obacunone fluctuated surprisingly little in these compounds until about the 300th residue, when it peaked at .3 nm. 1BYG-Beta-sitosterol and 1BYG-Sitosterol have peaks at .32 and .4 nm, respectively with substantial variations in

position before residue 300. Their fluctuation trajectories followed the same pattern as 1BYG-Ponatinib.

The radius of gyration (Rg) is a measure of a complex system's stability in terms of the structural compactness of the molecular dynamics trajectory (Shahbaaz et al., 2019). Rg can also be used to confirm the complex's stable folding during the simulation. If Rg values are relatively consistent throughout the simulation, the protein structure is considered stable (Ghasemi et al., 2016). A higher Rg value indicates that the protein is more labile, while a lower value indicates that the protein is more stable. In Figure 9C, Rg values of 1BYG-Obacunone were lower and more stable than those of others, this leads in less structural bias and higher stability in the obacunone complexes during simulations.

In addition, we analyzed the number of hydrogen bonds for trajectories lasting 10 ns. Figure 9D shows the hydrogen bonding interaction of obacunone, beta-sitosterol and sitosterol at a distance of 3.5 Å (35 nm). The maximum number of hydrogen bonds found for obacunone, beta-sitosterol, and sitosterol were 5, 2, and 3, respectively. Over the course of 10 ns, beta-sitosterol, and sitosterol formed fewer hydrogen bonds with 1BYG, indicating poor binding stability. Obacunone formed five hydrogen bonds with 1BYG between 5 and 10 ns, and the number of hydrogen bonds formed was the greatest throughout the simulation process, ensuring better stability. Furthermore, obacunone has a similar density of hydrogen bonding conditions to ponatinib (Figure 9E), which explains its superior binding activity.

4 Conclusion

In this study, some potential active compounds such as Obacunone, Beta-sitosterol, Sitosterol, and others were screened using network pharmacology, and SRC was clearly identified as the most promising target of EF in the treatment of liver cancer. The top three signaling pathways are the EGFR signaling pathway, the ErbB signaling pathway, PD-L1 expression, and the PD-1 checkpoint pathway. In addition, the 2-dimensional QSAR pharmacophore model identified 24-methyl-31-norlanost-9(11)-enol, Obacunone, Beta-sitosterol, Sitosterol, Fordimine, Rutalinidine, Berberine, Ethanone, Gravacridoneshlirine, and Goshuyamide I as highly active compounds. The 3D-QSAR model we created revealed that Pharmacophore 01 had two hydrophobic features and one hydrogen bond acceptor feature. Heatmap and ROC analysis results revealed that pharmacophore 01 possessed high selectivity as well as the ability to distinguish between active and inactive compounds. The molecular docking results confirmed the active compounds' stable binding to SRC. Furthermore, MD simulations demonstrated the stability of Obacunone, Beta-sitosterol, and Sitosterol in dynamic systems and highlighted Obacunone's prominent binding capacity. The pharmacophore

model proposed in this study provides theoretical support for further screening of effective Chinese medicine compounds against liver cancer and compound structural optimization. Meanwhile, for the optimal compounds, additional pharmacodynamic and pharmacological studies will be conducted to clarify their mechanism of action for the treatment of liver cancer.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

P-YC: Select topic, research proposal design, data analysis, and paper writing. L-TH: Syntax optimization, review, and editing.

Funding

This work was supported by the Funds for Key Program for Traditional Chinese Medicine of Hubei University of Chinese Medicine (2022ZZXZ004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.1060500/full#supplementary-material>

References

- Adelusi, T. I., Abdul-Hammed, M., Idris, M. O., Oyedele, Q. K., and Adedotun, I. O. (2021). Molecular dynamics, quantum mechanics and docking studies of some Keap1 inhibitors - an insight into the atomistic mechanisms of their antioxidant potential. *Heliyon* 7 (6), e07317. doi:10.1016/j.heliyon.2021.e07317
- Ahmed, N., Anwar, S., and Thet Htar, T. (2017). Docking based 3D-QSAR study of tricyclic guanidine analogues of batzelladine K as anti-malarial agents. *Front. Chem.* 5, 36. doi:10.3389/fchem.2017.00036
- Al-Sha'er, M. A., Basheer, H. A., and Taha, M. O. (2022). Discovery of new PKN2 inhibitory chemotypes via QSAR-guided selection of docking-based pharmacophores. *Mol. Divers.* doi:10.1007/s11030-022-10434-410.1007/s11030-022-10434-4
- Anwanwan, D., Singh, S. K., Singh, S., Saikam, V., and Singh, R. (2020). Challenges in liver cancer and possible treatment approaches. *Biochim. Biophys. Acta Rev. Cancer* 1873 (1), 188314. doi:10.1016/j.bbcan.2019.188314
- Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., McLauchlan, H., et al. (2007). The selectivity of protein kinase inhibitors: A further update. *Biochem. J.* 408 (3), 297–315. doi:10.1042/bj20070797
- Cao, W., Chen, H. D., Yu, Y. W., Li, N., and Chen, W. Q. (2021). Changing profiles of cancer burden worldwide and in China: A secondary analysis of the global cancer statistics 2020. *Chin. Med. J. Engl.* 134 (7), 783–791. doi:10.1097/cm9.0000000000001474
- Ceppi, P., Rapa, I., Lo Iacono, M., Righi, L., Giorcelli, J., Pautasso, M., et al. (2012). Expression and pharmacological inhibition of thymidylate synthase and Src kinase in nonsmall cell lung cancer. *Int. J. Cancer* 130 (8), 1777–1786. doi:10.1002/ijc.26188
- Chen, P. Y., Yuan, C., Hong, Z. C., Zhang, Y., Ke, X. G., Yu, B., et al. (2021). Revealing the mechanism of "Huai Hua San" in the treatment of ulcerative colitis based on network pharmacology and experimental study. *J. Ethnopharmacol.* 281, 114321. doi:10.1016/j.jep.2021.114321
- Chen, Y., Chen, X., Ding, X., and Wang, Y. (2019). Afatinib, an EGFR inhibitor, decreases EMT and tumorigenesis of Huh-7 cells by regulating the ERK-VEGF/MMP9 signaling pathway. *Mol. Med. Rep.* 20 (4), 3317–3325. doi:10.3892/mmr.2019.10562
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8 (4), S11. doi:10.1186/1752-0509-8-s4-s11
- Cho, J. H., Lee, R. H., Jeon, Y. J., Shin, J. C., Park, S. M., Choi, N. J., et al. (2015). Role of transcription factor Sp1 in the 4-O-methylhonokiol-mediated apoptotic effect on oral squamous cancer cells and xenograft. *Int. J. Biochem. Cell Biol.* 64, 287–297. doi:10.1016/j.biocel.2015.05.007
- Ditty, M. J., and Ezhilarasan, D. (2021). β -sitosterol induces reactive oxygen species-mediated apoptosis in human hepatocellular carcinoma cell line. *Avicenna J. Phytomed* 11 (6), 541–550. doi:10.22038/ajp.2021.17746
- Dong, L., Shi, H., Ji, G., and Wu, D. (2010). Effects of Coptis chinensis and Evodia rutaecarpa water extract on DMH-induced precancerous lesion of colon. *Zhongguo Zhong Yao Za Zhi* 35 (9), 1185–1188. doi:10.4268/cjmm.20100923
- Dumaual, C. M., Sandusky, G. E., Soo, H. W., Werner, S. R., Crowell, P. L., and Randall, S. K. (2012). Tissue-specific alterations of PRL-1 and PRL-2 expression in cancer. *Am. J. Transl. Res.* 4 (1), 83–101.
- Dwivedi, N., Mishra, B. N., and Katoh, V. M. (2011). 2D-QSAR model development and analysis on variant groups of anti-tuberculosis drugs. *Bioinformation* 7 (2), 82–90. doi:10.6026/97320630007082
- El-Serag, H. B., Marrero, J. A., Rudolph, L., and Reddy, K. R. (2008). Diagnosis and treatment of hepatocellular carcinoma. *Gastroenterology* 134 (6), 1752–1763. doi:10.1053/j.gastro.2008.02.090
- Elekofehinti, O. O., Iwaloye, O., Molehin, O. R., and Famusiwa, C. D. (2021). Identification of lead compounds from large natural product library targeting 3C-like protease of SARS-CoV-2 using E-pharmacophore modelling, QSAR and molecular dynamics simulation. *Silico Pharmacol.* 9 (1), 49. doi:10.1007/s40203-021-00109-7
- Fathi, M. A. A., Abd El-Hafeez, A. A., Abdelhamid, D., Abbas, S. H., Montano, M. M., and Abdel-Aziz, M. (2019). 1, 3, 4-oxadiazole/chalcone hybrids: Design, synthesis, and inhibition of leukemia cell growth and EGFR, Src, IL-6 and STAT3 activities. *Bioorg Chem.* 84, 150–163. doi:10.1016/j.bioorg.2018.11.032
- Ferreira, L. G., Dos Santos, R. N., Oliva, G., and Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules* 20 (7), 13384–13421. doi:10.3390/molecules200713384
- Ghasemi, F., Zomorodipour, A., Karkhane, A. A., and Khorramizadeh, M. R. (2016). *In silico* designing of hyper-glycosylated analogs for the human coagulation factor IX. *J. Mol. Graph Model* 68, 39–47. doi:10.1016/j.jmgm.2016.05.011
- Hajalsiddiq, T. T. H., Osman, A. B. M., and Saeed, A. E. M. (2020). 2D-QSAR modeling and molecular docking studies on 1H-Pyrazole-1-carbothioamide derivatives as EGFR kinase inhibitors. *ACS Omega* 5 (30), 18662–18674. doi:10.1021/acsomega.0c01323
- He, W. Q., Gao, X., Gao, L., Ma, Y., Sun, D., and Sun, J. (2021). Contrasting trends of primary liver cancer mortality in Chinese mongol and non-mongol. *Asian Pac J. Cancer Prev.* 22 (9), 2757–2763. doi:10.31557/apjcp.2021.22.9.2757
- Hiscox, S., and Nicholson, R. I. (2008). Src inhibitors in breast cancer therapy. *Expert Opin. Ther. Targets* 12 (6), 757–767. doi:10.1517/14728222.12.6.757
- Hu, C. Y., Wu, H. T., Su, Y. C., Lin, C. H., Chang, C. J., and Wu, C. L. (2017). Evodiamine exerts an anti-hepatocellular carcinoma activity through a WWOX-dependent pathway. *Molecules* 22 (7), 1175. doi:10.3390/molecules22071175
- Huang, S. Y., and Zou, X. (2010). Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* 11 (8), 3016–3034. doi:10.3390/ijms11083016
- Imran, S., Taha, M., Ismail, N. H., Kashif, S. M., Rahim, F., Jamil, W., et al. (2015). Synthesis of novel flavone hydrazones: *In-vitro* evaluation of α -glucosidase inhibition, QSAR analysis and docking studies. *Eur. J. Med. Chem.* 105, 156–170. doi:10.1016/j.ejmech.2015.10.017
- Jiang, L., He, Y., Luo, G., Yang, Y., Li, G., and Zhang, Y. (2016). Discovery of potential novel microsomal triglyceride transfer protein inhibitors via virtual screening of pharmacophore modelling and molecular docking. *Mol. Simul.* 42, 1223–1232. doi:10.1080/08927022.2016.1149701
- Jin, Q., Cheng, M., Xia, X., Han, Y., Zhang, J., Cao, P., et al. (2021). Down-regulation of MYH10 driven by chromosome 17p13.1 deletion promotes hepatocellular carcinoma metastasis through activation of the EGFR pathway. *J. Cell Mol. Med.* 25 (24), 11142–11156. doi:10.1111/jcmm.17036
- Kassebaum, N. J., Bertozzi-Villa, A., Coggeshall, M. S., Shackelford, K. A., Steiner, C., Heuton, K. R., et al. (2014). Global, regional, and national levels and causes of maternal mortality during 1990–2013: A systematic analysis for the global burden of disease study 2013. *Lancet* 384 (9947), 980–1004. doi:10.1016/s0140-6736(14)60696-6
- Li, B., Yin, X., Cen, B., Duan, W., Lin, G., Wang, X., et al. (2022). High value-added application of natural forest product α -pinene: Design, synthesis and 3D-QSAR study of novel α -campholenic aldehyde-based 4-methyl-1, 2, 4-triazole-thioether compounds with significant herbicidal activity. *Nat. Prod. Res.* 1–6. doi:10.1080/14786419.2022.211717610.1080/14786419.2022.2117176
- Li, M., and Wang, C. (2020). Traditional uses, phytochemistry, pharmacology, pharmacokinetics and toxicology of the fruit of tetradium ruticarpum: A review. *J. Ethnopharmacol.* 263, 113231. doi:10.1016/j.jep.2020.113231
- Liao, W. Q., Li, B., Li, L., and Zhao, J. N. (2014). Study on molecular mechanism of Euodiae Fructus on liver toxicity in MICE. *Zhongguo Zhong Yao Za Zhi* 39 (24), 1–17.
- Liu, B., and Bai, C. (2020). Regulatory mechanisms of coicis semen on bionetwork of liver cancer based on network pharmacology. *Biomed. Res. Int.* 2020, 1. doi:10.1155/2020/5860704
- Luo, X., Zheng, E., Wei, L., Zeng, H., Qin, H., Zhang, X., et al. (2021). The fatty acid receptor CD36 promotes HCC progression through activating Src/PI3K/AKT axis-dependent aerobic glycolysis. *Cell Death Dis.* 12 (4), 328. doi:10.1038/s41419-021-03596-w
- Ma, X., Qiu, Y., Zhu, L., Zhao, Y., Lin, Y., Ma, D., et al. (2020). NOD1 inhibits proliferation and enhances response to chemotherapy via suppressing SRC-MAPK pathway in hepatocellular carcinoma. *J. Mol. Med. Berl.* 98 (2), 221–232. doi:10.1007/s00109-019-01868-9
- Martin, M. W., Newcomb, J., Nunes, J. J., McGowan, D. C., Armistead, D. M., Boucher, C., et al. (2006). Novel 2-aminopyrimidine carbamates as potent and orally active inhibitors of Ick: Synthesis, SAR, and *in vivo* antiinflammatory activity. *J. Med. Chem.* 49 (16), 4981–4991. doi:10.1021/jm060435i
- Meng, X. Y., Zhang, H. X., Mezei, M., and Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided Drug Des.* 7 (2), 146–157. doi:10.2174/157340911795677602
- Meng, Y., Zhao, Q., An, L., Jiao, S., Li, R., Sang, Y., et al. (2021). A TNFR2-hnRNPK Axis promotes primary liver cancer development via activation of YAP signaling in hepatic progenitor cells. *Cancer Res.* 81 (11), 3036–3050. doi:10.1158/0008-5472.Can-20-3175
- Miethe, C., Torres, L., Zamora, M., and Price, R. S. (2021). Inhibition of PI3K/Akt and ERK signaling decreases visfatin-induced invasion in liver cancer cells. *Horm. Mol. Biol. Clin. Investig.* 42 (4), 357–366. doi:10.1515/hmbci-2021-0011
- Mo, C. F., Li, J., Yang, S. X., Guo, H. J., Liu, Y., Luo, X. Y., et al. (2020). IQGAP1 promotes anoikis resistance and metastasis through Rac1-dependent ROS accumulation and activation of Src/FAK signalling in hepatocellular carcinoma. *Br. J. Cancer* 123 (7), 1154–1163. doi:10.1038/s41416-020-0970-z

- Muratov, E. N., Bajorath, J., Sheridan, R. P., Tetko, I. V., Filimonov, D., Poroikov, V., et al. (2020). QSAR without borders. *Chem. Soc. Rev.* 49 (11), 3525–3564. doi:10.1039/d0cs00098a
- Ni, Q., Chen, Z., Zheng, Q., Xie, D., Li, J. J., Cheng, S., et al. (2020). Epithelial V-like antigen 1 promotes hepatocellular carcinoma growth and metastasis via the ERBB-PI3K-AKT pathway. *Cancer Sci.* 111 (5), 1500–1513. doi:10.1111/cas.14331
- O'Reilly, M., Cleasby, A., Davies, T. G., Hall, R. J., Ludlow, R. F., Murray, C. W., et al. (2019). Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov. Today* 24 (5), 1081–1086. doi:10.1016/j.drudis.2019.03.009
- Park, S. Y., Park, C., Park, S. H., Hong, S. H., Kim, G. Y., Hong, S. H., et al. (2017). Induction of apoptosis by ethanol extract of *Evodia rutaecarpa* in HeLa human cervical cancer cells via activation of AMP-activated protein kinase. *Biosci. Trends* 10 (6), 467–476. doi:10.5582/bst.2016.01170
- Rampogu, S., Son, M., Baek, A., Park, C., Rana, R. M., Zeb, A., et al. (2018). Targeting natural compounds against HER2 kinase domain as potential anticancer drugs applying pharmacophore based molecular modelling approaches. *Comput. Biol. Chem.* 74, 327–338. doi:10.1016/j.compbiolchem.2018.04.002
- Ren, G., Guo, J. H., Feng, C. L., Ding, Y. W., Dong, B., Han, Y. X., et al. (2022). Berberine inhibits carcinogenesis through antagonizing the ATX-LPA-LPAR2-p38-leptin axis in a mouse hepatoma model. *Mol. Ther. Oncolytics* 26, 372–386. doi:10.1016/j.omto.2022.08.001
- Ren, Z., and Schaefer, T. S. (2002). ErbB-2 activates Stat3 alpha in a Src- and JAK2-dependent manner. *J. Biol. Chem.* 277 (41), 38486–38493. doi:10.1074/jbc.M112438200
- Roskoski, R., Jr. (2015). Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacol. Res.* 94, 9–25. doi:10.1016/j.phrs.2015.01.003
- Schenone, S., Manetti, F., and Botta, M. (2007). SRC inhibitors and angiogenesis. *Curr. Pharm. Des.* 13 (21), 2118–2128. doi:10.2174/138161207781039580
- Shahbaaz, M., Nkaule, A., and Christoffels, A. (2019). Designing novel possible kinase inhibitor derivatives as therapeutics against *Mycobacterium tuberculosis*: An *in silico* study. *Sci. Rep.* 9 (1), 4405. doi:10.1038/s41598-019-40621-7
- Shan, Q. Y., Sang, X. N., Hui, H., Shou, Q. Y., Fu, H. Y., Hao, M., et al. (2020). Processing and polyherbal formulation of tetradium ruticarpum (A. Juss.) hartley: Phytochemistry, pharmacokinetics, and toxicity. *Front. Pharmacol.* 11, 133. doi:10.3389/fphar.2020.00133
- Shang, N., Wang, H., Bank, T., Perera, A., Joyce, C., Kuffel, G., et al. (2019). Focal adhesion kinase and β -catenin cooperate to induce hepatocellular carcinoma. *Hepatology* 70 (5), 1631–1645. doi:10.1002/hep.30707
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., et al. (2022). David: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 50 (W1), W216–W221. doi:10.1093/nar/gkac194
- Sia, D., Villanueva, A., Friedman, S. L., and Llovet, J. M. (2017). Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology* 152 (4), 745–761. doi:10.1053/j.gastro.2016.11.048
- Taha, M. O., Al-Sha'er, M. A., Khanfar, M. A., and Al-Nadaf, A. H. (2014). Discovery of nanomolar phosphoinositide 3-kinase gamma (PI3K γ) inhibitors using ligand-based modeling and virtual screening followed by *in vitro* analysis. *Eur. J. Med. Chem.* 84, 454–465. doi:10.1016/j.ejmech.2014.07.056
- Tian, Q., Miller, E. G., Ahmad, H., Tang, L., and Patil, B. S. (2001). Differential inhibition of human cancer cell proliferation by citrus limonoids. *Nutr. Cancer* 40 (2), 180–184. doi:10.1207/s15327914nc402_15
- Tong, Z., Li, M., Wang, W., Mo, P., Yu, L., Liu, K., et al. (2015). Steroid receptor coactivator 1 promotes human hepatocellular carcinoma progression by enhancing wnt/ β -catenin signaling. *J. Biol. Chem.* 290 (30), 18596–18608. doi:10.1074/jbc.M115.640490
- Verma, J., Khedkar, V. M., and Coutinho, E. C. (2010). 3D-QSAR in drug design--a review. *Curr. Top. Med. Chem.* 10 (1), 95–115. doi:10.2174/156802610790232260
- Wang, H., Wang, X., Li, X., Zhang, Y., Dai, Y., Guo, C., et al. (2012). QSAR study and the hydrolysis activity prediction of three alkaline lipases from different lipase-producing microorganisms. *Lipids Health Dis.* 11, 124. doi:10.1186/1476-511x-11-124
- Wang, H. Y., Cao, Z. X., Li, L. L., Jiang, P. D., Zhao, Y. L., Luo, S. D., et al. (2008). Pharmacophore modeling and virtual screening for designing potential PLK1 inhibitors. *Bioorg. Med. Chem. Lett.* 18 (18), 4972–4977. doi:10.1016/j.bmcl.2008.08.033
- Xia, Z., and Tang, Z. (2021). Network pharmacology analysis and experimental pharmacology study explore the mechanism of gambogic acid against endometrial cancer. *ACS Omega* 6 (16), 10944–10952. doi:10.1021/acsomega.1c00696
- Yang, F., Shi, L., Liang, T., Ji, L., Zhang, G., Shen, Y., et al. (2017). Anti-tumor effect of evodiamine by inducing Akt-mediated apoptosis in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* 485 (1), 54–61. doi:10.1016/j.bbrc.2017.02.017
- Yang, J., Zhang, X., Liu, L., Yang, X., Qian, Q., and Du, B. (2021). c-Src promotes the growth and tumorigenesis of hepatocellular carcinoma via the Hippo signaling pathway. *Life Sci.* 264, 118711. doi:10.1016/j.lfs.2020.118711
- Yoon, J. Y., Jeong, H. Y., Kim, S. H., Kim, H. G., Nam, G., Kim, J. P., et al. (2013). Methanol extract of *Evodia lepta* displays Syk/Src-targeted anti-inflammatory activity. *J. Ethnopharmacol.* 148 (3), 999–1007. doi:10.1016/j.jep.2013.05.030
- You, H., Meng, K., and Wang, Z. Y. (2018). The ER- α 36/EGFR signaling loop promotes growth of hepatocellular carcinoma cells. *Steroids* 134, 78–87. doi:10.1016/j.steroids.2018.02.007
- Zhao, P. W., Zhang, J. W., Liu, Y., Liu, Y., Liu, J. W., and Huang, J. Z. (2020). SRC-1 and Twist1 are prognostic indicators of liver cancer and are associated with cell viability, invasion, migration and epithelial-mesenchymal transformation of hepatocellular carcinoma cells. *Transl. Cancer Res.* 9 (2), 603–612. doi:10.21037/tcr.2019.11.56
- Zhao, Z., He, X., Han, W., Chen, X., Liu, P., Zhao, X., et al. (2019). Genus tetradium L.: A comprehensive review on traditional uses, phytochemistry, and pharmacological activities. *J. Ethnopharmacol.* 231, 337–354. doi:10.1016/j.jep.2018.11.035



OPEN ACCESS

EDITED BY

Pedro Vaz,
Champalimaud Foundation, Portugal

REVIEWED BY

Kang-Nan Wang,
Shandong University, China
Qiu-Juan Ma,
Henan University of Traditional Chinese
Medicine, China

*CORRESPONDENCE

Y. J. Peng,
✉ hunterpyj2016@163.com

SPECIALTY SECTION

This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 11 July 2022

ACCEPTED 14 December 2022

PUBLISHED 09 January 2023

CITATION

Lin XY, Sun SH, Liu YT, Shi QQ, Lv JJ and
Peng YJ (2023), Thiophene and
diaminobenzo- (1,2,5-thiadiazol)- based
DAD-type near-infrared fluorescent probe
for nitric oxide: A theoretical research.
Front. Chem. 10:990979.
doi: 10.3389/fchem.2022.990979

COPYRIGHT

© 2023 Lin, Sun, Liu, Shi, Lv and Peng. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Thiophene and diaminobenzo- (1,2,5-thiadiazol)- based DAD-type near-infrared fluorescent probe for nitric oxide: A theoretical research

X. Y. Lin¹, S. H. Sun¹, Y. T. Liu¹, Q. Q. Shi¹, J. J. Lv¹ and Y. J. Peng^{2,3*}

¹College of Public Health, Jinzhou Medical University, Jinzhou, China, ²College of Bio informational Engineering, Jinzhou Medical University, Jinzhou, China, ³College of Physics, Nankai University, Tianjin, China

A near-infrared fluorescent probe (LS-NO) for the real-time detection of nitric oxide (NO) in inflammatory bowel disease (IBD) was developed recently. The probe used oligoglycol morpholine-functionalized thiophene as strong electron donors and diaminobenzene (1,2,5-thiadiazole) as a weak electron acceptor and NO trapping group. It could detect exogenous and endogenous NO in the lysosomes of living cells with high sensitivity and specificity. To further understand the fluorescent mechanism and character of the probes LS-NO and LS-TZ (after the reaction of the probe LS-NO with NO), the electron transfer in the excitation and emitting process within the model molecules DAD-NO and DAD-TZ was analyzed in detail under the density functional theory. The calculation results indicated the transformation from diaminobenzene (1,2,5-thiadiazole) as a weak electron acceptor to triazolo-benzo-(1,2,5-thiadiazole) as a strong electron acceptor made LS-NO an effective “off-on” near-infrared NO fluorescent probe.

KEYWORDS

fluorescent probe, nitric oxide, inflammatory bowel disease, density functional theory, electron transfer

1 Introduction

Nitric oxide (NO) is a metastable free radical molecule which plays an important role in signal transduction and regulation in cardiovascular, immune, respiratory, gastrointestinal, and central nervous systems, and other physiological systems (Wang et al., 2002; Cosby et al., 2003). Recent studies have demonstrated that NO disorders are associated with a number of human diseases, including atherosclerosis, immune diseases, neurodegenerative diseases, cancer, and inflammatory bowel disease (Bogdan, 2001; Mel et al., 2011; Fmedsci, 2016). Moreover, increasing the NO concentration in the intestine is closely related to IBD (Kamalian et al., 2020). However, due to the lack of an effective method for the real-time detection of NO in the intestinal tract, the mechanism between NO and the pathogenesis of IBD remains unclear. Therefore, there is an urgent need to develop NO imaging probes with high sensitivity, high specificity, and high spatial and temporal resolution for the real-time detection of NO *in vivo* so as to further improve the diagnosis and treatment of IBD (Weissleder and Ntziachristos, 2003; Sasaki et al., 2005; Yu et al., 2012; Vegesna et al., 2013; Zhang et al., 2017). A near-infrared (NIR) fluorescent probe has more advantages in non-invasive imaging *in vivo*, which can further enhance the penetration of deep tissue and improve the signal-to-noise ratio (Antaris et al., 2016; Hong et al., 2017; Liu et al., 2021a; Wang et al., 2021; Xu et al., 2021). However, the current small-molecule fluorescent probes used for NO detection still have shortcomings such as short wavelength (<700 nm) and poor water solubility, especially in deep tissue and disease animal

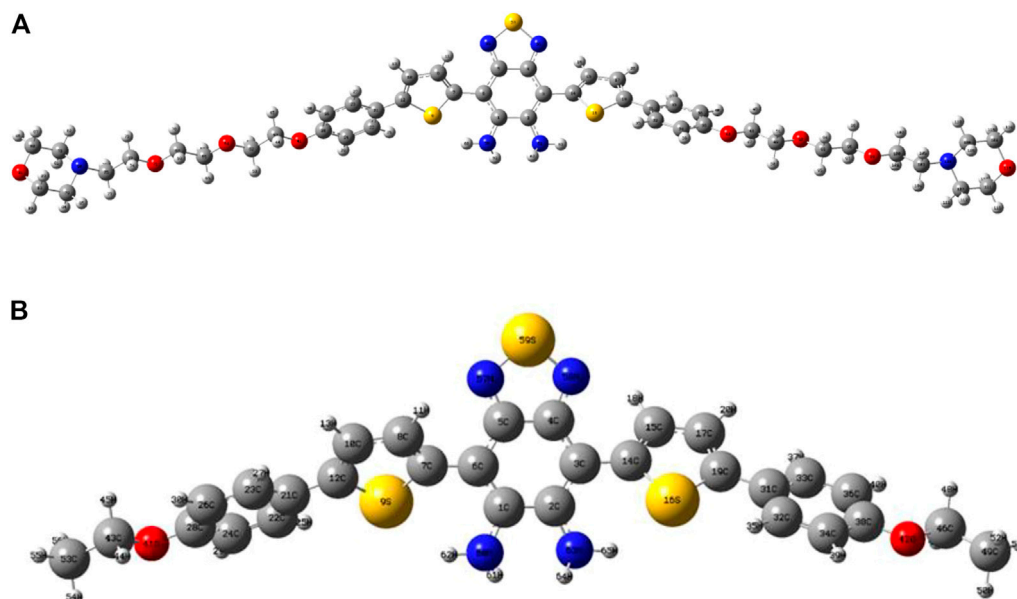


FIGURE 1
Optimized structures of (A) LS-NO and (B) DAD-NO.

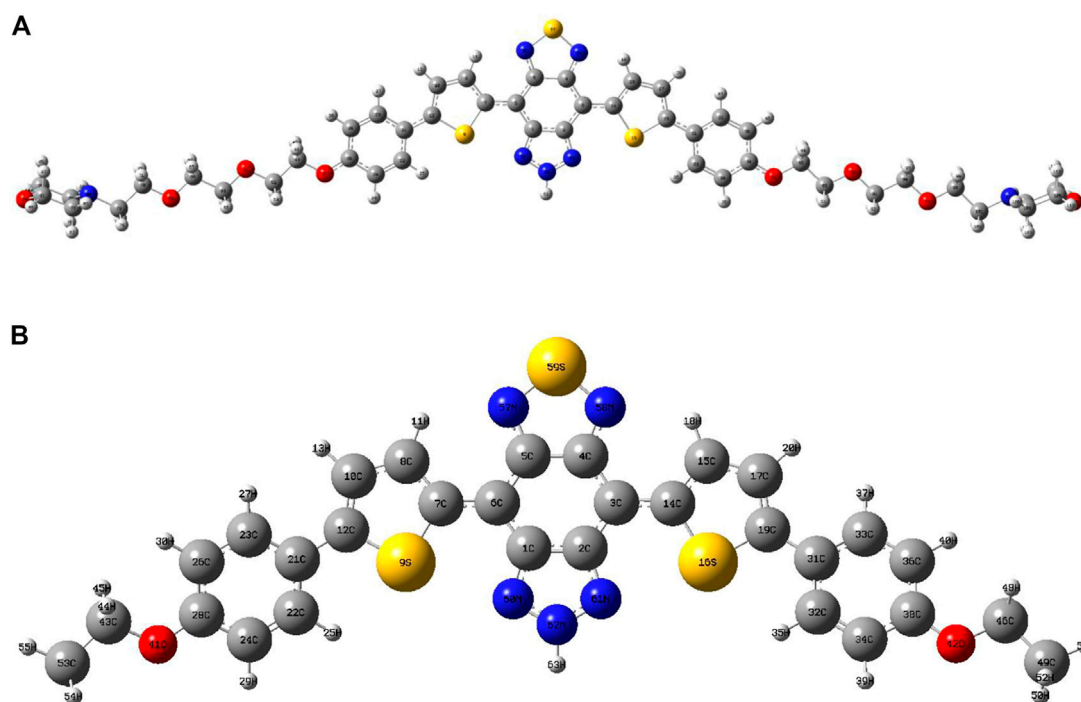


FIGURE 2
Optimized structures of (A) LS-TZ and (B) DAD-TZ.

models. It is still a great challenge to apply these probes to the real-time imaging detection of NO *in vivo* (Izumi et al., 2009; Li et al., 2016; Zhang et al., 2018).

A near-infrared fluorescent probe (LS-NO) for the real-time detection of NO in inflammatory bowel disease (IBD) was

developed by Liu et al. (2021b). The probe used oligoglycol morpholine-functionalized thiophene as strong electron donors and diaminobenzene (1,2,5-thiadiazole) as a weak electron acceptor and NO trapping group. After the specific reaction of the probe with NO, the weak electron acceptor group diaminobenzene (1,2,5-thiadiazole)

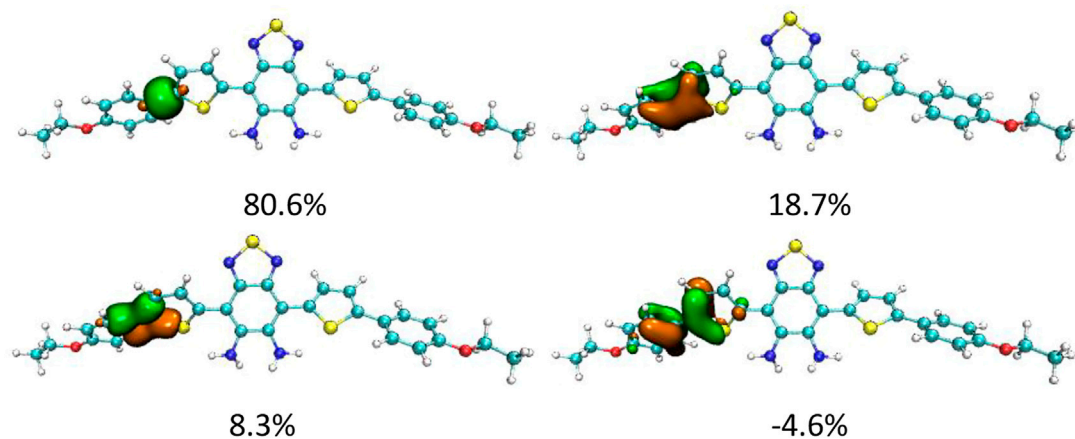


FIGURE 3
NAdO distribution of the α -related C–C bonds in DAD-NO.

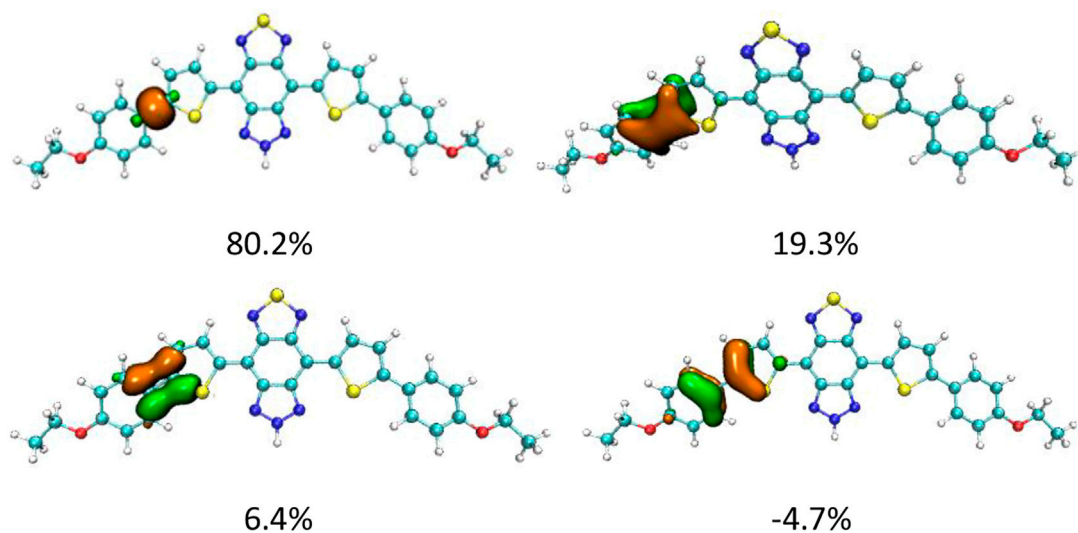


FIGURE 4
NAdO distribution of the α -related C–C bonds in DAD-TZ.

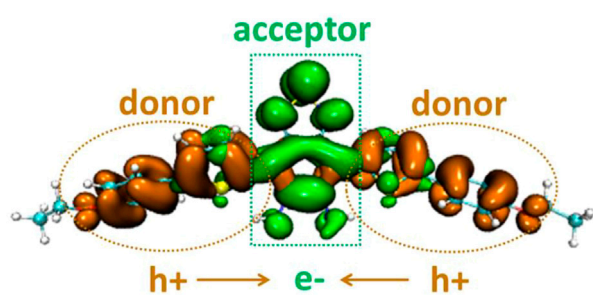


FIGURE 5
Electron distribution difference between S_0 and S_1 of DAD-NO (orange and green in the isosurface map represent the hole and electron distribution in the excitation process).

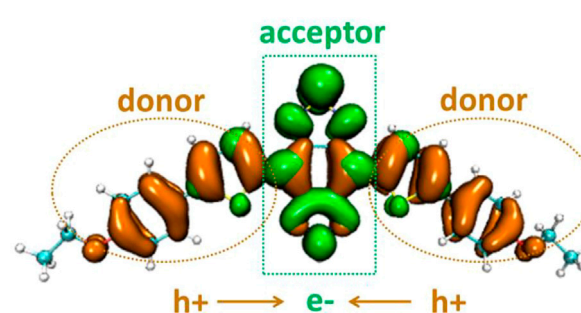
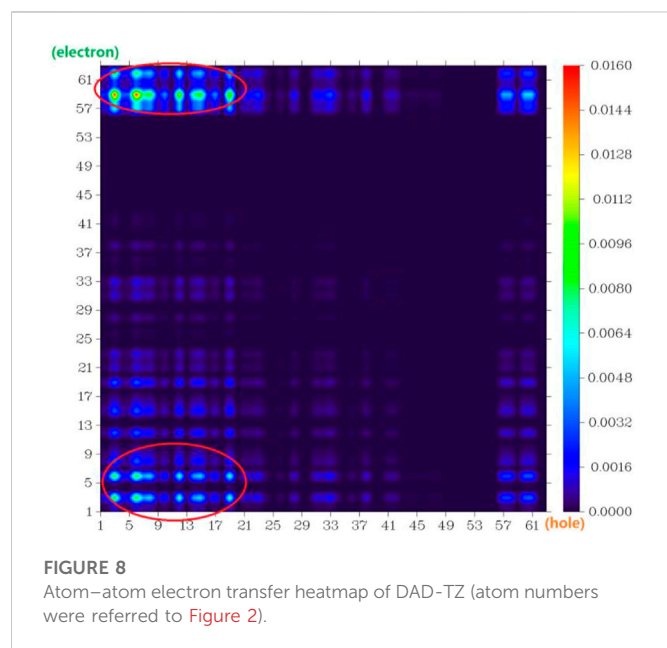
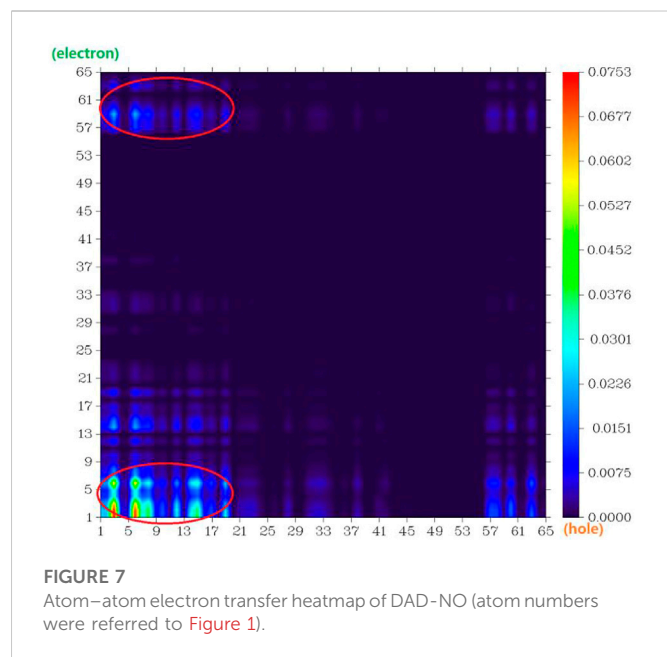


FIGURE 6
Electron distribution difference between S_0 and S_1 of DAD-TZ.



was transformed into the strong electron acceptor triazolo-benzo-(1,2,5-thiadiazole). By using the enhanced intramolecular charge transfer mechanism, the probe exhibited “off-on”-type

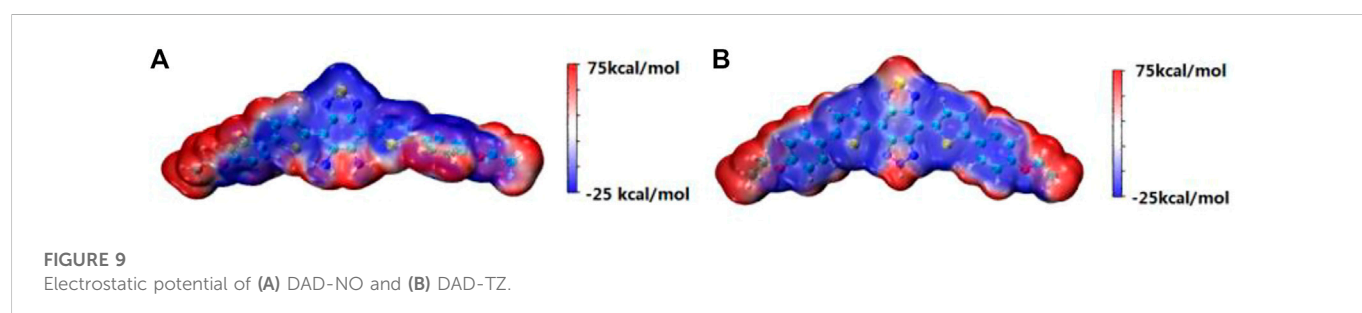
near-infrared absorption and emission at 700 and 750/800 nm, respectively. In addition, LS-NO showed good water solubility and optical stability. It can detect exogenous and endogenous NO in the lysosomes of living cells with high sensitivity and specificity. This work suggested that LS-NO was promising as a diagnostic probe for the real-time detection of NO in IBD and may also facilitate inflammatory stool detection. To further understand the fluorescent mechanism and character of the probes LS-NO and LS-TZ (after the reaction of the probe LS-NO with NO), the electron transfer in the excitation and emitting process within the probe model molecules DAD-NO and DAD-TZ was analyzed in detail under the density functional theory. The calculation results indicated the transformation from diaminobenzene (1,2,5-thiadiazole) as the weak electron acceptor to triazolo-benzo-(1,2,5-thiadiazole) as the strong electron acceptor made LS-NO an effective “off-on” NIR NO fluorescent probe.

2 Methods

The ORCA 5.0.1 ([Neese, 2018](#)) software was used to perform optimization and vibrational frequency analysis on the S_0 structures of the model probes DAD-NO and DAD-TZ under PBE0/def2-TZVP with D3 dispersion correction ([Adamo and Barone, 1999](#); [Weigend and Ahlrichs, 2005](#); [Grimme et al., 2011](#)), and then single-point energy and TDDFT calculation under wB2GP-PLYP/def2-TZVP so as to obtain the free energy with high precision ([Goerigk and Grimme, 2014](#); [Casanova-Páez et al., 2019](#); [Casanova-Paez and Goerigk, 2020](#); [Peng et al., 2021](#); [Liu et al., 2022](#)). The functional PBE0-D3(BJ) and wB2GP-PLYP used for structure optimization and TDDFT calculation of such organic probe molecules were verified to be proper ([Ali et al., 2020](#); [Bremond et al., 2021](#); [Santra and Martin, 2022](#)). The optimized S_1 structures of DAD-NO and DAD-TZ were obtained under a combination of wB2GP-PLYP/def2-TZVP to analyze the emitting wavelength in the excitation and radiation process of the probe. All the figures were rendered by means of VMD 1.9.3 software ([Humphrey et al., 1995](#)) and the analyses were conducted using the Multiwfn 3.7 code ([Lu and Chen, 2012](#)).

3 Results and discussion

The optimized structures of probes LS-NO and LS-TZ with corresponding model probes DAD-NO and DAD-TZ are depicted in [Figures 1, 2](#), respectively. In order to focus solely on the electron donor and acceptor parts in the probe, and to reduce computational time, the two terminal groups in the probes (LS-



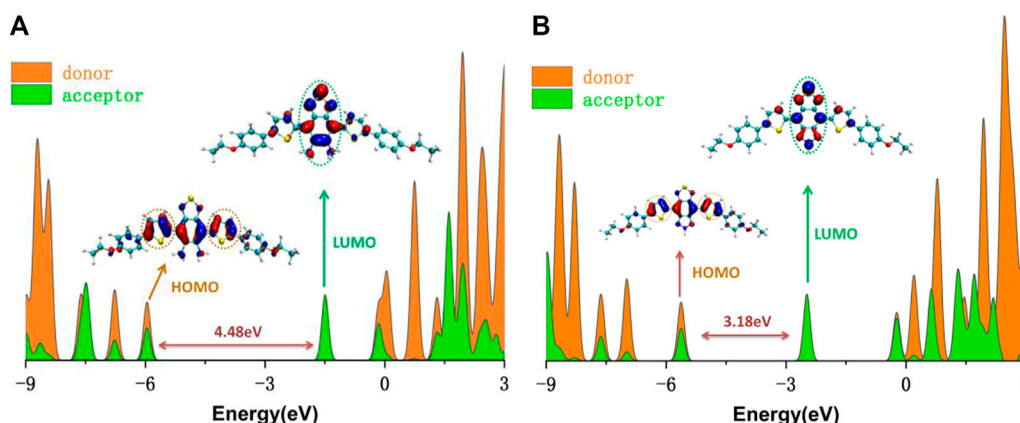


FIGURE 10
DOS of (A) DAD-NO and (B) DAD-TZ and the corresponding HOMO-LUMO energy gap calculated at the wb2GP-PLYP/def2-TZVP level in the gas.

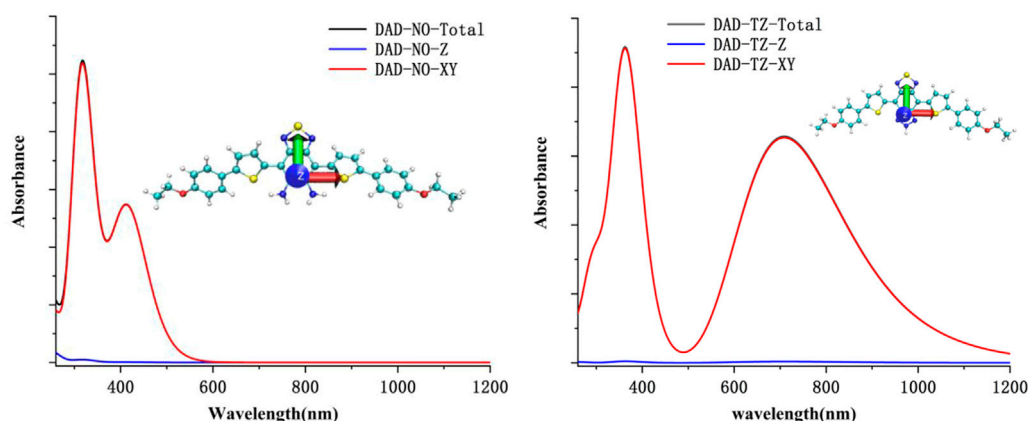


FIGURE 11
Calculated UV-Vis spectrum of DAD-NO and DAD-TZ in aqueous solution.

NO and LS-TZ) were cut to ethyl as shown in the model probes (DAD-NO and DAD-TZ).

The structures of the probes (as shown in Figures 1, 2) show that the DAD-NO had a more twisted structure than DAD-TZ, while the weak electron acceptor diaminobenzene (1,2,5-thiadiazole) was replaced by triazolo-benzo-(1,2,5-thiadiazole) as a strong electron acceptor.

Although DAD-NO had a more twisted structure than DAD-TZ, the $\alpha(\beta)$ -related C-C bonds shown in Figures 1, 2 were all typical C-C single bonds and had a similar natural adaptive orbital (NAO) distribution (Zhang et al., 2020). The details of the NAO distribution of the α -related C-C bonds in DAD-NO and DAD-TZ are shown in Figures 3, 4, respectively. The details of the NAO distribution of β -related C-C bonds in DAD-NO and DAD-TZ are given in Supplementary Figures S1, S2, respectively, for reference.

Figures 3, 4 show that the α -related C-C bond was a typical single C-C bond with main contribution from the localized sigma bond—about 80% in both DAD-NO and DAD-TZ. The second

large contribution (about 19%) came from the pi bond which delocalized to the neighbor carbon atoms unlike the sigma bond. There were two other NAOs which consist of P orbitals of carbon atoms with parallel and opposite phases, respectively, as shown in Figures 3, 4. The last two NAOs displayed non-negligible contributions to the C-C bond with positive and negative values, respectively.

The electron distribution difference between the first excitation state and ground state of DAD-NO and DAD-TZ was obtained using Multiwfn 3.7 and depicted in Figures 5, 6, respectively. The electron donor-acceptor-donor character in the probes could be clearly seen from the electron transfer process (from hole “h+” to electron “e−” as shown in Figures 5, 6). The electron acceptor parts in the probes mainly consist of N-contained and central hex-atomic rings, while the oxygen, five-membered and hex-atomic rings on the two sides contributed the donor parts in the probe.

The atom-atom electron transfer heatmap in the electron excited process of DAD-NO (Figure 7) and DAD-TZ (Figure 8) clearly

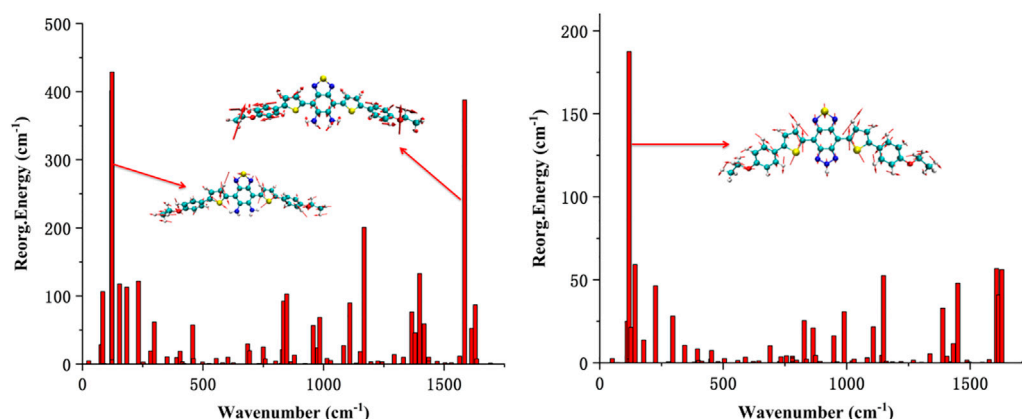


FIGURE 12
Calculated reorganization energy from each normal mode of DAD-NO and DAD-TZ.

indicated the obvious electron transfer from the donor parts to acceptor parts. The electrostatic potential of DAD-NO and DAD-TZ is shown in [Figure 9](#). It could be also seen that the replacement of weak electron acceptor diaminobenzene (1,2,5-thiadiazole) by strong electron acceptor triazolo-benzo-(1,2,5-thiadiazole) made the electron acceptor part of DAD-TZ take a larger electrostatic potential value than DAD-NO.

To clarify the contribution of the donor and acceptor parts in DAD-NO and DAD-TZ to the density of electronic states, the DOS of DAD-NO and DAD-TZ and the corresponding HOMO-LUMO energy gap calculated at the wB2GP-PLYP/def2-TZVP level in the gas are depicted in [Figure 10](#). It was obvious that the donor part's contribution to the HOMO exceeded that of the acceptor part, while the opposite situation happened within the LUMO. The replacement of diaminobenzene (1,2,5-thiadiazole) by triazolo-benzo-(1,2,5-thiadiazole) led to a smaller HOMO-LUMO energy gap which made the near-infrared fluorescence generation enhanced in DAD-TZ than in DAD-NO.

To understand the quantificational change in the energy and spectrum between DAD-NO and DAD-TZ, the UV-Vis spectrum of the probes in aqueous solution (a mixture of DMF/water with a volume ratio of 50/50) was analyzed using TDDFT calculation under the wB2GP-PLYP/def2-TZVP method. The calculated results are shown in [Figure 11](#). As shown in [Figure 11](#), there were absorption peaks located within the red and blue channels for DAD-TZ but only absorption peaks located within blue channels for DAD-NO which was consistent with the experimental results ([Liu et al., 2021b](#)). In addition, it could be clearly shown that the energy absorbance mainly located inside the probe molecular plane (XY plane) and the energy absorbance along the perpendicular direction to the molecular plane (Z axis) were almost negligible. This conclusion was consistent with the reorganization energy analysis between the ground and first excited states of DAD-NO and DAD-TZ in [Figure 12](#). It could be clearly shown that the reorganization energy of DAD-NO was bigger than that of DAD-TZ, while the direction of the norm modes with most contribution were both parallel to the molecular plane in the two probes.

4 Conclusion

The geometric and electronic structures of the ground and excited states of an effective “off-on” NIR NO fluorescent model probe DAD-NO (DAD-TZ) were analyzed under the density functional theory in detail. The calculated results indicated that the transformation from diaminobenzene (1,2,5-thiadiazole) as a weak electron acceptor in DAD-NO to triazolo-benzo-(1,2,5-thiadiazole) as a strong electron acceptor in DAD-TZ made DAD-NO an effective “off-on” NIR NO fluorescent probe with high sensitivity and specificity. The electrostatic potential and density of electronic state analysis also suggested the changing of the electron acceptor part within DAD-NO, and DAD-TZ was the structural origin of the switch on/off of the NIR fluorescence in the probes. Energy absorbance mainly located inside the probe molecular plane (XY plane) and energy absorbance along the perpendicular direction to the molecular plane (Z axis) were almost negligible. All these theoretical results would provide an insight for designing new effective probes with similar functions.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#); further inquiries can be directed to the corresponding author.

Author contributions

XL and SS collected the data; YL, QS, and JL contributed analytic tools and analyzed data; YP designed the research and prepared the manuscript.

Funding

This work was supported by the Undergraduate Innovation and Entrepreneurship Training Program of Jin Zhou Medical University

(2019055) and the Natural Science Foundation of Liaoning Province (2022-MS-389, JYTQN201923, and 20180550512).

Acknowledgments

Min Feng from Nankai University was appreciated for using GaussView 5.0 to draw **Figures 1, 2**.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Adamo, C., and Barone, V. (1999). Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* 110, 6158–6170. doi:10.1063/1.478522
- Ali, A., Rafiq, M. I., Zhang, Z., Cao, J., Geng, R., Zhou, B., et al. (2020). TD-DFT benchmark for UV-visible spectra of fused-ring electron acceptors using global and range-separated hybrids. *PCCP* 22, 7864–7874. doi:10.1039/d0cp00060d
- Antaris, A. L., Chen, H., Cheng, K., Sun, Y., Hong, G., Qu, C., et al. (2016). A small-molecule dye for NIR-II imaging. *Nat. Mater.* 15, 235–242. doi:10.1038/nmat4476
- Bogdan, C. (2001). Nitric oxide and the immune response. *Nat. Immunol.* 2, 907–916. doi:10.1038/ni1001-907
- Bremond, E., Ottochian, A., Perez-Jimenez, A. J., Ciofini, I., Scalmani, G., Frisch, M. J., et al. (2021). Assessing challenging intra- and inter-molecular charge-transfer excitations energies with double-hybrid density functionals. *J. Comput. Chem.* 42, 970–981. doi:10.1002/jcc.26517
- Casanova-Páez, M., Dardis, M. B., and Goerigk, L. (2019). ω B2PLYP and ω B2GPPLYP: The first two double-hybrid density functionals with long-range correction optimized for excitation energies. *J. Chem. Theory Comput.* 15, 4735–4744. doi:10.1021/acs.jctc.9b00013
- Casanova-Páez, M., and Goerigk, L. (2020). Assessing the Tamm–Dancoff approximation, singlet–singlet, and singlet–triplet excitations with the latest long-range corrected double-hybrid density functionals. *J. Chem. Phys.* 153, 064106. doi:10.1063/5.0018354
- Cosby, K., Partovi, K. S., Crawford, J. H., Patel, R. P., Reiter, C. D., Martyr, S., et al. (2003). Nitrite reduction to nitric oxide by deoxyhemoglobin vasodilates the human circulation. *Nat. Med.* 9, 1498–1505. doi:10.1038/nm954
- Fmedsci, P. J. B. F. (2016). *J. Allergy Clin. Immunol.* 138, 16
- Goerigk, L., and Grimme, S. (2014). Double-hybrid density functionals. *Wiley Interdiscip. Reviews-Computational Mol. Sci.* 4, 576–600. doi:10.1002/wcms.1193
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi:10.1002/jcc.21759
- Hong, G., Antaris, A. L., and Dai, H. (2017). Near-infrared fluorophores for biomedical imaging. *Nat. Biomed. Eng.* 1, 0010–0019. doi:10.1038/s41551-016-0010
- Humphrey, W., Dalke, A., and Schulten, K. K. (1995). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Izumi, S., Urano, Y., Hanaoka, K., Terai, T., and Nagano, T. (2009). A simple and effective strategy to increase the sensitivity of fluorescence probes in living cells. *J. Am. Chem. Soc.* 131, 10189–10200. doi:10.1021/ja902511p
- Kamalian, A., Asl, M. S., Dolatshahi, M., Afshari, K., Abdolghaffari, A. H., Roudsari, N. M., et al. (2020). Interventions of natural and synthetic agents in inflammatory bowel disease, modulation of nitric oxide pathways. *World J. Gastroenterol.* 26, 3365–3400. doi:10.3748/wjg.v26.i24.3365
- Li, Z., Liu, Z., Zeng, L., Feng, W., and Mao and Zhiqiang (2016). *Chem. Sci.* 35, 2235
- Liu, L.-Y., Zhao, Y., Zhang, N., Wang, K.-N., Tian, M., Pan, Q., et al. (2021). Ratiometric fluorescence imaging for the distribution of nucleic acid content in living cells and human tissue sections. *Anal. Chem.* 93, 1612–1619. doi:10.1021/acs.analchem.0c04064
- Liu, S., Zhu, Y., Wu, P., and Xiong, H. (2021). Highly sensitive D–A–D-type near-infrared fluorescent probe for nitric oxide real-time imaging in inflammatory bowel disease. *Anal. Chem.* 93, 4975–4983. doi:10.1021/acs.analchem.1c00281
- Liu, Y.-L., Huang, H., and Peng, Y.-j. (2022). Fluorescent probe for simultaneous detection of human serum albumin and sulfite: A theoretical analysis. *J. Mol. Struct.* 1255, 132441. doi:10.1016/j.molstruc.2022.132441
- Lu, T., and Chen, F. (2012). Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592. doi:10.1002/jcc.22885
- Mel, A. D., Murad, F., and Seifalian, A. M. (2011). *Chem. Rev.* 111, 5742–5767. doi:10.1021/cr200008n
- Neese, F. (2018). *Wiley Interdisciplinary Rev. Comput. Mol. Sci.* 8, e1327.
- Peng, Y. J., Huang, H., and Wang, C. J. (2021). DFT investigation on electronic structure, chemical bonds and optical properties of Cu₆(SR)₆ nanocluster. *Chem. Phys. Lett.* 780, 138898. doi:10.1016/j.cplett.2021.138898
- Santra, G., and Martin, J. M. L. (2022). *J. Phys. Chem. Lett.* 13, 3499–3506. doi:10.1021/acs.jpclett.2c00718
- Sasaki, E., Kojima, H., Nishimatsu, H., Urano, Y., Kikuchi, K., Hirata, Y., et al. (2005). Highly sensitive near-infrared fluorescent probes for nitric oxide and their application to isolated organs. *J. Am. Chem. Soc.* 127, 3684–3685. doi:10.1021/ja042967z
- Vegesna, G. K., Sripathi, S. R., Zhang, J., Zhu, S., He, W., Luo, F.-T., et al. (2013). Highly water-soluble BODIPY-based fluorescent probe for sensitive and selective detection of nitric oxide in living cells. *ACS Appl. Mater. Interfaces* 5, 4107–4112. doi:10.1021/am303247s
- Wang, K.-N., Liu, L.-Y., Mao, D., Xu, S., Tan, C.-P., Cao, Q., et al. (2021). A polarity-sensitive ratiometric fluorescence probe for monitoring changes in lipid droplets and nucleus during ferroptosis. *Angew. Chemie-International Ed.* 60, 15095–15100. doi:10.1002/anie.202104163
- Wang, P. G., Xian, M., Tang, X., Wu, X., Wen, Z., Cai, T., et al. (2002). Nitric oxide Donors: chemical activities and biological applications. *Chem. Rev.* 102, 1091–1134. doi:10.1021/cr000040l
- Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *PCCP* 7, 3297–3305. doi:10.1039/b508541a
- Weissleder, R., and Ntziachristos, V. (2003). Shedding light onto live molecular targets. *Nat. Med.* 9, 123–128. doi:10.1038/nm1013-123
- Xu, W., Wang, D., and Tang, B. Z. (2021). NIR-II AIEgens: A win-win integration towards bioapplications. *Angew. Chem. Int. Ed.* 60, 7476–7487. doi:10.1002/anie.202005899
- Yu, H., Xiao, Y., and Jin, L. (2012). A lysosome-targetable and two-photon fluorescent probe for monitoring endogenous and exogenous nitric oxide in living cells. *J. Am. Chem. Soc.* 134, 17486–17489. doi:10.1021/ja308967u
- Zhang, C., Hu, S. X., Liu, H. T., Yang, Y., and Zhang, P. (2020). Correction to “bonding properties and oxidation states of plutonium in Pu₂O_n (n = 1–8) molecules studied by using screened hybrid density functional theory”. *J. Phys. Chem. A* 124, 2513. doi:10.1021/acs.jpca.0c02144
- Zhang, P., Tian, Y., Liu, H., Ren, J., Wang, H., Zeng, R., et al. (2018). *In vivo* imaging of hepatocellular nitric oxide using a hepatocyte-targeting fluorescent sensor. *Chem. Commun.* 54, 7231–7234. doi:10.1039/c8cc03240h
- Zhang, R. R., Schroeder, A. B., Grudzinski, J. J., Rosenthal, E. L., Warram, J. M., Pinchuk, A. N., et al. (2017). Beyond the margins: Real-time detection of cancer using targeted fluorophores. *Nat. Rev. Clin. Oncol.* 14, 347–364. doi:10.1038/nrclinonc.2016.212

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.990979/full#supplementary-material>



OPEN ACCESS

EDITED BY

Lalith Perera,
National Institute of Environmental Health
Sciences (NIH), United States

REVIEWED BY

Zoran S. Markovic,
University of Kragujevac, Serbia
Renjith Thomas,
Mahatma Gandhi University, India
Nesimi Uludag,
Namik Kemal University, Türkiye

*CORRESPONDENCE

Nino Russo,
✉ nino.russo@unical.it

SPECIALTY SECTION

This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 13 December 2022

ACCEPTED 09 January 2023

PUBLISHED 24 January 2023

CITATION

Spiegel M, Ciardullo G, Marino T and
Russo N (2023), Computational
investigation on the antioxidant activities
and on the M^{pro} SARS-CoV-2 non-
covalent inhibition of isorhamnetin.
Front. Chem. 11:1122880.
doi: 10.3389/fchem.2023.1122880

COPYRIGHT

© 2023 Spiegel, Ciardullo, Marino and
Russo. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Computational investigation on the antioxidant activities and on the M^{pro} SARS-CoV-2 non-covalent inhibition of isorhamnetin

Maciej Spiegel^{1,2}, Giada Ciardullo¹, Tiziana Marino¹ and
Nino Russo^{1*}

¹Dipartimento di Chimica e Tecnologie Chimiche, Università della Calabria, Rende, Cosenza, Italy,

²Department of Pharmacognosy and Herbal Medicines, Wrocław Medical University, Wrocław, Poland

In the present work, we report a computational study on some important chemical properties of the flavonoid isorhamnetin, used in traditional medicine in many countries. In the course of the study we determined the acid-base equilibria in aqueous solution, the possible reaction pathways with the •OOH radical and the corresponding kinetic constants, the complexing capacity of copper ions, and the reduction of these complexes by reducing agents such as superoxide and ascorbic anion by using density functional level of theory Density Functional Theory. Finally, the non-covalent inhibition ability of the SARS-CoV-2 main protease enzyme by isorhamnetin was examined by molecular dynamics (MD) and docking investigation.

KEYWORDS

antioxidants, chemical equilibria, reaction mechanisms, kinetic constant, copper complexes, Fenton's reaction, main protease inhibition

1 Introduction

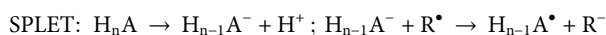
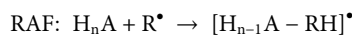
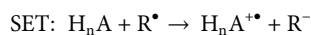
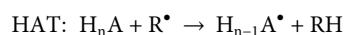
A commonly accepted definition of oxidative stress is an imbalance between the production of reactive free radicals and the ability of the organism to inactivate them before their excessive production becomes harmful (Sies, 1985; Sies et al., 2017). In other words, it is defined when the action of oxidizing agents is not effectively counteracted by molecules present in our body that have antioxidant capacities.

Oxidative stress damages cells and organs, and is a direct or indirect cause of many conditions, ranging from cancer to atherosclerosis, neurodegenerative disorders (e.g., Alzheimer's and others) and pulmonary diseases (Forman and Zhang, 2021). For these reasons, an enormous amount of scientific work in recent decades has concerned both the specific damage caused by oxidative stress and the chemical mechanisms of action underlying oxidative processes, as well as compounds capable of preventing them.

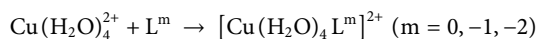
Considerable attention has been paid to understanding the structures, chemical-physical properties and mechanisms of action of both natural and synthetic antioxidant compounds (Leopoldini et al., 2011; Galano et al., 2016a; Apak et al., 2016). Many antioxidants originate from the plant kingdom (e.g., fruits, cereals, vegetables, and plants) and have been used in traditional medicinal systems of European, African, Asian, and American medicine for many centuries. In particular, flavonoid compounds of natural origin have proven to be powerful antioxidant systems, suitable for combating oxidative stress. Their actions are multidirectional, but the most important can be outlined as:

1) direct scavenging of free radicals, also known as primary antioxidantation, which essentially consists of a reaction between a powerful free radical and an antioxidant with the formation of a

new less active radical, and reaction can proceed until the formation of a neutral specie (Milenković et al., 2020). This reaction can occur through different mechanisms such as transferring a hydrogen atom or an electron or both from the antioxidant to the free radical (HAT and SET), or sequential proton-loss electron transfer (SPLET), or by bonding the radical to the structure of the antioxidant (RAF) (Galano et al., 2016b; Spiegel, 2022):



2) the capacity of the compound to chelate copper and iron ions responsible for the production of free radicals through the Fenton's reaction. It is also referred to as a secondary antioxidant effect and is also relevant to the possible treatment of Alzheimer's disease based on complexation of these metals (Sharma et al., 2013; Spiegel et al., 2022a):



In addition, some antioxidants are effective inhibitor of a number of enzymes implicated in various diseases, including SARS-CoV-2 (Yu et al., 2012; Puttaswamy et al., 2020; Xu et al., 2020).

In this article, we report a theoretical (DFT) and computations (MD and docking) investigation that can contribute to a better understanding of the mechanisms of action of isorhamnetin antioxidant flavonoid (H₄Is), which contains hydroxyl groups in its backbone, which are the main structural elements on which the antioxidant capacity of polyphenols rely. Isorhamnetin is present in the juice of Chinese sea buckthorn (Pengfei et al., 2009), and can also be extracted from the fruits, flowers and leaves of *Ginkgo biloba*, *A. roxburghii*, and *Hippophae rhamnoides* L. (Teng and Luan, 2016; Gong et al., 2020). Furthermore, the great importance of isorhamnetin is due to the fact that it is a metabolite of quercetin, converted by the enzyme catechol-O-methyltransferase. This means that isorhamnetin concentrations in human plasma high, even when only quercetin is taken (Manach et al., 1997).

In order to highlight the performance of theoretical methods in reproducing and predicting antioxidant properties and related mechanisms in environments simulating physiological ones (aqueous and lipid), the paper will examine: 1) the acid-base equilibria in water solvent, 2) the reactions between the [•]OOH radical and isorhamnetin, 3) the chelating capacity, and 4) the inhibition process of the SARS-CoV-2 main protease enzyme.

2 Methods and computational details

All the computations were performed using Density Functional Theory (DFT) implemented in Gaussian 09 code (Frisch et al., 2014) and following the QM-ORSA (Quantum Mechanics-Based Test for Overall Free Radical Scavenging Activity) computational protocol (Galano and Alvarez-Idaboy, 2013). We used the M06-2X (Zhao et al., 2006) exchange-correlation functional, the 6-311+G (d,p) basis set and the solvation model based on density (SMD) (Marenich et al., 2009), previously tested and applied to a series of molecules with antioxidant properties (Belcastro et al., 2006; Alberto et al., 2013; Ngo et al., 2019; Castaneda-Arriaga et al., 2020; Parise et al., 2021a; Reina et al., 2021; Spiegel et al., 2021; Spiegel, 2022).

Pentylethanoate ($\epsilon = 4.7$) and water ($\epsilon = 78.4$) was selected to simulate the physiological relevant environments. The geometries of minima (reactants and products) and transition states (TS) located along the considered reaction pathways were optimized and characterized by computing vibrational frequencies and establishing the intrinsic reaction coordinates (IRC).

For the open-shell systems, the unrestricted procedure was used. The pK_a values and molar fractions for neutral and charged species were determined according to the methodology proposed and tested earlier (Galano et al., 2016a).

To compute kinetic rate constants conventional transition state theory ((Evans and Polanyi, 1935; Truhlar et al., 1996) was used, and for reactions close to the diffusion limit, the Collins–Kimball theory was applied (Collins and Kimball, 1949). For SET reactions, energy barriers were calculated adopting the Marcus theory (Marcus, 1993). The intrinsic reactivity indices such as bond dissociation energy (BDE), ionization potential (IP), proton affinity (PA) and proton desorption energy (PDE) were estimated under the adiabatic approximation with the following values of solvation enthalpies of H⁺ ($\Delta H(\text{H}^+) = 1,055.7$ kJ/mol) and electron ($\Delta H(\text{e}^-) = 77.5$ kJ/mol. (Markovic et al., 2016). To visualize the structures the MarvinSketch version 21.15.0 software (ChemAxon) was used.

The initial unbound structure of the SARS-CoV-2 main protease (M^{pro}) has been obtained using the crystal structure of main protease bound to non-covalent inhibitor (PDB code 6W63) (<https://www.rcsb.org/structure/6W63>, 2020), removing the inhibitor from the crystallographic structure and adding the hydrogen atoms by using H++ (Anandakrishnan et al., 2012) and to calculate the protonation states of titratable residues at pH 7.4. Both the protonation states of the His41 catalytic residue neutral, with the hydrogen on the N_δ and N_ε side chain have been considered. 300 ns MDs were performed, by using AMBER16 code (Case et al., 2017) and the FF14SB force field (Maier et al., 2021), for the two protonation states of the unbound M^{pro} considering a solvated orthorhombic box with a buffer of 10 Å, using TIP3P water model and the following other conditions: integration step of 2 fs coupling SHAKE algorithm; NPT ensemble at 1 bar pressure using the Berendsen barostat (Berendsen et al., 1998) with a time constant $\tau_p = 2.0$ ps. The Particle mesh Ewald summation method (Darden et al., 1998) has been employed for the electrostatic potential long-range interactions with a 12 Å cutoff distance. In order to select different representative conformations of the system, root-mean square deviation (RMSD) based clustering of the whole trajectory has been performed using the agglomerative bottom-up approach available in Amber16 tools. After removing overall rotations and translations by RMS fitting the Ca atoms' positions of the trajectory, the average linkage clustering algorithm has been applied, identifying 10 representative conformations of the protein. The complete MD analysis of the unbound M^{pro} has been reported in our previous study (Parise et al., 2021b).

The representative structure of the unbound M^{pro}, having the hydrogen on the N_δ of His41, and with the highest percentage of population was considered for the molecular docking approach performed by using AutoDock Vina (version 4.2) (Trott and Olson, 2010). The ligand and the SARS-CoV-2 M^{pro} models were processed using the AutoDock tools (ADT) to obtain the PDBQT (Protein Data Bank, Partial Charge (Q), and Atom Type (T)) coordinate files containing the information, namely polar hydrogen atoms, partial charges, correct atom types, and information on the articulation of flexible molecules. In particular, Gasteiger-Marsili

charges were loaded in ADT. The substrate has been docked into a section of the crystal representing the minimal unit of ligand ($37.5 \times 15.2 \times 37.6$ Å) adopted in the simulations. The dimension box of $40 \times 40 \times 40$ Å has been chosen to abundant cover the active site.

Before MD simulations of the $M^{Pro}:H_4Iso/M^{Pro}:H_3Iso^-$ complexes, it was necessary to obtain H_4Iso and H_3Iso^- parameters. At this purpose gas-phase geometry optimization has been carried out using B3LYP/6-31G*. Atomic charges were derived by fitting the electrostatic potential according to the Merz–Singh–Kollman scheme, (Bayly et al., 1993), using the RESP procedure. Antechamber and parmchk modules of Amber16 (Case et al., 2017) have been employed to generate preparatory files to perform molecular mechanics (MM) relaxation of the complexes. 4 Na^+ counter ions were added to neutralize the system for $M^{Pro}:H_4Iso$ and 5 Na^+ for $M^{Pro}:H_3Iso^-$. The production step of 100 ns for the both complexes ($M^{Pro}:H_4Iso$ and $M^{Pro}:H_3Iso^-$) was performed using the same procedure of the apo-form simulation.

RMSD-based clustering of the entire trajectories was performed according to the relaxed complex scheme (RCS) protocol implemented in Amber 16 (Case et al., 2017) to provide a sampled and energetically accessible conformational ensemble. After removing the overall rotations and translations by RMS fitting of the positions of the C α atoms of the trajectory, the average binding clustering algorithm implemented in cpptraj was applied to identify 10 clusters of representative conformations of the protein, described in Supplementary Table S4.

The resulting MD trajectories were used to assess the magnitude of structural changes in terms of root mean square deviation (RMSD), propensity of a given residue or region to shift, and root mean square fluctuation (RMSF).

The binding free energies between the ligand (H_4Iso or H_3Iso^-) and M^{Pro} were calculated by solving the linearised Poisson-Boltzman equation using the MM-PBSA (Molecular mechanics-Poisson Boltzman surface area) method, as implemented in the Amber code 16 (Case et al., 2017). The igb flag value of five associated with a salt concentration of 0.1 M was used. For the calculations, 100 frames of each MD trajectory over the last 50 ns were analysed.

3 Results and discussion

3.1 Chemical equilibria in water

In understanding the chemical behavior of substances in aqueous solution, it is fundamental to know the relative acid-base equilibria.

Sometimes the very low solubility in water of various antioxidants makes experimental measurements difficult. Modern computational chemistry offers a suitable alternative to such situations (Galano et al., 2016b). Figure 1 shows possibly present neutral and charged species of isorhamnetin in water solution, reported along with their pK_a values. The distribution diagram of isorhamnetin species as a function of pH is given in the supporting information section (Supplementary Figure S1).

The lowest pK_a value (7.27) was associated with the deprotonation of the –OH group located at the C₇ position, followed by C_{4'} (8.78), C₃ (10.36), and C₅ (12.26). Also, for structurally similar scutellarein, chrysin and quercetin, deprotonation at the C₇ position is preferred. Another important data useful for studying antioxidant properties in water is the molar distribution at physiological pH. Figure 1 evidences that the neutral (41.58%) and mono-anion (56.08%) species are found in the highest molar fractions. Contrary to these, the di-anion form is present at much lower, yet still non-negligible amount (2.34%). Therefore, these three species must be considered in the investigation.

3.2 Thermodynamic descriptors

Some indication on the occurrence of possible reaction pathways between radicals and antioxidant compounds can be obtained from adiabatic energy computations of certain molecular indicators, such as:

- 1) ionization energy (IP):

$$IP = \Delta H(Cx - OH^+) + \Delta H(e^-) - \Delta H(Cx - OH)$$

- 2) proton affinity (PA):

$$PA = \Delta H(Cx - O^-) + \Delta H(H^+) - \Delta H(Cx - OH)$$

- 3) bond dissociation energy (BDE):

$$BDE = \Delta H(Cx - O\cdot) + \Delta H(H\cdot) - \Delta H(Cx - OH)$$

- 4) proton desorption energy (PDE):

$$PDE = \Delta H(Cx - O^-) + \Delta H(H^+) - \Delta H(Cx - OH^+).$$

The obtained values are collected in Table 1.

In the pentylethanoate solvent, commonly chosen to mimic a lipid-like environment, the BDE values indicate that the preferred site of dehydrogenation is the one involving the –OH group at the C₃ position. The same behavior is observed for all the species present in

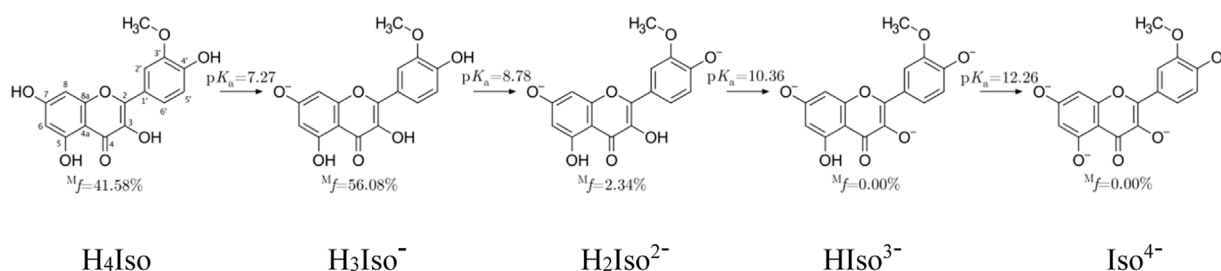


FIGURE 1

Dissociation constants and molar fractions for isorhamnetin species at pH = 7.4.

water. The ionization potential in water decrease when moving from neutral to deprotonated forms, as do PA and PDE. Similar values and trends were also found in previous work done at the DFT level, but using a different exchange-correlation functional (B3LYP) (Thong et al., 2019).

3.3 Reactions in lipid-like and aqueous environments

The computed Gibbs free energies of reaction (ΔG) and activation (ΔG^\ddagger) for the considered mechanisms—HAT, SET, and RAF—of the reaction between the $\cdot\text{OOH}$ radical and isorhamnetin in the two considered environments are shown in Table 2. For RAF processes,

we report only those with ΔG values less than 10 kcal/mol, since they are relevant from both thermodynamic and kinetic viewpoints.

From Table 2, we see that the more favored thermodynamic process is the hydrogen atom transfer from the C_3 position of neutral and charged isorhamnetin species to the $\cdot\text{OOH}$ radical in both lipid-like and aqueous environments. In particular, the latter cases provides a noteworthy observation on how the exergonic character of the process increases with the transition from neutral to mono- and di-anion forms. Also, the HAT processes from the C_4' position show negative Gibbs free energies, with a trend similar to HAT at the C_3 site. The ΔG values for the mechanisms of radical addition and electron transfer reactions allows us to hypothesize that they can occur in both considered environments. Concerning the activation energies of the HAT- C_3 and HAT- C_4' processes in water

TABLE 1 IP, BDE, PA, and PDE values calculated for neutral and charged isorhamnetin species in water and pentylethanoate solvents. All values are in kcal/mol.

Species	Solvent	OH position	IP	BDE	PA	PDE
H_4Iso	Pentylethanoate	C_3	131.3	73.8	61.0	9.2
		C_4'		75.6	60.9	10.9
		C_5		89.4	63.9	24.7
		C_7		85.2	53.6	20.5
	Water	C_3	118.4	75.8	32.6	7.2
		C_4'		77.3	32.7	8.7
		C_5		86.3	33.1	17.7
		C_7		87.3	29.0	18.7
H_3Iso^-	Water	C_3	109.8	72.9	35.3	12.8
		C_4'		75.9	33.8	15.8
		C_5		84.7	39.1	24.7
$\text{H}_2\text{Iso}^{2-}$	Water	C_3	91.9	66.0	38.7	23.9
		C_5		79.4	40.9	37.3

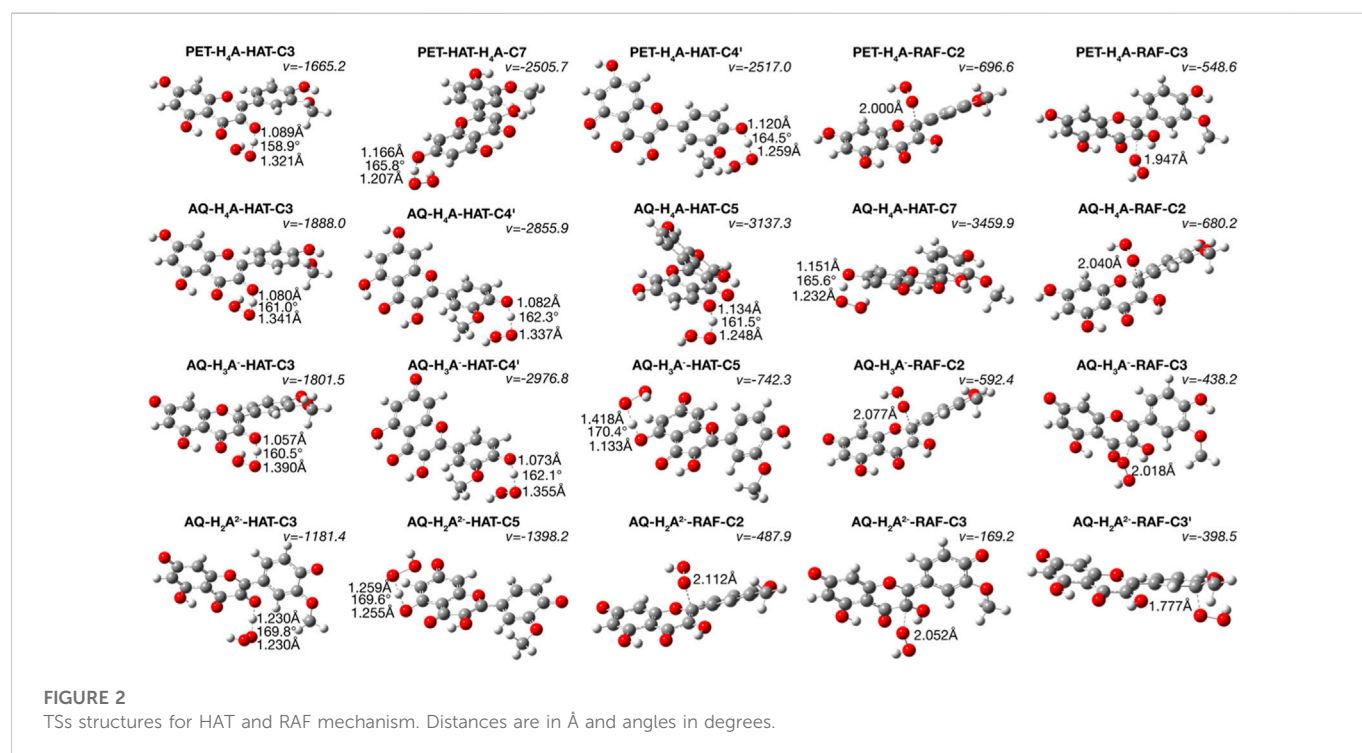
TABLE 2 Gibbs free energies of reaction (ΔG) and activation (ΔG^\ddagger) at 298.15 K in aqueous solution and pentylethanoate (indicated by $^{\text{PE}}$ apex) for neutral and charged isorhamnetin species. All values are in kcal/mol.

Mechanism	$\text{H}_4\text{Iso}^{\text{PE}}$		H_4Iso		H_3Iso^-		$\text{H}_2\text{Iso}^{2-}$	
	ΔG	ΔG^\ddagger	ΔG	ΔG^\ddagger	ΔG	ΔG^\ddagger	ΔG	ΔG^\ddagger
HAT- C_3	-2.3	20.1	-3.3	18.6	-6.3	17.4	-13.1	-3.3
HAT- C_4'	-0.5	21.3	-1.8	20.4	-3.2	19.7		
HAT- C_5	13.2	—	7.2	28.6	5.6	36.4	0.3	27.2
HAT- C_7	9.0	25.3	8.2	26.0				
RAF-OOH- C_2	2.8	22.5	1.4	20.7	0.2	18.4	1.8	18.0
RAF-OOH- C_3	9.5	20.8	10.2	—	6.7	18.4	5.9	14.1
RAF-OOH- C_3'							8.7	31.6
SET	—	—	28.8	53.5	20.3	26.4	2.3	7.6

TABLE 3 Rate constants (k)* and branching ratios (Γ) of the reaction between isorhamnetin and $^{\bullet}\text{OOH}$ radical in 1:1 ratio, computed at 298.15 K, for the different species present in solution at pH = 7.4

Mechanism	$\text{H}_4\text{Iso}^{\text{PE}}$		H_4Iso		H_3Iso^-		$\text{H}_2\text{Iso}^{2-}$	
	k ($\text{M}^{-1} \text{s}^{-1}$)	$\Gamma(\%)$	k ($\text{M}^{-1} \text{s}^{-1}$)	$\Gamma(\%)$	k ($\text{M}^{-1} \text{s}^{-1}$)	$\Gamma(\%)$	k ($\text{M}^{-1} \text{s}^{-1}$)	$\Gamma(\%)$
HAT- C_3	8.06×10^0	31.9	1.48×10^2	39.7	6.80×10^2	50.3	2.49×10^9	25.5
HAT- C_4'	1.71×10^1	67.6	2.52×10^2	60.3	6.59×10^2	48.7		
HAT- C_5			9.34×10^{-4}	0.0	5.76×10^{-13}	0.0	2.22×10^{-5}	0.0
HAT- C_7	3.42×10^{-3}	0.0	1.36×10^{-2}	0.0				
RAF- C_2	8.21×10^{-3}	0.0	1.67×10^{-1}	0.0	7.12×10^0	0.0	1.24×10^1	0.0
RAF- C_3	1.18×10^{-1}	0.5			6.03×10^0	0.0	7.16×10^3	0.0
RAF- C_3'							1.23×10^{-9}	0.0
SET	—	—	3.89×10^{-27}	0.0	7.14×10^{-9}	0.0	7.27×10^9	74.5
Total	2.53×10^1	3.73×10^2	1.35×10^3	9.76×10^9				
Overall		1.55×10^2	7.58×10^2	2.28×10^8				

*The sum of the individual rate constants from the studied reaction routes was computed as “Total,” while “Overall” is the sum of the rate constants for the different species present in solution at pH = 7.4.



solvent, we observe that the corresponding values decrease with the increasing ionic character of the species.

The rate constants and the branching ratios for the considered reaction mechanisms are reported in Table 3, while the transition state structures for HAT and RAF mechanisms are depicted in Figure 2.

In the lipid-like phase, the total rate constant is 2.53×10^1 and indicates that the process taking place mainly follows the HAT mechanism from the C_4' -OH position. Due the presence of different charged and neutral species, on the other hand, the

situation in water solution is different. In fact, for the H_4Iso form the hydrogen abstraction process at the C_4' position is kinetically favored ($k = 2.52 \times 10^2$), followed by the one at the C_3 site (1.48×10^2). The total rate constant take into account the branching ratio that is 31.9% and 67.6% for C_3 and C_4' site, respectively. In the mono-anion species, the antioxidant properties are essentially due to the SET mechanism as indicated by the k (8.20×10^9) and the molar fraction values (74.5%). The same mechanism is favored by the H_3Iso^- -form, but now, although the SET still gives the highest

kinetic constant, the HAT mechanism from the C3 position also contributes considerably ($\Gamma = 25.5\%$) to the total kinetic constant value. Considering the overall kinetic constant values, Table 3 clearly indicate that the isorhamnetin molecule has a high potential to scavenge $\cdot\text{OOH}$ radical ($k = 4.60 \times 10^9$), essentially owing to the mono- and di-anionic forms, underlining the importance of considering all species present at physiological pH.

Inspecting the TSs structures (Figure 2) of the attack of $\cdot\text{OOH}$ radical on one of the hydroxyl hydrogens provides reliable data on the breakage of the isorhamnetin-OH bonds and the formation of a new one with the radical, observed as the generation of H_2O_2 molecule. The presence of imaginary frequencies confirm this phenomenon.

A rationalization of the kinetic behaviors can be done also considering the electron spin densities of the radicals obtained after the abstraction of the hydrogens in the different positions. As shown in Supplementary Figure S2, the abstraction of a proton in C₃ position induces an electron spin delocalization that involve the entire molecular structure, in both neutral and charged species and in both considered solvents, that stabilize the radicals. On the

contrary, the hydrogen loss in position C₅ gives a radical in which the electronic spin density is more concentrated in one side of the structure.

A comparison with some other antioxidants of similar structure, studied previously using very similar or identical computational protocols, is possible by examining the data reported in Table 4.

In pentylethanoate, the scavenging activity of isorhamnetin against the $\cdot\text{OOH}$ radical is relatively low and lower than that of the other compounds tested except for scutellarin (Spiegel et al., 2022b). On the contrary, in aqueous solution, its kinetic constant takes on a very high value ($4.60 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$), comparable to that of quercetin ($k = 8.11 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$) (Castaneda-Arriaga et al., 2020) and daphnetin (1.51×10^7) (Boulebd and Khodja, 2021) and about five orders of magnitude higher than the corresponding value of Trolox ($8.96 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$) (Alberto et al., 2013), which is generally used as a comparison to determine the antioxidant power of a molecule. On the other hand, isorhamnetin is a less efficient scavenger with respect to [4-(benzo[d]thiazol-2-yl)-2-((4,7-dimethyl-1,4,7-triazonan-1-yl)-methyl)-6-methoxyphenol] (L1 in Table 4) in the lipid-like phase, but more efficient in water solution (Spiegel et al., 2022a).

TABLE 4 k_{overall} values of Isorhamnetin and other structurally similar antioxidant against the OOH radical. Value are in $\text{M}^{-1} \text{ s}^{-1}$.

Molecules	k_{overall} (lipid-like)	k_{overall} (water)
Isorhamnetin	2.53×10^1	4.60×10^9
Scutellarin	3.57×10^{-6}	7.09×10^5
Scutellarein	1.06×10^3	2.23×10^5
Trolox	3.40×10^3	8.96×10^4
L1	1.89×10^4	1.02×10^5
Quercetin		8.11×10^9
Daphnetin		1.51×10^7

3.4 Copper chelating ability

The computations of the Gibbs energies (ΔG_f) for the following reactions:

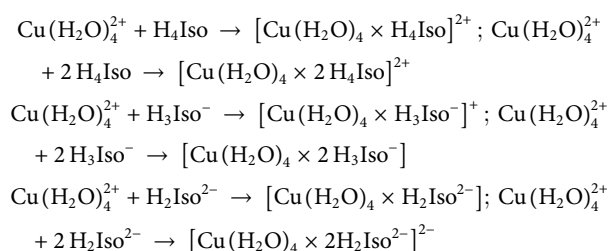


TABLE 5 Gibbs formation energies (ΔG_f), their differences ($\Delta\Delta G_f$), both in kcal/mol, and kinetic constants (K_f , $\sum K_f$, K_f^{II} , and K_f^{app} in $\text{M}^{-1} \text{ s}^{-1}$) for the different coordination sites of isorhamnetin with Cu(II) ion in 1: 1 ratio.

Coordination site	ΔG_f	$\Delta\Delta G_f$	K_f	$\sum K_f$	K_f^{II}
H_4Iso (41.58%)					
$\text{C}_3\text{-C}_4$	1.7		5.87×10^{-2}	9.57×10^2	3.98×10^9
C_3C_4	-1.3		9.47×10^6		
C_4C_5	1.8		4.88×10^{-2}		
H_3Iso^- (56.08%)					
$\text{C}_3\text{-C}_4$	0.7		2.83×10^{-1}	1.79×10^2	1.00×10^2
C_3C_4	-3.1		1.76×10^2		
C_4C_5	-0.4		2.01×10^6		
$\text{H}_2\text{Iso}^{2-}$ (2.34%)					
$\text{C}_3\text{-C}_4$	-16.7		1.95×10^{12}	1.95×10^{12}	4.56×10^{10}
C_3C_4	-4.7		2.98×10^3		
C_4C_5	-4.2		1.21×10^3		
$K_f^{\text{app}} = 4.56 \times 10^{10}$					

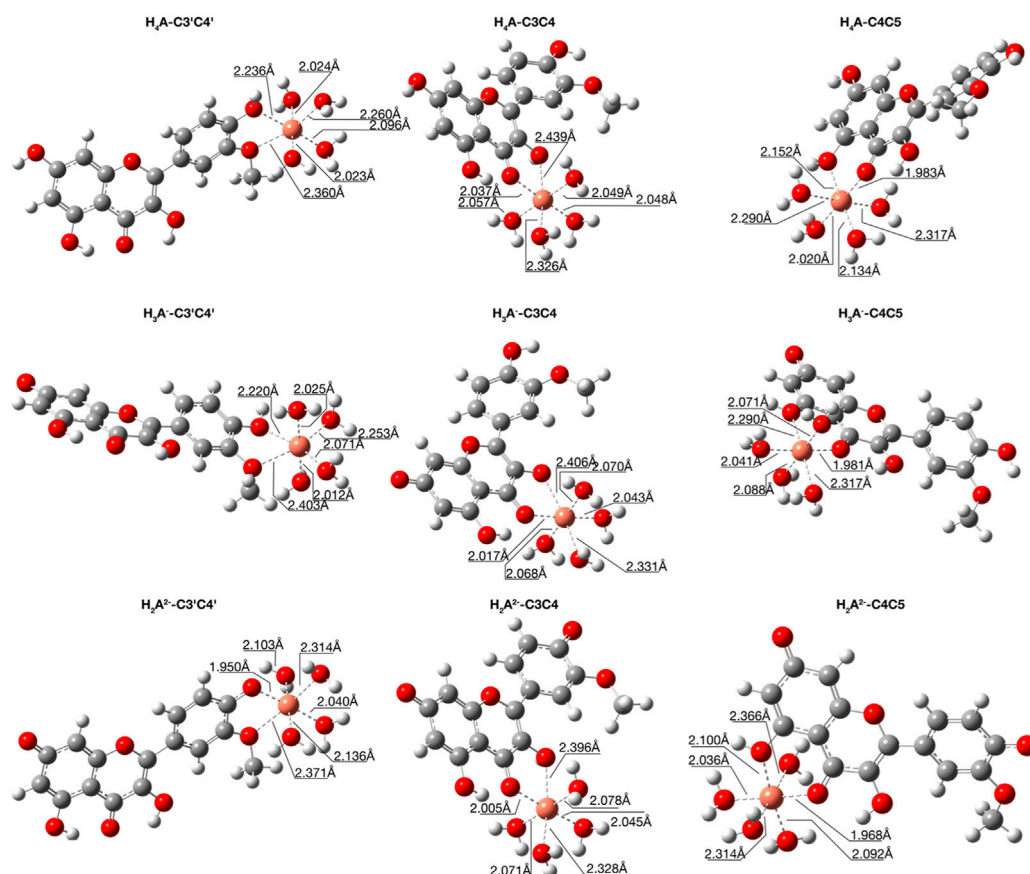


FIGURE 3

Structures for the most stable 1:1 complexes between Cu^{2+} and the neutral and charged species of isorhamnetin in water solvent.

TABLE 6 The Gibbs formation energies (ΔG_f), their differences ($\Delta\Delta G_f$), both in kcal/mol, and kinetic constants (K_f , $\sum K_f$, K_f^{II} and K_f^{aPP} in $\text{M}^{-1} \text{s}^{-1}$) for the different coordination sites of isorhamnetin with Cu (II) ion in 1:2 ratio.

Coordination site	ΔG_f	$\Delta\Delta G_f$	K_f	$\sum K_f$	K_f^{II}
H ₄ Iso (41,58%)					
C ₃ C ₄ '	1.5		8.36×10^{-2}	2.12×10^5	8.82×10^4
C ₃ C ₄	-7.3		2.12×10^5		
C ₄ C ₅	-2.1		3.54×10^4		
H ₃ Iso ⁻ (56.08%)					
C ₃ C ₄ '					
C ₃ C ₄	-12.2		9.64×10^8		
C ₄ C ₅	-9.7		1.38×10^7		
H ₂ Iso ²⁻ (2.34%)					
C ₃ C ₄ '				1.73×10^{25}	4.06×10^{23}
C ₃ C ₄	-14.9		8.96×10^{10}		
C ₄ C ₅	-9.3		7.23×10^6		
$K_f^{aPP} = 4.06 \times 10^{23}$					

made it possible to establish the Cu^{2+} chelating power of isorhamnetin. The relative apparent equilibrium constants (K^{app}) were calculated using the following expressions (Perez-Gonzalez et al., 2020):

$$K^{\text{app}} = \sum K_i^{\text{II}}$$

$$K_i^{\text{II}} = \sum K_f^{\text{II}} f$$

where K_i^{II} equals $\sum K_f$ multiplied by the molar fraction, f , of the species under consideration at pH 7.4; $\sum K_f$ is the sum of K_f for all possible complexation sites; and K_f ($K_f = e^{-\Delta G_f/RT}$) represents each reaction pathway that contributes to the chelation process.

We considered three different chelating sites, which include oxygens at the $\text{C}_3\text{-C}_4$, C_3C_4 and C_4C_5 positions. The results for a 1:1 M ratio are reported in Table 5, and the related structures are displayed in Figure 3. For neutral and mono-anionic isorhamnetin, the most stable complex is the one with copper being coordinated at the $\text{C}_3\text{-C}_4$ site, while for $\text{H}_2\text{Iso}^{2-}$ the preferred one is C_3C_4 . In all cases, the thermodynamic stability of the complexes increases with the ionic character of the ligand.

The computed $\sum K_f$ reflects the obtained ΔG_f , with that for Iso^{2-} being the greatest among all the considered species. Taking into account the molar fraction of the ligand at physiological pH, the apparent equilibrium constant becomes equal $4.56 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$, implying that all species present in the water solvent equilibrium must be taken into account to obtain a reliable outcomes. Comparison with other systems treated at the same level of theory reveals that the copper chelating power of isorhamnetin is lower than that of scutellarin and scutellarein, as the K_f^{app} values of these systems are 4.77×10^{20} and 1.29×10^{12} , respectively (Spiegel et al., 2022b).

From Figure 3, where the optimized structures and main geometrical parameters are reported, we note the distorted tetrahedral topology around the Cu^{2+} ion, which is coordinated with two ligand oxygens and two water molecules with Cu-O distances that range from about 1.9 Å to 2.1 Å. The other two water molecules act as a micro-solvation sphere.

We also explored what happens when the Cu^{2+} ion interacts with isorhamnetin at molar ratios of 1:2. The results are summarized in Table 6. For all considered ligand species, the preferred coordination site is C_3C_4 , and the most stable complex is the one in which the copper cation is coordinated by two H_2A^{2-} forms ($\Delta G_f = -14.9 \text{ kcal/mol}$). From Figure 4, it can be proven that the structural topologies are different: the coordination with two charged ligands results in a butterfly-like structure, with the coordinated antioxidant oxygens distances that are different. As in the case of the 1:1 ratio, three coordinate H_2O strongly interact with the copper center.

Comparing the K_f^{app} value obtained for the 2:1 chelates with that computed for those at a 1:1 M ratio, it can be seen that the formers are associated with higher complexation power. Given previous studies that considered a 1:1 complexes of antioxidants with copper, we note that the chelating power of Iso ($K_f^{\text{app}} = 4.56 \times 10^{10}$) is lower than that of scutellarin and scutellarein, which show K_f^{app} values of 4.77×10^{20} and $1.29 \times 10^{12} \text{ M}^{-1} \text{ s}^{-1}$, respectively (Thong et al., 2019).

In order to explore the spectroscopic changes upon Cu^{2+} complexation, we calculated the excitation energies for both the isolated isorhamnetin and its complexes with copper (see Supplementary Table S1). The lowest energy transition (S_1) for the bare molecule, characterized by high oscillator strength and being of HOMO- > LUMO in nature (more than 90%), undergo a sensible bathochromic shift starting from H_4Iso (325.4 nm), to

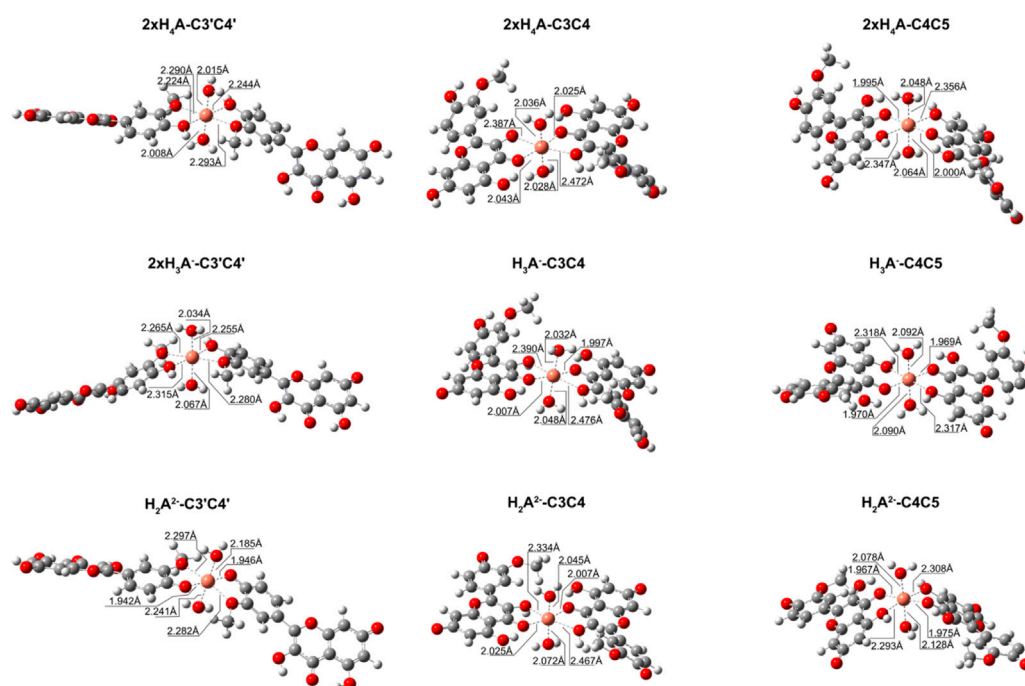


FIGURE 4

Structures for the most stable 1:2 complexes between Cu^{2+} and the neutral and charged species of isorhamnetin.

TABLE 7 The standard enthalpy ($\Delta_r H^\circ$, kcal/mol), Gibbs free energy of reaction ($\Delta_r G^\circ$, kcal/mol), reorganization energy (λ , kcal/mol), Gibbs free energy of activation ($\Delta_r G^\ddagger$, kcal/mol), diffusion rate constant (k_D , $M^{-1} s^{-1}$), TST thermal rate constant (k_T , $M^{-1} s^{-1}$), diffusion-corrected apparent rate constant (k_{app} , $M^{-1} s^{-1}$) calculated at 298.15 K for the redox reaction between the copper complexes and two reducing agents ($O_2^{\bullet -}$ and Asc $^-$) in water.

Species	Position	$\Delta_r H^\circ$	$\Delta_r G^\circ$	λ	$\Delta_r G^\ddagger$	k_D	k_{app}
$[isorhamnetin \bullet Cu(H_2O)_2]^{2+} + O_2^{\bullet -} \rightarrow [isorhamnetin \bullet Cu(H_2O)_2]^+ + O_2$							
		41.9	36.5	34.1	36.5	7.70×10^9	1.01×10^{-14}
H ₄ Iso	C ₃ C _{4'}	31.5	29.9	34.2	30.0	8.31×10^9	6.02×10^{-10}
	C ₃ C ₄	37.6	35.4	35.0	35.4	8.35×10^9	6.93×10^{-14}
	C ₄ C ₅	37.4	34.3	32.4	34.4	8.37×10^9	4.27×10^{-13}
H ₃ Iso $^-$	C ₃ C _{4'}	32.3	30.0	35.4	30.2	8.39×10^9	4.47×10^{-10}
	C ₃ C ₄	38.5	34.8	37.3	34.9	8.37×10^9	1.77×10^{-13}
	C ₄ C ₅	39.6	35.5	37.8	35.5	8.26×10^9	5.54×10^{-14}
H ₂ Iso $^{2-}$	C ₃ C _{4'}	42.3	36.9	41.3	37.0	8.41×10^9	4.54×10^{-15}
	C ₃ C ₄	40.4	35.9	37.6	36.0	8.36×10^9	2.89×10^{-14}
	C ₄ C ₅	37.8	39.1	32.6	39.5	8.56×10^9	7.80×10^{-17}
		29.3	23.0	29.7	23.41	7.43×10^9	4.45×10^{-5}
H ₄ Iso	C ₃ C _{4'}	18.9	16.4	29.8	17.9	7.50×10^9	4.64×10^{-1}
	C ₃ C ₄	25.0	21.9	30.6	22.5	7.52×10^9	1.93×10^{-4}
	C ₄ C ₅	24.9	20.9	28.1	21.3	7.52×10^9	1.36×10^{-3}
H ₃ Iso $^-$	C ₃ C _{4'}	19.7	16.6	31.0	18.2	7.53×10^9	2.58×10^{-1}
	C ₃ C ₄	25.9	21.4	33.0	22.4	7.52×10^9	2.34×10^{-4}
	C ₄ C ₅	27.0	22.0	33.4	23.0	7.49×10^9	8.75×10^{-5}
H ₂ Iso $^{2-}$	C ₃ C _{4'}	19.7	16.6	43.8	20.8	7.53×10^9	3.38×10^{-3}
	C ₃ C ₄	25.9	21.4	34.3	22.6	7.52×10^9	1.65×10^{-4}
	C ₄ C ₅	27.0	22.0	31.9	22.8	7.59×10^9	1.27×10^{-4}

H₃Iso $^-$ (343.2 nm) and H₂Iso $^{2-}$ (373.6 nm). After Cu $^{2+}$ complexation, the peak that undergoes the major shift at higher wavelengths is related to the C₅C₆ coordination site in all the species and different molar ratios. In particular, the highest wavelength (410 nm) is observed for the C₅C₆ coordination site with the isorhamnetin mono-anion in the 1:1 metal-ligand molar ratio. The excited energies in the complexes continue to be characterized by HOMO- > LUMO transitions.

3.5 Reduction mechanism with O $_2^{\bullet -}$ and Asc $^-$

The pro-oxidant activity can be theoretically evaluated by studying the Cu(II)- > Cu(I) reduction reaction of the given metal-antioxidant complex (Fenton's reaction) that leads to the production of $\bullet OH$ radicals. For this purpose, the reduction reaction of the studied complexes (in molar ratios of 1:1 and 1:2) with the reducing agents present in the physiological environment, such as superoxide (O $_2^{\bullet -}$) and ascorbic acid (Asc $^-$), were studied. Thermodynamic and kinetic results are reported in [Table 7](#) (1:1 M ratio complexes) and [Supplementary Table S2](#) (1:2 M ratio complexes), along with the

results of an analogous reactions but with a solvated copper ion only, which were taken as a reference.

In all the studied systems, the reaction with the superoxide is more exergonic than the corresponding one with the ascorbic acid anion, as also noted previously for the pyridoxal antioxidant ([Ngo et al., 2022](#)).

Regarding the reaction with O $_2^{\bullet -}$, we found its feasibility to be greatest among neutral and mono-anionic forms of isorhamnetin with the C₃C_{4'}, and then the C₄C₅ coordination sites occupied. The relative kinetic constants indicate that the reduction reaction of these kind of complexes favors the reduction of the copper ion to a lower oxidation state.

Turning to the study of what happens when the Asc anion is used to reduce the copper ion in the Fenton's reaction, we first note that the Gibbs reaction energies are much less exothermic than for the reactions with the superoxide ion. Indeed, they fall in the range of 16.4–22.0 kcal/mol for the complexes with molar ratio 1:1 and 15.3–29.0 kcal/mol (excluding the value of 53.8 kcal/mol found for the C₃C_{4'} coordination site) for the complexes with a stoichiometry of 1:2.

From the values of the kinetic constants, it can be seen that the formation of complexes promotes the process of copper reduction

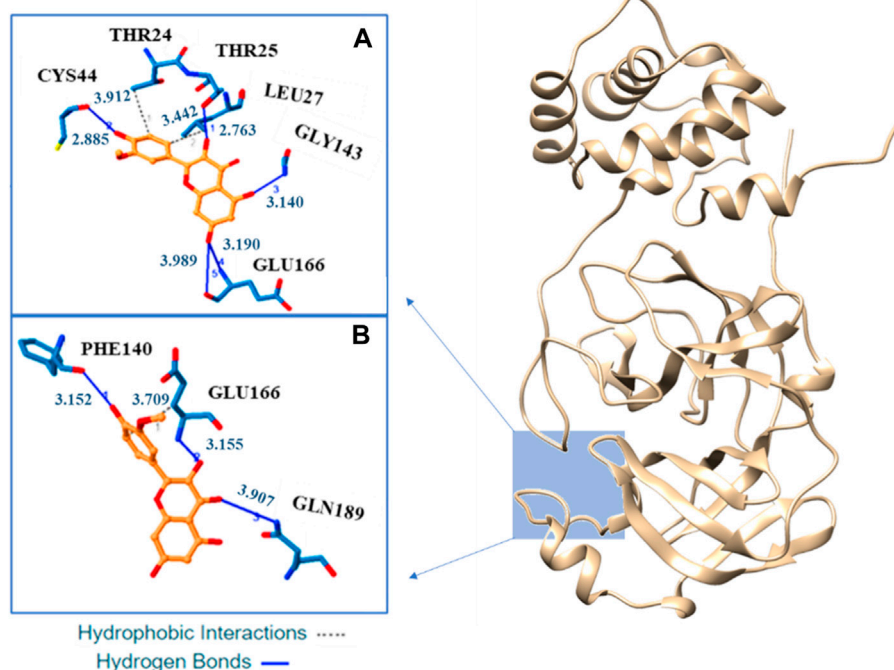


FIGURE 5

Three-dimensional representation of the best docking pose for H_4Iso (A) and H_3Iso^- (B) species. The residues that have crucial contacts with the compounds are shown in the square windows (A,B), for H_4Iso and H_3Iso respectively.

towards lower oxidation states. In general, we can support the fact that the presence of ascorbic acid causes a greater pro-oxidant hazard than the reaction with the superoxide radical does, as previously noted for other antioxidants (Ngo et al., 2022).

3.6 Inhibitory activity

Flavonoids have been proven to be effective inhibitors of several enzymes involved in several biological and medical processes. In particular, isorhamnetin has been proposed as a promising inhibitor of the cyclooxygenase-2 (Seo et al., 2014), lactate dehydrogenase, adenosine diphosphate and other enzymes. Recently, some flavonoids have been proposed as non-covalent inhibitors of the main protease (M^{pro}) protein which plays an important role in SARS-CoV-2 main protease enzyme infection (Abian et al., 2020; Puttaswamy et al., 2020; Rizzuti et al., 2021).

M^{pro} explains its action in the cleavage of polyproteins at many sites generating non-structural proteins, relevant in the replication process of the virus (Anand et al., 2003), such as endo- and exoribonuclease as well as RNA polymerase. So, it is an important target for the development of new anticoronavirus therapeutic agents (Yang et al., 2005; Pillaiyar et al., 2016). From a structural point of view, SARS-CoV-2 M^{pro} is an homodimer and each protomer is characterized by three domains connected by a loop region. The catalytic center contain two crucial residues (His41 and Cys145) and catalyze the cleaving the of the polyprotein, translated from the viral RNA at different positions, generating proteins that

contribute in the arresting process of the viral replication cycle (Ullrich and Nitsche, 2020).

Since the global health emergency generated by coronavirus disease 2019 is still in progress, we considered interesting to verify whether the isorhamnetin, coming from plants spread in various continents, can be a substrate capable to inhibit the M^{pro} .

Our docking analysis shows as isorhamnetin and its mono-anionic form establish many hydrophobic interactions (HI) and hydrogen bonds (H-bond) with different aminoacids in the enzyme catalytic pocket (see Figure 5 and Supplementary Table S3). For H_4Iso two HI are established with the Thr24 and Leu27 while for H_3Iso^- only the interaction with Glu166 is present. Five H-bonds, with Cys44, Thr25, Gly143, and Glu166, are present in the binding of isorhamnetin with M^{pro} . In the case of H_3Iso^- they involve the Phe140, Glu166, and Gln189. In both the systems their lengths range from 2.91 to 3.99 Å (see Figure 5).

Albeit with different amino acid residues, the binding energy values for both the neutral and anionic forms (Supplementary Table S3) result to be very close. The found value for the best pose (-6.1 kcal/mol) well agrees with that proposed for quercetin throughout isothermal titration calorimetry that result to be -7.6 kcal/mol (Abian et al., 2020) and with the computed binding energy between eugenol and M^{pro} (Rizzuti et al., 2021). Furthermore, our computed binding energy is slightly higher than that obtained for Ginkgetin (-9.5 kcal/mol), Delphinidin (-9.4 kcal/mol), Cyanidin 3,5-diglucoside, (-9.4 kcal/mol) and Amentofavone (-9.7 kcal/mol) antioxidants with similar structures (Puttaswamy et al., 2020). Finally, a comparison can be made also with the binding energies of quercetin and its anion with furin protein that

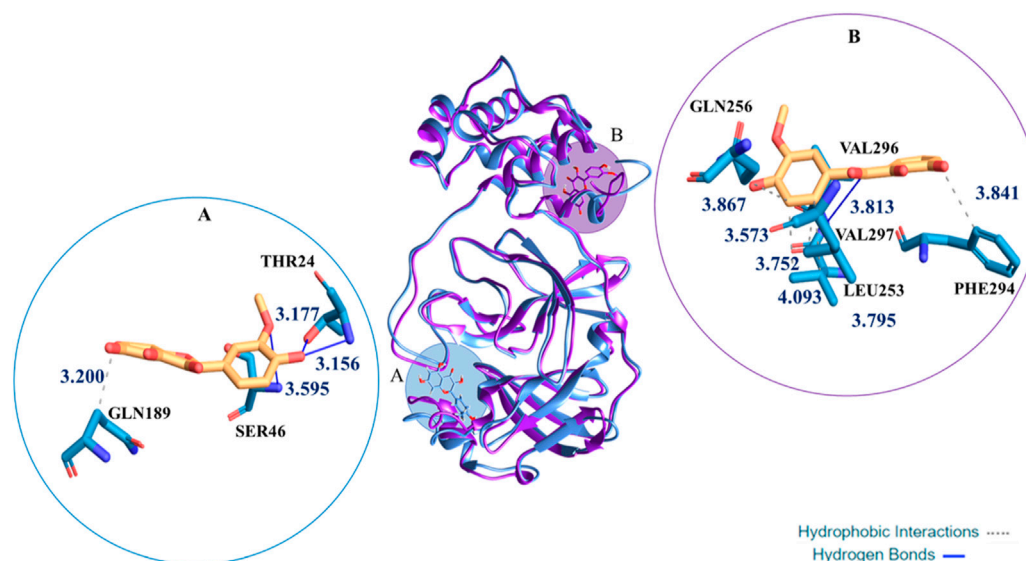


FIGURE 6

Superimposition of the most representative cluster from 100 ns of MD for the $M^{pro}:H_4Iso$ and $M^{pro}:H_3Iso^-$. In round windows complexes three-dimensional representation of the clustered structures with the residues that have crucial contacts with the compounds, in (A) (catalytic site) the H_4Iso , in (B) (site #3) the H_3Iso^- one.

result to be -7.8 and -7.7 kcal/mol, respectively (Milanovic et al., 2021).

The molecular docking examination induced us to perform MD simulations with the aim to observe the dynamic behaviour of the complexes $M^{pro}:H_4Iso$ and $M^{pro}:H_3Iso$. The analysis of their MD trajectories indicated a different behaviour of the two tested molecules, despite starting from the docked pose where both the ligands were located in proximity of catalytic site. During the simulation time in correspondence of ~ 20 ns, the anionic form goes to a distal area from the catalytic pocket generating a different RMSD trend (Supplementary Figure S3) from that obtained in the case of $M^{pro}:H_4Iso$. In particular, the H_3Iso is placed in a site defined by Ile213, Pro252-Leu253, Gln256, Val296-Val297, Cys300-Gly302 residues. The Supplementary Figure S4 related to RMSF value evidenced a different fluctuation in the related region. This site corresponds to the named site #3, one of the six allosteric sites experimentally proposed (Douangamath et al., 2020; El BabaLutowski et al., 2020; Günther et al., 2021) and then *in silico* observed (Alzyoud et al., 2022).

In fact, the M^{pro} enzyme includes several pockets on its surface believed important for its catalytic activity; some of them exist in distal areas from the main catalytic pocket. The site #3 located at the dimer interface showed poor druggability due to its very small and shallow cavity that is significantly less hydrophilic. The protease movement during the MD of $M^{pro}:H_3Iso^-$ makes more exposed such distal sites, making them more accessible. This does not take place during the MD of $M^{pro}:H_4Iso$ as it can be evinced from a Figure 6 where the superposition of the most representative structure of the complexes $M^{pro}:H_4Iso$ and $M^{pro}:H_3Iso$ is reported. Furthermore the ligand affinity to the M^{pro} , evaluated in terms of ΔG binding, calculated *via* MMPBSA method proposed also the neutral form, H_4Iso , as the most thermodynamically favorable. (see Supplementary Table S5) This outcome better

clarifies the different nature of the interactions generated by the two ligands in the two different sites.

The possible coexistence of the two forms (neutral and anionic), at physiological pH, could be helpful to enhance the antiviral response of the isorhamnetin molecule providing provide a good (natural) starting point for lead optimization chemistry to deactivate the SARS CoV-2 M^{pro} .

4 Conclusion

In the present work, we have shown, by examining the isorhamnetin molecule, how quantum mechanics methods based on density functional theory can provide useful and reliable information on geometric and electronic structures, chemical equilibria in solution, and reaction mechanisms of antioxidant systems. The results show that.

- isorhamnetin in aqueous solution and at physiological pH (7.4) exists in neutral, mono- and di-anionic forms (molar fractions equal 41.58%, 56.08%, and 2.34%, respectively) rendering them important to consider in calculations aiming to provide reliable kinetics of the reactivity towards $^{\bullet}OOH$ radical;
- it is important to consider different molar ratios (1:1, 1:2) to determine the chelating capacity towards copper ions;
- the reduction process of the obtained complexes with the reducing agents present in the physiological environment (superoxide anion and ascorbic acid anion) are essential for predicting their chemical behavior in Fenton's reactions;
- isorhamnetin has good non-covalent inhibitory potency towards M^{pro} , which is a pharmacological target for the treatment of SARS-CoV-2 infection.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.rcsb.org/> - PDB code 6W63.

Author contributions

NR conceived of the presented idea and supervised the project. MS and GC performed the computations. NR, TM, and MS verified and analysed the results. NR wrote the manuscript. GC, MS, TM, and NR contributed to the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

The authors thank the Dipartimento di Chimica e Tecnologie Chimiche of Università della Calabria for their financial support. The computing resources and the related technical support used for this work have been provided by: the supercomputer Marconi 100 from ISCRA (project code: IsC92) for computer simulations and by resources provided by Poznan Supercomputing and Networking Center (grant number 467).

References

- Abian, O., Ortega-Alarcon, D., Jimenez-Alesanco, A., Ceballos-Laita, L., Vega, S., Reyburn, H. T., et al. (2020). Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *Int. J. Biol. Macromol.* 164, 1693–1703. doi:10.1016/j.jbiomac.2020.07.235
- Alberto, M. E., Grand, A., Russo, N., and Galano, A. (2013). A physicochemical examination of the free radical scavenging activity of Trolox: Mechanism, kinetics and influence of the environment. *Phys. Chem. Chem. Phys.* 15, 4642. doi:10.1039/c3cp43319f
- Alzayoud, L., Ghattas, M. A., and Atatreh, N. (2022). Allosteric binding sites of the SARS-CoV-2 main protease: Potential targets for broad-spectrum anti-coronavirus agents. *Drug Des. Dev. Ther.* 16, 2463–2478. doi:10.2147/dddt.s370574
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: Basis for design of anti-SARS drugs. *Science* 300, 1763–1767. doi:10.1126/science.1085658
- Anandkrishnan, R., Aguilar, B., and Onufriev, A. V. (2012). H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, 537–541. doi:10.1093/nar/gks375
- Apak, R., Özyürek, M., Kubilay, G., and Çapanoglu, E. (2016). Antioxidant activity/capacity measurement. I. Classification, physicochemical principles, mechanisms, and electron transfer (ET)-Based assays. *J. Agric. Food Chem.* 64, 997–1027. doi:10.1021/acs.jafc.5b04739
- Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* 97, 10269–10280. doi:10.1021/j100142a004
- Belcastro, M., Marino, T., Russo, N., and Toscano, M. (2006). Structural and electronic characterization of antioxidants from marine organisms. *Theor. Chem. Acc.* 115, 361–369. doi:10.1007/s00214-006-0077-5
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1998). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Boulebd, H., and Khodja, I. A. (2021). A detailed DFT-based study of the free radical scavenging activity and mechanism of daphnetin in physiological environments. *Phytochemistry* 189, 112831. doi:10.1016/j.phytochem.2021.112831
- Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., et al. (2017). *AMBER 2017*. San Francisco: University of California.
- Castaneda-Arriaga, R., Marino, T., Russo, N., Alvarez-Idaboy, J. R., and Galano, A. (2020). Chalcogen effects on the primary antioxidant activity of chrysin and quercetin. *New J. Chem.* 44, 9073–9082. doi:10.1039/d0nj01795g
- Collins, F. C., and Kimball, G. E. (1949). Diffusion-controlled reaction rates. *J. Colloid. Sci.* 4, 425–437. doi:10.1016/0095-8522(49)90023-9
- Darden, T., York, D., and Pedersen, L. (1998). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi:10.1063/1.464397
- Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., et al. (2020). Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* 11 (1), 5047. doi:10.1038/s41467-020-18709-w
- El-Baba, T. J. A., Lutowski, C., Kantsadi, A. L., Malla, T. R., John, T., Mikhailov, V., et al. (2020). Allosteric inhibition of the SARS-CoV-2 main protease: Insights from mass spectrometry based assays. *Angew. Chem. Int. Ed.* 59, 23544–23548. doi:10.1002/anie.202010316
- Evans, M. G., and Polanyi, M. (1935). Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* 31, 875–882. doi:10.1039/tf9353100875
- Forman, H. J., and Zhang, H. (2021). Targeting oxidative stress in disease: Promise and limitations of antioxidant therapy. *Nat. Rev.* 20, 689–709. doi:10.1038/s41573-021-00233-1
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2014). *Gaussian 09, revision D.01*.
- Galano, A., Perez-Gonzalez, A., Castaneda-Arriaga, R., Munoz-Rugeles, L., Mendoza-Sarmiento, G., Romero-Silva, A., et al. (2016b). Empirically fitted parameters for calculating pKa values with small deviations from experiments using a simple computational strategy. *J. Chem. Inf. Model.* 56, 1714–1724. doi:10.1021/acs.jcim.6b00310
- Galano, A., and Alvarez-Idaboy, J. R. (2013). A computational methodology for accurate predictions of rate constants in solution: Application to the assessment of primary antioxidant activity. *J. Comput. Chem.* 34, 2430–2445. doi:10.1002/jcc.23409
- Galano, A., Mazzone, G., Alvarez-Diduk, R., Marino, T., Alvarez-Idaboy, J. R., and Russo, N. (2016a). Food antioxidants: Chemical insights at the molecular level. *Annu. Rev. Food Sci. Technol.* 7, 335–352. doi:10.1146/annurev-food-041715-033206
- Gong, G., Guan, Y.-Y., Zhang, Z.-L., Rahman, K., Wang, S.-J., Zhou, S., et al. (2020). Isorhamnetin: A review of pharmacological effects. *Biomed. Pharmacother.* 128, 110301. doi:10.1016/j.biopha.2020.110301
- Günther, S., Reinke, P. Y. A., Fernández-García, Y., Lieske, J., Lane, T. J., Ginn, H. M., et al. (2021). X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* 7 (6542), 372642–372646. doi:10.1126/science.abf7945
- Leopoldini, M., Russo, N., and Toscano, M. (2011). The molecular basis of working mechanism of natural polyphenolic antioxidants. *Food Chem.* 125, 288–306. doi:10.1016/j.foodchem.2010.08.012
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2021). ff14SB, Available at: <https://pubs.acs.org/doi/full/10.1021/acs.jctc.5b00255>, (accessed 13 January 2021).
- Manach, C., Morand, C., Demigne, C., Texier, O., Regerat, F., and Remesy, C. (1997). Bioavailability of rutin and quercetin in rats. *FEBS Lett.* 409, 12–16. doi:10.1016/s0014-5793(97)00467-5

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2023.1122880/full#supplementary-material>

- Marcus, R. A. (1993). Electron transfer reactions in chemistry: Theory and experiment (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* 32, 1111–1121. doi:10.1002/anie.199311113
- Marenich, A. V., Cramer, C. J., and Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 113, 6378–6396. doi:10.1021/jp810292n
- Marković, Z., Tošović, J., Milenković, D., and Marković, S. (2016). Revisiting the solvation enthalpies and free energies of the proton and electron in various solvents. *Comput. Theor. Chem.* 1077, 11–17. doi:10.1016/j.comptc.2015.09.007
- Milanovic, Z. B., Antonijević, M. R., Amic, A. D., Avdović, E. H., Dimic, D. S., Milenkovic, D. A., et al. (2021). Inhibitory activity of quercetin, its metabolite, and standard antiviral drugs towards enzymes essential for SARS-CoV-2: The role of acid–base equilibria. *RSC Adv.* 11, 2838–2847. doi:10.1039/d0ra09632f
- Milenković, D. A., Dimić, D. S., Avdović, E. H., Amić, A. D., Dimitrić Marković, J. M., and Marković, Z. S. (2020). Advanced oxidation process of coumarins by hydroxyl radical: Towards the new mechanism leading to less toxic products. *Chem. Eng. J.* 395, 124971. doi:10.1016/j.cej.2020.124971
- Ngo, T. C., Nguyen, T. H., and Dao, D. Q. (2019). Radical scavenging activity of natural-based cassaine diterpenoid amides and amines. *J. Chem. Inf. Model.* 59, 766–776. doi:10.1021/acs.jcim.8b00847
- Ngo, T. C., Truong, D. H., Nguyen, T. T. N., Quang, D. T., and Dao, D. Q. (2022). On the free radical scavenging and metallic ion chelating activities of pyridoxal - could the pro-oxidant risk be competitive? *Phytochemistry* 192, 113176. doi:10.1016/j.phytochem.2022.113176
- Parise, A., RomeoRusso, I. N., and Marino, T. (2021b). The Se-S bond formation in the covalent inhibition mechanism of SARS-CoV-2 main protease by ebselen-like inhibitors: A computational study. *Int. J. Mol. Sci.* 22 (18), 9792. doi:10.3390/ijms22189792
- Parise, A., De Simone, B. C., Marino, T., Toscano, M., and Russo, N. (2021a). Quantum mechanical predictions of the antioxidant capability of moracin C isomers. *Front. Chem.* 9, 666647. doi:10.3389/fchem.2021.666647
- Pengfei, L., Tiansheng, D., Xianglin, H., and Jianguo, W. (2009). Antioxidant properties of isolated isorhamnetin from the sea buckthorn marc. *Plant Foods Hum. Nutr.* 64, 141–145. doi:10.1007/s11130-009-0116-1
- Pérez-González, A., Prejanò, M., Russo, N., Marino, T., and Galano, A. (2020). Capsaicin, a powerful •OH-inactivating ligand. *Antioxidants* 9, 1247. doi:10.3390/antiox9121247
- Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., and Jung, S.-H. (2016). An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* 59, 6595–6628. doi:10.1021/acs.jmedchem.5b01461
- Puttaswamy, H., Gowtham, H. G., Ojha, M. D., Yadav, A., Choudhri, G., Raguraman, V., et al. (2020). *In silico* studies evidenced the role of structurally diverse plant secondary metabolites in reducing SARS-CoV-2 pathogenesis. *Sci. Rep.* 10, 20584. doi:10.1038/s41598-020-77602-0
- Reina, M., Guzman-Lopez, E. G., Romeo, I., Marino, T., Russo, N., and Galano, A. (2021). Computationally designed p-coumaric acid analogs: Searching for neuroprotective antioxidants. *New J. Chem.* 45, 14369–14380. doi:10.1039/d1nj01235e
- Rizzuti, B., Ceballos-Laita, L., Ortega-Alarcon, D., Jimenez-Alesanco, A., Vega, S., Grande, F., et al. (2021). Sub-micromolar inhibition of SARS-CoV-2 3CLpro by natural compounds. *Pharmaceuticals* 14, 892. doi:10.3390/ph14090892
- Seo, K., Yang, J. H., Kim, S. C., Ku, S. K., Ki, S. H., and Shin, S. M. (2014). The antioxidant effects of isorhamnetin contribute to inhibit COX-2 expression in response to inflammation: A potential role of HO-1. *Inflammation* 37, 712–722. doi:10.1007/s10753-013-9789-6
- Sharma, A. K., Pavlova, S. T., Kim, J., Kim, J., and Mirica, L. M. (2013). The effect of Cu²⁺ and Zn²⁺ on the Aβ42 peptide aggregation and cellular toxicity. *Metallomics* 5, 1529–1536. doi:10.1039/c3mt00161j
- Sies, H., Berndt, C., and Jones, D. P. (2017). Oxidative stress. *Annu. Rev. Biochem.* 86, 715–748. doi:10.1146/annurev-biochem-061516-045037
- Sies, H. (1985). *Oxidative stress*, 1–8. Cambridge: Academic Press.
- Spiegel, M. (2022). Current trends in computational quantum chemistry studies on antioxidant radical scavenging activity. *J. Chem. Inf. Model.* 62 (11), 2639–2658. doi:10.1021/acs.jcim.2c00104
- Spiegel, M., Gamian, A., and Sroka, Z. (2021). Antiradical activity of beetroot (*Beta vulgaris* L.) betalains. *Molecules* 26, 2439. doi:10.3390/molecules26092439
- Spiegel, M., Marino, T., Prejanò, M., and Russo, N. (2022a). Antioxidant and copper-chelating power of new molecules suggested as multiple target agents against Alzheimer's disease. A theoretical comparative study. *Phys. Chem. Chem. Phys.* 24, 16353–16359. doi:10.1039/d2cp01918c
- Spiegel, M., Marino, T., Prejanò, M., and Russo, N. (2022b). Primary and secondary antioxidant properties of scutellarin and scutellarein in water and lipid-like environments: A theoretical investigation. *J. Mol. Liq.* 366, 120343. doi:10.1016/j.molliq.2022.120343
- Teng, D., and Luan, X. (2016). Research progress of isorhamnetin in pharmacodynamics. *Clin. J. Tradit. Chin. Med.* 28, 593–596.
- Thong, N. M., Vo, Q. V., Huyen, T. L., Bay, M. V., Tuan, D., and Cam Nam, P. (2019). Theoretical study for exploring the diglycoside substituent effect on the antioxidative capability of isorhamnetin extracted from *Annoecochilus roxburghii*. *ACS Omega* 4 (12), 14996–15003. doi:10.1021/acsomega.9b01780
- Trott, O., and Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Truhlar, D. G., Garrett, B. C., and Klippenstein, S. J. (1996). Current status of transition-state theory. *J. Phys. Chem.* 100, 12771–12800. doi:10.1021/jp953748q
- Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30, 127377. doi:10.1016/j.bmcl.2020.127377
- Xu, Z., Yang, L., Zhang, X., Zhang, Q., Yang, Z., Liu, Y., et al. (2020). Discovery of potential flavonoid inhibitors against COVID-19 3CL proteinase based on virtual screening strategy. *Front. Mol. Biosci.* 7, 556481. doi:10.3389/fmolb.2020.556481
- Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., et al. (2005). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* 3, e324. doi:10.1371/journal.pbio.0030324
- Yu, M.-S., Lee, J., Lee, J. M., Kim, Y., Chin, Y. W., Jee, J.-G., et al. (2012). Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP3. *Bioorg. Med. Chem. Lett.* 22, 4049–4054. doi:10.1016/j.bmcl.2012.04.081
- Zhao, Y., Schultz, N. E., and Truhlar, D. G. (2006). Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* 2, 364–382. doi:10.1021/ct0502763



OPEN ACCESS

EDITED BY

Renjith Thomas,
Mahatma Gandhi University, India

REVIEWED BY

Ernesto Chigo,
Meritorious Autonomous University of
Puebla, Mexico
Jose Luis Cabellos,
Polytechnic University of Tapachula,
Mexico
Slavko Radenković,
University of Kragujevac, Serbia

*CORRESPONDENCE

Kang-Ning Li,
✉ knli@nxu.edu.cn

SPECIALTY SECTION

This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 13 January 2023

ACCEPTED 07 February 2023

PUBLISHED 16 February 2023

CITATION

Li L-K, Ma Y-Q, Li K-N, Xie W-L and
Huang B (2023), Structural and electronic
properties of H₂, CO, CH₄, NO, and NH₃
adsorbed onto Al₁₂Si₁₂ nanocages using
density functional theory.
Front. Chem. 11:1143951.
doi: 10.3389/fchem.2023.1143951

COPYRIGHT

© 2023 Li, Ma, Li, Xie and Huang. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Structural and electronic properties of H₂, CO, CH₄, NO, and NH₃ adsorbed onto Al₁₂Si₁₂ nanocages using density functional theory

Liu-Kun Li¹, Yan-Qiu Ma¹, Kang-Ning Li^{1*}, Wen-Li Xie² and
Bin Huang³

¹Ningxia Key Laboratory of Intelligent Sensing for the Desert Information, School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan, China, ²Basic Education Department, Guangdong Ocean University, Yangjiang, China, ³Environmental Monitoring Site of Ningxia Ningdong Energy and Chemical Industry Base, Yinchuan, China

In this study, the adsorption of gases (CH₄, CO, H₂, NH₃, and NO) onto Al₁₂Si₁₂ nanocages was theoretically investigated using density functional theory. For each type of gas molecule, two different adsorption sites above the Al and Si atoms on the cluster surface were explored. We performed geometry optimization on both the pure nanocage and nanocages after gas adsorption and calculated their adsorption energies and electronic properties. The geometric structure of the complexes changed slightly following gas adsorption. We show that these adsorption processes were physical ones and observed that NO adsorbed onto Al₁₂Si₁₂ had the strongest adsorption stability. The E_g (energy band gap) value of the Al₁₂Si₁₂ nanocage was 1.38 eV, indicating that it possesses semiconductor properties. The E_g values of the complexes formed after gas adsorption were all lower than that of the pure nanocage, with the NH₃-Si complex showing the greatest decrease in E_g . Additionally, the highest occupied molecular orbital and the lowest unoccupied molecular orbital were analyzed according to Mulliken charge transfer theory. Interaction with various gases was found to remarkably decrease the E_g of the pure nanocage. The electronic properties of the nanocage were strongly affected by interaction with various gases. The E_g value of the complexes decreased due to the electron transfer between the gas molecule and the nanocage. The density of states of the gas adsorption complexes were also analyzed, and the results showed that the E_g of the complexes decreased due to changes in the 3p orbital of the Si atom. This study theoretically devised novel multifunctional nanostructures through the adsorption of various gases onto pure nanocages, and the findings indicate the promise of these structures for use in electronic devices.

KEYWORDS

nanocage, Al₁₂Si₁₂, equilibrium geometries, stability, electronic properties

1 Introduction

Since Kroto et al. (1985) discovered (C₆₀) fullerene, nanomaterials have gradually become a popular research topic due to their unique properties and wide range of applications (Goroff, 1996; Diederich and Gómez-López, 1999; Kataura et al., 2002; Martín, 2006). Instead of using only C to construct fullerenes, researchers have turned

to fullerenes made from other elements, especially elements from group III of the periodic table and the group V elements that are adjacent to C. These fullerenes include $B_{12}N_{12}$, $Al_{12}N_{12}$, $B_{12}P_{12}$, and $Al_{12}P_{12}$ (Beheshtian et al., 2012a; Beheshtian et al., 2012b; Beheshtian et al., 2012c; Rad and Ayub, 2016; Hussain et al., 2020a). In addition, researchers have used elements from groups II and VI to construct similar fullerene structures, such as $Be_{12}O_{12}$, $Mg_{12}O_{12}$, $Ca_{12}O_{12}$, and all-boron fullerene (Kakemam and Peyghan, 2013; Zhai et al., 2014; Rezaei-Sameti and Abdoli, 2020; Ahsan et al., 2021). Other fullerenes are made from transition elements and oxygen element (de Oliveira et al., 2015). Nanocages with the general formula $(XY)_n$, where n is the number of atoms, are more popular among researchers because they are more stable (Strout, 2000). Many types of nanostructures exist, among which inorganic nanostructures have attracted the attention of researchers due to their extremely high stability and asymmetric charge distribution (Dai, 2002; Tasis et al., 2006; Bindhu et al., 2016). The study of fullerene structures is a critical branch of nanotechnology because of their applications in electronic devices, special materials, and environmental processes. Shakerzadeh et al. (2015) studied the properties of alkali metals (Li, Na, and K) interacting with $Be_{12}O_{12}$ and $Mg_{12}O_{12}$ nanoclusters. Doping with alkali metals can significantly improve the non-linear optical response of nanocages. Beheshtian et al. (2012c) have studied the adsorption of NO and CO by $Al_{12}N_{12}$. Due to the different changes in the E_g (energy band gap) of NO and CO when adsorbed, resulting in different changes in electrical conductivity, $Al_{12}N_{12}$ clusters may selectively detect NO molecules when CO molecules are present. Vergara-Reyes et al. (2021) studied the adsorption of NO on the $C_{36}N_{24}$ fullerene, providing a possible carrier/protector of nitric oxide and thus fulfill its correct biological functions. Escobedo et al. (2019) conducted the effect of chemical order on the structural and physicochemical properties of $B_{12}N_{12}$ fullerene. Bautista Hernandez et al. (2022) found that the $(TiO_2)_{19}$ cluster is a good candidate for storing various gases, and can also be used as a hydrogen storage device.

For China in particular and the rest of the world in general, fossil fuel use must be scaled back, and hydrogen, which has the highest calorific value per unit mass of fuels and produces only water during combustion, is an attractive alternative (Yilanci et al., 2009; Züttel et al., 2010; Mazloomi and Gomes, 2012). Technologies for capturing pollutants and greenhouse gases from the atmosphere will also play a key role in our fight against climate change (Cazorla-Amorós et al., 1996; Zhang et al., 2008; Mastalerz et al., 2011).

Oku et al. (2004) synthesized the first inorganic nanocages in 2004 before the later emergence of $B_{12}N_{12}$, $Al_{12}N_{12}$, $B_{12}P_{12}$, and $Al_{12}P_{12}$ nanocages. Yarovsky and Goldberg, (2005) studied the adsorption of H_2 by Al_{13} clusters using DFT (density functional theory). Thereafter, Felício-Sousa et al. (2021) examined the adsorption properties of H_2 , CO, CH_4 , and CH_3OH on Fe_{13} , Co_{12} , Ni_{13} , and Cu_{13} clusters using an *ab initio* investigation. Tan et al. (2022) studied clusters of different numbers of Al atoms. Yang et al. (2018) predicted the properties of electron redundant Si_nN_n fullerenes. Metal oxide nanocages are also a popular area of research; Otaibi et al. conducted a theoretical study on the adsorption of glycoluril by $Mg_{12}O_{12}$ nanocages (Al-

Otaibi et al., 2022). Some nanocages exhibit unique electronic properties after doping with alkali metals, and these have also been widely studied (Ahsin and Ayub, 2021). We replaced the N atoms in $Al_{12}N_{12}$ with Si atoms to investigate a more diverse range of materials than those in the literature. In this study, we adopted DFT to analyze the properties of $Al_{12}Si_{12}$, such as its stability, after the adsorption of CH_4 , CO, H_2 , NO, and NH_3 .

2 Computational methodology

An $Al_{12}Si_{12}$ nanocage was selected as the model adsorbent. Geometry optimization was performed using hybrid functional DFT (B3LYP) (Becke, 1993) implemented in Gaussian 09 (Frisch et al., 2004). B3LYP is a suitable and widely accepted functional for nanoclusters (Chen et al., 2009; Hussain et al., 2020b). The mixed basis set 6-31G (d, p) was used. Gas molecules CH_4 , CO, H_2 , NO, and NH_3 were adsorbed onto $Al_{12}Si_{12}$ nanocages using the same method. Two adsorption sites on the $Al_{12}Si_{12}$ nanocage were considered. Vibration frequencies were also calculated at equivalent levels to verify that all stationary points corresponded to true minima on the potential energy surface. Geometry optimization was conducted, and D_{ads} (distance of adsorption), E_{ads} (adsorption energy), E_{HOMO} (energy of the highest occupied molecular orbital), E_{LUMO} (energy of lowest unoccupied molecular orbital), DOS (density of states), and Q_T (Mulliken charge transfer) were calculated to determine the adsorption mechanism.

The stability of the $Al_{12}Si_{12}$ nanocage was examined in terms of E_{coh} (cohesive energy), which can be determined by calculating the average energy difference of each atom before and after bonding using the following equation:

$$E_{coh} = \frac{E_{total} - 12(E_{Al} + E_{Si})}{24}$$

where E_{Al} and E_{Si} are the energies of non-interacting Al and Si atoms, respectively, and E_{total} is the energy of the $Al_{12}Si_{12}$ nanocage.

E_{ads} is defined as follows:

$$E_{ads} = E_{adsorbate@nanocage} - E_{adsorbate} - E_{nanocage}$$

Here, $E_{adsorbate@nanocage}$, $E_{adsorbate}$, and $E_{nanocage}$ are the total energies of an adsorbate adsorbed onto the pure nanocage, of the adsorbate, and of the pure nanocage, respectively. A negative value of E_{ads} corresponds to exothermic adsorption. The more negative the adsorption value, the stronger the adsorption capacity. The DOS was generated in Multiwfn (O'Boyle et al., 2008).

3 Results and discussion

3.1 Geometrical characteristics

3.1.1 $Al_{12}Si_{12}$ structure

The optimized structures of the bare nanocage comprising eight hexagons and six tetragon rings are given in Figure 1. In this nanocage, two non-equivalent bonds exist: one between the tetragon and hexagon ring, and the other between the two hexagonal rings, represented as b_{64} and b_{66} , respectively. The

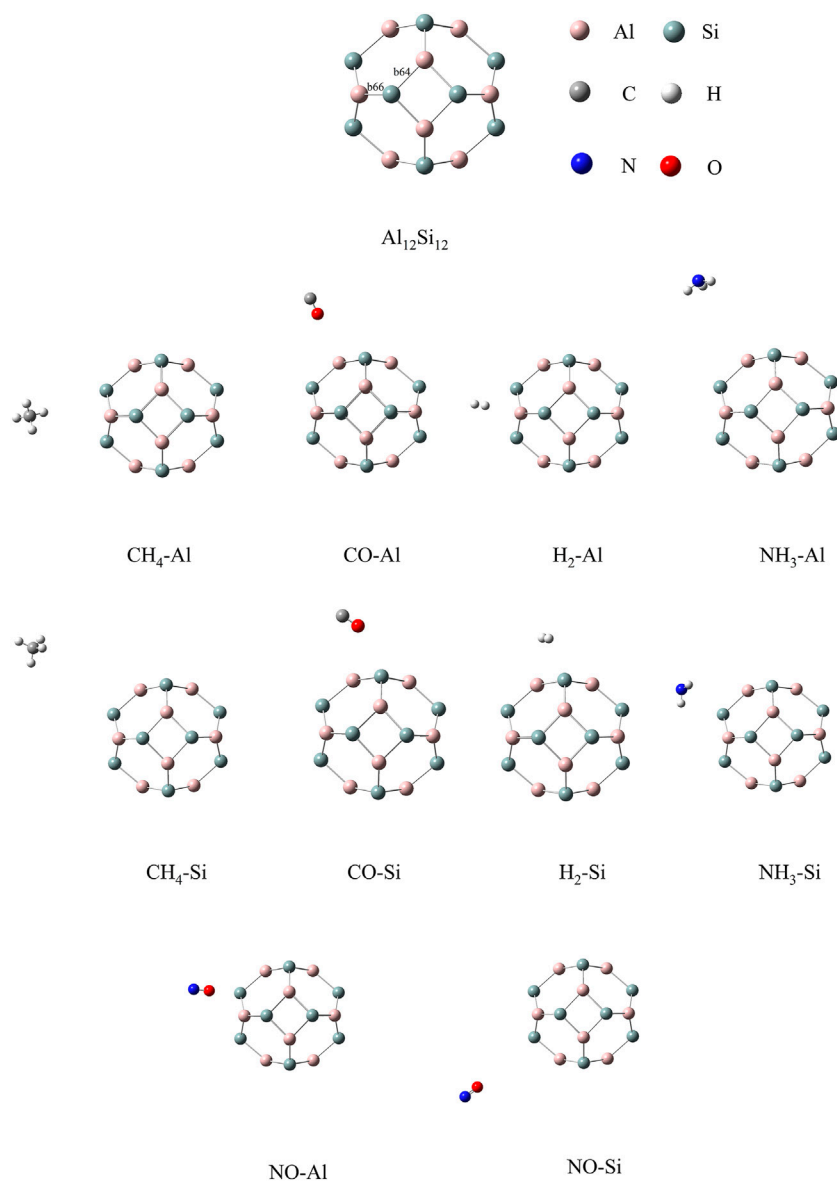


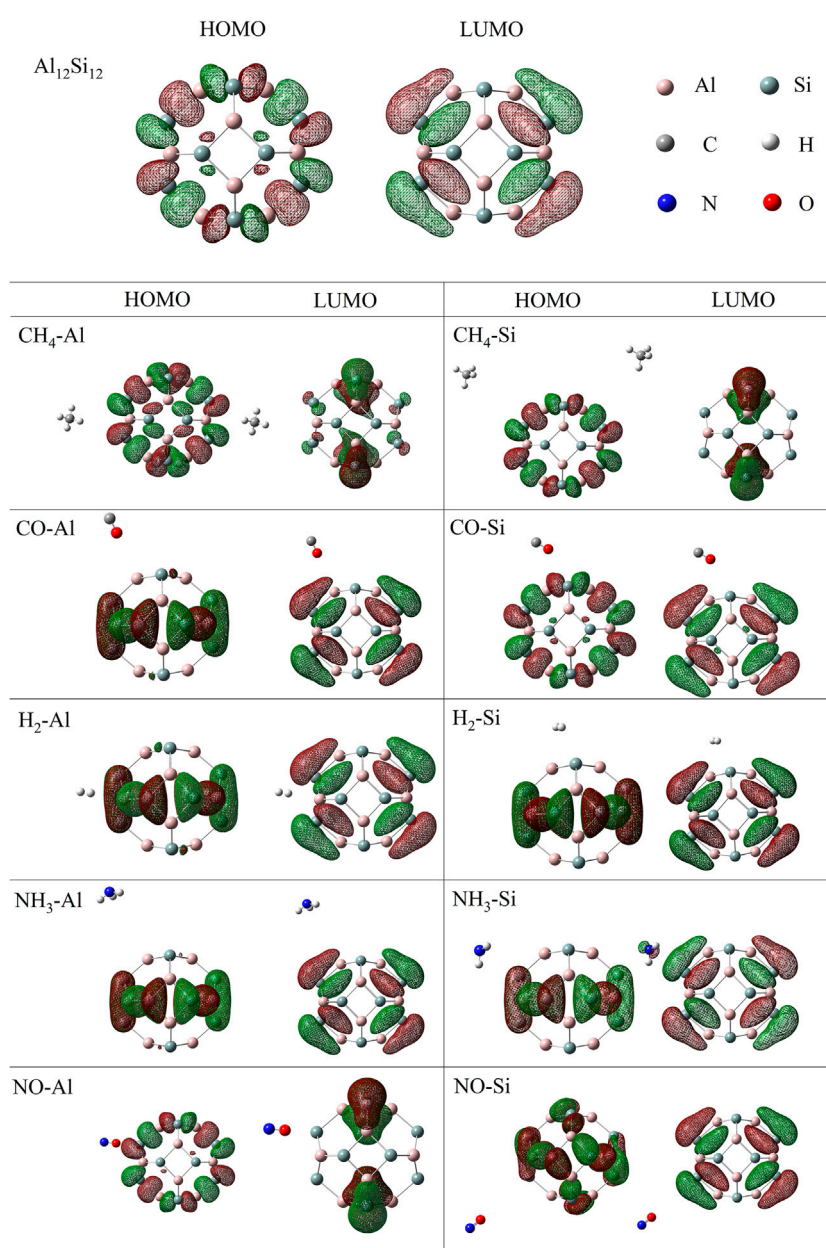
FIGURE 1
Stable configurations of nanocage complexes.

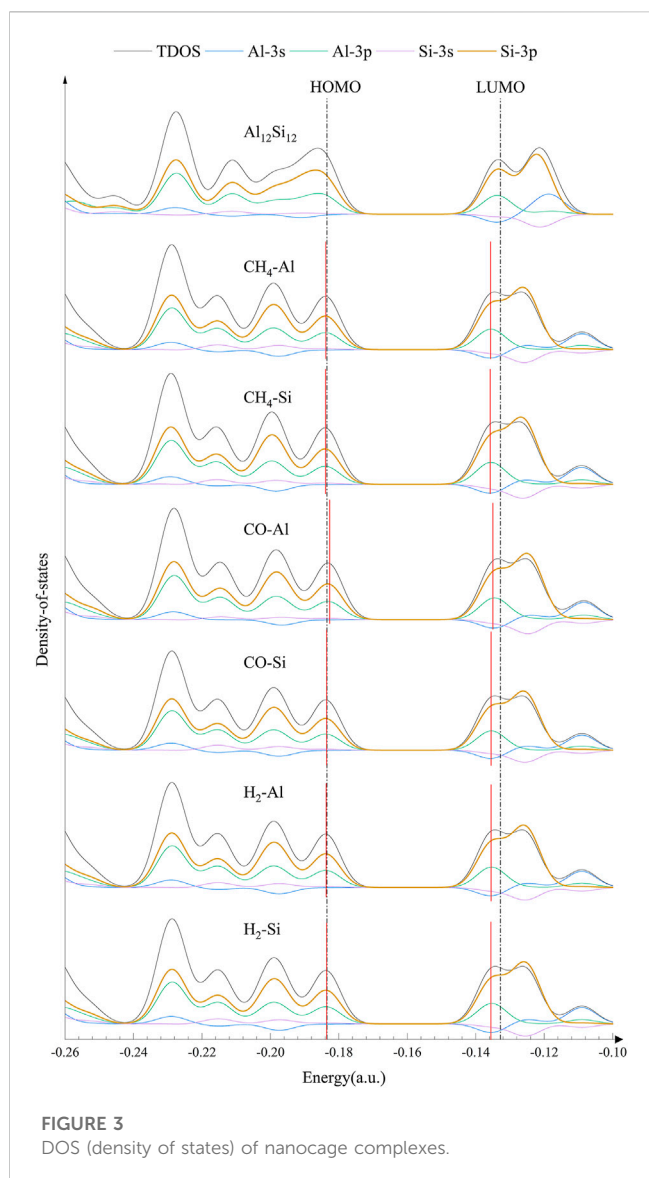
TABLE 1 Bond length (b_{66} and b_{64}), distance of adsorption (D_{ads}) and energy of adsorption (E_{ads}).

Complexes	b_{66} (Å)	b_{64} (Å)	D_{ads} (Å)	E_{ads} (kcal/mol)	Complexes	b_{66} (Å)	b_{64} (Å)	D_{ads} (Å)	E_{ads} (kcal/mol)
$\text{Al}_{12}\text{Si}_{12}$	2.43	2.47	—	—	$\text{Al}_{12}\text{Si}_{12}$	2.43	2.47	—	—
$\text{CH}_4\text{-Al}$	2.47	2.43	5.76	−1.00	$\text{CH}_4\text{-Si}$	2.47	2.43	7.36	−1.02
CO-Al	2.47	2.43	4.51	−1.05	CO-Si	2.47	2.43	4.39	−0.96
$\text{H}_2\text{-Al}$	2.47	2.43	4.52	−12.37	$\text{H}_2\text{-Si}$	2.47	2.43	3.92	−12.34
NO-Al	2.47	2.43	3.94	−15.35	NO-Si	2.47	2.43	5.58	−15.18
$\text{NH}_3\text{-Al}$	2.47	2.43	6.11	−0.55	$\text{NH}_3\text{-Si}$	2.47	2.42	3.48	−1.23

TABLE 2 The Energy of HOMO (E_{HOMO}), energy of LUMO (E_{LUMO}), energy band gap (E_g), change of E_g of nanocage upon the adsorption process (ΔE_g), Mulliken charge transfer (Q_T).

Complexes	E_{LUMO} (eV)	E_{HOMO} (eV)	E_g (eV)	ΔE_g (eV)	Q_T (e)	Complexes	E_{LUMO} (eV)	E_{HOMO} (eV)	E_g (eV)	ΔE_g (eV)	Q_T (e)
$\text{Al}_{12}\text{Si}_{12}$	-3.61	-4.99	1.38	—	—	$\text{Al}_{12}\text{Si}_{12}$	-3.61	-4.99	1.38	—	—
$\text{CH}_4\text{-Al}$	-3.69	-5.00	1.31	-0.06	-0.0022	$\text{CH}_4\text{-Al}$	-3.69	-5.00	1.31	-0.07	-0.0008
CO-Al	-3.67	-4.97	1.30	-0.08	0.0480	CO-Al	-3.69	-5.00	1.31	-0.07	0.0001
$\text{H}_2\text{-Al}$	-3.69	-4.99	1.30	-0.08	0.0197	$\text{H}_2\text{-Al}$	-3.69	-5.00	1.31	-0.07	0.0224
NO-Al	-3.70	-4.99	1.29	-0.09	0.0190	NO-Al	-3.68	-5.00	1.32	-0.06	0.0021
$\text{NH}_3\text{-Al}$	-3.74	-5.05	1.31	-0.07	-0.0021	$\text{NH}_3\text{-Al}$	-3.76	-4.99	1.23	-0.15	0.1177

**FIGURE 2**
HOMO and LUMO distributions of nanocage complexes.



lengths of b_{64} and b_{66} are 2.47 and 2.43 Å, respectively. The calculated E_{coh} of $\text{Al}_{12}\text{Si}_{12}$ is -3.11 eV with zero point energy (ZPE) correction (Buelna-García et al., 2022). The calculated E_{coh} of each atom is closer to the reported E_{coh} of the most stable inorganic analogue, $\text{B}_{12}\text{N}_{12}$. The experimental and theoretical reported E_{coh} of $\text{B}_{12}\text{N}_{12}$ are -4.00 (Pokropivny et al., 2000) and -6.06 eV (Pokropivny and Ovsyannikova, 2007), respectively. As indicated by these results, the $\text{Al}_{12}\text{Si}_{12}$ nanocage is highly stable.

3.1.2 $\text{Al}_{12}\text{Si}_{12}$ –gas molecule structure

CH_4 , CO , H_2 , and NH_3 were selected as the target adsorbates. We performed sufficient structural optimization of the $\text{Al}_{12}\text{Si}_{12}$ nanocage containing CH_4 , CO , H_2 , NO , and NH_3 molecules. For each gas molecule, two different adsorption sites above the Al and Si atoms on the nanocage surface were considered. Each gas is discussed with respect to adsorption at each of the two nanocage positions. The bond length (b_{64} and b_{66}), D_{ads} value (defined as the center to center distance between the atoms of the nanocage and gas molecule that are closest to each other), and

E_{ads} value are listed in Table 1. According to previous studies, the threshold E_{ads} at which physisorption becomes chemisorption is approximately 23.00 kcal/mol (1.00 eV) (Jouypazadeh and Farrokhpour, 2018; Nagarajan and Chandiramouli, 2019; Swetha et al., 2020). D_{ads} must be < 2.00 Å for chemisorption to occur (Contreras et al., 2014). As shown in Table 1, the adsorption processes were all physical, and NO adsorbed above the Al atom of the nanocage had the strongest adsorption stability.

The calculated stable configurations (local minima) are summarized in Figure 1. When the gas was adsorbed onto the Al or Si atom of the $\text{Al}_{12}\text{Si}_{12}$ nanocage, slight local structural deformation of both the molecule and the nanocage occurred. The b_{66} and b_{64} bonds and angles remained almost unchanged. The adsorption that induced greatest change was that of NH_3 onto the Si atom; it changed the lengths of b_{66} and b_{64} by 0.005 and 0.006 Å, respectively. The smallest change was caused by the adsorption of NO onto the Al atom; it changed the lengths of b_{66} and b_{64} by 0.0003 and 0.0204 Å, respectively.

In the Al atom adsorption group, the E_{ads} of CH_4 -Al, CO -Al, NH_3 -Al, NO -Al, and H_2 -Al were -1.00 , -1.05 , -0.55 , -15.35 , and -12.37 kcal/mol, respectively. The adsorption stability of the NO and H_2 analytes was much better than that of the others. The D_{ads} values of NO -Al and H_2 -Al were 3.94 and 4.52 Å, respectively. For NO -Al, the N–O bond length increased from 1.06 Å in isolated NO to 1.16 Å in the adsorbed state. When $\text{Al}_{12}\text{Si}_{12}$ adsorbed H_2 , the H–H bond length increased from 0.6 Å in isolated H_2 to 0.74 Å in the adsorbed state. The D_{ads} between the gas molecule and the Al atom of the nanocage were 5.76 and 4.51 Å for CH_4 -Al and CO -Al, respectively. The C–H bonds of CH_4 -Al and the C–O bond of CO -Al also became slightly longer. The lengths of the four C–H bonds of CH_4 -Al increased by 0.02269, 0.02286, 0.02290, and 0.02313 Å. The length of the C–O bond of CO -Al increased by 0.02 Å. The D_{ads} of NH_3 -Al was 6.11 Å. The length of the three N–H bonds in the NH_3 molecule increased by 0.015636, 0.01599, and 0.01542 Å. Among these complexes, the three N–H bonds of NH_3 -Al had the smallest increase in length.

In the Si atom adsorption group, the E_{ads} of CH_4 -Si, CO -Si, NH_3 -Si, NO -Si, and H_2 -Si were -1.02 , -0.96 , -1.23 , -15.18 , and -12.34 kcal/mol, respectively. The D_{ads} values of CH_4 -Si, CO -Si, NH_3 -Si, NO -Si, and H_2 -Si were 7.36, 4.39, 3.48, 5.58, and 3.92 Å, respectively. In the Si atom adsorption group of complexes, both nanocages and gas molecules changed slightly. The length of the four C–H bonds increased by 0.02253, 0.02268, 0.02284, and 0.02315 Å. The length of the C–O bond increased by 0.02214 Å. The three N–H bonds increased in length by 0.01504, 0.01483, and 0.01485 Å. The lengths of the N–O bond and H–H bond increased by 0.09682, and 0.14372 Å, respectively. Among these complexes, the lengths of the three N–H bonds of NH_3 -Si increased by the least, but they increased by more than those of the three N–H bonds of NH_3 -Al in the Al group.

3.2 Electronic properties

3.2.1 Mulliken charge and energy band gap analysis

Detailed information including the E_{HOMO} and E_{LUMO} , Q_{T} , and the ΔE_{g} (change in E_{g} of nanocage upon adsorption) values is listed in Table 2.

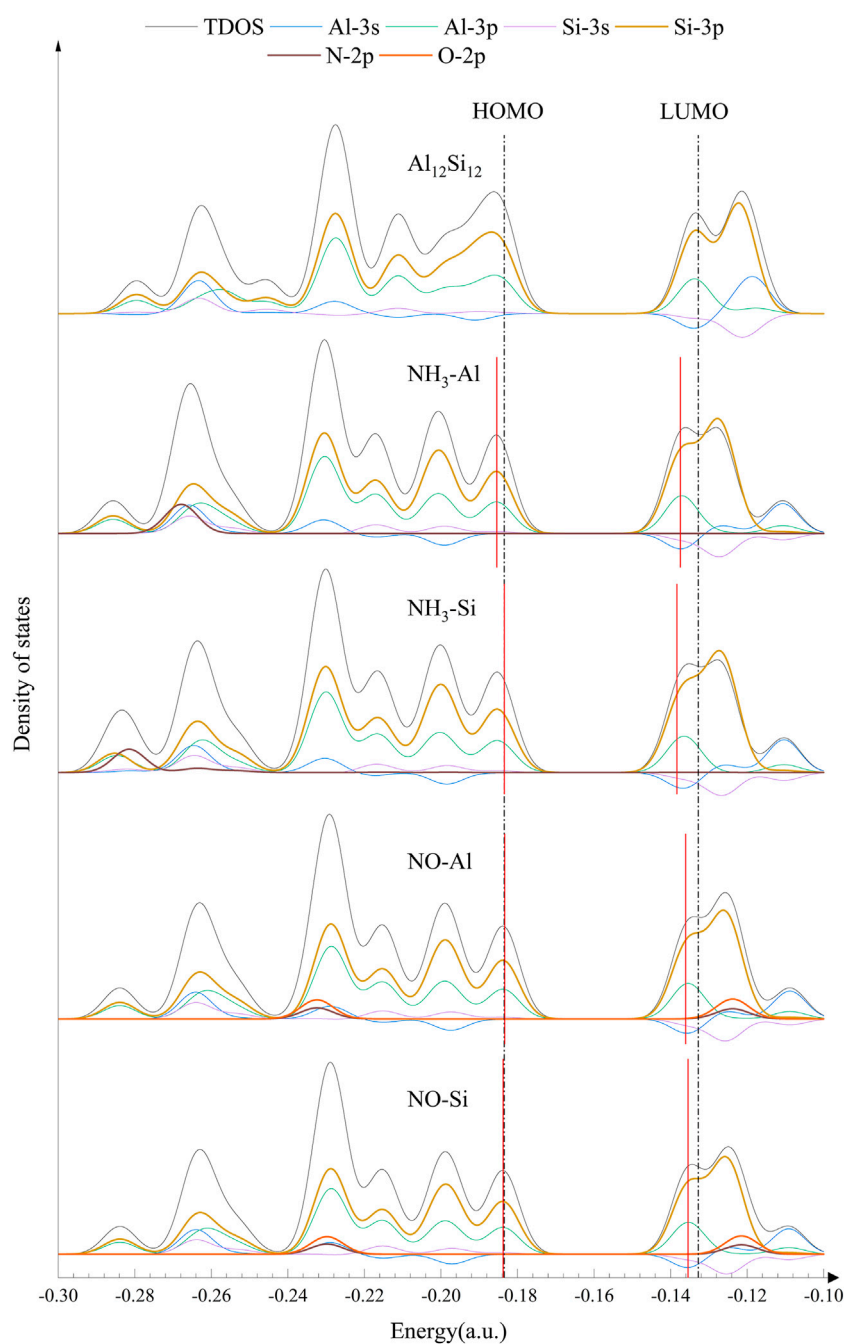


FIGURE 4

DOS (density of states) of nanocage complexes.

In the Al atom adsorption group, the Mulliken charge analysis shows that the number of transferred electrons in CH₄-Al, CO-Al, NH₃-Al, NO-Al, and H₂-Al were -0.0020, -0.0480, -0.0021, 0.0190, and 0.0197 |e|, respectively. The E_g of CH₄-Al, CO-Al, NH₃-Al, NO-Al, and H₂-Al are 1.31, 1.30, 1.31, 1.29, and 1.30 eV, respectively. The E_g of these complexes was lower than that of the pure nanocage, and this decrease was largest in the NO complex, at 0.09 eV, and smallest in the CH₄ complex, at 0.06 eV.

In the Si atom adsorption group, the Mulliken charge analysis demonstrated that the number of transferred electrons in CH₄-Si,

CO-Si, NH₃-Si, NO-Si, and H₂-Si were -0.0008, 0.0001, 0.1177, 0.0021, and 0.0224 |e|, respectively. The E_g of CH₄-Si, CO-Si, NH₃-Si, NO-Si, and H₂-Si was 1.31, 1.31, 1.23, 1.31, and 1.31 eV, respectively. The E_g values of these complexes were also lower than that of the pure nanocage, with the NH₃ complex exhibiting the greatest decrease, at 0.15 eV, and the NO complex exhibiting is the smallest decrease at 0.06 eV.

As demonstrated in Table 2, the E_{LUMO} values of these complexes were slightly lower than those of the pure nanocage, but their E_{HOMO} values did not vary much from those of the pure

nanocage. The adsorption stability of the Si adsorption group of complexes is similar to that of the Al adsorption group. However, the adsorbates in the Si adsorption group transferred fewer electrons than the corresponding adsorbates in the Al adsorption group, except for the H_2 and NH_3 analytes. The N atom gained electrons, and the three H atoms lost electrons in the NH_3 -Al adsorption. However, the N atom and the three H atoms all lost electrons in the NH_3 -Si adsorption due to the difference in adsorption position. The adsorption stability of NH_3 -Si was higher than that of NH_3 -Al due to the short adsorption distance and the large number of transferred electrons in the NH_3 -Si adsorption.

3.2.2 Frontier molecular orbitals and DOS

To further analyze the effect of gas molecules on the electronic properties of nanocages, the HOMO (highest occupied molecular orbital), LUMO (lowest unoccupied molecular orbital) and DOS of isolated and complexed nanocages were analyzed. The HOMO and LUMO distributions of $Al_{12}Si_{12}$ from a vertical perspective are shown in Figure 2. The HOMO primarily serves as an electron donor, and the LUMO primarily serves as an electron acceptor (Dai, 2002). For the HOMO of the pure $Al_{12}Si_{12}$ nanocage, the positive and negative areas alternate between Al and Si atoms (in Figure 2 the red area is negative, the green area is positive). However, for the LUMO of the pure $Al_{12}Si_{12}$ nanocage, the positive and negative areas alternated between the inside and outside of the nanocage and the HOMO and LUMO were distributed on the same plane. The obtained frontier molecular orbital energies (E_{HOMO} and E_{LUMO}) and the calculated, E_g value of the nanocages are shown in Table 2. The E_g of $Al_{12}Si_{12}$ was approximately 1.38 eV, demonstrating its semiconductor properties.

The HOMO and LUMO plots of the complexes are displayed in Figure 2. The lower the E_g value, the more easily a molecule was excited. In the Al adsorption group, the HOMO and LUMO of complexes were almost unchanged. After gas adsorption, the distribution positions of the HOMO and LUMO changed from being on the same plane to being perpendicular to each other. In the Si adsorption group, the HOMO and LUMO of CO-Si remained on the same plane after adsorption. This may be due to the low number of electron transfers. In NH_3 -Si, the LUMO distribution was in the NH_3 molecule. This may be because NH_3 -Si has the shortest adsorption distance and the greatest charge transfer. In NO-Si, part of the HOMO is transferred to the quadrilateral ring opposite the top and bottom. This is most likely because the HOMO energy level of NO-Si is lower than that of NO-Al.

For further confirmation of the electronic behavior of these complexes, DOS analyses were performed (shown in Figures 3, 4). The orbitals in the lower energy region of $Al_{12}Si_{12}$ were mainly the 3p orbitals of the Si and Al atoms. At approximately -0.26 a.u., 3s orbitals of both the Al atoms and the Si atoms were mainly present. The HOMO-LUMO energy level was composed mainly of the 3p orbitals of the Si and Al atoms, whereas the 3s orbitals of Al atoms were also present near the LUMO energy level.

The effect of the gas molecular orbital on the complex orbital when the nanocage adsorbed some gases was not significant (atomic orbitals that do not contribute to the total orbitals are not depicted in Figures 3, 4). We divided the complexes into two groups: The first including gas molecular orbitals that did not contribute to the total orbitals of the complexes, and the second including the gas

molecular orbitals that contributed to the total orbitals of the complexes. In the first group, after gas adsorption, a new peak at 0.2 a.u. appeared, resulting in a slight decrease in the HOMO of the complexes (except for CO-Al). In addition, due to the adsorption of the gas molecule, the 3s orbital of the Al atom produced a small peak at -0.12 a.u. (shown by the blue curve), causing the peak at -0.11 a.u. to shift to the right slightly. This may be the reason for the decrease in LUMO energy level. In the second group, after gas adsorption, a new peak also appeared at 0.2 a.u.. The 3p orbital of the N atom in NH_3 -Al peaked at -0.27 a.u., whereas it peaked at -0.28 a.u. in NH_3 -Si (depicted by the brown curve). The 3p orbitals of the N and O atoms in NO-Al and NO-Si had peaks at -0.23 and -0.12 a.u. (the brown curve and the orange curve), respectively. Moreover, the peak value of the 3p orbital of the O atom was greater than that of the N atom. However, compared with the first group, the new peaks in the second group caused almost no decrease in HOMO and LUMO energy levels. The peak generated at 0.2 a.u. of the 3p orbital of the Si and Al atoms was the main reason for the reduction in HOMO energy level.

4 Conclusion

DFT calculations were performed to investigate the equilibrium geometries, stabilities, and electronic properties of gases adsorbed onto $Al_{12}Si_{12}$ nanocages. The E_g value of $Al_{12}Si_{12}$ was 1.38 eV, demonstrating the semiconductor properties of the nanocage. We performed structural optimization of $Al_{12}Si_{12}$ nanocages with adsorbed CH_4 , CO, H_2 , NO, and NH_3 molecules to study their adsorption energies, equilibrium geometry, and electronic properties. Two adsorption positions on the nanocage were studied. The adsorption energy of these complexes demonstrated that the adsorption of these gases by $Al_{12}Si_{12}$ nanocages is physical. Our findings reveal that the stabilities of the complexes are as follows: $Al_{12}Si_{12}$ -NO > $Al_{12}Si_{12}$ - H_2 > $Al_{12}Si_{12}$ - NH_3 > $Al_{12}Si_{12}$ - CH_4 > $Al_{12}Si_{12}$ -CO. The geometric structure of the nanocage changed slightly following adsorption of molecules. The E_g of the complexes was lower than that of the pure nanocage due to the electron transfer between the gas molecules and the nanocage. The more electrons were transferred, the greater the decrease in E_g . Except for the 3p orbitals of the N and O atoms, the orbitals of the gas molecules did not contribute to the total orbitals of the complexes. The 3p orbitals of the Si atoms in the $Al_{12}Si_{12}$ nanocage are the main reason for the change in HOMO-LUMO energy levels detected.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, L-KL and K-NL; methodology, K-NL; software, W-LX; validation, Y-QM; formal analysis, L-KL; investigation, BH; resources, BH; data curation, K-NL; writing original draft preparation, L-KL; writing—review and editing, K-NL; visualization, W-LX;

supervision, K-NL; project administration, BH; funding acquisition, K-NL. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China, grant number 42167016. The Natural Science Foundation of Ningxia Province in China, grant number 2022AAC03123. The Key Research and Development Program of Ningxia Province in China, grant number 2020BEB04003, and the fifth batch of the Ningxia Youth Science and Technology Talents Project, grant number NXTJGC147.

References

- Ahsan, A., Khan, S., Gilani, M. A., and Ayub, K. (2021). Endohedral metallofullerene electrides of $\text{Ca}_{12}\text{O}_{12}$ with remarkable nonlinear optical response. *RSC Adv.* 11 (3), 1569–1580. doi:10.1039/d0ra08571e
- Ahsin, A., and Ayub, K. (2021). Extremely large static and dynamic nonlinear optical response of small superalkali clusters NM_3M^+ ($\text{M}, \text{M}^+=\text{Li}, \text{Na}, \text{K}$). *J. Mol. Graph. Model.* 109, 108031. doi:10.1016/j.jmgm.2021.108031
- Al-Otaibi, J. S., Mary, Y. S., Mary, Y. S., Acharjee, N., and Churchill, D. G. (2022). Theoretical study of glycoluril by highly symmetrical magnesium oxide $\text{Mg}_{12}\text{O}_{12}$ nanostructure: Adsorption, detection, SERS enhancement, and electrical conductivity study. *J. Mol. Model.* 28 (10), 332. doi:10.1007/s00894-022-05332-3
- Bautista Hernandez, A., Chigo Anota, E., Severiano Carrillo, F., Vázquez Cuchillo, O., and Salazar Villanueva, M. (2022). *In-silico* study of the adsorption of H_2 , CO and CO_2 chemical species on $(\text{TiO}_2)_n$ $n=15-20$ clusters: The $(\text{TiO}_2)_{19}$ case as candidate promising. *J. Mol. Graph. Model.* 117, 108316. doi:10.1016/j.jmgm.2022.108316
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98 (7), 5648–5652. doi:10.1063/1.464913
- Beheshtian, J., Bagheri, Z., Kamfiroozi, M., and Ahmadi, A. (2012). A comparative study on the $\text{B}_{12}\text{N}_{12}$, $\text{Al}_{12}\text{N}_{12}$, $\text{B}_{12}\text{P}_{12}$ and $\text{Al}_{12}\text{P}_{12}$ fullerene-like cages. *J. Mol. Model.* 18 (6), 2653–2658. doi:10.1007/s00894-011-1286-y
- Beheshtian, J., Kamfiroozi, M., Bagheri, Z., and Peyghan, A. A. (2012). $\text{B}_{12}\text{N}_{12}$ Nanocage as potential sensor for NO_2 Detection. *Chin. J. Chem. Phys.* 25 (1), 60–64. doi:10.1088/1674-0068/25/01/60-64
- Beheshtian, J., Peyghan, A. A., and Bagheri, Z. (2012). Selective function of $\text{Al}_{12}\text{N}_{12}$ nano-cage towards NO and CO molecules. *Comput. Mat. Sci.* 62, 71–74. doi:10.1016/j.commatsci.2012.05.041
- Bindhu, M. R., Umadevi, M., Kavin Micheal, M., Arasu, M. V., and Abdullah Al-Dhabi, N. (2016). Structural, morphological and optical properties of MgO nanoparticles for antibacterial applications. *Mat. Lett.* 166, 19–22. doi:10.1016/j.matlet.2015.12.020
- Buelna-García, C. E., Castillo-Quevedo, C., Quiroz-Castillo, J. M., Paredes-Sotelo, E., Cortez-Valadez, M., Martín-del-Campo-Solis, M. F., et al. (2022). Relative populations and IR spectra of Cu_{38} cluster at finite temperature based on DFT and statistical thermodynamics calculations. *Front. Chem.* 10, 841964. doi:10.3389/fchem.2022.841964
- Cazorla-Amorós, D., Alcañiz-Monge, J., and Linares-Solano, A. (1996). Characterization of activated carbon fibers by CO_2 adsorption. *LANGMUIR* 12 (11), 2820–2824. doi:10.1021/la960022s
- Chen, L., Zhou, G.-Q., Xu, C., Zhou, T., and Huo, Y. (2009). Structural and electronic properties of hydrated MgO nanotube clusters. *J. Mol. Struct-Theochem.* 900 (1), 33–36. doi:10.1016/j.theochem.2008.12.019
- Contreras, M. L., Cortés-Arriagada, D., Villarroel, I., Alvarez, J., and Rozas, R. (2014). Evaluating the hydrogen chemisorption and physisorption energies for nitrogen-containing single-walled carbon nanotubes with different chiralities: A density functional theory study. *Struct. Chem.* 25 (4), 1045–1056. doi:10.1007/s11224-013-0377-z
- Dai, H. (2002). Carbon nanotubes: Opportunities and challenges. *Surf. Sci.* 500 (1), 218–241. doi:10.1016/s0039-6028(01)01558-8
- de Oliveira, O. V., Pires, J. M., Neto, A. C., and Divino dos Santos, J. (2015). Computational studies of the $\text{Ca}_{12}\text{O}_{12}$, $\text{Ti}_{12}\text{O}_{12}$, $\text{Fe}_{12}\text{O}_{12}$ and $\text{Zn}_{12}\text{O}_{12}$ nanocage clusters. *Chem. Phys. Lett.* 634, 25–28. doi:10.1016/j.cplett.2015.05.069
- Diederich, F., and Gómez-López, M. (1999). Supramolecular fullerene chemistry. *Chem. Soc. Rev.* 28 (5), 263–277. doi:10.1039/a804248i
- Escobedo, A., Bautista-Hernandez, A., Camacho-García, J. H., Cortes-Arriagada, D., and Chigo-Anota, E. (2019). Effect of chemical order in the structural stability and physicochemical properties of $\text{B}_{12}\text{N}_{12}$ fullerenes. *Sci. Rep.-UK.* 9, 16521. doi:10.1038/s41598-019-52981-1
- Felicio-Sousa, P., Andriani, K. F., and Da Silva, J. L. F. (2021). *Ab initio* investigation of the role of the d-states occupation on the adsorption properties of H_2 , CO, CH_4 and CH_3OH on the Fe_{13} , Co_{13} , Ni_{13} and Cu_{13} clusters. *Phys. Hys. Chem. Chem. Phys.* 23 (14), 8739–8751. doi:10.1039/d0cp06091g
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2004). *Gaussian 03, revision D.01*. Wallingford, CT: Gaussian, Inc.
- Goroff, N. S. (1996). Mechanism of fullerene formation. *Acc. Chem. Res.* 29 (2), 77–83. doi:10.1021/ar950162d
- Hussain, S., Hussain, R., Mehboob, M. Y., Chatha, S. A. S., Hussain, A. I., Umar, A., et al. (2020). Adsorption of phosgene gas on pristine and copper-decorated $\text{B}_{12}\text{N}_{12}$ nanocages: A comparative DFT study. *ACS Omega* 5 (13), 7641–7650. doi:10.1021/acsomega.0c00507
- Hussain, S., Shahid Chatha, S. A., Hussain, A. I., Hussain, R., Mehboob, M. Y., Gulzar, T., et al. (2020). Designing novel Zn-decorated inorganic $\text{B}_{12}\text{P}_{12}$ nanoclusters with promising electronic properties: A step forward toward efficient CO_2 sensing materials. *ACS Omega* 5 (25), 15547–15556. doi:10.1021/acsomega.0c01686
- Jouypazadeh, H., and Farrokhpour, H. (2018). DFT and TD-DFT study of the adsorption and detection of sulfur mustard chemical warfare agent by the C 24, C 12 Si 12, Al 12 N 12, Al 12 P 12, Be 12 O 12, B 12 N 12 and Mg 12 O 12 nanocages. *J. Mol. Struct.* 1164, 227–238. doi:10.1016/j.molstruc.2018.03.051
- Kakemam, J., and Peyghan, A. A. (2013). Electronic, energetic, and structural properties of C- and Si-doped $\text{Mg}_{12}\text{O}_{12}$ nano-cages. *Comput. Mat. Sci.* 79, 352–355. doi:10.1016/j.commatsci.2013.06.036
- Kataura, H., Maniwa, Y., Abe, M., Fujiwara, A., Kodama, T., Kikuchi, K., et al. (2002). Optical properties of fullerene and non-fullerene peapods. *Appl. Phys. A* 74 (3), 349–354. doi:10.1007/s003390201276
- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F., and Smalley, R. E. (1985). C_{60} : Buckminsterfullerene. *Nature* 318 (6042), 162–163. doi:10.1038/318162a0
- Martin, N. (2006). New challenges in fullerene chemistry. *Chem. Commun.* (20), 2093–2104. doi:10.1039/b601582b
- Mastalerz, M., Schneider, M. W., Oppel, I. M., and Presley, O. (2011). A salicylbisimine cage compound with high surface area and selective CO_2/CH_4 adsorption. *Angew. Chem. Int. Ed.* 50 (5), 1046–1051. doi:10.1002/anie.201005301
- Mazloomi, K., and Gomes, C. (2012). Hydrogen as an energy carrier: Prospects and challenges. *Renew. Sust. Energy Rev.* 16 (5), 3024–3033. doi:10.1016/j.rser.2012.02.028
- Nagarajan, V., and Chandiramouli, R. (2019). Acrylonitrile vapor adsorption studies on armchair arsenene nanoribbon based on DFT study. *Appl. Surf. Sci.* 494, 1148–1155. doi:10.1016/j.apsusc.2019.07.214
- O'Boyle, N. M., Tenderholt, A. L., and Langner, K. M. (2008). cclib: A library for package-independent computational chemistry algorithms. *J. Comput. Chem.* 29 (5), 839–845. doi:10.1002/jcc.20823
- Oku, T., Nishiwaki, A., and Narita, I. (2004). Formation and atomic structure of $\text{B}_{12}\text{N}_{12}$ nanocage clusters studied by mass spectrometry and cluster calculation. *Sci. Technol. Adv. Mat.* 5 (5), 635–638. doi:10.1016/j.stam.2004.03.017

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Pokropivny, V. V., and Ovsyannikova, L. I. (2007). Electronic structure and the infrared absorption and Raman spectra of the semiconductor clusters C₂₄, B₁₂N₁₂, Si₁₂C₁₂, Zn₁₂O₁₂, and Ga₁₂N₁₂. *Phys. Solid State* 49 (3), 562–570. doi:10.1134/s1063783407030328
- Pokropivny, V. V., Skorokhod, V. V., Oleinik, G. S., Kurdyumov, A. V., Bartnitskaya, T. S., Pokropivny, A. V., et al. (2000). *J. Solid State Chem.* 154 (1), 214.
- Rad, A. S., and Ayub, K. (2016). Ni adsorption on Al₁₂P₁₂ nano-cage: A DFT study. *J. Alloys Compd.* 678, 317–324. doi:10.1016/j.jallcom.2016.03.175
- Rezaei-Sameti, M., and Abdoli, S. K. (2020). The capability of the pristine and (Sc, Ti) doped Be₁₂O₁₂ nanocluster to detect and adsorb of mercaptopyridine molecule: A first principle study. *J. Mol. Struct.* 1205, 127593. doi:10.1016/j.molstruc.2019.127593
- Shakerzadeh, E., Tahmasebi, E., and Shamlouei, H. R. (2015). The influence of alkali metals (Li, Na and K) interaction with Be₁₂O₁₂ and Mg₁₂O₁₂ nanoclusters on their structural, electronic and nonlinear optical properties: A theoretical study. *Synth. Mater.* 204, 17–24. doi:10.1016/j.synthmet.2015.03.008
- Strout, D. L. (2000). Structure and stability of boron Nitrides: isomers of B₁₂N₁₂. *J. Phys. Chem. A* 104 (15), 3364–3366. doi:10.1021/jp994129a
- Swetha, B., Nagarajan, V., Soltani, A., and Chandiramouli, R. (2020). Novel gamma arsenene nanosheets as sensing medium for vomiting agents: A first-principles research. *Comput. Theor. Chem.* 1185, 112876. doi:10.1016/j.comptc.2020.112876
- Tan, L.-P., Die, D., and Zheng, B.-X. (2022). Growth mechanism, electronic properties and spectra of aluminum clusters. *Spectrochim. Acta. A* 267, 120545. doi:10.1016/j.saa.2021.120545
- Tasis, D., Tagmatarchis, N., Bianco, A., and Prato, M. (2006). Chemistry of carbon nanotubes. *Chem. Rev.* 106 (3), 1105–1136. doi:10.1021/cr050569o
- Vergara-Reyes, H. N., Acosta-Alejandro, M., and Chigo-Anota, E. (2021). Quantum-mechanical assessment of the adsorption of nitric oxide molecules on the magnetic carbon nitride (C₃₆N₂₄)– fullerene. *Struct. Chem.* 32, 1775–1786. doi:10.1007/s11224-021-01736-8
- Yang, H., Song, Y., Zhang, Y., and Chen, H. (2018). Prediction of the electron redundant SinNn fullerenes. *Phys. E* 99, 208–214. doi:10.1016/j.physe.2018.02.010
- Yarovsky, I., and Goldberg, A. (2005). DFT study of hydrogen adsorption on Al₁₃clusters. *Mol. Simul.* 31 (6–7), 475–481. doi:10.1080/08927020412331337041
- Yilanci, A., Dincer, I., and Ozturk, H. K. (2009). A review on solar-hydrogen/fuel cell hybrid energy systems for stationary applications. *Prog. Energy Combust. Sci.* 35 (3), 231–244. doi:10.1016/j.pecs.2008.07.004
- Zhai, H.-J., Zhao, Y.-F., Li, W.-L., Chen, Q., Bai, H., Hu, H.-S., et al. (2014). Observation of an all-boron fullerene. *Nat. Chem.* 6, 727–731. doi:10.1038/nchem.1999
- Zhang, W. J., Rabiei, S., Bagreev, A., Zhuang, M. S., and Rasouli, F. (2008). Study of NO adsorption on activated carbons. *Appl. Catal. B-Environ.* 83 (1), 63–71. doi:10.1016/j.apcatb.2008.02.003
- Züttel, A., Remhof, A., Borgschulte, A., and Friedrichs, O. (2010). Hydrogen: The future energy carrier. *Philos. Trans. R. Soc. A* 368, 3329–3342. doi:10.1098/rsta.2010.0113



OPEN ACCESS

EDITED BY
Miquel Solà,
University of Girona, Spain

REVIEWED BY
Francesc Viñes,
University of Barcelona, Spain
Sergio Posada Pérez,
Universitat de Girona, Spain

*CORRESPONDENCE
Stephen A. Miller,
✉ miller@chem.ufl.edu

SPECIALTY SECTION
This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal
Frontiers in Chemistry

RECEIVED 13 December 2022
ACCEPTED 17 January 2023
PUBLISHED 16 February 2023

CITATION
Miller SA (2023), The location of the
chemical bond. Application of long
covalent bond theory to the structure
of silica.
Front. Chem. 11:1123322.
doi: 10.3389/fchem.2023.1123322

COPYRIGHT
© 2023 Miller. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

The location of the chemical bond. Application of long covalent bond theory to the structure of silica

Stephen A. Miller *

The George and Josephine Butler Laboratory for Polymer Research, Department of Chemistry, University of Florida, Gainesville, FL, United States

Oxygen is the most abundant terrestrial element and is found in a variety of materials, but still wanting is a universal theory for the stability and structural organization it confers. Herein, a computational molecular orbital analysis elucidates the structure, stability, and cooperative bonding of α -quartz silica (SiO_2). Despite geminal oxygen-oxygen distances of 2.61–2.64 Å, silica model complexes exhibit anomalously large O–O bond orders (Mulliken, Wiberg, Mayer) that increase with increasing cluster size—as the silicon-oxygen bond orders decrease. The average O–O bond order in bulk silica computes to 0.47 while that for Si–O computes to 0.64. Thereby, for each silicate tetrahedron, the six O–O bonds employ 52% (5.61 electrons) of the valence electrons, while the four Si–O bonds employ 48% (5.12 electrons), rendering the O–O bond the most abundant bond in the Earth’s crust. The isodesmic deconstruction of silica clusters reveals cooperative O–O bonding with an O–O bond dissociation energy of 4.4 kcal/mol. These unorthodox, long covalent bonds are rationalized by an excess of O $2p$ –O $2p$ bonding versus anti-bonding interactions within the valence molecular orbitals of the SiO_4 unit (48 vs. 24) and the Si_6O_6 ring (90 vs. 18). Within quartz silica, oxygen $2p$ orbitals contort and organize to avoid molecular orbital nodes, inducing the chirality of silica and resulting in Möbius aromatic Si_6O_6 rings, the most prevalent form of aromaticity on Earth. This long covalent bond theory (LCBT) relocates one-third of Earth’s valence electrons and indicates that non-canonical O–O bonds play a subtle, but crucial role in the structure and stability of Earth’s most abundant material.

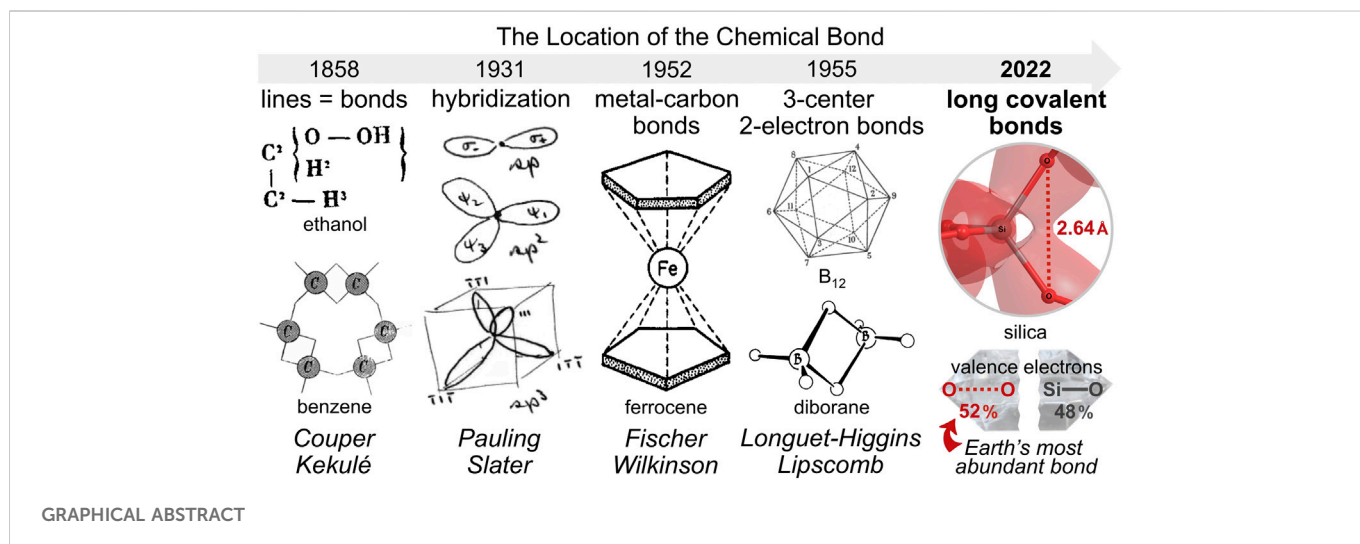
KEYWORDS

bonding, molecular structure, resonance, aromaticity, computational chemistry

1 Introduction

Silica (SiO_2) constitutes about 59% of the Earth’s crust (Clarke and Washington, 1924) and its abundance is largely explained by its stability. By mass, silica is 53% oxygen, the most abundant element on Earth. A universal connection between oxygen and the material stability it bestows has not been described. For years, Pauling argued that a “molecule is stabilized by . . . resonance, its energy being less than the energy corresponding to any one of the structures among which it resonates.” (Pauling, 1946) “It is this extra stability of the system. . . that is called the *resonance energy*.” (Pauling, 1960) Although Pauling never quantified this value for silica (Torgunrud et al., 2020) his detailed resonance description of “low quartz”, illustrated in Figure 1, distinctly implies that it enjoys a resonance energy (Pauling, 1980).

The 50% ionic/50% covalent character of silica was ardently defended by Pauling and determined by the electronegativity difference of $\Delta x = 1.7$ between oxygen ($x = 3.5$) and silicon ($x = 1.8$) (Pauling, 1952). With the premise of 50% ionic character and only silicon-oxygen single bonds, resonance hybrid averaging cannot achieve the charges of +1 on silicon and –0.5 on each oxygen, as limited by the electroneutrality postulate (Newton and Gibbs,



1980). However, Pauling's silicon-oxygen double bonds introduce negative charge on silicon and positive charge on oxygen, allowing for resonance averaging to Si^{+1} and $\text{O}^{-0.5}$ when the ionic character is 50%. Since each silicon atom forms two single bonds and two double bonds, Pauling concluded that the Si-O bond order in silica is $3/2$, with the double bonds accommodated by sp^3d^2 orbitals of silicon (Pauling, 1980). While a topic of considerable debate (Janes and Oldfield, 1986), recent experimental and computational results indicate that O $2p$ -Si $3d$ π -bonding is minimal in silica (Silvi et al., 1997; Garvie et al., 2000; Gibbs et al., 2009) and that Si-O bond orders are less than unity (Newton and Gibbs, 1980; Noritake and Kawamura, 2015; Torgunrud et al., 2020), contradicting Pauling's resonance scheme and stated bond order. Without silicon-oxygen double bonds, the 50% covalent component defaults to the single silicon-oxygen bonds of the canonical resonance form. Resonance averaging that is 50% canonical and 50% ionic yields a net +2 charge on silicon and an obvious violation of the electroneutrality postulate.

Herein, an alternative resonance formulation for silica is presented that ameliorates these incongruities. Non-canonical, long covalent oxygen-oxygen bonds are proposed and rationalized—ostensibly necessitated—by a molecular orbital analysis of silica model complexes, thereby demystifying the structure, bonding, and stability of Earth's most abundant material.

2 Results and discussion

2.1 Resonance hybrids with long bonds

The ionic component of Pauling's 1980 resonance scheme for silica (Figure 2A) employs oxidized silicon(IV) and reduced oxygen(-II); the covalent component employs reduced silicon(-II) and oxidized oxygen(I). An alternative covalent resonance formulation is proposed in Figure 2B, which assumes the same oxidation states, but a different arrangement of bonds. Instead of arguably discounted silicon-oxygen double bonds, long oxygen-oxygen bonds averaging 2.63 \AA are enlisted. Each silicon-oxygen interaction comprises a 50% single bond and 50% no-bond (Roberts et al., 1950) resonance formulation; hence the predicted Si-O bond order is $1/2$.

Each geminal oxygen-oxygen interaction comprises a 33% single bond and 67% no-bond resonance formulation; hence the predicted O-O bond order is $1/3$. Note that this resonance scheme involves an equal number of Si-O ($4 = 4 \times 2 \times 1/2$) and O-O ($4 = 6 \times 2 \times 1/3$) valence electrons.

2.2 Bond order analysis

This non-canonical O-O bond can be interrogated by a Mulliken bond order analysis (Mulliken, 1955a), which “characterizes the accumulation of the electrons in the region between the chemically bonded atoms, and is a very useful quantity often characterizing well the bond strength” (Mayer, 2007) and bond covalency (Mulliken, 1955b; Segall et al., 1996). The canonical silica structure employs 8 of its 16 valence electrons ($= 4 + 6 + 6$) to make four Si-O single bonds, leaving 4 lone pairs of non-bonding electrons on the oxygen atoms. Computed at the DFT/B3LYP/6-311++G** level of theory (Shao et al., 2015; Levien et al., 1980)¹, silicic acid ($\text{Si}(\text{OH})_4$) appears nearly canonical with four Si-O bond orders averaging 1.03 (range = 1.01–1.04) and six oxygen-oxygen bond orders averaging an inconspicuous 0.05 (range = 0.04–0.06). Accordingly, each silicic acid drafts 8.21 valence electrons for Si-O bonding and 0.55 electrons for O-O bonding, totaling 8.77 valence bonding electrons and leaving 7.23 as non-bonding.

However, for the silicate at the center of a large 29-silicon cluster ($\text{Si}_{29}\text{O}_{76}\text{H}_{36}$), where structural regularity is high and edge effects are low, the Si-O bond orders are substantially less than unity at 0.63, 0.63, 0.68, and 0.68, averaging 0.66 and requiring only 5.25 electrons. The O-O bond orders are conspicuously greater than zero at 0.38, 0.38, 0.43, 0.43, 0.63, and 0.63, averaging 0.48 and requiring 5.78 electrons—with these six O-O bonds occupying the edges of a silicate tetrahedron, as shown in Figure 3A. Thus, the central silicate of the Si_{29} cluster employs 11.03 valence bonding electrons, leaving

¹ Diffuse functionals (++) are employed since they better account for long range interactions and periodic bonding while universally providing lower electronic energies versus confined functionals by about 0.21 kcal/mol of valence electrons. For example, electron confinement costs 136.1 kcal/mol for the $\text{Si}_{29}\text{O}_{76}\text{H}_{36}$ cluster (Si_{29}) and about 3.4 kcal/mol of SiO_2 .

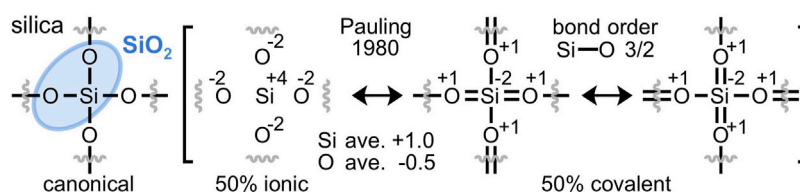


FIGURE 1

Pauling's resonance hybrid model of α -quartz silica (SiO_2) is 50% ionic and 50% covalent. Silicon-oxygen double bonds formally reduce silicon and oxidize oxygen, yielding averaged Si^{+1} and $\text{O}^{-0.5}$ atomic charges and Si-O bond orders of $3/2$ (Pauling, 1980). This resonance model is contradicted by modern computational investigations that discount silicon-oxygen double bonds in silica.

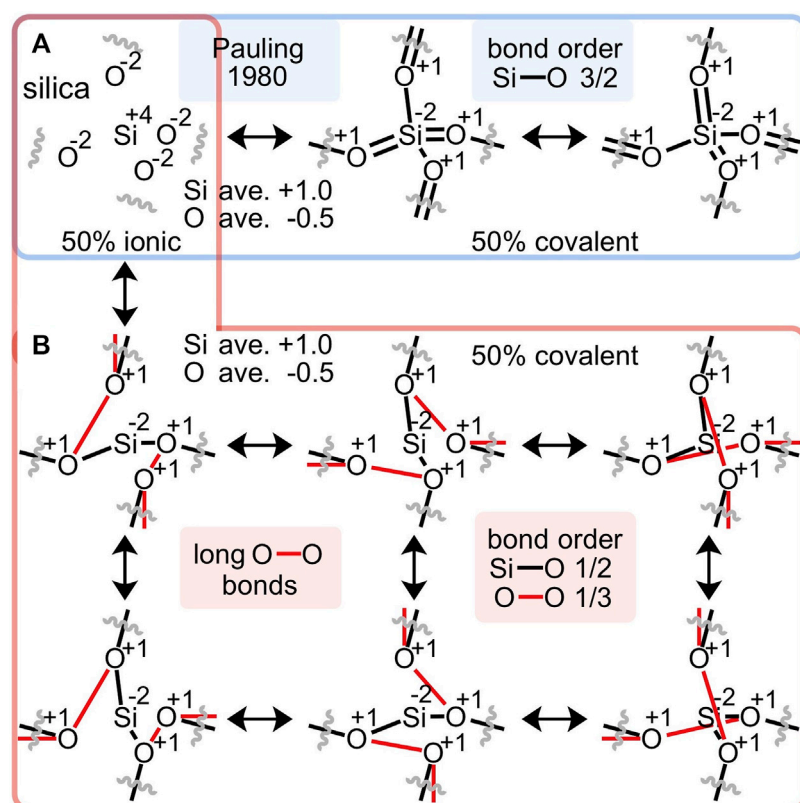


FIGURE 2

An alternative to Pauling's resonance scheme for silica (A) proposes long oxygen-oxygen bonds, ranging from 2.61 to 2.64 Å (B). This alternative conforms to Pauling's 50% ionic/50% covalent apportionment, enlists the same average atomic charges and oxidation states, avoids unwarranted Si-O double bonds, suggests Si-O bond orders of $1/2$ and geminal O-O bond orders of $1/3$ and thus, employs equal numbers of Si-O and O-O valence electrons.

4.97 electrons as non-bonding. Moreover, this central silicate enlists 5.23 (= 5.78–0.55) *additional* electrons for oxygen-oxygen bonding versus silicic acid. The electrostatic charge on silicon is +0.94e and the charges on the oxygen atoms are –0.51, –0.51, –0.54, and –0.54e, averaging –0.52e. The computed bond orders and atomic charges mirror the simplistic resonance hybrid averaging of Figure 2B.

Figure 3B plots the computed central silicate Si-O and O-O bond orders versus the number of silicon atoms in model quartz silica clusters ranging from *Si1* (silicic acid) to *Si35*. As the model complex is enlarged from *Si1* to *Si5*, the average O-O bond order increases about 0.05 for each additional silicate unit appended to the central silicate. Thereafter, the O-O bond order increases about 0.01 for each additional silicate—reaching 0.48 for *Si29* and 0.47 for *Si35* (presently the computational limit). The

average Si-O bond order decreases with increasing cluster size—albeit with less regularity—reaching 0.66 for *Si29* and 0.59 for *Si35*. Note also that the Si 3d atomic orbital contribution to the valence molecular orbitals is consistently below 1% (*Si1*, 0.8%; *Si5*, 0.7%; *Si29*, 0.4%), further bolstering the premise that Si-O π bonding is minimal in silica.

For these same model clusters, Figure 3C plots the percentage of valence electrons belonging to the Si-O and O-O bonds of the central silicate unit. The O-O valence electron percentage climbs steeply to 36% (*Si5*) as the first shell of four silicates is added to the central silicate. The second shell accommodates twelve additional silicates, but their addition has a less dramatic effect. With 2–7 second-shell silicates (*Si8*–*Si18*), the O-O valence electron percentage spans 37%–43%. With 8–12 second-shell silicates (*Si21*–*Si35*), the percentage spans 51%–54%. These plots affirm a

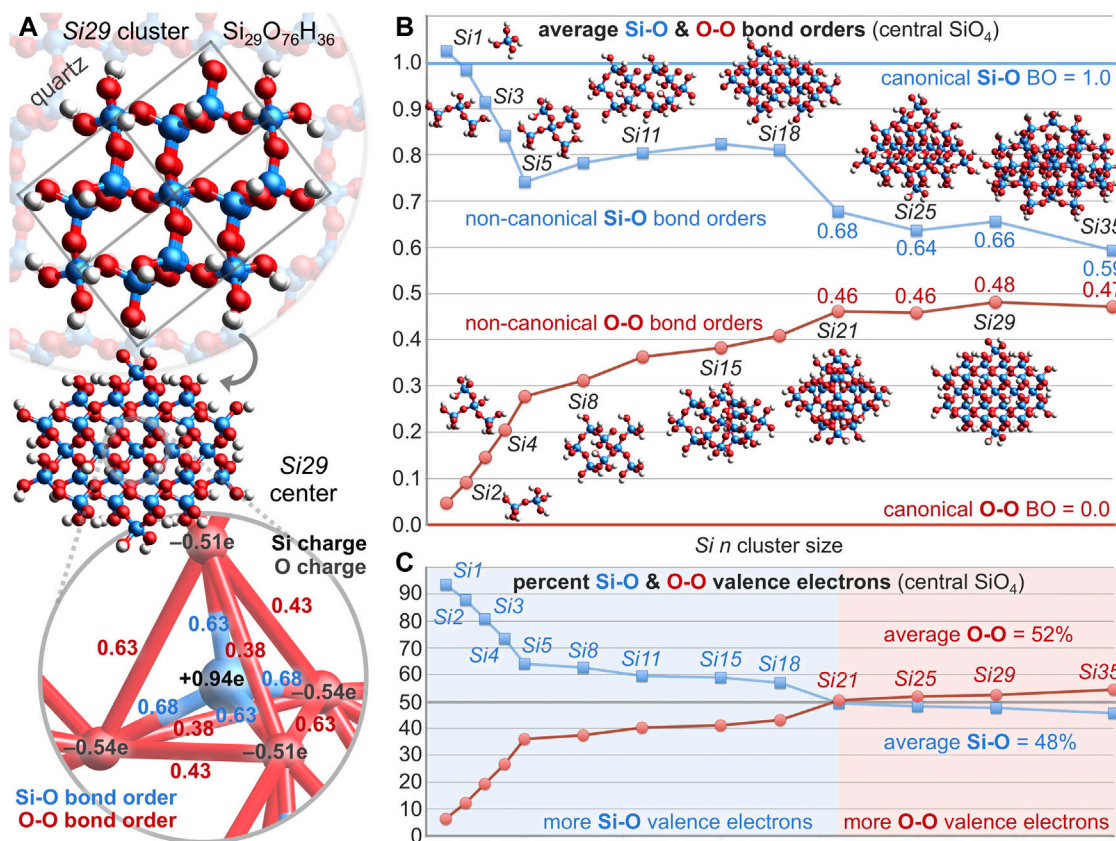


FIGURE 3

(A) For the central silicate of the quartz Si₂₉ cluster, the Si-O bond orders average 0.66 (blue) and the O-O bond orders average 0.48 (red). (B) For the central silicate of clusters ranging from Si1 to Si35, average bond orders deviate substantially from the canonical bond order of 1.0 for Si-O (blue) and 0.0 for O-O (red). (C) For the central silicate of large clusters (Si21–Si35), more valence electrons belong to O-O bonds (52%) than Si-O bonds (48%)—indicating that the O-O bond is the most prevalent bond on Earth.

coherent dependence of bond order and valence electron apportionment on the size and structure of the silica model complexes. The averaged values for the largest group (Si21–Si35) indicate that only 48% of silica's valence bonding electrons are allocated to silicon-oxygen bonds, while 52% are allocated to oxygen-oxygen bonds—insofar as the central silicate faithfully represents bulk silica. By this measure, *the oxygen-oxygen bond in silica is the most abundant bond on Earth*, either in number (six O-O bonds vs. four Si-O bonds) or by valence electron count (5.61 O-O electrons vs. 5.12 Si-O electrons, averaged over Si21–Si35). Long O-O bond ubiquity is further suggested by a similar analysis of alumina (Al₂O₃) (Kirfel and Eichhorn, 1990), the second most abundant material in the Earth's crust (15%) (Clarke and Washington, 1924). In this case, maximum O-O bond orders increase from 0.13 to 0.20 to 0.32 for geminal oxygen atoms 2.72 Å apart in Al₈O₁₂, Al₁₂O₁₈, and Al₁₆O₂₄ clusters, respectively (see the Supplementary Material).

2.3 Molecular orbital analysis

Generally, atoms are not covalently bonded unless they experience an excess of bonding interactions versus anti-bonding interactions and thereby, “a concentration of charge between the two nuclei.” (Feynman, 1939) Similarly, Hoffmann stated succinctly that “positive overlap implies stabilization or bonding.” (Hoffmann, 1971) So, is there

charge concentration and positive overlap between oxygen atoms of silica that account for the non-canonical O-O bond orders of Figure 3B?

The repeat unit of silica can be considered as SiO₂ or Si(½O)₄ and the four oxygen atoms of the latter can be modeled by Si(OSi)₄ to provide the molecular orbitals (MOs) of Figure 4A. Depicted are sixteen valence MOs built primarily from the four oxygen 2s orbitals and the twelve oxygen 2p orbitals (IsoValue = 0.01 [e/bohr]^{0.5}). The oxygen character is above 70% for eleven of these MOs, averages 71%, and exceeds the valence electron percentage contributed by oxygen, 55%.² The remaining six valence MOs (HOMO-5 through HOMO, not shown) are built primarily from silicon atomic orbitals and have a silicon character averaging 86%—well exceeding the valence electron percentage contributed by silicon, 45%. Clearly, there is disproportional and poor silicon/oxygen atomic orbital mixing and the oxygen-rich MOs rationalize a high degree of oxygen-oxygen covalency.

The six valence O-O interactions in each Si(OSi)₄ MO can be categorized as bonding or anti-bonding³. These are illustrated and tabulated in Figure 4A. Of the 96 O-O interactions, 61 are bonding

2 For Si(OSi)₄, oxygen contributes 55% (= [4x6]/[4x6 + 5x4]) of the 44 valence electrons and silicon contributes 45% (= [5x4]/[4x6 + 5x4]).

3 For mathematical simplicity, apparent non-bonding interactions are counted as anti-bonding interactions.

and 35 are anti-bonding; there are 26 excess bonding interactions. A *bonding excess* parameter can be defined as $be = (\text{bonding} - \text{anti-bonding}) / (\text{bonding} + \text{anti-bonding})$ and computes to $be_{\text{O-O}} = 27\%$ in $\text{Si}(\text{OSi})_4$. Among just the 24 O 2s–O 2s interactions, 13 are bonding and 11 are anti-bonding; there are 2 excess bonding interactions and $be_{2s} = 8\%$. Among just the 72 O 2p–O 2p interactions, 48 are bonding and 24 are anti-bonding; there are 24 excess bonding interactions and $be_{2p} = 33\%$. The greater *bonding excess* for the O 2p subset illuminates the importance of O 2p atomic orbitals vs. O 2s atomic orbitals for long covalent bonding. Collectively, there is a definitive excess of bonding interactions providing the net positive overlap that allows—or perhaps mandates—oxygen-oxygen covalent bonding. Accordingly, the O–O bond orders are decidedly greater than zero (0.07, 0.11, 0.13, 0.13, 0.14, 0.14) and average 0.12. The Si–O bond orders are decidedly less than unity (0.88, 0.88, 0.80, 0.79) and average 0.84.

The proclamation that “there are no chemical bonds, only bonded interactions,” (Gibbs et al., 2009) compels scrutiny of the Figure 4A MOs for electron density associated with O–O interactions. Figure 4B illustrates HOMO-*n* slices intersecting the upper right O–O axis of the O_4 tetrahedron (bond order = 0.14). These planes contain Si or are perpendicular to that—depending on which shows greater overlap. Eleven of these sixteen interactions are bonding and five are anti-bonding; locally, this O–O relationship has $be_{\text{O-O}} = 38\%$. Of the bonding interactions, nine utilize σ overlap and two utilize π overlap. Of the anti-bonding interactions, four are σ^* and one is π^* . The HOMO-13 through HOMO-9 slices reveal considerable O–O covalency, arising from three σ interactions and two π interactions. These MOs average an oxygen atomic orbital contribution of 76%—well above the 55% predicted by purely proportional silicon/oxygen atomic orbital mixing. This excess oxygen MO character portends high O–O covalency (bond order > canonical value of 0) and low Si–O covalency (bond order < canonical value of 1). The *bonding excess* and the abundance of O–O electron density, implied by Ψ^2 of the HOMO-*n* slices, are further evidence of long bonded interactions in locations not previously described.

Every atom in quartz silica is part of the $\text{Si}(\frac{1}{2}\text{O})_4$ repeat unit. Additionally, every atom belongs to a twelve-membered ring of alternating silicon and oxygen atoms, Si_6O_6 . It is illustrative to first consider the bonding arrangement of a geometry-optimized Si_6O_6 ring, which adopts a planar, D_{6h} conformation. Figure 5A depicts the twenty-four valence MOs built primarily from the six oxygen 2s orbitals and the eighteen oxygen 2p orbitals (IsoValue = 0.005 [e/bohr^3] $^{0.5}$). Of the 144 O–O interactions, 68 are bonding and 76 are anti-bonding; there are 8 excess anti-bonding interactions. Hence, the *bonding excess* parameter is negative: $be = -6\%$. Accordingly, the computed O–O bond orders are also negative at -0.05 (Mulliken, 1955c). Note that the HOMO-16, HOMO-13, HOMO-12, HOMO-11, HOMO-10, and HOMO-7 mimic the six π MOs of benzene (also D_{6h}), built from six C 2p_z atomic orbitals (Hoffmann, 1963). They possess the same number of π nodal planes (0, 1, 1, 2, 2, 3), mandating that half are bonding and half are anti-bonding. Naturally, there is no net π bonding when these six MOs are filled.

However, the Si_6O_6 ring of quartz silica is *not* planar. Instead, it contorts to a chiral, C_2 -symmetric saddle shape, yielding a dramatically different array of oxygen-based valence molecular orbitals, as illustrated in Figure 5B. Of the 144 O–O interactions, 108 are bonding and 36 are anti-bonding; there are 72 excess bonding interactions. Hence, the *bonding*

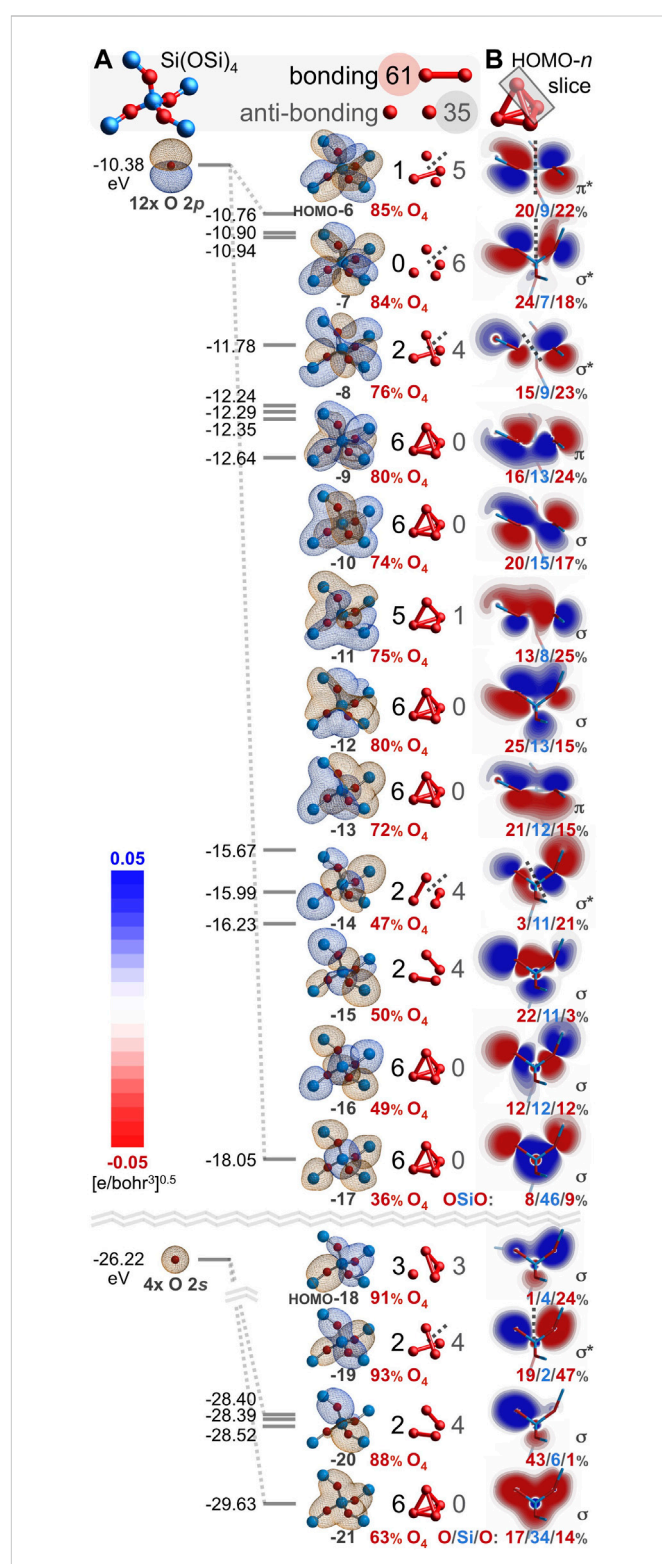


FIGURE 4 Oxygen-based valence molecular orbitals (IsoValue = 0.01 [e/bohr^3] $^{0.5}$) for $\text{Si}(\text{OSi})_4$ excised from quartz silica (A). Bonding and anti-bonding O–O interactions are tabulated along with the oxygen atomic orbital contribution (red). Slices for each HOMO-*n* along the upper right O–O axis reveal excess bonding (11; σ and π) vs. anti-bonding (5; σ^* and π^*) interactions (B). The corresponding O/Si/O atomic orbital contributions reveal the greatest O–O covalency for HOMO-13 through HOMO-9.

excess parameter is large and positive: $be = 50\%$. Among just the 108 O $2p$ -O $2p$ interactions, 90 are bonding, 18 are anti-bonding, and $be_{2p} = 67\%$. Accordingly, the computed O-O bond orders are positive: 0.04, 0.04, 0.06, 0.06, 0.07, and 0.07, averaging 0.05. For comparison, the Si_6O_6 ring embedded in a larger $\text{Si}_{20}\text{O}_{24}$ cluster possesses 373 bonding and 203 anti-bonding O-O interactions, yielding a substantial be of 30% and O-O bond orders ranging from 0.13 to 0.33 and averaging 0.23. While there is no net oxygen-oxygen bonding in the planar D_{6h} Si_6O_6 ring, such long bonds prevail in the non-planar rings of silica.

How can the deplanarization/contortion described in Figure 5 eliminate 40 anti-bonding O-O interactions and yield net O-O bonding? Nearly all of the eliminated nodes are lost from the HOMO-17 through HOMO-6; these twelve MOs derive mostly from the O $2p_y$ and O $2p_z$ atomic orbitals. Figure 6 compares and contrasts the O $2p_y$ MOs from the D_{6h} Si_6O_6 ring with those from the silica-based Si_6O_6 ring. The MOs for the planar ring have 18 bonding and 18 anti-bonding O-O interactions (Figure 6A). After warping to the chiral, saddle-shaped ring adopted by silica, O-O overlap greatly increases, resulting in 34 bonding and only 2 anti-bonding O-O interactions (Figure 6B). Inspection of these silica-based molecular orbitals reveals bonding, node-free, contiguous π MOs built mostly from oxygen atomic orbitals (78%–91%, averaging 84%) and minimally from silicon atomic orbitals (9%–22%, averaging 16%). Arguably, this constitutes a modified case of homoconjugation, which is defined as π overlap separated by a *single* non-conjugating atom or group (McNaught and Wilkinson, 1997). In the present case, the annular π -system spans *multiple* non-conjugating silicon atoms, being built from atomic orbitals of relatively distant (2.61–2.64 Å) and alternating oxygen atoms. Such an arrangement that nonetheless yields bonding molecular orbitals is herein defined as *alternohomoconjugation* (Hoffmann et al., 1970).

The HOMO-17 of Figure 6B is a twist-free π MO composed mostly (83%) of six O $2p_y$ orbitals pointing towards the middle of the Si_6O_6 silica ring. However, the HOMO-13 and HOMO-12 are contiguous π MOs with two half-twists and thus, have a topological linking number of $L_k = 2$ (Rappaport and Rzepa, 2008). Moreover, these MOs twist in opposite directions, resulting in quasi-enantiomorphic MOs (but not strictly enantiomorphic because the Si_6O_6 ring itself is chiral). The HOMO-11 and HOMO-8 are also contiguous π MOs, but with four half-twists and thus, a linking number of $L_k = 4$. These too are quasi-enantiomorphic because they twist in opposite directions. The HOMO-6 is apparently not a contiguous π MO because of a nodal surface bisecting the ring. Molecular orbitals with $L_k = 2$ or 4 are Möbius aromatic, although such twisting generally results in a smaller resonance energy (stabilization) compared to non-twisted (Hückel) analogues with $L_k = 0$ (Rzepa, 2005). For the HOMO-13, this Möbius aromaticity is further visualized by the six contiguous O $2p_y$ -O $2p_y$ π bonding interactions of Figure 6C; these HOMO-13 slices are oblique but nonetheless lack a disruptive nodal plane. Figure 6 focuses on O $2p_y$ atomic orbitals, but there are also MOs built from O $2p_z$ atomic orbitals (Figure 5B and Figure 7B) that are Möbius aromatic: HOMO-15 and HOMO-14 with $L_k = 2$; HOMO-10 and HOMO-9 with $L_k = 4$. In total, eight of the 24 oxygen-based valence MOs are Möbius aromatic (HOMO-15 through HOMO-8). The HOMO-17 (O $2p_y$) and HOMO-16 (O $2p_z$) are twist-free π MOs with $L_k = 0$ and are thus Hückel aromatic.

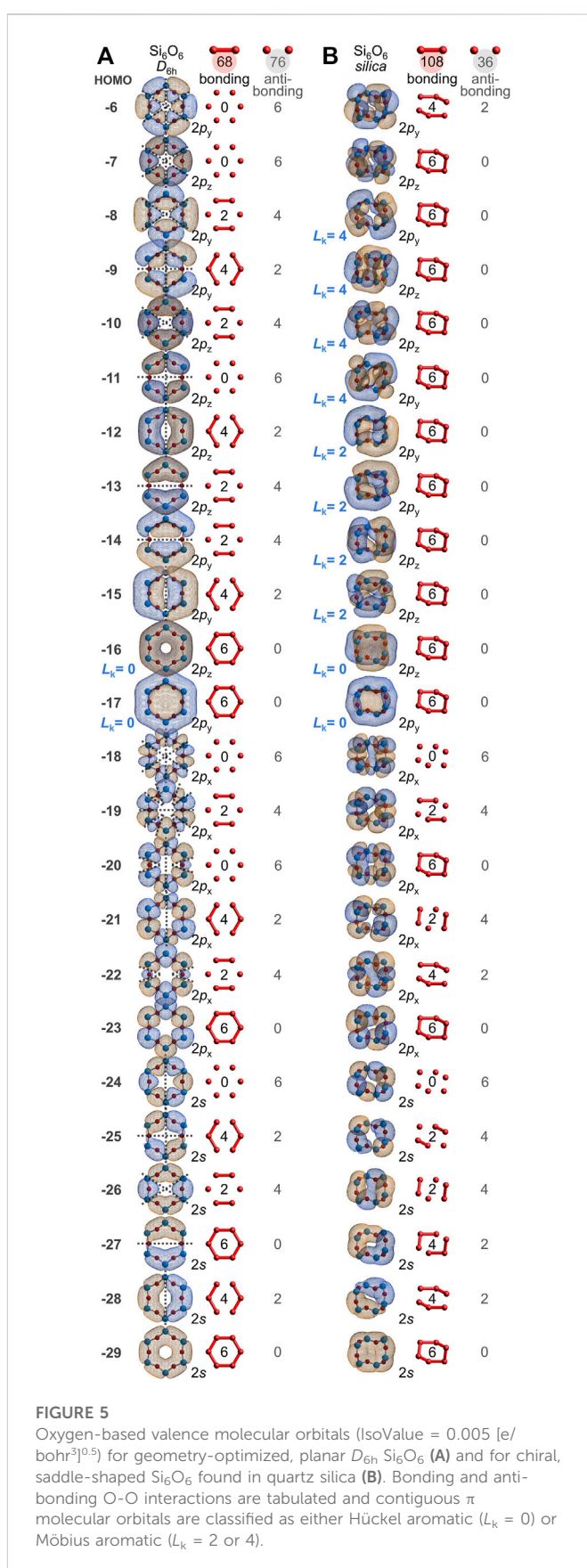


FIGURE 5
Oxygen-based valence molecular orbitals (IsoValue = 0.005 [e/bohr^3]) for geometry-optimized, planar D_{6h} Si_6O_6 (A) and for chiral, saddle-shaped Si_6O_6 found in quartz silica (B). Bonding and anti-bonding O-O interactions are tabulated and contiguous π molecular orbitals are classified as either Hückel aromatic ($L_k = 0$) or Möbius aromatic ($L_k = 2$ or 4).

Altogether, the Si_6O_6 ring of silica has ten total oxygen-based, contiguous π MOs (Figure 5B, HOMO-17 through HOMO-8), while the D_{6h} Si_6O_6 ring has only two (Figure 5A, HOMO-17 and

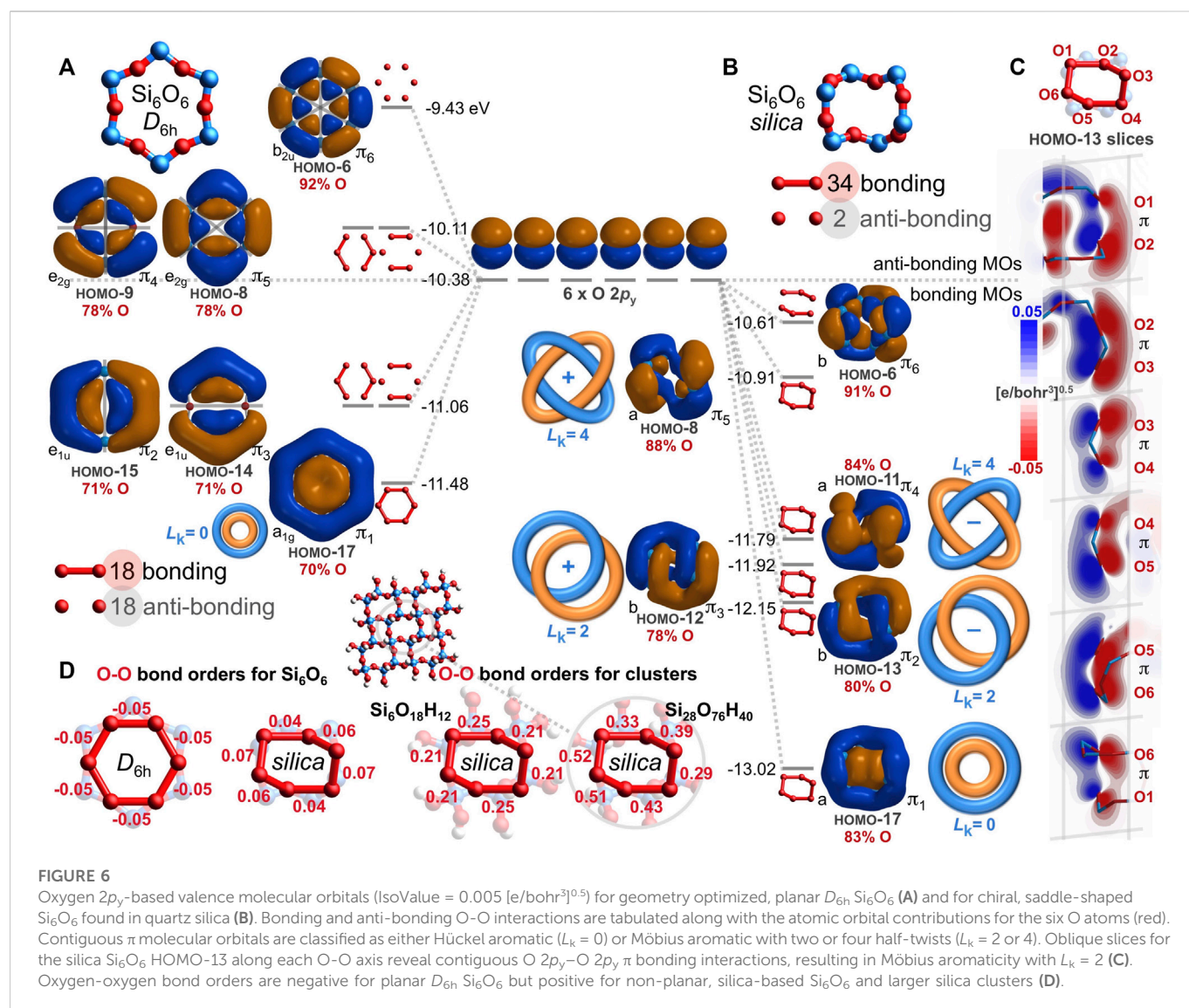


FIGURE 6

Oxygen $2p_y$ -based valence molecular orbitals (IsoValue = 0.005 $[\text{e}/\text{bohr}^3]^{0.5}$) for geometry optimized, planar D_{6h} Si_6O_6 (A) and for chiral, saddle-shaped Si_6O_6 found in quartz silica (B). Bonding and anti-bonding O-O interactions are tabulated along with the atomic orbital contributions for the six O atoms (red). Contiguous π molecular orbitals are classified as either Hückel aromatic ($L_k = 0$) or Möbius aromatic with two or four half-twists ($L_k = 2$ or 4). Oblique slices for the silica Si_6O_6 HOMO-13 along each O-O axis reveal contiguous O $2p_y$ -O $2p_y$ π bonding interactions, resulting in Möbius aromaticity with $L_k = 2$ (C). Oxygen-oxygen bond orders are negative for planar D_{6h} Si_6O_6 but positive for non-planar, silica-based Si_6O_6 and larger silica clusters (D).

HOMO-16). This difference largely explains the disparity in bonding/anti-bonding O-O interactions and thus, the negative O-O bond orders for the planar Si_6O_6 ring (-0.05) and the positive O-O bond orders (averaging 0.05) for the saddle-shaped Si_6O_6 ring based on silica (Figure 6D). These long bond orders are larger for the model silicate ring $\text{Si}_6\text{O}_{18}\text{H}_{12}$ (averaging 0.22) and larger still for the central ring of $\text{Si}_{28}\text{O}_{76}\text{H}_{40}$ (averaging 0.41). This molecular orbital analysis attests that the long covalent O-O bonds within quartz silica are significantly attributable to aromaticity, rendered by alternation of the oxygen atoms *via* uninterrupted O $2p_y$ or O $2p_z$ overlap. Moreover, this analysis claims Möbius aromaticity is the most prevalent kind of aromaticity on Earth since 25% of silica's valence electrons belong to this category⁴.

2.4 Molecular orbital analysis of silica vs. sulfur

α -Sulfur is a stable chalcogen allotrope that consists of S_8 rings with sulfur-sulfur bonding (Rettig and Trotter, 1987). Compared to the O-O bonds of silica (average = 2.63 Å), the S-S bonds are shorter (average = 2.05 Å) and sulfur shares no electrons with an electropositive element. Hence, the S-S Mulliken bond orders are typical of canonical single bonds, ranging from 1.13 to 1.18. A valence molecular orbital analysis of cyclic S_8 (Figure 7A) reveals that its S-S bonds exist because of p atomic orbital overlap and not s atomic orbital overlap—akin to the O-O bonds of the silica Si_6O_6 ring. For the eight S $3s$ -based valence MOs, there are 32 bonding and 32 anti-bonding S-S interactions ($be_{3s} = 0\%$). For the sixteen S $3p$ -based valence MOs, there are 128 bonding and 0 anti-bonding S-S interactions ($be_{3p} = 100\%$). Collectively, $be = 67\%$, allowing for net S-S bonding. Furthermore, as shown in Figure 7A and Supplementary Figure S5, ten of the sixteen $3p$ -based S_8 MOs are aromatic, including: Hückel aromaticity for HOMO-12 ($3p_y$) and HOMO-7 ($3p_z$), with $L_k = 0$; Möbius aromaticity for HOMO-9 ($3p_y$), HOMO-8 ($3p_y$), HOMO-6 ($3p_z$), HOMO-5 ($3p_z$), HOMO-2 ($3p_z$), and HOMO ($3p_z$), with $L_k = 2$; and Möbius aromaticity for HOMO-4 ($3p_y$) and HOMO-3 ($3p_y$), with $L_k = 4$. Of the

⁴ Oxygen contributes $(6 + 6)/(6 + 6 + 4) = 75\%$ of silica's valence electrons and 8 of the 24 oxygen-based valence MOs are Möbius aromatic for the Si_6O_6 model. Accordingly, $25\% (= 75\% \times [8/24])$ of silica's valence electrons are Möbius aromatic. Analogously, only $6\% (= 75\% \times [2/24])$ of silica's valence electrons are Hückel aromatic.

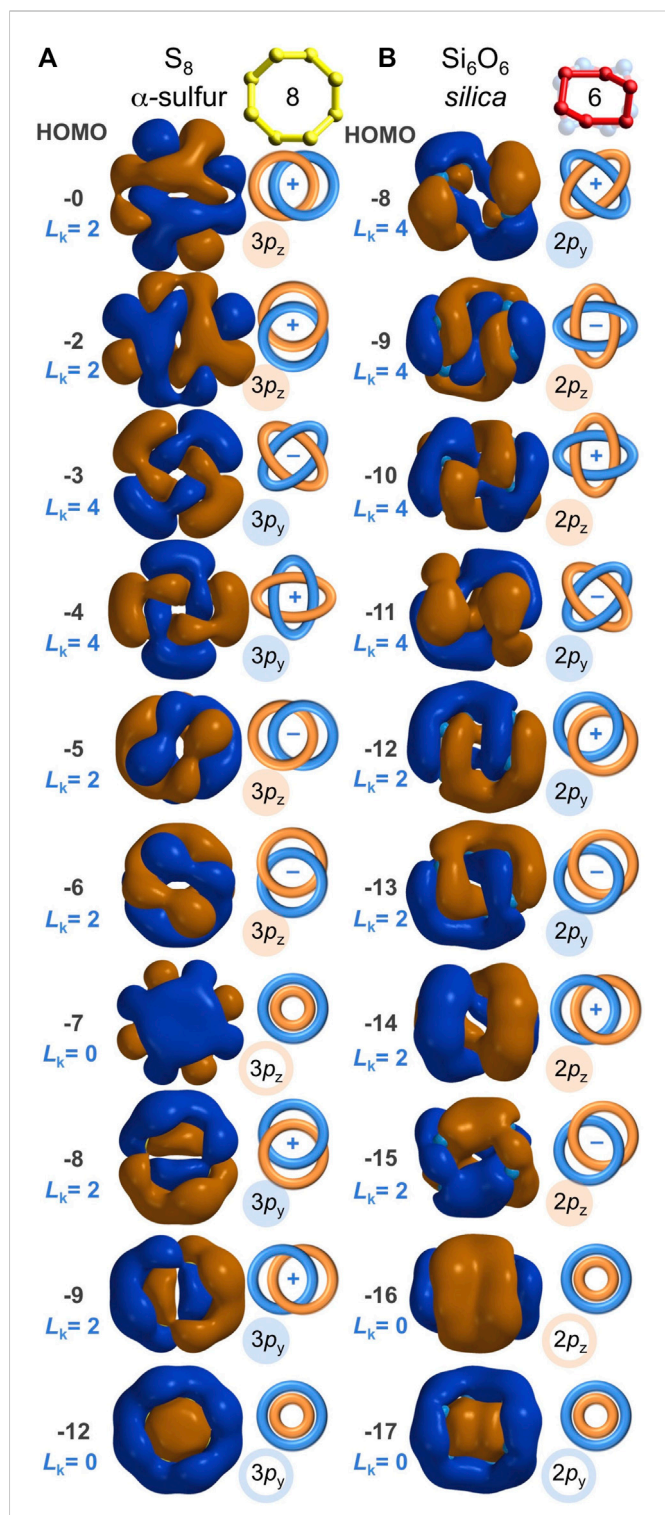


FIGURE 7

Analogous sulfur-based and oxygen-based valence molecular orbitals (IsoValue = 0.005 [e/bohr³]^{0.5}) for α -sulfur S_8 (A) and for Si_6O_6 found in quartz silica (B). These molecular orbitals are built from contiguous p orbitals with a bonding excess of $be_{3p} = 100\%$ and $be_{2p} = 100\%$, respectively, resulting in both Hückel ($L_k = 0$) and Möbius ($L_k = 2$ or 4) aromaticity. Their resemblance warrants equal expectations for covalent S-S bonding in sulfur and covalent O-O bonding in silica.

24 valence MOs of S_8 , eight are Möbius aromatic and two are Hückel aromatic—the same aromatic MO representation found in the silica Si_6O_6 ring (Figure 7B). The chalcogen-chalcogen bonding in S_8 has never been

disputed. The chalcogen-chalcogen bonding in silica's Si_6O_6 ring is equally evident and valid, given the molecular orbital and bonding parallels between S_8 and the O_6 ring embedded within silica.

2.5 Electron density analysis

The “longest O-O bond in any known molecule” was reported for HOON, studied in a supersonic molecular beam. Its O-O distance was measured to be 1.91 Å (Jabłoński, 2012; Crabtree et al., 2013; Chen et al., 2016). A computational analysis (Figure 8A) shows its O-O bond order is 0.41, its minimum O-O electron density (ED, core plus valence) is 0.09 e/bohr³, and its minimum O-O ionization potential (IP) is 14.1 eV. This last parameter is an index of the MO energy, with a greater IP implying greater MO stability and more bonding between the atoms. The same computation (Figure 8B) for the terminal oxygen atoms (2.16 Å) of ozone (O_3) (Hay et al., 1975) reveals a bond order of 0.59, a minimum O-O ED of 0.22 e/bohr³, and a minimum O-O IP of 25.5 eV. For an O-O bond (2.61 Å) in the center of the Si_{29} cluster, the same computation (Figure 8C) reveals a bond order of 0.63, a minimum O-O ED of 0.05 e/bohr³, and a minimum O-O IP of 16.8 eV. Which one of these three long O-O bonds can claim the most bonding? A bond order comparison suggests the least O-O bonding in the canonical bond of HOON (0.41) and more O-O bonding in the long, non-canonical bonds of ozone (0.59) and silica (0.63).

Bonding can also be gauged by the valence electron density (e/bohr³) along each O-O axis, as plotted in Figure 9A. Excluding core electron density⁵ the average valence electron density computes to $VED_{ave} = 0.46, 0.44$, and 0.35 e/bohr³ for HOON, O_3 , and the Si_{21} cluster, respectively. The valence electron density can also be parameterized according to the area under the curves of Figure 9A. With units of e/bohr² ($= [e/bohr^3] \times [bohr]$), this parameter measures valence electron accumulation, being a *projection* viewed down the O-O bond axis. This *bond axis valence electron density projection* computes to $VED_{proj} = 1.63, 1.79$, and 1.73 e/bohr², respectively, revealing the largest value for ozone, but the smallest value for HOON. Consequently, there are more valence electrons along a central O-O bond axis in Si_{21} than along that of HOON, indicating a new longest O-O bond of 2.64 Å belonging to quartz silica.

For comparison, analysis of a Si-O bond (1.61 Å) in the center of the Si_{29} cluster (Figure 8C) reveals a bond order of 0.63, a minimum Si-O ED (core plus valence) of 0.15 e/bohr³, and a minimum Si-O IP of 18.9 eV. For the Si_{21} cluster, the average Si-O valence electron density of $VED_{ave} = 0.33$ e/bohr³ (Figure 9A) is similar to that of its O-O bond ($VED_{ave} = 0.35$ e/bohr³)⁶. However, the shortness of the Si-O bond renders a minimal projection parameter of only $VED_{proj} = 0.99$ e/bohr², which is markedly smaller than that of the long O-O bond of silica (1.73 e/bohr²). This difference can be further visualized by the three-dimensional valence electron density surface

⁵ Valence bond path computations were performed with the *Chemissian* software program via Spartan Output files for HOON, O_3 , and the Si_{21} and Si_{29} silica clusters. See the Supplementary Material. *Chemissian*; version 4.67 www.chemissian.com.

⁶ As expected, this same analysis for the short, canonical O-O bond of ozone (1.26 Å) indicates metrics of greater bonding: Bond order = 1.44; minimum O-O ED = 0.47 e/bohr³; minimum O-O IP = 28.5 eV; $VED_{ave} = 0.68$ e/bohr³; $VED_{min} = 0.23$ e/bohr³; and $VED_{proj} = 1.60$ e/bohr².

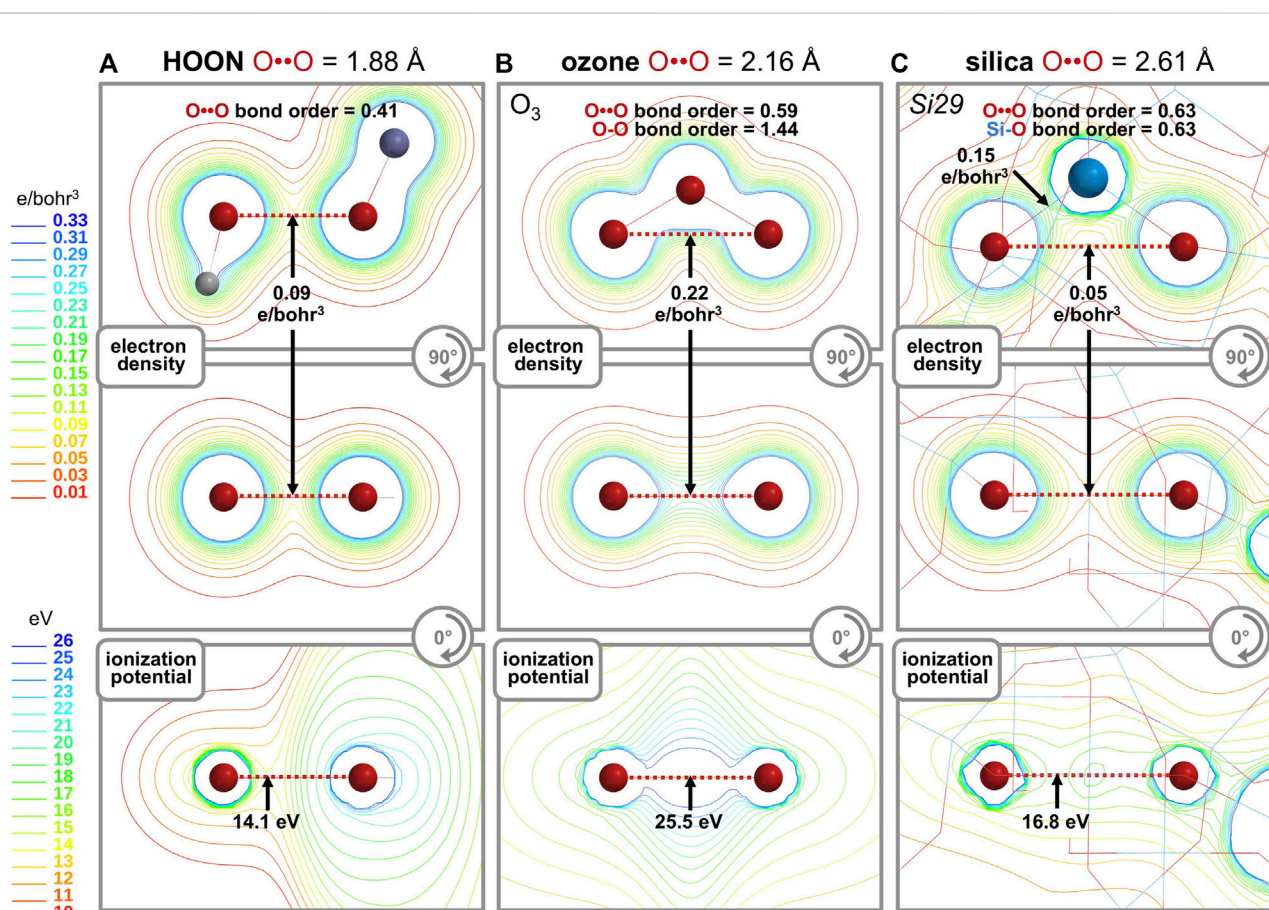


FIGURE 8

Computational analyses of HOON (A), ozone (B), and quartz silica (C) compare the O-O bond order, minimum O-O electron density (units of e/bohr^3 , core plus valence), and minimum O-O ionization potential (units of eV). Two orthogonal contour maps are shown for the electron density and a single contour map is shown for the ionization potential.

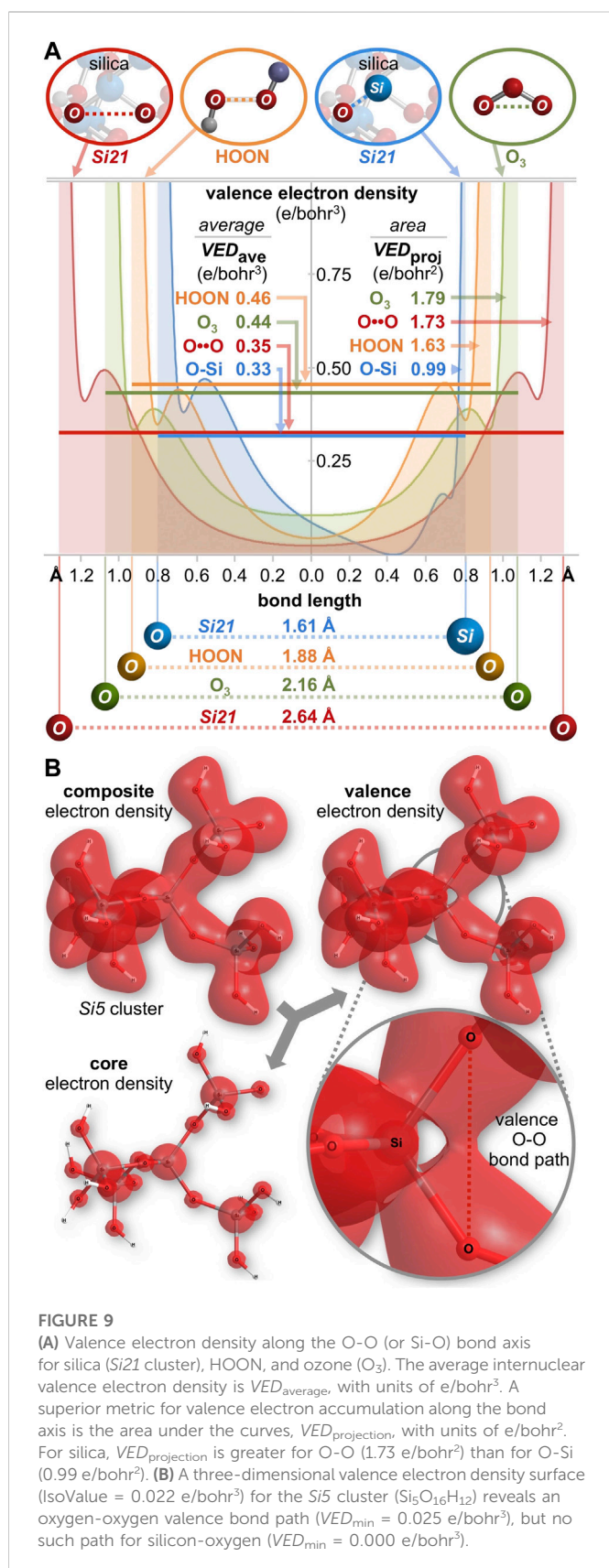
(IsoValue = $0.022 e/\text{bohr}^3$) of the Si_5 cluster ($\text{Si}_5\text{O}_{16}\text{H}_{12}$), shown in Figure 9B. There is a clear oxygen-oxygen valence bond path along the O-O axis with a minimum valence electron density of $\text{VED}_{\min} = 0.025 e/\text{bohr}^3$ at the midpoint of this long bond. However, there is no corresponding silicon-oxygen valence bond path along the Si-O axis since the VED_{\min} reaches $0.000 e/\text{bohr}^3$ approximately 0.4 \AA from the silicon atom—corresponding to the outer nodal sphere of the Si $3s$ atomic orbital. This electron density analysis comports with the Mulliken population analysis (orbital-based bond orders) and further corroborates the supposition that silica's valence electrons are not confined to their canonical locations.

2.6 Mulliken bond orders vs subsequent methods

Although Mulliken's overlap population method "permits one to identify chemically bonded atoms," it does not provide integral bond multiplicity values (corresponding to single, double, or triple bonds) (Mayer, 2003) and it is accompanied by other various objections (Reed et al., 1985). Historically, bond orders have been calculated via several generalized methods, including those described by Mulliken (1955b), Weinhold and Landis (2005), Wiberg (1968), Löwdin (1970), Mayer (1983), and Bader (1990)—although "none of them is the 'right' one."

(Lewars, 2008) A comprehensive comparison is beyond the scope of this manuscript. However, a preliminary analysis demonstrates that the Mulliken, Wiberg, and Mayer methods yield comparably large O-O bond orders within silica clusters. These computed bond orders vary by only ± 0.01 for a given silica cluster (see Supplementary Table S32) (Gorelsky and Lever, 2001)⁷. Furthermore, multicenter oxygen-based bond order indices have been computed and are proportional to silica cluster size. For the large Si_{11} – Si_{35} silica clusters, the 3-center 2-electron ($3c2e$) bond order index ranges from $I_{\text{OOO}} = 0.319$ to 0.383 ; these values exceed that of $I_{\text{BHB}} = 0.253$ computed for diborane (B_2H_6) as well as the theoretical⁷ and computed value (Bochicchio et al., 1998) of $8/27 \approx 0.296$ for H_3^+ . For the same large clusters, the 4-center 2-electron ($4c2e$) bond order index ranges from $I_{\text{OOOO}} = 0.110$ to 0.209 ; although such 4-center bonds are rare (de Giambiagi et al., 1997; Ponc and Mayer, 1997), these values mostly exceed that for B_4 ($I_{\text{BBBB}} = 0.118$) (Sannigrahi and Kar, 1999) and the value of

⁷ Wiberg bond order, Mayer bond order, multicenter Mayer-type bond order indices, and Mayer atomic valencies were calculated with the AOMix software package via the Spartan Output files for various silica cluster complexes. AOMix: Program for molecular orbital analysis. See the Supplementary Material. Gorelsky, S. I. version 6.94. <http://www.sg-chem.net/>.



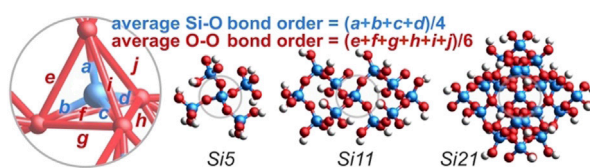
$I_{\text{O} \cdots \text{O}} = 0.209$ might be the largest 4c2e index reported to date. For comparison, multicenter bonding involving oxygen and silicon is uncommon within the silica clusters. There are no computed 3c2e

$I_{\text{O} \cdots \text{Si}}$ values greater than 0.02 and only three clusters have 4c2e $I_{\text{O} \cdots \text{Si}}$ values greater than 0.01, ranging from 0.023 to 0.141. This multicenter bonding analysis highlights the electron delocalization among oxygen atoms at the expense of silicon atom participation. An additional informative metric is the atomic valence (AV). For silicic acid (Si1), the Mulliken and Mayer atomic valencies are essentially canonical with $\text{Si}_{\text{AV}} \approx 4.33$ and $\text{O}_{\text{AV}} \approx 2.05$. However, in accord with the valence electron apportionment reported in Figure 3C, the Mulliken and Mayer atomic valencies invert for large silica clusters reaching, for example, $\text{Si}_{\text{AV}} \approx 3.14$ and $\text{O}_{\text{AV}} \approx 4.35$ for Si₂₉—confirming the high atomic valence of oxygen markedly exceeding the canonical value of 2. The Löwdin method routinely yields smaller O-O bond orders near 0.10, invariant to the silica cluster size or the location within the cluster. The Natural Bond Order (NBO) method fails with all silica clusters because of their electronic delocalization (Goodman and Sauers, 2007). The NBO method also cannot determine long bond orders in the following molecules: the terminal oxygen atoms of ozone (Hay et al., 1975); the transannular sp^2 hybridized carbons of norbornadiene (Hoffmann et al., 1970); and the transannular sulfur atoms of cyclic S₄N₄ (Bridgeman et al., 2001). For silica, a substantial Mulliken O-O bond order is required to accord with the resonance scheme of Figure 2B (BO = 1/3). Without such long covalent bonds, Earth's most abundant material fails to have a suitable resonance formulation. Not only did Mulliken deem his method “a good measure of covalent bonding,” (Mulliken, 1955b) he championed its “obvious” advantages for the interrogation of resonance and delocalized structures (Mulliken, 1955a).

2.7 Mulliken bond orders vs. computational methods and basis sets

Mulliken populations are often “unduly sensitive to basis set.” (Reed et al., 1985) Table 1 reports the Mulliken bond order sensitivity for the Si₅, Si₁₁, and Si₂₁ silica clusters subjected to various computational methods and basis sets. From the center of each cluster, the four Si-O bond orders and the six O-O bond orders are averaged to provide the values of Table 1. For the 6-311++G** basis set (used throughout this study), the bond order results are largely invariant to the computational method. Si-O bond orders are substantially less than unity and O-O bond orders are substantially greater than zero—converging as the cluster size increases and mimicking the plot for B3LYP/6-311++G** in Figure 3B. For the B3LYP method, basis set variation yields a broader range of bond orders. Pople-style basis sets (6-...) substantiate the claim of oxygen-oxygen bonding, although O-O bond orders (0.04–0.62) generally decrease with increasing basis set size. Dunning-style basis sets (cc-...) yield smaller O-O bond orders (0.00–0.30) and Ahlrichs/Weigend-style basis sets (def2-...) yield the broadest range of O-O bond orders (0.00–1.59). Note that the majority of computed Si-O bond orders are significantly below unity, arguing against the simplistic Lewis structure for silica having solely single Si-O bonds; negative Si-O bond orders are possibly an artifact of small cluster size and/or an indication of anti-bonding (Mulliken, 1955c). Without diffuse functionals in the basis set (see Supplementary Table S30), the Si-O bond order is near unity and the O-O bond order is near zero, yielding the simple Lewis structure having no option for resonance.

TABLE 1 Average computed Mulliken bond orders for the central SiO₄ unit within the *Si5* (Si₅O₁₆H₁₂), *Si11* (Si₁₁O₃₂H₂₀), and *Si21* (Si₂₁O₅₆H₂₈) clusters. Four bond order values are averaged for Si-O and six bond order values are averaged for O-O. (Not all computations converged).



Method	Basis set	Si5		Si11		Si21	
		Si-O	O-O	Si-O	O-O	Si-O	O-O
B3LYP	6-311++G**	0.74	0.28	0.80	0.36	0.68	0.46
B3LYP-D3	6-311++G**	0.74	0.28	0.80	0.36	0.68	0.46
HF	6-311++G**	0.65	0.33	0.77	0.41	0.63	0.50
ωB97X	6-311++G**	0.74	0.26	0.80	0.34	0.70	0.42
ωB97X-D	6-311++G**	0.76	0.27	0.83	0.35	0.71	0.44
ωB97X-V	6-311++G**	0.82	0.23	0.91	0.30	0.78	0.38
MP2	6-311++G**	0.65	0.33	0.78	0.41		
EDF2	6-311++G**	0.75	0.27	0.81	0.36	0.68	0.46
M06-2X	6-311++G**	0.74	0.27	0.75	0.35	0.59	0.46
VV10	6-311++G**	0.79	0.25	0.81	0.33	0.67	0.43
B97-D2	6-311++G**	0.80	0.27	0.89	0.34	0.73	0.44
ωB97X-V	6-311++G(2df,2p)	1.23	0.10	1.38	0.16	1.25	0.23
B3LYP	6-31+G	-0.13	0.52	0.55	0.36	0.79	0.13
B3LYP	6-31++G**	0.01	0.57	0.83	0.42	1.22	0.13
B3LYP	6-311++G(2d,p)	0.37	0.34	0.37	0.53	0.28	0.62
B3LYP	6-311++G(2df,2p)	1.25	0.10	1.38	0.18	1.24	0.25
B3LYP	6-311++G(3df,2p)	1.06	0.04	0.94	0.12		
B3LYP	aug-cc-pVDZ	-0.48	0.00	-0.27	0.06	-0.01	0.30
B3LYP	aug-cc-pVTZ	0.41	0.22	0.53	0.07		
B3LYP	aug-cc-pVQZ	0.43	0.12				
B3LYP	def2-SVPD	-2.61	1.59	0.89	0.63	0.50	0.21
B3LYP	def2-TZVPD	0.80	0.01	0.63	0.11		
B3LYP	def2-QZVPD	1.16	0.00				

2.8 Computed oxygen-oxygen bond dissociation energy for silica

If oxygen-oxygen bonding in silica is real, then perhaps an apt computational analysis could reveal silica's O-O bond dissociation energy (BDE). Conventionally, the reaction of **Figure 10A** between Si(OH)₄ and 3 SiH₄ is an isodesmic (Ponomarev and Takhistov, 1997) and homodesmotic (Wheeler, 2012) reaction since 4 Si-O bonds, 12 Si-H bonds, and 4 O-H bonds are maintained along with no hybridization changes. Unconventionally, however, six long O-O bonds in Si(OH)₄ are broken to yield 4 molecules of HOSiH₃ for which O-O bonding is not possible. The computed enthalpy of this reaction is $\Delta H = +27.0$ kcal/mol; dividing by six suggests a geminal O-O BDE value of 4.5 kcal/mol. Concerns about hydrogen

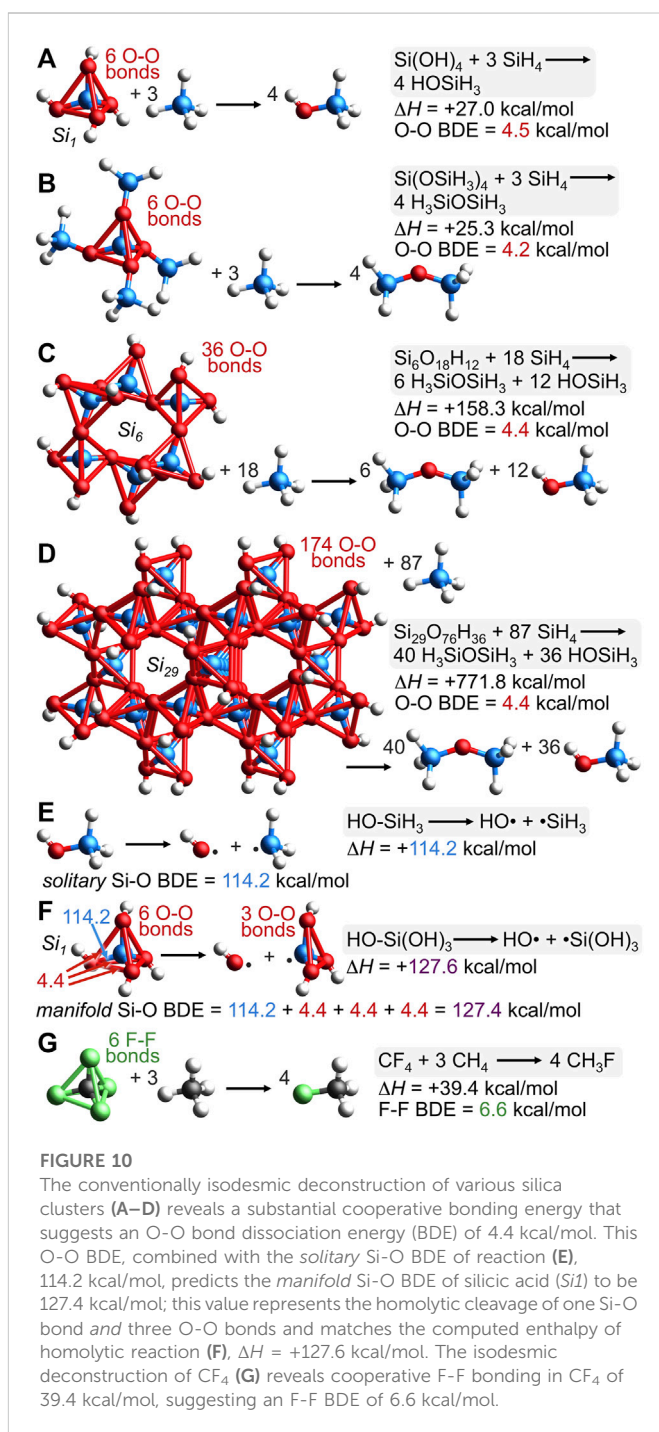
bonding within Si(OH)₄ are eliminated by the reaction of **Figure 10B** (also conventionally isodesmic and homodesmotic) because it has no possibility of hydrogen bonding, but has a similar reaction enthalpy of $\Delta H = +25.3$ kcal/mol and an O-O BDE value of 4.2 kcal/mol. Oxygen-oxygen BDE values for larger silica clusters, *Si6* and *Si29*, are computed *via* the conventionally isodesmic reactions of **Figures 10C, D**. In both cases, the O-O BDE values compute to 4.4 kcal/mol. Note that Si(OSiH₃)₄, the *Si6* cluster, and the *Si29* cluster possess siloxane oxygens (Si-O-Si)—thus requiring the production of H₃SiOSiH₃ to maintain the number and kind of bonds. If there were no cooperative bonding among the oxygen atoms of Si(OH)₄, Si(OSiH₃)₄, or the silica clusters, then these isodesmic reaction enthalpies (**Figures 10A–D**) should be zero. However, the cooperative stabilization is substantial and the

energetic cost for breaking a long O-O bond in silica averages to 4.4 kcal/mol.

Figure 10D illustrates the conventionally isodesmic deconstruction of the Si_{29} cluster. The dendriform array of 174 O-O bonds stabilizing this cluster engenders a new term to describe the pervasive O-O bonding that supplements the conventional Si-O bonding: *dendrivalent*. The 8 bonds emanating from a single oxygen atom (six to oxygen and two to silicon) simultaneously depict the bonds of multiple resonance forms, none being more than trivalent at oxygen (Figure 2B). This number of bonds will not alarm anyone fluent in resonance formulations. Such multivalent bonding is commonplace, as in the 6 bonds emanating from carbon in the resonance hybrids of the carbonate dianion (3 sigma and 3 pi) or the 9 bonds emanating from carbon in the resonance hybrids of polycyclic aromatic hydrocarbons or graphite (3 sigma, 3 pi, and arguably 3 transannular bonds of the Dewar benzene type (Sorella et al., 2014; Pauling and Wheland, 1933);⁸ see Supplementary Figures S19–S21 for these multiple resonance hybrids). Inspection of valence electron density surfaces for Si_5 or Si_{21} reveals clear valence bond paths between oxygen and all six of its oxygen neighbors (Figure 9B; Supplementary Figures S14–S17). As noted above, the outer spherical node of the $\text{Si } 3s$ orbital precludes silicon-oxygen valence bond paths (Figure 9, $\text{VED}_{\min} = 0.000 \text{ e/bohr}^3$).

While silica employs more O-O valence electrons than Si-O valence electrons according to Figure 3C, the greater O-O distance ensures that the O-O BDE of 4.4 kcal/mol is much less than that of Si-O, which computes to 114.2 kcal/mol according to the homolytic bond cleavage of HO-SiH_3 , depicted in Figure 10E. Furthermore, this *solitary* Si-O BDE in HO-SiH_3 can be used to predict the *manifold* Si-O BDE in HO-Si(OH)_3 , which should cost 114.2 kcal/mol plus three times the aforementioned O-O BDE value since such a reaction breaks one Si-O bond and three O-O bonds, as depicted in Figure 10F. The prediction is $[114.2 + 4.4 + 4.4 + 4.4] \text{ kcal/mol}$ or 127.4 kcal/mol, which is very close to the enthalpy computed for the homolytic bond cleavage of HO-Si(OH)_3 , $\Delta H = +127.6 \text{ kcal/mol}$.

The similarity of these values indicates that a consistent and predictive bond energy additivity scheme can be formulated by proper inclusion of both long and short covalent bonds. For example, the greater C-F bond strength in CF_4 (130.5 kcal/mol) versus that in CH_3F (109.9 kcal/mol) has been rationalized by electronegativity and atomic charge effects (Lemal, 2004). However, a bond energy additivity scheme, analogous to that for silica in Figure 10, provides a simpler explanation. The computational isodesmic reaction of $\text{CF}_4 + 3 \text{ CH}_4 \rightarrow 4 \text{ CH}_3\text{F}$ has $\Delta H = +39.4 \text{ kcal/mol}$ (Figure 10G), suggesting an F-F BDE of 6.6 kcal/mol since this reaction breaks six long (2.17 Å), unconventional F-F bonds. Separately, the *manifold* C-F BDE in F-CF_3 , which represents one C-F bond and three F-F bonds, is predicted by adding the experimental *solitary* C-F BDE in F-CH_3 to three F-F BDE values: $[109.9 + 6.6 + 6.6 + 6.6] \text{ kcal/mol} = 129.7 \text{ kcal/mol}$. This result is strikingly close to the experimental value of 130.5 kcal/mol and does not invoke nebulous arguments about electronegativity/atomic charge (Lemal, 2004), negative hyperconjugation (Hine, 1963; Reed and von Ragué Schleyer, 1987), or Coulombic interactions (Wiberg and Rablen, 1993), but elaborates on double bond/no-bond resonance formulations (Dolbier et al., 1982) by adding six



new resonance hybrids, each with a long F-F bond, as in $[\text{F-C-F}][\text{F-F}]$ (explicitly drawn in Supplementary Figure S22). A valence molecular orbital analysis corroborates the F-F bond in this unconventional resonance hybrid for CF_4 since $be_{2s} = 0\%$ and $be_{2p} = 39\%$, yielding a composite *bonding excess* greater than zero at $be_{\text{F-F}} = 29\%$, which is rather close to $be_{\text{O-O}} = 27\%$ for the silicic acid model $\text{Si}(\text{OSi})_4$ (Figure 4), or $be_{\text{O-O}} = 21\%$ for silicic acid itself. To recapitulate: CF_4 and silica clusters have a substantial excess of bonding versus anti-bonding F-F or O-O interactions that mandate an F-F or O-O bond dissociation energy of non-zero magnitude. While these long BDE values are small compared to those of short bonds, there are six per tetrahedron and thus, they stabilize CF_4 by 39.4 kcal/mol and silica

⁸ Herein, the Mulliken bond order for the central transannular sigma bond (Dewar type) computes to 0.07 for coronene or 0.11 for hexabenzocoronene.

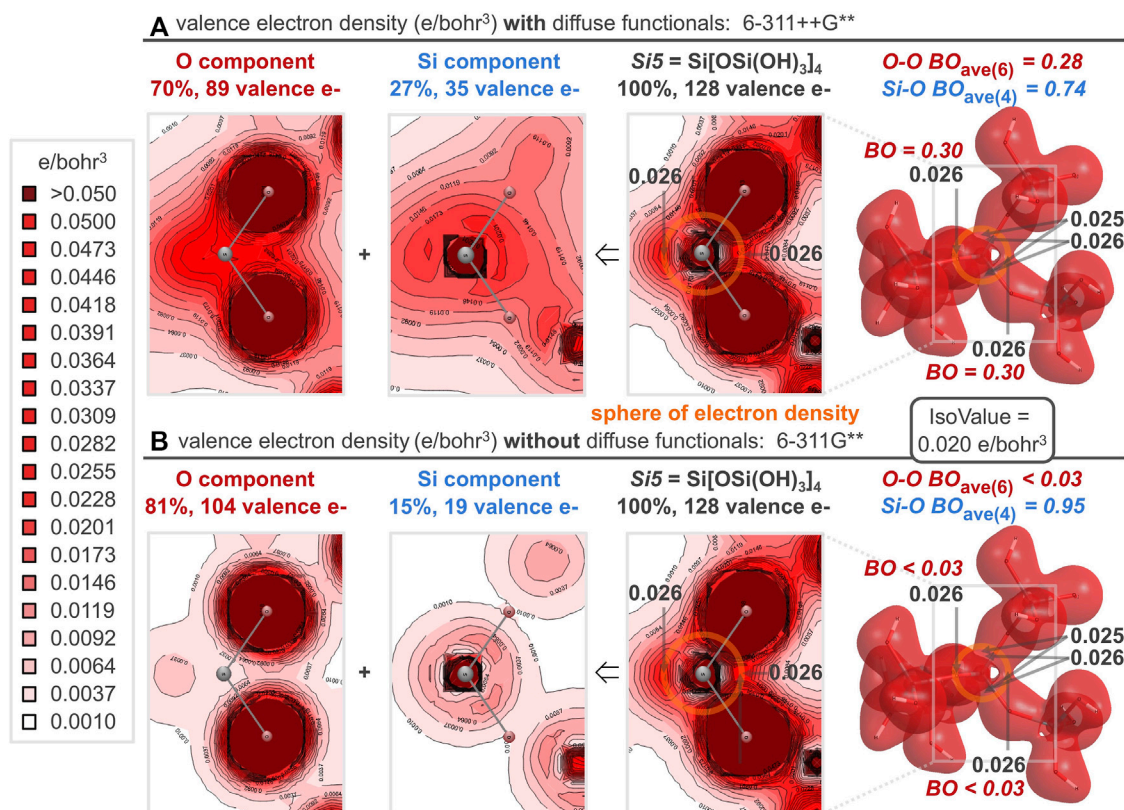


FIGURE 11

For the Si5 silica cluster, the computed valence electron density distribution *with* diffuse functionals (A) is nearly identical to that *without* diffuse functionals (B), despite differential employment of the constituent oxygen and silicon atomic orbitals. This differential accounting of each electron's orbital origin results in non-canonical (A) or canonical (B) O-O and Si-O bond orders. Nonetheless, both computational methods show a spherical node between silicon and oxygen, but a definitive accumulation of electron density between all geminal oxygens, resulting in a *sphere of electron density* (orange circle) responsible for the cooperative bonding in silica.

clusters by 26.4 kcal/mol of SiO₂—thereby augmenting the measured strength of *manifold* C-F and Si-O bonds to values considerably greater than those of *solitary* C-F and Si-O bonds.

2.9 Ball-and-stick model vs cooperative sphere model

For silica, it is evident that silicon-oxygen bond dissociation energies are cooperative (Figures 10E, F) and there are no valence bond paths between oxygen and silicon because of an intervening nodal surface (Figures 9A, B). Hence, a ball-and-stick bonding model—germane to water, methane, and many other molecules—is too simplistic to accurately describe the bonding in silica. For Si[OSi(OH)₃]₄ (the Si5 silica cluster), the oxygen and silicon contributions to the valence electron density can be functionally separated into the 2D plots of Figure 11A⁹.

Electron density from the silicon 3s and 3p atomic orbitals accumulates about 0.8 Å from silicon, coinciding with the midpoint of the geminal O-O axes, where oxygen 2s and 2p atomic orbitals also accumulate electron density. Thus, silicon and oxygen are perfectly matched to create a *sphere of electron density* 0.8 Å from silicon, about halfway along the silicon-oxygen axes, which are 1.61 Å long; Figure 11A locates this sphere (in orange) just beyond the outer spherical node of the Si 3s atomic orbital, which encircles silicon 0.4 Å away (where VED = 0 in Figure 9A). Because of this nodal surface, the silicon-oxygen bond of silica does not merit a canonical line between the nuclei; instead, bonding is better defined by a *cooperative sphere model*, where the valence electron density accumulates in a spherical region 0.8 Å from silicon and 0.8–1.3 Å from oxygen.

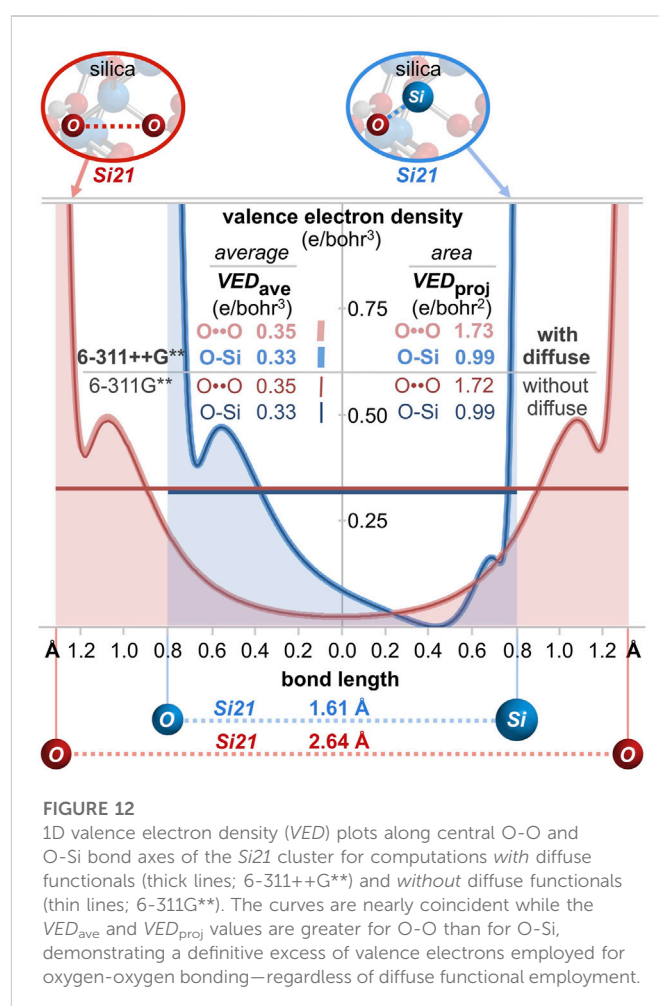
Between each pairwise O-O interaction of the Si5 central silicate, the valence electron density bond paths reach minima of 0.026, 0.026, 0.026, 0.026, 0.025, and 0.025 e/bohr^3 (Figure 11A). Notably, the same analysis computed *without* diffuse functionals (Figure 11B) yields the exact same minima between each pairwise O-O interaction and generates indistinguishable valence electron density 3D surfaces and 2D plots. Both diffuse and non-diffuse computational methods aspire to the same valence electron density distribution—with equivalent accumulation of electron density between all geminal oxygen atoms and a nodal surface between canonically bonded silicon and oxygen atoms. However, the two computational methods employ oxygen and silicon atomic orbitals differently to

⁹ The valence electron density of Si5 was parsed to the oxygen valence electron density component by setting all silicon and hydrogen coefficients to zero in the Spartan Output (.txt) file prior to the *Analyze MO compositions* function in *Chemission*. The silicon valence electron density component was similarly computed after setting all oxygen and hydrogen coefficients to zero.

build the valence molecular orbitals. With diffuse functionals, the valence molecular orbitals are 70% oxygen-based and 27% silicon-based. This allocation is more balanced than the computation without diffuse functionals, for which the valence molecular orbitals are 81% oxygen-based and 15% silicon-based. (In both cases, about 3% of the valence molecular orbitals are hydrogen-based.) A purely canonical model intervenes with valence molecular orbitals that are 75% oxygen-based (96 e/128 e) and 16% silicon-based (20 e/128 e). Furthermore, with diffuse functionals, the oxygen component shows significant overlap of the oxygen atomic orbitals (Figure 11A). This results in substantially greater O-O bond orders prescribed by diffuse functionals, averaging 0.28 for the six O-O bonds of the Si5 central silicate, and commensurately smaller Si-O bond orders, averaging 0.74 for the four Si-O bonds of the Si5 central silicate. Without diffuse functionals, the oxygen component shows minimal overlap of the oxygen atomic orbitals (Figure 11B). This results in small O-O bond orders near the canonical value of zero (all below the computational threshold of 0.025) and Si-O bond orders near the canonical value of 1.0 (average = 0.95). Since the diffuse and confined computational models allocate electrons differently to atomic orbitals, they differ in the resultant bond orders. Nonetheless, the valence electron density is unmistakably similar, as exhibited by the 3D and 2D valence electron density plots of Figure 11 as well as the 1D plots of Figure 12 for the Si21 silica cluster. Specifically, the VED, VED_{ave} , and VED_{proj} along O-O or O-Si axes are essentially invariant to the employment of diffuse functionals. But, regardless of whether the computational model includes or excludes diffuse functionals, more valence electrons reside between oxygen atoms ($VED_{ave} = 0.35$; $VED_{proj} = 1.73$ or 1.72) than between silicon and oxygen atoms ($VED_{ave} = 0.33$; $VED_{proj} = 0.99$).

2.10 Overlap population density of states analysis

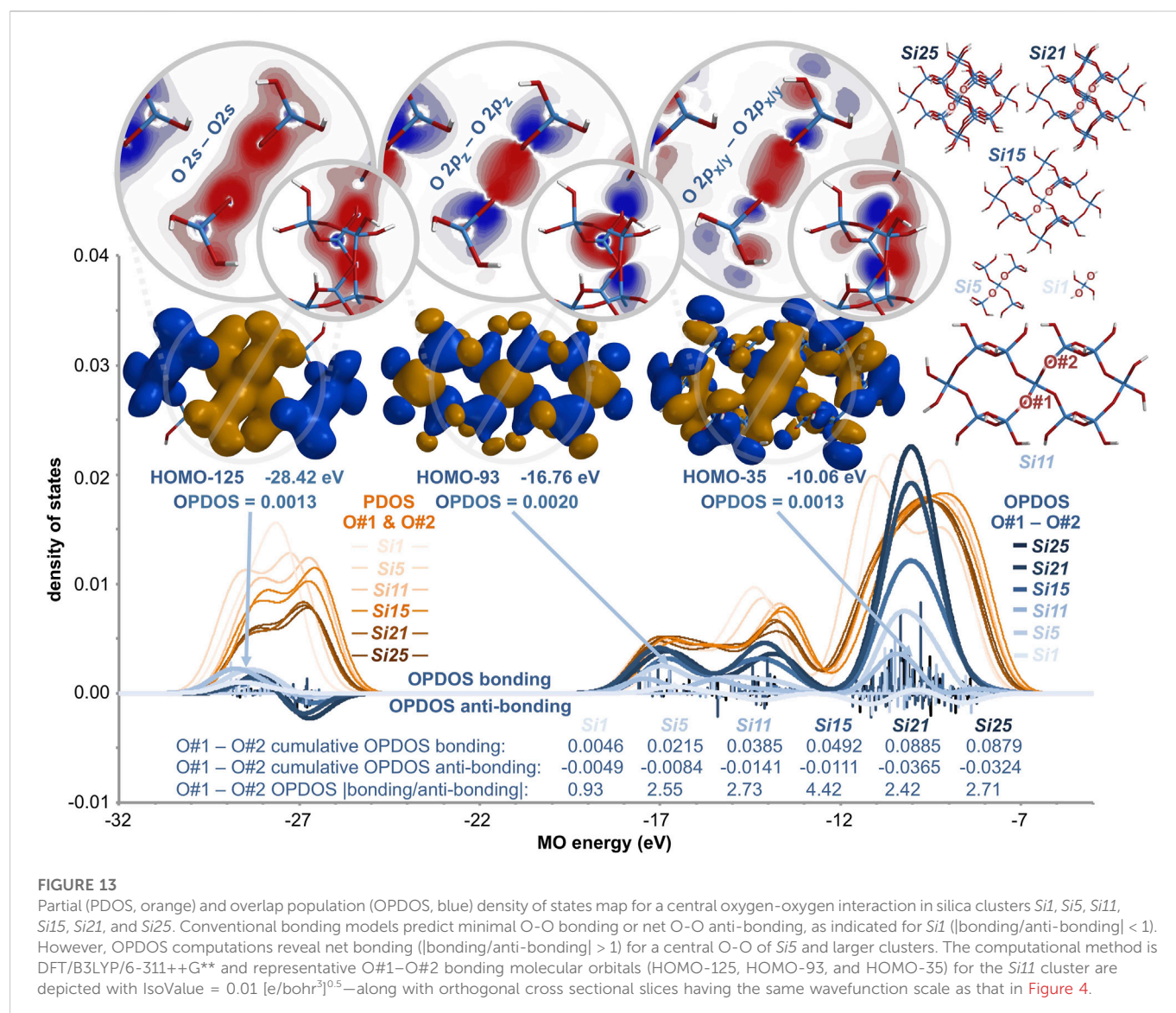
Additional bonding insight for silica is provided by the valence density of states map (Lu and Chen, 2012) shown in Figure 13. For a central oxygen-oxygen interaction in the silica clusters Si1, Si5, Si11, Si15, Si21, and Si25, the partial density of states (PDOS, orange thin lines) is plotted as a function of molecular orbital energy. Also plotted is the overlap population density of states (OPDOS) (Hughbanks and Hoffmann, 1983), which reveals the bonding states (positive) and anti-bonding states (negative) for a series of molecular orbitals—in analogy to the crystal orbital overlap population (COOP) (Grechnev et al., 2003) method applicable to extended solids. Classical bonding models for silica predict non-bonding or somewhat anti-bonding O-O interactions and thus, OPDOS values near zero as a function of MO energy. The generated O-O OPDOS curve for the Si1 cluster (thick light blue line) is relatively flat and indicates net anti-bonding, as measured by the sub-unity cumulative OPDOS [bonding/anti-bonding] quotient of 0.93. All extant bonding models predict similarly flat OPDOS curves for any O-O interaction, in any silica cluster, of any size. However, for clusters Si5 and larger, the curves are not flat and the OPDOS [bonding/anti-bonding] quotient exceeds unity, ranging from 2.42 to 4.42. For these larger clusters, the cumulative O-O OPDOS bonding parameter is proportional to the cluster size and the generated OPDOS curves are positive at nearly all MO energy levels. For the largest cluster, Si25, the



OPDOS curve closely tracks the PDOS curve and the cumulative OPDOS value (0.088) reaches 43% of the cumulative PDOS value (0.102 + 0.101).

In the molecular orbital region of -29 to -26 eV for all six clusters, the OPDOS curves (Figure 13, thick blue lines) are both positive and negative. This suggests that MOs built from O 2s atomic orbitals do not provide substantial net bonding. However, in the molecular orbital region of -18 to -8 eV, the OPDOS curves become increasingly positive with increasing silica cluster size. This suggests that O 2p atomic orbital overlap is largely responsible for the oxygen-oxygen bonding in silica. Exemplary molecular orbitals are shown in Figure 13 for the Si11 cluster. Oxygen atom #1 and oxygen atom #2 have σ bonding interactions via their 2s atomic orbitals in the HOMO-125, their $2p_z$ atomic orbitals in the HOMO-93, and their $2p_{x/y}$ atomic orbitals in the HOMO-35. Orthogonal cross sections of these MOs locate the accumulation of electron density along the O-O axis. Importantly, the OPDOS analysis shows that increasing cluster size augments the bonding for the O $2p_{x/y}$ -O $2p_{x/y}$ interactions (-12 to -8 eV) more than the O $2p_z$ -O $2p_z$ interactions (-18 to -12 eV).

What is the basis for this bonding augmentation? A detailed OPDOS analysis of all six central O-O valence interactions in the Si1-Si25 clusters (Table 2) reveals two effects. The first effect concerns the number of bonding vs anti-bonding interactions. Adding oxygen atoms to the periphery of a silica cluster decreases the nodal density between oxygen atoms in the center



of the silica cluster. The nodal density is inversely related to the OPDOS % *bonding excess* reported in Table 2. While the % *bonding excess* is somewhat invariant to cluster size for the O 2s MOs (–29 to –26 eV) and the O 2p_z MOs (–18 to –12 eV), this parameter increases correspondingly with cluster size for the O 2p_{x/y} MOs (–12 to –8 eV)—from –38% to 26%—fully consistent with the dominant growth of the OPDOS curve in the same region (Figure 13). The second effect concerns the *magnitude* of the bonding vs anti-bonding interactions. Adding oxygen atoms to the *periphery* of a silica cluster increases the cumulative OPDOS bonding parameter more than the cumulative anti-bonding OPDOS parameter for the oxygen atoms in the *center* of the silica cluster. Thus, the $|\text{bonding}/\text{anti-bonding}|$ quotient steadily increases from 0.76 to 2.42 with increasing cluster size (Table 2). (Note that the $|\text{bonding}/\text{anti-bonding}|$ quotient in Figure 13 represents only one of the six O-O interactions in the central silicate unit of the clusters—all of which are averaged to yield the values in Table 2.) It is clear that the oxygen-oxygen bonding metrics of a silicate unit are sensitive to the number of peripheral

silicate units. Assuredly, this effect tapers with cluster size, but the trends all indicate considerable covalent O-O bonding in bulk silica.

3 Conclusions and outlook

A molecular orbital analysis of α -quartz silica model complexes reveals that oxygen valence electrons abandon their canonically prescribed locations to form long covalent oxygen-oxygen bonds. Oblique arrangements of the oxygen atoms minimize molecular orbital nodes and maximize bonding interactions that are 2.61–2.64 Å apart, yet have Mulliken bond orders reaching 0.63 and averaging 0.47. This nodal minimization is inherently challenging for linear or planar molecules, but may prove widespread among atoms arranged in three dimensions—accomplished especially *via p* atomic orbitals. Thereby, the geminal oxygen atoms of silica bond cooperatively with a computed O-O bond dissociation energy of 4.4 kcal/mol

TABLE 2 Bonding vs anti-bonding interactions for the six central silicate oxygen-oxygen pairs according to overlap population density of states (OPDOS) analysis for silica clusters *Si1–Si25*.

cluster	<i>Si1</i>	<i>Si5</i>	<i>Si11</i>	<i>Si15</i>	<i>Si21</i>	<i>Si25</i>
# oxygen atoms	4	16	32	42	56	66
# valence MOs	16	64	128	168	224	264
# O-O interactions (+& –)	96	384	768	1,008	1,344	1,584
# O-O bonding (+)	43	219	462	623	826	991
2s	16	63	118	153	188	248
2p _z	12	60	129	163	218	245
2p _x and 2p _y	15	96	215	307	420	498
# O-O anti-bonding (–)	53	165	306	385	518	593
2s	8	33	74	99	148	148
2p _z	12	36	63	89	118	151
2p _x and 2p _y	33	96	169	197	252	294
OPDOS % bonding excess	–10	14	20	24	23	25
2s	33	31	23	21	12	25
2p _z	0	25	34	29	30	24
2p _x and 2p _y	–38	0	12	22	25	26
cumulative O-O OPDOS						
cumulative bonding	0.003	0.015	0.024	0.037	0.067	0.079
cumulative anti-bonding	–0.004	–0.008	–0.013	–0.017	–0.030	–0.032
bonding/anti-bonding	0.76	1.85	1.78	2.17	2.26	2.42
average Mulliken O-O BO	0.05	0.28	0.36	0.38	0.46	0.46

when they are suitably arranged and interspersed with the electropositive element silicon. Generally, as silica model complexes increase in size, canonical bonding paradigms decrease in accuracy and resonance hybrids increase in relevance. This comports with the original Hund-Mulliken molecular orbital theory, which states that the “best MOs . . . spread at least to some slight extent over all atoms.” (Mulliken, 1970) For Pauling, the concept of resonance had “its most important chemical applications” to “molecules to which no satisfactory single structure in terms of single bonds, double bonds, and triple bonds can be assigned.” (Pauling, 1946) Indeed, the structure, bonding, and energetic stability of silica cannot be fully understood without resonance hybrids involving long oxygen-oxygen bonds. This covalent bonding likely exists between other distant atoms and promises to impact the understanding of many materials and processes.

The hybridization theory of Pauling (Pauling, 1930; Pauling, 1931) and Slater (Slater, 1931) compelled chemists to reconsider the location of nearly *all* valence electrons. The curious bonding in ferrocene (Fischer and Pfab, 1952; Wilkinson et al., 1952), with multiple metal-carbon bonds, also provided a paradigm shift in our understanding of bonding and the location of electrons; however, this bonding arrangement pertains only to a class of esoteric, man-made organometallic sandwich compounds. The bonding in diborane (Lipscomb, 1973), with three-center two-electron bonds, provided another paradigm shift in our understanding of bonding arrangements in molecules; but this bonding motif is also uncommon, primarily ascribed to a limited number of main group and transition metal complexes (Longuet-Higgins and Roberts, 1955). The long covalent bond theory (LCBT) posited herein signals a new paradigm shift in the location of the chemical

bond. As evidenced by resonance formulations, bond orders (Mulliken, Wiberg, Mayer), multicenter bond order indices (3-center and 4-center), atomic valencies (Mulliken, Mayer), molecular orbitals (bonding > anti-bonding interactions), bonding analogies, electron density calculations, valence bond path calculations, overlap population density of states (OPDOS) analysis, and bonding energetics, it is clear that Nature builds materials with long covalent bonds—not just the short canonical bonds introduced by Couper (1858) and Kekulé (1866) whose primitive bonding model has somehow reigned over 160 years. These long bonds underpin an explanation for the contorted, chiral structure and energetic stability of α -quartz silica, wherein distant oxygen-oxygen bonding supplements conventional bonding. This bonding paradigm is abundant and pervasive since more than one-third of the 10^{49} valence electrons in the Earth’s crust are allocated to long covalent bonds¹⁰. Moreover, 10^{48} crustal valence electrons inhabit silica’s oxygen-based Möbius aromatic orbitals—manifestly the most prevalent sort of aromaticity. Astonishingly, LCBT implicates the oxygen-oxygen bond as the most abundant bond on Earth. While this study focuses on silica, future work will reveal the prevalence, energetics, and importance of long covalent bonds in a rather wide variety of materials—especially those with periodic structures, including ice, biopolymers, bone, and superconducting ceramics.

¹⁰ Silica constitutes 59% of the Earth’s crust (Clarke and Washington, 1924) and provides about 66% of crustal valence electrons. Since over half (52%) of these electrons are spent on oxygen-oxygen bonding, over one-third of crustal valence electrons participate in the long O-O covalent bonds of silica.

4 Experimental section

All calculations herein are performed with Spartan/Q-Chem according to DFT/B3LYP/6-311++G**, except where noted in Table 1, Figure 11B, and Figure 12 (Shao et al., 2015). Spartan Output files were processed with AOMix: Program for Molecular Orbital Analysis (version 6.94, written by S. I. Gorelsky) (Gorelsky and Lever, 2001)⁷ or with Chemissian (version 4.67, written by L. Skripnikov)⁵ to generate additional computational metrics, as noted. OPDOS plots were computed with Multiwfn (version 3.8, written by T. Lu and F. Chen) (Lu and Chen, 2012) via Gaussian 09 formatted checkpoint files (Frisch et al., 2009).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

References

- Bader, R. F. W. (1990). *Atoms in molecules: A quantum theory*. Oxford, UK: Oxford University Press. <https://global.oup.com/academic/product/atoms-in-molecules-9780198558651>.
- Bochicchio, R. C., Ponc, R., Lain, L., and Torre, A. (1998). On the physical meaning of bond indices from the population analysis of higher order densities. *J. Phys. Chem. A* 102, 7176–7180. doi:10.1021/jp981816d
- Bridgeman, A. J., Cavigliasso, G., Ireland, L. R., and Rothery, J. (2001). The Mayer bond order as a tool in inorganic chemistry. *J. Chem. Soc. Dalton Trans.* 2001, 2095–2108. The transannular S-S Mayer bond order (2.67 Å) in S₄N₄ reportedly ranges from 0.22 to 0.56, while its Mulliken bond order is herein computed to be 0.40. doi:10.1039/b102094n
- Chen, S., Xu, Z., and Li, J. (2016). The observation of oxygen-oxygen interactions in ice. *New J. Phys.* 18, 023052. A long “oxygen-oxygen interaction” in various pressurized ices (2.5–2.7 Å) has been proposed to explain vibrational spectra, but “a detailed mechanism is still unclear.” doi:10.1088/1367-2630/18/2/023052
- Clarke, F. W., and Washington, H. S. (1924). *The composition of the Earth's crust*. Washington, DC: USGS Professional Paper 127, Department of the Interior, United States Geological Survey, U.S. Government Printing Office. doi:10.3133/pp127
- Couper, A. S. (1858). Sur une nouvelle théorie chimique. *Ann. Chim. Phys.* 53, 469–489. <https://gallica.bnf.fr/ark:/12148/bpt6k34794n/f468.item>.
- Crabtree, K. N., Talipov, M. R., Martinez, O., Jr., O'Connor, G. D., Khursan, S. L., and McCarthy, M. C. (2013). Detection and structure of HOON: Microwave spectroscopy reveals an O–O bond exceeding 1.9 Å. *Science* 342, 1354–1357. doi:10.1126/science.1244180
- de Giambiagi, M. S., Giambiagi, M., and de Souza Fortes, M. (1997). Multicenter bonds, bond valence and bond charge apportionment. *J. Mol. Struct. (Theochem)* 391, 141–150. doi:10.1016/S0166-1280(96)04815-4
- Dolbier, W. R., Jr., Medinger, K. S., Greenberg, A., and Liebman, J. F. (1982). The thermodynamic effect of fluorine as a substituent: Vinylic CF₂ and CFH and allylic CF₂C. *Tetrahedron* 38, 2415–2420. doi:10.1016/0040-4020(82)87020-8
- Feynman, R. P. (1939). Forces in molecules. *Phys. Rev.* 56, 340–343. doi:10.1103/PhysRev.56.340
- Fischer, E. O., and Pfab, W. (1952). Cyclopentadien-Metallkomplexe, ein neuer Typ metallorganischer Verbindungen. *Z. Naturforsch. B* 7, 377–379. doi:10.1515/znb-1952-0701
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09, revision D.01*. Wallingford CT: Gaussian, Inc. <https://gaussian.com/g09citation/>.
- Garvie, L. A. J., Rez, P., Alvarez, J. R., Buseck, P. R., Craven, A. J., and Brydson, R. (2000). Bonding in alpha-quartz (SiO₂): A view of the unoccupied states. *Am. Mineral.* 85, 732–738. doi:10.2138/am-2000-5-611
- Gibbs, G. V., Wallace, A. F., Cox, D. F., Downs, R. T., Ross, N. L., and Rosso, K. M. (2009). Bonded interactions in silica polymorphs, silicates, and siloxane molecules. *Am. Mineral.* 94, 1085–1102. doi:10.2138/am.2009.3215
- Goodman, L., and Sauers, R. R. (2007). Diffuse functions in natural bond orbital analysis. *J. Comput. Chem.* 28, 269–275. doi:10.1002/jcc.20519
- Gorelsky, S. I., and Lever, A. B. P. (2001). Electronic structure and spectra of ruthenium diimine complexes by density functional theory and INDO/S. Comparison of the two methods. *J. Organomet. Chem.* 635, 187–196. doi:10.1016/S0022-328X(01)01079-8
- Grechnev, A., Ahuja, R., and Eriksson, O. (2003). Balanced crystal orbital overlap population—A tool for analysing chemical bonds in solids. *J. Phys. Condens. Matter* 15, 7751–7761. doi:10.1088/0953-8984/15/45/014
- Hay, P. J., Dunning, T. H., and Goddard, W. A., III (1975). Configuration interaction studies of O₃ and O₃⁺. Ground and excited states. *J. Chem. Phys.* 62, 3912–3924. Ozone reportedly has “weak bonding between the... terminal oxygen atoms.” doi:10.1063/1.430306
- Hine, J. (1963). Polar effects on rates and equilibria. VIII. Double bond-no bond resonance. *J. Am. Chem. Soc.* 85, 3239–3244. doi:10.1021/ja00903a041
- Hoffmann, R. (1963). An extended Hückel theory. I. Hydrocarbons. *J. Chem. Phys.* 39, 1397–1412. doi:10.1063/1.1734456
- Hoffmann, R., Heilbronner, E., and Gleiter, R. (1970). Interaction of nonconjugated double bonds. *J. Am. Chem. Soc.* 92, 706–707. An early version of *alternocojugation* described norbornadiene with alternating pi bonds (C2=C3 and C5=C6) and sp³ hybridized carbons (C1 and C4). Presently, a computational analysis of norbornadiene indicates long carbon-carbon bonds (C2–C6 and C3–C5) with a length of 2.49 Å and a Mulliken bond order of 0.19 via interaction of distant C 2p orbitals. doi:10.1021/ja00706a051
- Hoffmann, R. (1971). Interaction of orbitals through space and through bonds. *Acc. Chem. Res.* 4, 1–9. doi:10.1021/ar50037a001
- Hughbanks, T., and Hoffmann, R. (1983). Chains of trans-edge-sharing molybdenum octahedra: Metal-metal bonding in extended systems. *J. Am. Chem. Soc.* 105, 3528–3537. doi:10.1021/ja00349a027
- Jabloniński, M. (2012). Energetic and geometrical evidence of nonbonding character of some intramolecular Halogen...Oxygen and other Y...Y interactions. *J. Phys. Chem. A* 116, 3753–3764.

Funding

This research was supported by the National Science Foundation (CHE-1904768) and the University of Florida.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2023.1123322/full#supplementary-material>

- Longer oxygen-oxygen interactions in organic molecules (e.g., 2.71–2.90 Å in HO-(CR)₃=O) were studied computationally, but deemed “most likely” repulsive and presently are calculated to have small bond orders ranging from 0.03 to 0.06. doi:10.1021/jp300993b
- Janes, N., and Oldfield, E. (1986). Oxygen-17 NMR study of bonding in silicates: The d-orbital controversy. *J. Am. Chem. Soc.* 108, 5743–5753. doi:10.1021/ja00279a014
- Kekulé, A. (1866). *Lehrbuch der Organischen chemie*. Second Edition. Erlangen: Verlag von Ferdinand Enke. See page 496 for an early structure of benzene <https://books.google.com/books?id=9KMEAAAAYAAJ>.
- Kirfel, A., and Eichhorn, K. (1990). Accurate structure analysis with synchrotron radiation. The electron density in Al₂O₃ and Cu₂O. *Acta Cryst.* A46, 271–284. doi:10.1107/S0108767389012596
- Lemal, D. M. (2004). Perspective on fluorocarbon chemistry. *J. Org. Chem.* 69, 1–11. doi:10.1021/jo302556
- Levien, L., Prewitt, C. T., and Weidner, D. J. (1980). Structure and elastic properties of quartz at pressure. *Am. Mineral.* 65, 920–930. For all model complexes, silicon and oxygen atoms are restricted to the atomic coordinates of α-quartz silica. <https://pubs.geoscienceworld.org/msa/ammin/article-abstract/65/9-10/920/41195>.
- Lewars, E. G. (2008). “Pyramidal carbon (chapter 2),” in *Modeling marvels: Computational anticipation of novel molecules* (Netherlands: Springer Science+Business Media B.V.), 13–29. doi:10.1007/978-1-4020-6973-4
- Lipscomb, W. N. (1973). Three-center bonds in electron-deficient compounds. Localized molecular orbital approach. *Acc. Chem. Res.* 6, 257–262. doi:10.1021/ar50068a001
- Longuet-Higgins, H. C., and Roberts, M. D. V. (1955). The electronic structure of an icosahedron of boron atoms. *Proc. R. Soc. Lond. Ser. A* 230, 110–119. doi:10.1098/rspa.1955.0115
- Löwdin, P.-O. (1970). On the nonorthogonality problem. *Adv. Quantum Chem.* 5, 185–199. doi:10.1016/S0065-3276(08)60339-1
- Lu, T., and Chen, F. (2012). Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592. doi:10.1002/jcc.22885
- Mayer, I. (2007). Bond order and valence indices: A personal account. *J. Comput. Chem.* 28, 204–221. doi:10.1002/jcc.20494
- Mayer, I. (1983). Charge, bond order and valence in the *ab initio* SCF theory. *Chem. Phys. Lett.* 97, 270–274. doi:10.1016/0009-2614(83)80005-0
- Mayer, I. (2003). *Population analysis, bond orders, and valences (chapter 7) in Simple theorems, proofs, and derivations in quantum chemistry*. New York: Springer Science+Business Media, 227–249. doi:10.1007/978-1-4757-6519-9
- McNaught, A. D., and Wilkinson, A. (1997). IUPAC Gold Book definition for homoconjugation: The overlap of two π-systems separated by a non-conjugating group. *IUPAC Compendium of Chemical Terminology*. 2nd ed. Oxford: Blackwell Scientific Publications. (the “Gold Book”). Compiled by McNaught, A. D.; Wilkinson, A. Online version 2019, created by Chalk, S. J. <https://goldbook.iupac.org/terms/view/H02842>.
- Mulliken, R. S. (1955a). Electronic population analysis on LCAO–MO molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. *J. Chem. Phys.* 23, 1841–1846. Mulliken bond orders are more precisely termed *overlap populations*. Mulliken stated that “the advantages of overlap populations over bond orders as measures of resonance energy are obvious”. doi:10.1063/1.1740589
- Mulliken, R. S. (1955b). Electronic population analysis on LCAO–MO molecular wave functions. I. *J. Chem. Phys.* 23, 1833–1840. doi:10.1063/1.1740588
- Mulliken, R. S. (1955c). Electronic population analysis on LCAO–MO molecular wave functions. IV. Bonding and antibonding in LCAO and valence-bond theories. *J. Chem. Phys.* 23, 2343–2346. While negative bond orders have no physical meaning, Mulliken interpreted negative overlap populations to signify antibonding between atoms. Negative Mulliken bond orders generally translate to Wiberg and Mayer bond orders of zero, as they do here for D_{6h} Si₆O₆. doi:10.1063/1.1741877
- Mulliken, R. S. (1970). The path to molecular orbital theory. *Pure Appl. Chem.* 24, 203–216. doi:10.1351/pac197024010203
- Newton, M. D., and Gibbs, G. V. (1980). *Ab initio* calculated geometries and charge distributions for H₄SiO₄ and H₆Si₂O₇ compared with experimental values for silicates and siloxanes. *Phys. Chem. Miner.* 6, 221–246. doi:10.1007/BF00309858
- Noritake, F., and Kawamura, K. (2015). The nature of Si–O–Si bonding via molecular orbital calculations. *J. Comput. Chem. Jpn.* 14, 124–130. doi:10.2477/jccj.2015-0009
- Pauling, L. (1930). “Eigenfunctions for chemical bonds,” in *Linus Pauling, the Nature of the Chemical Bond, A Documentary History*. Special Collections & Archives Research Center, Oregon State University Libraries. <http://scarc.library.oregonstate.edu/coll/pauling/bond/notes/sci3.002.2.html>.
- Pauling, L. (1952). Interatomic distances and bond character in the oxygen acids and related substances. *J. Phys. Chem.* 56, 361–365. According to Pauling’s original equation for partial ionic character = 1 – e^{–1/4(Δχ)²}, doi:10.1021/j150495a016
- Pauling, L. (1946). “Resonance,” in *Linus Pauling, the Nature of the Chemical Bond, A Documentary History*, 5–6. Special Collections & Archives Research Center, Oregon State University Libraries. <http://scarc.library.oregonstate.edu/coll/pauling/bond/notes/1946a.3.html>.
- Pauling, L. (1980). The nature of silicon-oxygen bonds. *Am. Mineral.* 65, 321–323. <https://pubs.geoscienceworld.org/msa/ammin/article-abstract/65/3-4/321/41136>.
- Pauling, L. (1960). *The nature of the chemical bond*. Third Edition. Ithaca, New York: Cornell University Press, 12. <https://books.google.com/books?id=L-1K9HmKmUUC&pg=PA12>.
- Pauling, L. (1931). The nature of the chemical bond. Application of results obtained from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules. *J. Am. Chem. Soc.* 53, 1367–1400. doi:10.1021/ja01355a027
- Pauling, L., and Wheland, G. W. (1933). The nature of the chemical bond. V. The quantum-mechanical calculation of the resonance energy of benzene and naphthalene and the hydrocarbon free radicals. *J. Chem. Phys.* 1, 362–374. doi:10.1063/1.1749304
- Ponec, R., and Mayer, I. (1997). Investigation of some properties of multicenter bond indices. *J. Phys. Chem. A* 101, 1738–1741. doi:10.1021/jp962510e
- Ponomarev, D. A., and Takhistov, V. V. (1997). What are isodesmic reactions? *J. Chem. Educ.* 74, 201–203. doi:10.1021/ed074p201
- Rappaport, S. M., and Rzepa, H. S. (2008). Intrinsically chiral aromaticity. Rules incorporating linking number, twist, and writhe for higher-twist Möbius annulenes. *J. Am. Chem. Soc.* 130, 7613–7619. Because of opposite twisting directions, one MO has L_k = +2 and one MO has L_k = –2. doi:10.1021/ja710438j
- Reed, A. E., and von Ragué Schleyer, P. (1987). The anomeric effect with central atoms other than carbon. I. Strong interactions between nonbonded substituents in polyfluorinated first- and second-row hydrides. *J. Am. Chem. Soc.* 109, 7362–7373. doi:10.1021/ja00258a020
- Reed, A. E., Weinstock, R. B., and Weinhold, F. (1985). Natural population analysis. *J. Chem. Phys.* 83, 735–746. doi:10.1063/1.449486
- Rettig, S. J., and Trotter, J. (1987). Refinement of the structure of orthorhombic sulfur, α-S₈. *Acta Cryst.* C43, 2260–2262. doi:10.1107/S0108270187088152
- Roberts, J. D., Webb, R. L., and McElhill, E. A. (1950). The electrical effect of the trifluoromethyl group. *J. Am. Chem. Soc.* 72, 408–411. doi:10.1021/ja01157a111
- Rzepa, H. S. (2005). Möbius aromaticity and delocalization. *Chem. Rev.* 105, 3697–3715. doi:10.1021/cr030092l
- Sannigrahi, A. B., and Kar, T. (1999). Some remarks on multi-center bond index. *Chem. Phys. Lett.* 299, 518–526. doi:10.1016/S0009-2614(98)01267-6
- Segall, M. D., Shah, R., Pickard, C. J., and Payne, M. C. (1996). Population analysis of plane-wave electronic structure calculations of bulk materials. *Phys. Rev. B* 54, 16317–16320. doi:10.1103/PhysRevB.54.16317
- Shao, Y., Gan, Z., Epifanovsky, E., Gilbert, A. T. B., Wormit, M., Kussmann, J., et al. (2015). Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* 113, 184–215. doi:10.1080/00268976.2014.952696
- Silvi, B., Savin, A., and Wagner, F. R. (1997). “The nature of silicon-oxygen bonds in silica polymorphs,” in *Modelling of minerals and silicated materials*. Editors B. Silvi and P. D’Arco (Dordrecht: Springer), Vol. 15. Topics in Molecular Organization and Engineering. doi:10.1007/0-306-46933-2_7
- Slater, J. C. (1931). Directed valence in polyatomic molecules. *Phys. Rev.* 37, 481–489. doi:10.1103/PhysRev.37.481
- Sorella, S., and Zen, A. (2014). “The new resonating valence bond method for *ab-initio* electronic simulations,” in *Many-electron approaches in physics, chemistry and mathematics*. Editors V. Bach and L. Delle Site (Switzerland: Springer International Publishing), 377–392. doi:10.1007/978-3-319-06379-9_21
- Torgunrud, J. L., Faria, A. J., and Miller, S. A. (2020). Thermodynamics of silica depolymerization with alcohols. *Polyhedron* 187, 114562. A resonance energy of 16.0 kcal/mol of SiO₂ has been estimated via computational hydrolysis of Si[OSi(OH)₃]₄. doi:10.1016/j.poly.2020.114562
- Weinhold, F., and Landis, C. R. (2005). *Valency and bonding: A natural bond orbital donor-acceptor perspective*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511614569
- Wheeler, S. E. (2012). Homodesmotic reactions for thermochemistry. *WIREs Comput. Mol. Sci.* 2, 204–220. doi:10.1002/wcms.72
- Wiberg, K. B. (1968). Application of the Pople-Santry-Segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* 24, 1083–1096. doi:10.1016/0040-4020(68)88057-3
- Wiberg, K. B., and Rablen, P. R. (1993). Origin of the stability of carbon tetrafluoride: Negative hyperconjugation reexamined. *J. Am. Chem. Soc.* 115, 614–625. doi:10.1021/ja00055a034
- Wilkinson, G., Rosenblum, M., Whiting, M. C., and Woodward, R. B. (1952). The structure of iron biscyclopentadienyl. *J. Am. Chem. Soc.* 74, 2125–2126. doi:10.1021/ja01128a527



OPEN ACCESS

EDITED BY

Minmin Xu,
Soochow University, China

REVIEWED BY

Xuechao Yu,
Chinese Academy of Sciences (CAS),
China
Laigui Hu,
Fudan University, China

*CORRESPONDENCE

Yassine Madoune,
✉ madouneyassine001@csu.edu.cn

[†]These authors share first authorship

SPECIALTY SECTION

This article was submitted to Theoretical and Computational Chemistry, a section of the journal Frontiers in Chemistry

RECEIVED 27 December 2022

ACCEPTED 22 February 2023

PUBLISHED 03 March 2023

CITATION

Madoune Y, Yang D, Ahmed Y, Al-Makeen MM and Huang H (2023), PVD growth of spiral pyramid-shaped WS₂ on SiO₂/Si driven by screw dislocations. *Front. Chem.* 11:1132567. doi: 10.3389/fchem.2023.1132567

COPYRIGHT

© 2023 Madoune, Yang, Ahmed, Al-Makeen and Huang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PVD growth of spiral pyramid-shaped WS₂ on SiO₂/Si driven by screw dislocations

Yassine Madoune^{1*†}, DingBang Yang^{1†}, Yameen Ahmed², Mansour M. Al-Makeen¹ and Han Huang¹

¹Hunan Key Laboratory of Super-microstructure and Ultrafast Process, School of Physics and Electronics, Central South University, Changsha, China, ²Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada

Atomically thin layered transition metal dichalcogenides (TMDs), such as MoS₂ and WS₂, have been getting much attention recently due to their interesting electronic and optoelectronic properties. Especially, spiral TMDs provide a variety of candidates for examining the light-matter interaction resulting from the broken inversion symmetry, as well as the potential new utilization in functional optoelectronic, electromagnetic and nanoelectronics devices. To realize their potential device applications, it is desirable to achieve controlled growth of these layered nanomaterials with a tunable stacking. Here, we demonstrate the Physical Vapor Deposition (PVD) growth of spiral pyramid-shaped WS₂ with ~200 μm in size and the interesting optical properties via AFM and Raman spectroscopy. By controlling the precursors concentration and changing the initial nucleation rates in PVD growth, WS₂ in different nanoarchitectures can be obtained. We discuss the growth mechanism for these spiral-patterned WS₂ nanostructures based on the screw dislocations. This study provides a simple, scalable approach of screw dislocation-driven (SDD) growth of distinct TMD nanostructures with varying morphologies, and stacking.

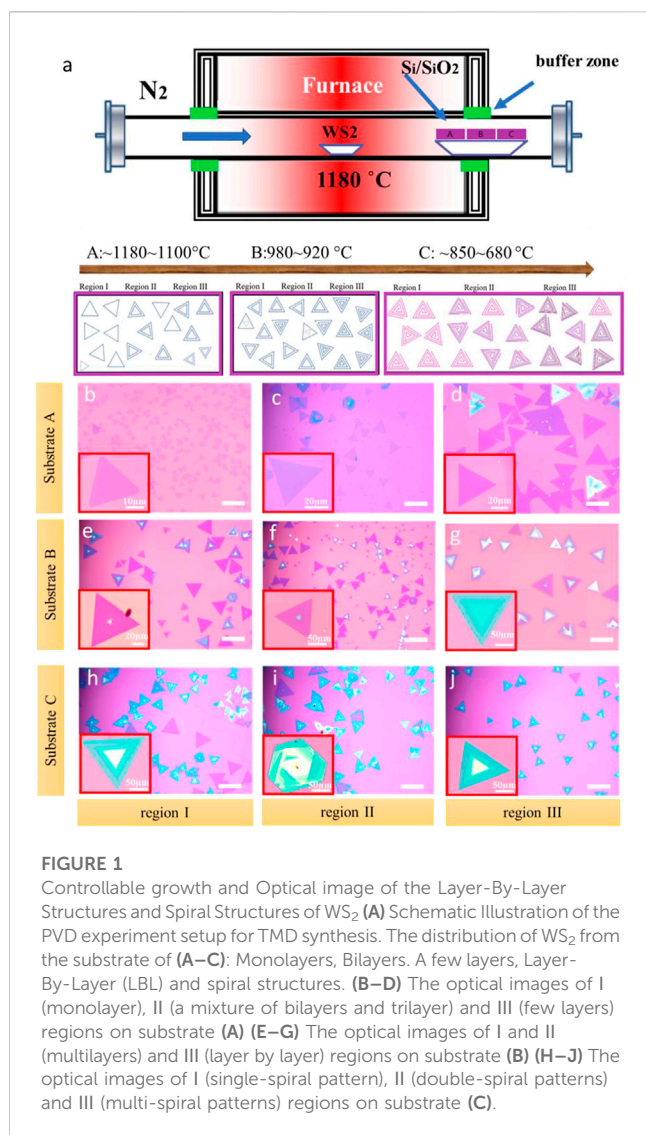
KEYWORDS

spiral patterns, tungsten disulfide, physical vapor deposition, Raman spectra, screw-dislocation driven growth

Introduction

Two-dimensional (2D) layered transition metal dichalcogenides (TMDs) have emerged as promising candidates for optoelectronic devices due to a plentiful choice of materials (Mak et al., 2010) (Xiao et al., 2022) (Yang et al., 2017). Multilayer TMDs display unique optoelectronic properties dependent on the layer number, such as a transition from the indirect band gap in multilayer to the direct band gap in monolayer (Manzeli et al., 2017). The different stacking modes also have critical effects on the properties of TMDs (Fan et al., 2017; Sarma et al., 2019) (Ci et al., 2022). It appeared that polarization enhancement with two petals along staggered stacking direction in 3R MoS₂ (Shi et al., 2017). The 3R-like TMDs few layers and spiral structures also show high degree of SHG polarization at room temperature due to the inversion symmetry breaking (Ma et al., 2021). Furthermore, bilayer or multilayer TMDs materials with different twisted angles always exhibit a different interlayer coupling to tune the properties of TMDs (Wang et al., 2019).

Spiral TMDs structures display more complex stacking than 2H stacking multilayer structures such as continuously supertwisted stacking, richer physical, and chemical effects (Ci et al., 2022) (Tong et al., 2022). Screw dislocation is a typical line defect for the materials



and the growth driven by spiral dislocation is a classic crystal growth mode at relatively low supersaturation (Cain et al., 2016) (Forticaux et al., 2015) (Meng et al., 2013). Many methods have been developed for the fabrication of spiral TMDs, including chemical vapor deposition (CVD) and physical vapor deposition (PVD) (Barman et al., 2019) (Wang et al., 2019) (Hao et al., 2016) (Fan et al., 2018). The hybrid spiral-like MoS₂ crystals with distinctive electrostatic properties have been synthesized through increasing the growth temperature to 1,000°C by CVD method (Hao et al., 2016). The spiral WS₂ with non-linear optical effects, a high value of valley polarization could be prepared by CVD using a mixture of WO₃, and S as precursors (Barman et al., 2019) (Fan et al., 2017). Compared with CVD method, PVD method is simpler, pollution-free as well as can synthesize spiral TMDs which have higher hardness, better thermal stability, and more stable chemical properties. The controllable growth of layer by layer and spiral WS₂ with ~30 μm in size has been realized by PVD. It reveals that the number of screw dislocations, orientation of new layer determine the morphologies, and stacking behaviors of the complex spiral nanostructures combined with SHG (Fan et al., 2018). However, it is difficult to

synthesize high quality spiral WS₂ in specific areas, the size of spiral WS₂ is too small obtained by these methods (~70 μm), limiting their applications on non-linear optical effects or electrocatalytic hydrogen evolution, etc. It is still a significant challenge to controllable fabrication for large area, and size of spiral TMDs.

Here, we report a controllable growth of large size (~200 μm) spiral patterned WS₂ with different stackings and elucidate their formation mechanisms. By tuning PVD temperature and precursor supersaturation, we demonstrated the controllable growth of layer by layer (LBL) structure and a series of spiral structures including single, double, and multi-spiral patterns. As the temperature decreasing, the complexity of spiral structures gradually increases. By related atomic force microscopy (AFM) with Raman spectroscopy measurements, we reveal how multiple dislocations, how the orientation of screw dislocations can affect the stacking behaviors, and formation mechanisms of spiral WS₂.

Materials and methods

To synthesize WS₂ nanoflakes, a ceramic boat filled with WS₂ (ALDRICH, 99%) powder (0.4 g) was placed at the heating center of a 2-inch quartz tube (2-inch diameter, 60 cm length), a three-piece of SiO₂/Si (300 nm, 1.5 × 1.5 cm) was placed at 6 cm (A), 8 cm (B), and 10 cm (C) downstream from the center of the tube furnace. N₂ flow was passed into the reaction chamber at a rate of 300 sccm for 10 min to ensure a clean environment for sample synthesis (anhydrous and oxygen-free environment). N₂ flow of 60 sccm was applied from the source to the substrate during the heating process (target temperature 1,180°C in 70 min) and held for 12 min, then cooled down the PVD tube furnace. To ensure the temperature difference, the three substrates were placed in different areas, substrate A inside the tube with a temperature of 1,180°C, substrate C at the end of the estuary, which is outside of the oven chamber, with a temperature of 850°C–680°C, and the substrate B in the middle of the estuary with a temperature of 920°C–980°C. A gradient in the temperature difference is illustrated in Figure 1A the red color gradually changed from the middle to the edge from dark to light as an indication of the temperature change.

Results and discussion

We design a PVD method to prepare monolayer (ML), few layers (FL), layer by layer (LBL), and spiral WS₂, as shown in Figure 1A. The distances between substrates A, B, and C and WS₂ powder (heating center) are 6 cm, 8 cm, and 10 cm, respectively, where the growth temperature can be controlled well. The growth temperature of substrate A was 1,180°C which was inside the tube, that of substrate C was ~680°C–850°C which was at the end of the estuary, nearly outside the oven chamber, and that of substrate B was ~920°C–980°C which was in the middle of the estuary. Figures 1B–D show that many monolayers and a few layers WS₂ were synthesized on substrate A with a high temperature ranging from 1,100°C to 1,180°C. The region I of substrate A displays the monolayer WS₂ with regular triangle and an average size of 30 μm–40 μm as shown in lower-left corner of Figure 1B. The region II of substrate A contains a mixture of bi/trilayer WS₂ with an

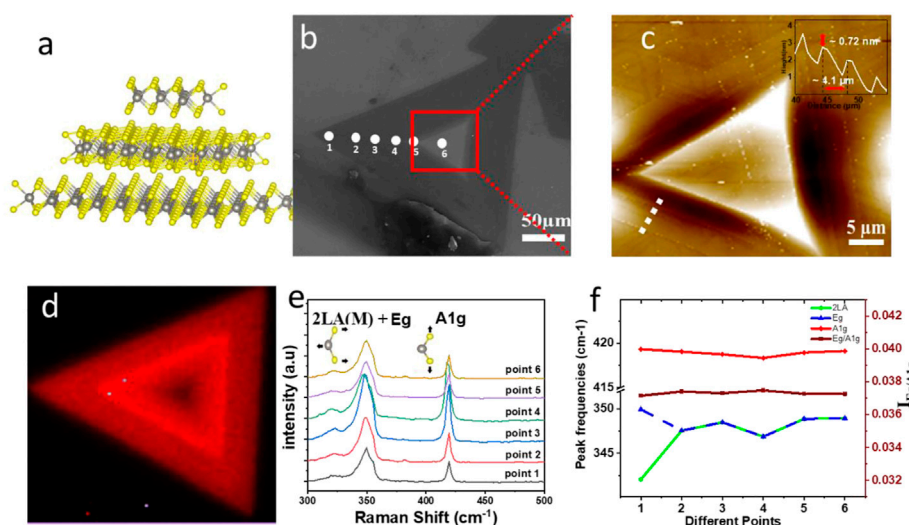


FIGURE 2

Layer by Layer (LBL) growth. (A) Illustrations of the crystal LBL structure. (B) SEM image of WS_2 LBL structure. (C) AFM image corresponding to the marked red square in SEM image. (D) Raman intensity mapping of E_{2g}^1 mode (348 cm^{-1}). (E,F) Raman spectra and peak positions (A_{1g} , E_{2g}^1 , $2LA$) as well as intensity ratios of $I_{E_{2g}^1}/I_{A_{1g}}$ of six points from edge to center, as marked in SEM image.

average size of about $50\text{ }\mu\text{m}$ – $60\text{ }\mu\text{m}$ (Figure 1C). The region III of substrate A, as shown in Figure 1D, contains multilayers with an average size of about $70\text{ }\mu\text{m}$ – $90\text{ }\mu\text{m}$ where the temperature in this area reaches approximately $1,100^\circ\text{C}$, lower than the I region of substrate A. Multilayers and LBL structures are synthesized on substrate B with a temperature ranging from 920°C to 980°C , as shown in Figures 1E–G. For substrate B, the number of layers increases from multilayers in region I to AA or AB stacking in region II with an increase in size from $90\text{ }\mu\text{m}$ to $100\text{ }\mu\text{m}$ to $120\text{ }\mu\text{m}$ – $150\text{ }\mu\text{m}$. The LBL structures are synthesized in the region III on substrate B with an estimated size of $180\text{ }\mu\text{m}$ – $220\text{ }\mu\text{m}$, which suggests a flat triangular top on a larger and thinner flake. As the temperature decreasing further, the spiral patterns of WS_2 were observed on substrate C. The structures in region I (Figure 1H) present a single spiral pattern with an average size of $\sim 220\text{ }\mu\text{m}$ where the temperature is approximately 850°C . Compared to region I, region II (Figure 1I) has double spiral patterns with temperatures between 850°C and 680°C . The region III (Figure 1J) exhibits more intricate multi-spiral patterns with a temperature near 680°C . The average size of double spiral and multi-spiral patterns is $\sim 200\text{ }\mu\text{m}$. Compared with LBL growth above, the size of WS_2 obtained by spiral growth did not have an obvious difference with the decreasing temperature.

We found that a monolayer is formed at a high temperature with small sizes, that the sizes, the number of layers increase little by little as the temperature decreases to form a LBL, and SDD structures in the end (Morin et al., 2011). With the temperature decreasing, the LBL and the size of spiral WS_2 reaches $\sim 220\text{ }\mu\text{m}$ with a total area over $0.3\text{ cm}^2 \times 0.3\text{ cm}^2$, larger than the largest obtained by CVD techniques at low temperature ($\sim 70\text{ }\mu\text{m}$) (Yin et al., 2015) (Zhao and Jin, 2020).

The PVD-deposited 2H- WS_2 nanoplates have a variety of morphologies, including triangles, truncated triangles, hexagons, etc (Figures 1B–J). The surface free energy of the relevant crystal

face affects the development rate of the crystal face in 2D crystals, where the surface free energy corresponds to the crystal's edge free energy. When the free energy is high, the crystal edge is exceedingly unstable. At this point, the free electrons in the free state will be swiftly absorbed to form the edge in a stable state of low free energy, causing the edge of high free energy to develop fast and the edge of low free energy to grow slowly (Park et al., 2017; Wang et al., 2018). As a consequence of the fast growth behavior, the crystal face with high free energy will ultimately get smaller or vanish. On the contrary, the slow-growing crystal face will eventually become the biggest (Li et al., 2018). As a result, the final crystal form will be connected to the development rate of various kinds of edge terminals, which is further influenced by the precursor volume ratio, which directly influences the growth rate of distinct crystal planes. We employed a constant flow in a quartz tube to retain the precursor to establish a stable air flow environment, which guarantees that the WS_2 vapor is thoroughly volatilized and evenly dispersed throughout the reaction process. As a result, the concentration distribution of WS_2 precursors (volatilization and diffusion) may be regarded as the primary factor influencing the various form morphologies (Yan et al., 2023).

Figure 2A shows the illustrations of LBL WS_2 structures. There are many possible WS_2 stacking configurations but two of the most important are the AB-stacking and the AA-stacking configurations, as shown in Supplementary Figure S2B (Rudenko et al., 2021). Figure 2B shows the SEM images of high-coverage LBL structures in substrate B of region III with a size of $200\text{ }\mu\text{m}$. Figure 2C shows that the AFM image corresponds to the marked red square in Figure 2B, which confirms the LBL growth of WS_2 . The line profile in Figure 2C indicates that the height increases gradually from edge to center and the height of each step is $\sim 0.72\text{ nm}$, consistent with monolayer WS_2 . Figure 2D shows the Raman intensity mapping of E_{2g}^1 mode (348 cm^{-1}). It can be observed that the Raman intensity of LBL structure from the edge to the center first increases then starts to

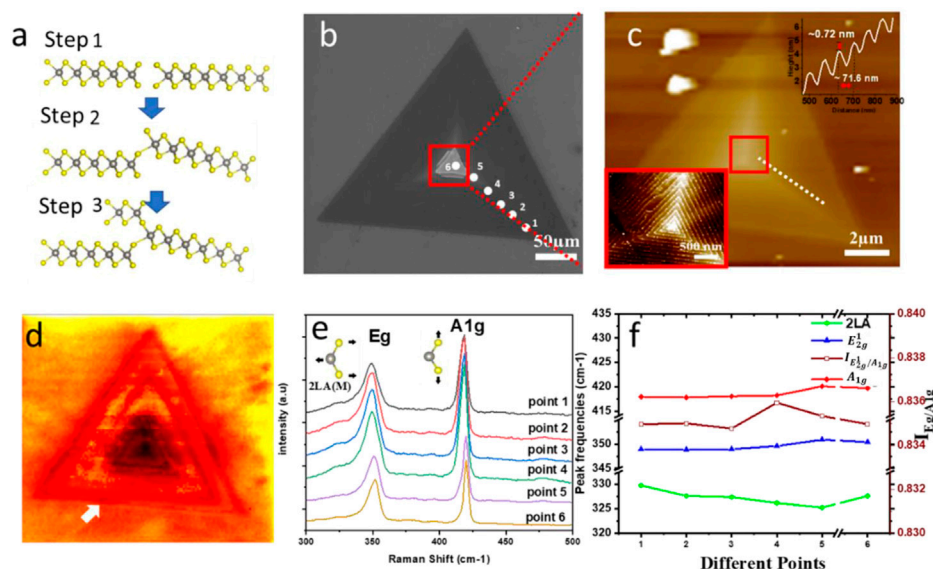


FIGURE 3

Screw dislocation growth of single spiral. (A) Illustrations of the single spiral structure (B) SEM image of single spiral structure of WS₂ (C) AFM images corresponding the marked areas in SEM image. (D) Raman intensity mapping of E_{2g} mode (350 cm⁻¹). (E,F) Intensity ratios and peak frequencies of WS₂ Raman modes. (E) Frequencies of the (2LA, E_{2g}¹, A_{1g}) Raman modes as a function of thickness (different positions mean different thicknesses) for λ_{exc} = 633. The error bars correspond to the standard deviations and Each point represents an average over six different positions (E) Thickness-dependent intensity ratios of I_{E_{2g}¹}/I_{A_{1g}} (E,F) of six points from edge to center, as marked in SEM image.

decrease, and the corresponding number of layers gradually increases. The WS₂ Raman spectra related to the number layers in Figure 2E confirm the above Raman intensity mapping results (Zeng et al., 2013; Qian et al., 2018). The number layers of materials and the intensity of incident light in the scattering process both affect the Raman intensity as mentioned in previous reports (Shao et al., 2020) (Diep et al., 2019). The local electrical field will be significantly less than the incident electrical field for thick WS₂ materials, which is known as the local field effect for high reflective index TMDCs. Therefore, a faint Raman signal will be seen in thick WS₂ materials. Furthermore, the local field impact is minimal for extremely thin WS₂ layers. As a result, there may be a maximum Raman intensity at a certain height in WS₂. According to a recent study, the first few layers of WS₂ (bilayers to five layers) have the largest Raman intensity which gradually decrease after the bulk starts to form (Li et al., 2013).

Figure 2F shows that both A_{1g} and E_{2g}¹ peaks position is not shifted despite the thickness increases (the height of difference between the first and the sixth points is ~24 nm). This solidification in the two modes is in good agreement with the previous results for MoS₂ (Yan et al., 2015), and the higher restoring strength brought on by the vdW interactions produced between the layers. We observe that the error in calculating the frequency shift of both the 2LA(M) and E_{2g}¹ modes rise due to their close proximity. A clear frequency relationship with the number of layers cannot be established since the change in frequency is of the order of the error bar, which is also impacted by the fitting procedure. A shift in dielectric screening with the number of layers is also anticipated for WS₂ (Jin et al., 2010). Stronger dielectric screening of the long-range coulomb interactions between the effective charges in thicker samples may be the

reason for the E_{2g}¹ mode's anomalous behavior. We also looked at how the number of layers affected the relative intensity of the strongest Raman signals. With the number of layers increasing, there are no discernible variations in the intensity ratio I_{E_{2g}¹}/I_{A_{1g}} (Berkdemir et al., 2013).

The anisotropic pyramidal growth is continued in this fashion (formed) by self-perpetuating steps of SDD layers. Under mild supersaturation circumstances, screw dislocations form step edges (slipped planes) in the bottom layer (Daher et al., 2018). Because of the large concentration of precursors, the unintentional rising of a grain boundary caused by varying growth rates of multiple edge terminations begins spiral growth, which is schematically shown as step 1 in Figure 3A. The spiral development is also catalyzed by the uneven surface of the substrate, which is created by the partial etching of SiO₂/Si using piranha and plasma treatment. The unsaturated sulfur edges in the slipping plane operate as nucleation sites for additional precursor atom addition, resulting in the formation of the second layer on top of the bottom layer, as illustrated in Figure 3A (steps 2, 3). The growth process continues up to an interface or a nodal point, generating a spiral after the slip plane is produced with vertically mismatched edges (Sarma et al., 2019).

Figure 3B shows a typical SEM image of a single spiral WS₂ with ~220 μm, which indicates the atomically flat surface of WS₂. The zoom-in AFM image of the red square region of the SEM image is shown in Figure 3C, revealing that each step height of a single spiral of WS₂ is ~0.72 nm, consistent with the height of monolayer WS₂, with an estimated height of ~25 nm (from the edge to the center). When the single spiral WS₂ is projected onto the 2D basal plane, the layer size progressively decreases from the center to the edge, as shown in Supplementary Figure S4. The low-left corner of

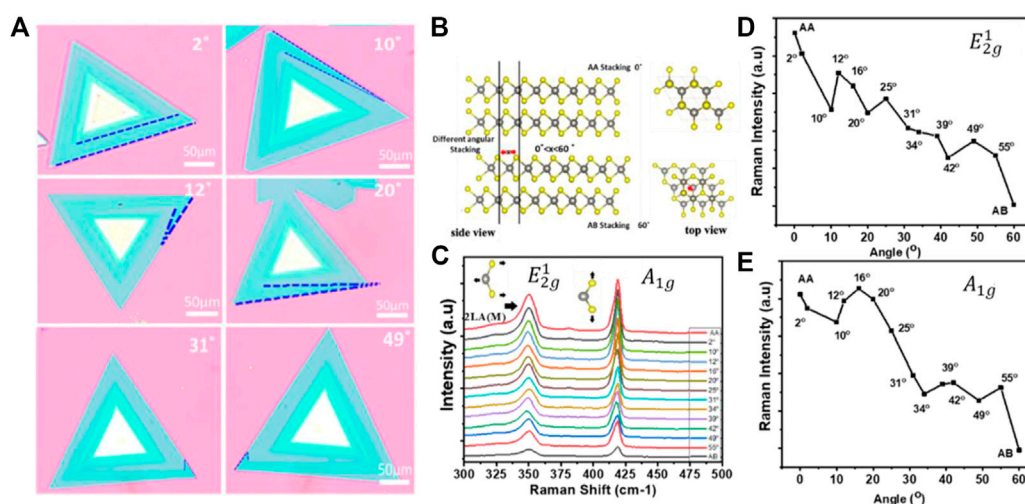


FIGURE 4

Twist angle-dependent responses bilayer WS₂ crystals with different twist angles. (A) Optical images and (B) atomic structure of different angles hetero-bilayer stacking WS₂ in a $2 \times 2 \times 1$ supercell from a side view (upper panel) and top view (lower panel), respectively. (C) Raman intensity at different twist angles on single spiral bilayer WS₂. (D,E) Twist angular dependence of Raman intensity (E_{2g}^1 mode A_{1g} mode).

Figure 3C; Supplementary Figure S5A as well as Supplementary Figure S5B display AFM images of a single spiral, double spiral, and multi-spiral with clear screw dislocation, respectively. The white lines in Supplementary Figure S5 (near the dislocation center) of the core serve as an example of the included angle, defined as the angle between the traces of the many dislocation spirals that share a common core. There may be two sets of included angles between screw dislocation spirals based on the three-fold symmetry of the monolayer TMD structure. The center structure of spiral nanoplates stays as triangular spiral traces whether the included angles are just (0°C, 360°C) or (120°C, 240°C), as shown in low-left corner of Figure 3C; Supplementary Figure S5. In contrast, the center of those spiral nanoplates shows hexagonal spiral traces when we inspect the second set of included angles of (60°C, 180°C, and 300°C) (Shearer et al., 2017). It should be noted that multi-spiral pattern nanostructures are far less often seen compared to single or double-spiral pattern structures.

Figure 3D shows the Raman intensity (350 cm⁻¹) mapping of the spiral structure, revealing a regular variety of E_{2g}^1 intensities in the region of single spiral WS₂ (Chen et al., 2017). Also, it can reveal that they are multilayer flakes, which corresponds with AFM measurement shown above. Figure 3E summarizes the single spiral WS₂ Raman spectra as a function of the thickness, the Raman spectra was measured six consecutive times at each of the 6 points indicated in Figure 3B to calculate the error bars of the Raman shift: $[2LA (\pm 0.8 \text{ cm}^{-1}), E_{2g}^1 (\pm 0.4 \text{ cm}^{-1}), A_{1g} (\pm 0.3 \text{ cm}^{-1})]$. The thicknesses, distances between the first and last point are ~25 nm and 150 μm, respectively, as shown in Supplementary Figure S4.

To determine the frequency dependence of the main WS₂ Raman peaks (E_{2g}^1 and A_{1g}), we fitted several Lorentzian peaks to each spectrum, as shown in Figure 3F. When the thickness increases, the E_{2g}^1 , A_{1g} modes exhibit minor redshift (Ci et al., 2022). The growing restoring force brought on by vdW contacts that have been developed between layers is compatible with the hardness of the A_{1g} mode.

However, the 2LM phonon mode also exhibits subtle blueshifts when the thickness increases (Ci et al., 2022). In screw dislocation (single, double, and multi-spiral patterns) WS₂ the proximity of 2LA (M) and E_{2g}^1 increases the error in determining the frequency offset for both modes. Especially since the number of layers is very large, it may reach more than 30 layers at least (in this sample, the thickness is ~25 nm, every layer ~0.72 nm). Due to the closeness of the layers to one another, it is a challenge to establish a clear frequency dependence with number of layers. (~71.6 nm shown in the inset of Figure 3C). A change in buffer shifting is also expected with the number of layers in WS₂. It may be caused by enhanced and a stronger dielectric sifting of long-range or by coulomb interactions between effective charges in bulk samples (Yin et al., 2014). We also studied the relative intensities of the strongest Raman peaks as a function of thicknesses (different points with different numbers of layers). The most intense features in the Raman spectrum correspond to the E_{2g}^1 and A_{1g} modes and the intensity ratio $I_{E_{2g}^1}/I_{A_{1g}}$ does not show major changes with the number of layers or thickness. It has been generalized by observing different samples of single spiral patterns (Supplementary Figure S6) (Berkdemir et al., 2013).

Figure 4A shows the Optical images of WS₂ bilayers with varying twist angles of 2°C, 10°C, 12°C, 20°C, 31°C, and 49°C. As a result of the steric repulsion effect, the interlayer coupling in the randomly twisted bilayer was lower than in LBL stacking (AA stacking) (Zhang et al., 2020). Spin-orbit coupling would find a wider range of uses in angle-dependent moiré excitons, spintronics, and valley electronics if the twist angles were varied. It is depicted by the atomic structure schematic of LBL, single spiral patterns hetero-bilayer WS₂ in a $2 \times 2 \times 1$ supercell from a top perspective (Ghatak et al., 2020) (lower panel), and side view (upper panel) in Figure 4B, respectively. The variable interlayer coupling may be further confirmed by the twist angle-dependent optical responses by accumulating Raman spectra at different angles (Figure 4C). The characteristic peak intensities (E_{2g}^1 and A_{1g}) of WS₂ in the Raman spectra steadily decrease periodically as the twist angle increasing, showing that the

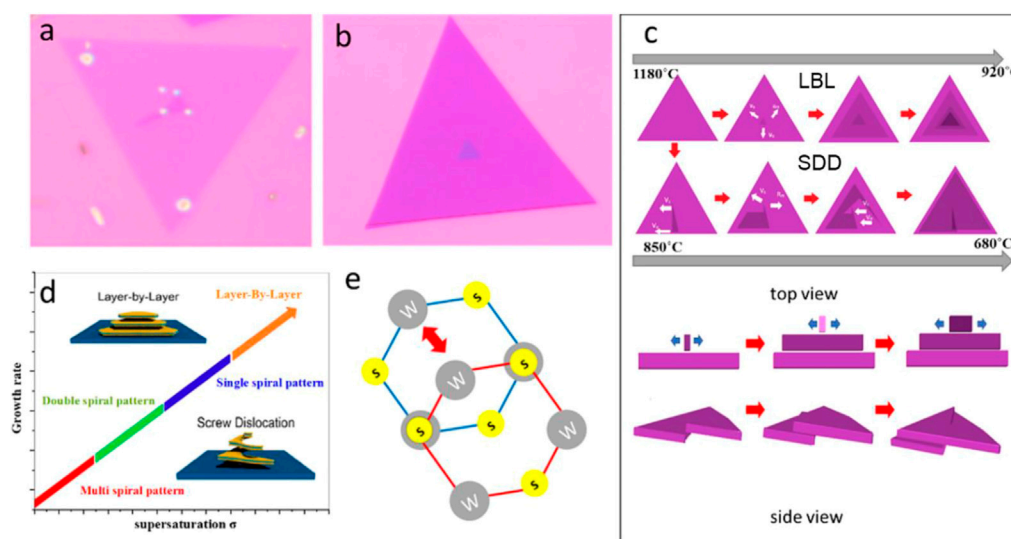


FIGURE 5

Mechanisms Formation of Layer by Layer and Dislocated WS_2 Spirals (A,B) Optical image of single spiral pattern and layer by layer WS_2 , respectively. (C) Schematic diagrams of Layer-By-Layer and screw dislocation propagation process. (D) Schematic illustrations of the growth rate of different dominant crystal growth modes as a function of supersaturation. (E) The tungsten atom in the top layer (red hexagon) positioned at the center of the bottom layer hexagon (blue).

mechanical coupling effect was typically less due to the stacking angle in Figure 4D, E (Shao et al., 2020).

Interlayer coupling interactions of SDD (single spiral) differ in randomly twisted bilayers compared to LBL due to the upper and lower layers of multi-spiral pattern nanostructures being far less often seen compared to single or double-spiral pattern structure saving different arrangements of atoms. The majority of 2H- WS_2 nanosheets with terraced or spiral structures are dispersed across the substrate, according to the morphological information of 2H- WS_2 nanosheets. So, the moiré superlattice from SDD stacking with different twist angles can change the electronic structure of 2D materials, giving them unusual transport properties like unusual superconductivity and insulating behavior. The interlayer exciton coupling can also be controlled, which makes it possible to study moiré excitons, spintronics, and valley electronics of spin-orbit coupling (Jin et al., 2019) (Seyler et al., 2019).

The optical images in Figures 5A, B show the onset of the formation of the single helical pattern, and the LBL, respectively. It is termed anisotropic growth as it is considered the defining feature of growth almost without exception in which cells grow faster in one direction than the other, which describes the condition when the rates of growth are uneven in all directions (Figure 5A). In contrast, when growth rates are the same in all directions, growth is isotropic (LBL in Figure 5B). Figure 5C shows the model diagram of terraced structure growth. The active adsorption atoms are shown by white arrows, and the dislocation core (v_c) depicts the axial growth rate, which is governed by the nucleation rate of new atomic layers. Outer edges (v_o) denote the lateral growth rate, as determined by the rate at which adsorbed atoms cling to the step's edge (Yin et al., 2015). The new surface layer progressively emerges throughout the growing phase of a terraced construction as the new atomic layer slowly nucleates on the top surface. Instead of a 1D structure growth, 2D flakes do so because the growth rate of normal to the surface (R_m) is

much slower than lateral step velocities (v_s). When this situation occurs repeatedly, layered terraced structures will grow layer by layer. It has been discovered that the number of terraced nanosheets would directly depend on the disparity between axial and lateral development speeds. Furthermore, we translate the number of terraced into the v_c/v_o ratio, which represents the ratio of axial and lateral growth rates. Furthermore, the nucleation rate of new layers on the existing top layer determines the axial growth rate. Only when the current top terraced reaches the critical size, the nucleation of the new atomic layer occurs. Therefore, the axial growth rate can be expressed as (Tersoff et al., 1994): $v_c = hR_c v_o$, where h is the height of the freshly formed terraced and R_c is the crucial height of the top terraced at the time. According to the variations in supersaturation in the furnace, the ratios of v_c/v_o and supersaturation have a linear relationship, meaning that the ratio of v_c/v_o rises as supersaturation rises, which also predicts an increase in the number of terraced fields (Figure 5D). The concentration of adsorption atoms may build up, but the energy barrier at the step's edge prevents this from happening; therefore, the nucleation of new atoms must overcome this high energy barrier (Krishnamurthy et al., 1993). Therefore, High supersaturation is crucial during the growth process of terraced structures in order to encourage the nucleation of new atomic layers and provide LBL the ability to grow at a favorable crystal growth rate. When the substrate is heated to a high temperature, spiral structures make up the majority of the nanosheets. According to Figure 1A, the nanosheets display a straightforward triangular spiral pattern, which indicates that in contrast to the low-temperature region, the screw dislocation is formed by first forming a step edge at the bottom of the growth process as the nucleation position of a new layer. Then, by continuously adsorbing the precursor atoms, results in the growth of a second layer at the top of the bottom layer, which is faster than that of the first layer.

Furthermore, the spiral structures were developed using the SDD growth mode, forming pyramid-shaped WS₂ flakes. The Burton-Cabrera-Frank (BCF) crystal growth hypothesis provides an explanation for a number of these pyramidal growths that have been observed in 1D and 2D nanomaterials (Burton et al., 1951) (Fan et al., 2017), which states that the supersaturation of the local growth environment controls the crystal growth process. The degree of supersaturation is denoted as $\sigma = \ln(c/c_0)$, where c represents the local precursor concentration and c_0 represents the equilibrium precursor concentration. The furnace's temperature profile has an impact on both c and c_0 . Particularly, the deposition reaction's thermodynamic equilibrium constant is largely governed by the local temperature, determining the value of c_0 . The precursor's heating temperature and the precursor's movement all dynamically affect the value of c . The deposition temperature largely affects the local supersaturation, which also affects the growth rate of nanostructures. In this theory, the system's supersaturation drives various growth modes, including LBL and SDD growth (Woodruff, 2015). The nucleation will never be homogenous if the deposition is done on different substrates. There will always be certain defect sites that have a larger chemical potential than the rest of the substrate, which are more active for crystal development. The SDD growth mode is the most advantageous mode of development at low supersaturation, which results in spirally stacked WS₂ pyramidal structures. These self-replicating steps of SDD layers continue the anisotropic pyramidal development in this way. Step edges (slipped planes) are produced in the bottom layer due to screw dislocations that arise under mild supersaturation conditions. Due to the large concentration of precursors, spiral development is started by the unintentional rising of a grain boundary caused by the numerous edge terming.

According to the above hypothesis, the LBL structures of WS₂ tend to develop at the high-temperature deposition zone with a high supersaturation. In contrast, the spiral structures with screw dislocations are more likely to occur at low-temperature deposition zones with low supersaturation, where atoms may be added to the spiral step edges. Different atomic configurations are found in the top and bottom layers of the WS₂ spiral patterns, proving the existence of a unique stacking sequence in spiral domains (Morin et al., 2011; Sarma et al., 2019). The two layers are aligned, and a lattice vector translates the upper layer. The tungsten atom in the top layer is located in the middle of the bottom layer hexagon (Figure 5E). (Zhang et al., 2014) (Ghatak et al., 2020) Additionally, toward the lower temperature region, each spiral nanostructure contains more screw dislocations than the straightforward triangular single dislocation spirals that are more commonly found in low-temperature region, suggesting that lower supersaturation condition tends to induce more complex spiral structures that correspond to multi-spiral pattern structures (Yin et al., 2015).

Conclusion

In summary, we report the regulated growth of SDD WS₂ nanoflakes and disclosed the underlying growth mechanisms. Through controlling the reaction temperature, different types of WS₂ growth such as LBL and SDD growth modes can be obtained. LBL and spiral WS₂ with ~200 μm in size was obtained at ~ 920°C and 850°C–680°C respectively. The interlayer coupling was weaker in the randomly twisted bilayer and at 60°C stacking than at 0°C stacking (the interlayer coupling at 60°C

stacking was much weaker than the random angles.) due to the steric repulsion effect. SDD growth offers a straightforward, controllable approach to construct TMD nanostructures of various morphologies, and stackings. With a knowledge of the growth principles of stacked TMD materials, we anticipate that the comprehension of stacked TMD materials will provide attractive options for creating new materials for innovative functional nanodevices.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

MY: Conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization. DY: Validation, formal analysis, investigation, data curation, writing—review and editing, visualization. YA: Formal analysis, investigation, visualization, writing—review and editing. MA-M: Formal analysis, investigation, visualization, writing—review and editing. HH: Conceptualization, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, supervision, project administration, Funding acquisition.

Acknowledgments

The authors thank Mr. Yongsong Wang, Mr. Xiao Guo, Mr. Junjie Jiang and Miss Siwen for the useful discussions. We acknowledge project supported by State Key Laboratory of Powder Metallurgy, Central South University, Changsha, China.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2023.1132567/full#supplementary-material>

References

- Barman, P. K., Sarma, P. V., Shaijumon, M. M., and Kini, R. N. (2019). High degree of circular polarization in WS₂ spiral nanostructures induced by broken symmetry. *Sci. Rep.* 9, 2784–2787. doi:10.1038/s41598-019-39246-7
- Berkdemir, A., Gutiérrez, H. R., Botello-Méndez, A. R., Perea-López, N., Elías, A. L., Chia, C. I., et al. (2013). Identification of individual and few layers of WS₂ using Raman Spectroscopy. *Sci. Rep.* 3, 1755–1758. doi:10.1038/srep01755
- Burton, W. K., Cabrera, N., and Frank, F. C. (1951). The growth of crystals and the equilibrium structure of their surfaces. *Philosophical Trans. R. Soc. Lond. Ser. A, Math. Phys. Sci.* 243, 299–358. doi:10.1098/rsta.1951.0006
- Cain, J. D., Shi, F., Wu, J., and Dravid, V. P. (2016). Growth mechanism of transition metal dichalcogenide monolayers: The role of self-seeding fullerene nuclei. *ACS Nano* 10, 5440–5445. doi:10.1021/acsnano.6b01705
- Chen, J., Zhao, X., Tan, S. J. R., Xu, H., Wu, B., Liu, B., et al. (2017). Chemical vapor deposition of large-size monolayer MoSe₂ 2 crystals on molten glass. *J. Am. Chem. Soc.* 139, 1073–1076. doi:10.1021/jacs.6b12156
- Ci, P., Zhao, Y., Sun, M., Rho, Y., Chen, Y., Grigoropoulos, C. P., et al. (2022). Breaking rotational symmetry in supertwisted WS₂ spirals via moiré magnification of intrinsic heterostrain. *Nano Lett.* 22, 9027–9035. doi:10.1021/acs.nanolett.2c03347
- Daher, F. B., Chen, Y., Bozorg, B., Clough, J., Jö Nsson, H., and Braybrook, S. A. (2018). Anisotropic growth is achieved through the additive mechanical effect of material anisotropy and elastic asymmetry. *Elife* 7, e38161. doi:10.7554/eLife.38161.001
- Diep, N. Q., Liu, C. W., Wu, S. K., Chou, W. C., Huynh, S. H., and Chang, E. Y. (2019). Screw-dislocation-driven growth mode in two dimensional GaSe on GaAs(001) substrates grown by molecular beam epitaxy. *Sci. Rep.* 9, 17781–17788. doi:10.1038/s41598-019-54406-5
- Fan, X., Jiang, Y., Zhuang, X., Liu, H., Xu, T., Zheng, W., et al. (2017). Broken symmetry induced strong nonlinear optical effects in spiral WS₂ nanosheets. *ACS Nano* 11, 4892–4898. doi:10.1021/acsnano.7b01457
- Fan, X., Zhao, Y., Zheng, W., Li, H., Wu, X., Hu, X., et al. (2018). Controllable growth and formation mechanisms of dislocated WS₂ spirals. *Nano Lett.* 18, 3885–3892. doi:10.1021/acs.nanolett.8b01210
- Forticaux, A., Dang, L., Liang, H., and Jin, S. (2015). Controlled synthesis of layered double hydroxide nanoplates driven by screw dislocations. *Nano Lett.* 15, 3403–3409. doi:10.1021/acs.nanolett.5b00758
- Ghatak, K., Kang, K. N., Yang, E. H., and Datta, D. (2020). Controlled edge dependent stacking of WS₂-WS₂ Homo- and WS₂-WSe₂ Hetero-structures: A Computational Study. *Sci. Rep.* 10, 1648–1711. doi:10.1038/s41598-020-58149-6
- Hao, S., Yang, B., and Gao, Y. (2016). Controllable growth and characterizations of hybrid spiral-like atomically thin molybdenum disulfide. *Phys. E Low. Dimens. Syst. Nanostruct.* 84, 378–383. doi:10.1016/j.physe.2016.07.026
- Jin, C., Regan, E. C., Yan, A., Iqbal Bakht Utama, M., Wang, D., Zhao, S., et al. (2019). Observation of moiré excitons in WSe₂/WS₂ heterostructure superlattices. *Nature* 567, 76–80. doi:10.1038/s41586-019-0976-y
- Jin, S., Bierman, M. J., and Morin, S. A. (2010). A new twist on nanowire formation: Screw-dislocation-driven growth of nanowires and nanotubes. *J. Phys. Chem. Lett.* 1, 1472–1480. doi:10.1021/jz100288z
- Krishnamurthy, M., Wassermeier, M., Williams, D. R. M., and Petroff, P. M. (1993). Periodic faceting on vicinal GaAs(110) surfaces during epitaxial growth. *Appl. Phys. Lett.* 62, 1922–1924. doi:10.1063/1.109545
- Li, H., Li, Y., Aljarb, A., Shi, Y., and Li, L. J. (2018). Epitaxial growth of two-dimensional layered transition-metal dichalcogenides: Growth mechanism, controllability, and scalability. *Chem. Rev.* 118, 6134–6150. doi:10.1021/acs.chemrev.7b00212
- Li, H., Lu, G., Wang, Y., Yin, Z., Cong, C., He, Q., et al. (2013). Mechanical exfoliation and characterization of single- and few-layer nanosheets of WSe₂, TaS₂, and TaSe₂. *Small* 9, 1974–1981. doi:10.1002/sml.201202919
- Ma, R., Sutherland, D. S., and Shi, Y. (2021). Harmonic generation in transition metal dichalcogenides and their heterostructures. *Mater. Today* 50, 570–586. doi:10.1016/j.mattod.2021.07.023
- Mak, K. F., Lee, C., Hone, J., Shan, J., and Heinz, T. F. (2010). Atomically thin MoS₂: A new direct-gap semiconductor. *Phys. Rev. Lett.* 105, 2–5. doi:10.1103/PhysRevLett.105.136805
- Manzeli, S., Ovchinnikov, D., Pasquier, D., Yazyev, O. V., and Kis, A. (2017). 2D transition metal dichalcogenides. *Nat. Rev. Mater.* 2, 17033. doi:10.1038/natrevmats.2017.33
- Meng, F., Morin, S. A., Forticaux, A., and Jin, S. (2013). Screw dislocation driven growth of nanomaterials. *Acc. Chem. Res.* 46, 1616–1626. doi:10.1021/ar400003q
- Morin, S. A., Forticaux, A., Bierman, M. J., and Jin, S. (2011). Screw dislocation-driven growth of two-dimensional nanoplates. *Nano Lett.* 11, 4449–4455. doi:10.1021/nl202689m
- Park, S. J., Pak, S. W., Qiu, D., Kang, J. H., Song, D. Y., and Kim, E. K. (2017). Structural and optical characterization of MoS₂ quantum dots defined by thermal annealing. *J. Lumin.* 183, 62–67. doi:10.1016/j.jlumin.2016.11.014
- Qian, Q., Zhang, Z., and Chen, K. J. (2018). Layer-dependent second-order Raman intensity of MoS₂ and WS₂: Influence of intervalley scattering. *Phys. Rev. B* 97, 1–9. doi:10.1103/PhysRevB.97.165409
- Rudenko, A. N., Katsnelson, M. I., and Gornostyrev, Y. N. (2021). Dislocation structure and mobility in the layered semiconductor InSe: A first-principles study. *2d Mater.* 8, 045028. doi:10.1088/2053-1583/ac207b
- Sarma, P., Kayal, A., Sharma, C. H., Thalakulam, M., Mitra, J., Shaijumon, M. M., et al. (2019). Electrocatalysis on edge-rich spiral WS₂ for hydrogen evolution. *ACS Nano* 13, 10448–10455. doi:10.1021/acsnano.9b04250
- Seyler, K. L., Rivera, P., Yu, H., Wilson, N. P., Ray, E. L., Mandrus, D. G., et al. (2019). Signatures of moiré-trapped valley excitons in MoSe₂/WSe₂ heterobilayers. *Nature* 567, 66–70. doi:10.1038/s41586-019-0957-1
- Shao, G., Xue, X. X., Liu, X., Zhang, D., Jin, Y., Wu, Y., et al. (2020). Twist angle-dependent optical responses in controllably grown WS₂ vertical homojunctions. *Chem. Mater.* 32, 9721–9729. doi:10.1021/acs.chemmater.0c03413
- Shearer, M. J., Samad, L., Zhang, Y., Zhao, Y., Puzetzy, A., Eliceiri, K. W., et al. (2017). Complex and noncentrosymmetric stacking of layered metal dichalcogenide materials created by screw dislocations. *J. Am. Chem. Soc.* 139, 3496–3504. doi:10.1021/jacs.6b12559
- Shi, J., Yu, P., Liu, F., He, P., Wang, R., Qin, L., et al. (2017). 3R MoS₂ with broken inversion symmetry: A promising ultrathin nonlinear optical device. *Adv. Mater.* 29, 1701486–1701489. doi:10.1002/adma.201701486
- Tersoff, J., Denier Van Der Gon, A. W., and Tromp, R. M. (1994). Critical island size for layer-by-layer growth. *Phys. Rev. Lett.* 72, 266.
- Tong, X., Zhao, Y., Zhuo, Z., Yang, Z., Wang, S., Liu, Y., et al. (2022). Dual-regulation of defect sites and vertical conduction by spiral domain for electrocatalytic hydrogen evolution. *Angew. Chem. - Int. Ed.* 61, e202112953. doi:10.1002/anie.202112953
- Wang, J., Cai, X., Shi, R., Wu, Z., Wang, W., Long, G., et al. (2018). Twin defect derived growth of atomically thin MoS₂ dendrites. *ACS Nano* 12, 635–643. doi:10.1021/acsnano.7b07693
- Wang, X., Yang, H., Yang, R., Wang, Q., Zheng, J., Qiao, L., et al. (2019). Weakened interlayer coupling in two-dimensional MoSe₂ flakes with screw dislocations. *Nano Res.* 12, 1900–1905. doi:10.1007/s12274-019-2456-y
- Woodruff, D. P. (2015). How does your crystal grow? A commentary on burton, cabrera and frank (1951) “the growth of crystals and the equilibrium structure of their surfaces. *Philosophical Trans. R. Soc. A Math. Phys. Eng. Sci.* 373, 230. doi:10.1098/rsta.2014.0230
- Xiao, Y., Qu, J., Luo, Z., Chen, Y., Yang, X., Zhang, D., et al. (2022). Van der Waals epitaxial growth and optoelectronics of a vertical MoS₂/WSe₂ p-n junction. *Front. Optoelectron.* 15, 41. doi:10.1007/s12200-022-00041-4
- Yan, J., Lian, S., Cao, Z., Du, Y., Wu, P., Sun, H., et al. (2023). CVD controlled preparation and growth mechanism of 2H-WS₂ nanosheets. *Vac.* 207, 111564. doi:10.1016/j.vacuum.2022.111564
- Yan, J., Xia, J., Wang, X., Liu, L., Kuo, J. L., Tay, B. K., et al. (2015). Stacking-dependent interlayer coupling in trilayer MoS₂ with broken inversion symmetry. *Nano Lett.* 15, 8155–8161. doi:10.1021/acs.nanolett.5b03597
- Yang, T., Zheng, B., Wang, Z., Xu, T., Pan, C., Zou, J., et al. (2017). Van der Waals epitaxial growth and optoelectronics of large-scale WSe₂/SnS₂ vertical bilayer p-n junctions. *Nat. Commun.* 8, 1906–1909. doi:10.1038/s41467-017-02093-z
- Yin, X., Shi, J., Niu, X., Huang, H., and Wang, X. (2015). Wedding cake growth mechanism in one-dimensional and two-dimensional nanostructure evolution. *Nano Lett.* 15, 7766–7772. doi:10.1021/acs.nanolett.5b04072
- Yin, X., Ye, Z., Chenet, D. A., Ye, Y., O'Brien, K., Hone, J. C., et al. (2014). Edge nonlinear optics on a MoS₂ atomic monolayer. *Sci.* (1979) 344, 488–490. doi:10.1126/science.1250564
- Zeng, H., Liu, G., Dai, J., Yan, Y., Zhu, B., He, R., et al. (2013). Optical signature of symmetry variations and spin-valley coupling in atomically thin tungsten dichalcogenides. *Sci. Rep.* 3, 1608. doi:10.1038/srep01608
- Zhang, L., Liu, K., Wong, A. B., Kim, J., Hong, X., Liu, C., et al. (2014). Three-dimensional spirals of atomic layered MoS₂. *Nano Lett.* 14, 6418–6423. doi:10.1021/nl502961e
- Zhang, L., Zhang, Z., Wu, F., Wang, D., Gogna, R., Hou, S., et al. (2020). Twist-angle dependence of moiré excitons in WS₂/MoSe₂ heterobilayers. *Nat. Commun.* 11, 5888–8. doi:10.1038/s41467-020-19466-6
- Zhao, Y., and Jin, S. (2020). Controllable water vapor assisted chemical vapor transport synthesis of WS₂-MoS₂ heterostructure. *ACS Mater. Lett.* 2, 42–48. doi:10.1021/acsmaterialslett.9b00415



OPEN ACCESS

EDITED BY

Nino Russo,
University of Calabria, Italy

REVIEWED BY

Isabella Romeo,
Magna Graecia University, Italy
Zoran S. Markovic,
University of Kragujevac, Serbia

*CORRESPONDENCE

Haroon Khan,
✉ hkdr2006@gmail.com

SPECIALTY SECTION

This article was submitted to Theoretical and Computational Chemistry, a section of the journal Frontiers in Chemistry

RECEIVED 20 February 2023

ACCEPTED 21 March 2023

PUBLISHED 13 April 2023

CITATION

Ahmad I, Kuznetsov AE, Pirzada AS, Alsharif KF, Daglia M and Khan H (2023), Computational pharmacology and computational chemistry of 4-hydroxyisoleucine: Physicochemical, pharmacokinetic, and DFT-based approaches. *Front. Chem.* 11:1145974. doi: 10.3389/fchem.2023.1145974

COPYRIGHT

© 2023 Ahmad, Kuznetsov, Pirzada, Alsharif, Daglia and Khan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Computational pharmacology and computational chemistry of 4-hydroxyisoleucine: Physicochemical, pharmacokinetic, and DFT-based approaches

Imad Ahmad¹, Aleksey E. Kuznetsov², Abdul Saboor Pirzada¹, Khalaf F. Alsharif³, Maria Daglia^{4,5} and Haroon Khan^{1*}

¹Department of Pharmacy, Abdul Wali Khan University Mardan, Mardan, Pakistan, ²Department of Chemistry, Universidad Tecnica Federico Santa Maria, Santiago, Chile, ³Department of Clinical Laboratory, College of Applied Medical Science, Taif University, Taif, Saudi Arabia, ⁴Department of Pharmacy, University of Naples Federico II, Naples, Italy, ⁵International Research Centre for Food Nutrition and Safety, Jiangsu University, Zhenjiang, China

Computational pharmacology and chemistry of drug-like properties along with pharmacokinetic studies have made it more amenable to decide or predict a potential drug candidate. 4-Hydroxyisoleucine is a pharmacologically active natural product with prominent antidiabetic properties. In this study, ADMETLab 2.0 was used to determine its important drug-related properties. 4-Hydroxyisoleucine is compliant with important drug-like physicochemical properties and pharma giants' drug-ability rules like Lipinski's, Pfizer, and GlaxoSmithKline (GSK) rules. Pharmacokinetically, it has been predicted to have satisfactory cell permeability. Blood-brain barrier permeation may add central nervous system (CNS) effects, while a very slight probability of being CYP2C9 substrate exists. None of the well-known toxicities were predicted *in silico*, being congruent with wet lab results, except for a "very slight risk" for respiratory toxicity predicted. The molecule is non ecotoxic as analyzed with common indicators such as bioconcentration and LC₅₀ for fathead minnow and *daphnia magna*. The toxicity parameters identified 4-hydroxyisoleucine as non-toxic to androgen receptors, PPAR-γ, mitochondrial membrane receptor, heat shock element, and p53. However, out of seven parameters, not even a single toxicophore was found. The density functional theory (DFT) study provided support to the findings obtained from drug-like property predictions. Hence, it is a very logical approach to proceed further with a detailed pharmacokinetics and drug development process for 4-hydroxyisoleucine.

KEYWORDS

4-hydroxyisoleucine, pharmacokinetics, computational chemistry, drug likeness, DFT

1 Introduction

4-Hydroxyisoleucine (4HIL) is an amino acid which has been purified from *Trigonella foenum-graecum* (fenugreek seeds; Fabaceae). Fenugreek is known for its antidiabetic properties in traditional medicine (Fowden et al., 1973; Abou El-Soud et al., 2007). 4-Hydroxyisoleucine increases insulin secretion *via* a direct effect on isolated islet cells both in rats and humans. The stimulating effect was strictly glucose dependent. Furthermore, it was shown that the insulinotropic effect is biphasic, without associated harm to other islet cells (α and δ -cells) and increase in insulin secretion by increasing glucose concentration (Sauvaire et al., 1998). 4-Hydroxyisoleucine reduces insulin resistance in the muscles and the liver by the activation of insulin receptor substrate-associated phosphoinositide-3 kinase activity. Reduction in body weight and plasma triglyceride and total cholesterol levels resulted from application of 4-hydroxyisoleucine. This plant-derived amino acid offers an attractive package for the treatment of the metabolic syndrome as a whole and also type 2 diabetes, obesity, and dyslipidaemias (Jetté et al., 2000; Narender et al., 2006). Keeping in mind the pharmacological importance of 4HIL, its efficacy needs to be evaluated before it moves to trials and/or is dismissed. Undesirable pharmacokinetic (PK) properties or unacceptable toxicity are the main causes of the failure of drug candidates at the clinical trial stage (Jia et al., 2020).

Since the concept of drug likeness was first proposed, it has become an important consideration in the selection of compounds with desirable bioavailability during the early phases of drug discovery. Over the past decade, online resources have facilitated drug-likeness studies in an economical and time-efficient manner. Michael J. Waring and colleagues compiled and analyzed the attrition of drug candidates from four pharma giants, AstraZeneca, Eli Lilly and Company, GlaxoSmithKline, and Pfizer. They reaffirmed that the control of physicochemical properties during structural optimization helps identify a compound of drug quality. Safety and toxicology are the largest sources of failure within the data set. However, none of the physicochemical descriptors correlate with preclinical toxicology outcomes (Waring et al., 2015).

While considering the various pharmacology effects in different diseased conditions, we calculated various parameters related to drug discovery for 4-hydroxyisoleucine using ADMETLab 2.0. in order to predict the drug-like properties and safety profile and thereby therapeutic potentials.

2 Methodology

Various virtual labs are authorized to share the pharmacokinetic profile of pharmacologically active drug candidates such as SwissADME (Daina et al., 2017), FAF-Drugs4 (Lagorce et al., 2017), ADMETLab (Xiong et al., 2021), admetSAR (Yang et al., 2019), pkCSM (Pires et al., 2015), and ProTox-II (Banerjee et al., 2018). However, ADMETLab was selected because it provides predictive spectrum of multiple drug-likeness parameters and most of ADMET properties. The updated version, ADMETLab 2.0, has overpowered all known shortcomings of the older version while keeping its valuable advantages.

This virtual lab calculates 88 ADMET-related properties, i.e., 17 physicochemical parameters, 13 medicinal chemistry properties, 23 ADME properties, 27 toxicity endpoints, and eight toxicophore rules (751 substructures). ADMETLab 2.0 is freely accessible at <https://admetmesh.scbdd.com>. The 2D structure of 4-hydroxyisoleucine was drawn in ADMETLab 2.0 (Table 1). The user interface also offers string-based search. The search string for 4-hydroxyisoleucine is CC(O)C(C)C(N)C(=O)O.

An oral bioavailability graph obtained from SwissADMET was calculated using physicochemical characteristics. This graph uses six parameters, i.e., lipophilicity, insolubility, size, unsaturation, polarity, and flexibility, and defines appropriate ranges for oral bioavailability. The GI absorption (HIA), BBB penetration, and being Pgp substrate were also predicted with SwissADMET. The result is depicted as “boiled egg” having two areas: one for HIA (white) and another for BBB penetration (yolk). The background is a gray zone, indicating neither GI absorption nor BBB penetration.

The human ether-a-go-go-related gene (hERG) K^+ channels blockade is linked with fatal cardiac arrhythmias. The hERG K^+ channel blockade was predicted with both ADMETLab 2.0 and pred-hERG 4.2. The pred-hERG 4.2, <http://predherg.labmol.com.br>, is a web tool for early detection of presumable hERG blockers and non-blockers, to predict cardiac toxicity.

DFT studies were performed with Gaussian 16 software (Frisch et al., 2016). We optimized the 4-hydroxyisoleucine molecule without any symmetry constraints, checking different possible isomers (*vide infra*), and then performed frequency calculations to verify that the optimized structures were true energy minima. All calculations were performed using the combination of the hybrid density functional B3LYP (Axel, 1993) with the triple zeta split-valence polarized basis set of Ahlrichs and co-workers Def2TZVP (Weigend and Ahlrichs, 2005; Weigend, 2006); the approach is furthermore referred to as B3LYP/Def2TZVP. We studied the 4-hydroxyisoleucine both in the gas phase (vacuum) and with the implicit effects from water (dielectric constant $\epsilon = 78.3553$) as a solvent taken into account. In the second type of calculations, we used the self-reliable IEF-PCM approach (Tomasi et al., 2005) using the UFF default model as implemented in Gaussian 16 software, with the electrostatic scaling factor $\alpha = 1.0$. The charge analysis was performed using the natural bond orbital (NBO) method as implemented in Gaussian 16 program (Reed et al., 1988), using the B3LYP/Def2TZVP approach with the implicit water effects. Frontier molecular orbitals (FMOs) and molecular electrostatic potential (MEP) were computed at the B3LYP/Def2TZVP level with the implicit water effects as well.

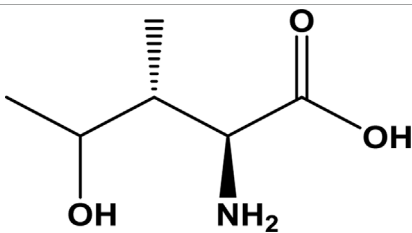
We consider the calculated structures of the seven isomers of 4-hydroxyisoleucine, structural parameters for the lowest-lying isomer, its NBO charges, FMOs, and MEP. Furthermore, we used the values of the energies of the lowest-lying isomer, HOMO and LUMO, to compute its global reactivity parameters (GRP) (Geerlings et al., 2003; Chakraborty et al., 2013; Jorio et al., 2019) (Eqs 1–6). Eqs 1, 2 were used to calculate the values of the ionization potential (IP) and electron affinity (EA):

$$IP = -E_{\text{HOMO}}, \quad (1)$$

$$EA = -E_{\text{LUMO}}. \quad (2)$$

For global hardness η and electronegativity X values, we used Eqs 3, 4:

TABLE 1 Structure, IUPAC name, and SMILES notation of 4-hydroxyisoleucine.

Ligand	4-Hydroxyisoleucine	
IUPAC	2-Amino-4-hydroxy-3-methylpentanoic acid	
SMILES notation	CC(O)C(C)C(N)C(=O)O	

$$\eta = \frac{[IP - EA]}{2} = -\frac{[E_{LUMO} - E_{HOMO}]}{2} \quad (3)$$

$$X = \frac{[IP - EA]}{2} = -\frac{[E_{LUMO} - E_{HOMO}]}{2} \quad (4)$$

Also, global electrophilicity ω value was calculated by Eq. 5:

$$\omega = \frac{\mu^2}{2\eta} \quad (5)$$

where $\mu = \frac{[E_{HOMO} + E_{LUMO}]}{2}$ is the chemical potential of the system.

Finally, the global softness σ value was computed with Eq. 6:

$$\sigma = \frac{1}{2\eta} \quad (6)$$

Open GL version of Molden 5.8.2 visualization program was used for the visualization of the structures and FMOs (Schaftenaar and Noordik, 2000), and Avogadro, version 1.1.1, was used to visualize the MEP maps (Hanwell et al., 2012).

3 Results

ADMETLab 2.0 offers results in different formats, even in a downloadable pdf format. The following parameters were predicted for 4-HIL.

3.1 Physicochemical properties

The widely accepted physicochemical property guidelines are the Lipinski's rule of five (RO5), which is satisfied by an ideal drug molecule. It offers a prediction of the drug likeness of a chemical compound having a biological activity and designed for oral route of administration. The molecular weight of 4-hydroxyisoleucine is 147.09 g/mol, which reflects its suitability for oral drug development. This provides therapeutic convenience of its formulation as oral dosage form. The optimal range is 100–600 approximately. The molecular volume to van der Waals volume ratio was calculated as 147.06. The number of hydrogen bond acceptors (nHA; optimal: 0–12) is 4. The number of hydrogen bond donors (nHD; optimal: 0–7) is also 4. 4-Hydroxyisoleucine has three rotatable bonds only (optimal: 0–11).

There are no rings or cyclic groups in 4-HIL (optimal: 0–6). There are four heteroatoms in 4-hydroxyisoleucine, which satisfy the range 1–15 given by ADMETLab. There is no formal charge on this molecule. A charge range of –4–4 is devised as standardized

parameter for a drug candidate. The number of rigid bonds may range from 0 to 30; 4HIL has a single rigid bond. This makes the molecule flexible enough to achieve a stable conformation at its target. In this continuity, the flexibility value is 3.0 obtained as ratio of nRot/nRig. The molecule has three stereo centers which is somewhat closer to the optimal value of ≤ 2 . A stereo center may achieve a different conformation at the active site. The topological polar surface area (TPSA) for 4-hydroxyisoleucine is 83.55.

The aqueous solubility (logS) is 0.06, falling within the range of –4–0.5 log mol/L. LogP value is the Log of octanol/water partition coefficient whose optimal range is 0–3. A higher logP value implies greater lipophilicity, and this depends upon molecular size, polarity, and hydrogen bonding. LogP value of 4-hydroxyisoleucine is –2.63 which reflects its partition into water compartment. This value is congruent to LogS value described previously. The value of logP at normal physiological pH 7.4 is logD. For 4-hydroxyisoleucine it is calculated as –0.93 just below the optimal range of 1–3. Physicochemical properties are summarized in Table 2.

3.2 Medicinal chemistry

The QED score (quantitative estimate of drug-likeness) is 0.49, which implies it is an unattractive molecule. An attractive drug has a QED score of >0.67; for unattractive molecules, the score is 0.49–0.67, while it is <0.34 for too complex molecules. The QED score is a drug-likeness score based on the calculated physicochemical properties of marketed oral drugs and published human data.

SAscore is implemented based on synthetic accessibility score, which is an estimated ease of synthesis of a drug-like molecule. Based on SAscore (3.59), 4-hydroxyisoleucine is easy to synthesize. Drug-like molecules having SAscore of ≥ 6 are difficult to synthesize, while those with a score of <6 are easy to synthesize. The Fsp³ value for 4-HIL is 0.83. Fsp³ is the fraction or number of sp³ carbon atoms out of the total carbon count. This parameter reflects the carbon saturation of molecules and characterizes the complexity of their spatial structure. The suitable value defined by ADMETLab for Fsp³ is >0.42, with 4-hydroxyisoleucine being compliant.

The optimum MCE-18 score limit given by ADMETLab 2.0 is ≥ 45 . For 4-HIL MCE-18 score is 4.0, which is extremely low. NP score reflects the natural product-likeness score. This score ranges from –5 to 5; the higher the score, the higher the probability of being a natural product. Natural products are more likely to be safe, with reduced risk of toxicity. NP score for 4-hydroxyisoleucine is 0.99, which is in this safe range.

TABLE 2 Physicochemical properties and medicinal chemistry parameters of 4-hydroxy isoleucine calculated with ADMETLab 2.0.

Physicochemical properties		Medicinal chemistry	
Parameter	Value	Parameter	Value
Molecular weight	147.09 g/mol	QED	0.49
van der Waals volume	147.06	SAscore	3.59
nHA	4	Fsp ³	0.83
nHD	4	MCE-18	4.0
nRot	3	NP score	0.99
nRing	0	Lipinski rule	Accepted
MaxRing	0	Pfizer rule	Accepted
nHet	4	GSK rule	Accepted
fChar	0	Golden Triangle	Rejected
nRig	1	PAINS	0 alerts
Flexibility = nRot/nRig	3.0	ALARM NMR	0 alerts
Stereo centers	3	BMS	0 alerts
TPSA	83.55 Å ²	Chelator rule	0 alerts
logS	0.06 log mol/L		
logP	−2.63		
logD	−0.94		

nHA, no. of hydrogen bond acceptors; nHD, no. of hydrogen bond donors; nRot, no. of rotatable bonds; nRing, no. of rings; MaxRing, no. of atoms in the biggest ring; nHet, no. of heteroatoms; fChar, formal charge; nRig, no. of rigid bonds; TPSA, total polar surface area; logS, aqueous solubility; logP, octanol–water partition coefficient; logD, distribution coefficient; QED, quantitative estimate of drug-likeness; SAscore, synthetic accessibility score; Fsp³, no. of sp³-hybridized carbon/total carbon count; MCE-18, medicinal chemistry evolution; NPscore, natural product-likeness score; GSK, GlaxoSmithKline; and PAINS, pan assay interference compounds.

*Unless otherwise specified, unit of measurement is a number/score explained in the respective text.

ADMETLab 2.0 has described Lipinski's rule as follows: $MW \leq 500$, $\log P \leq 5$, $Hacc \leq 10$, and $Hdon \leq 5$. These parameters are already described individually under physicochemical properties. All these parameters are met by 4-HIL and thus the widely accepted Lipinski rule is satisfied. In the case where two of these properties are out of range, there is a possibility of poor absorption or permeability. Deviation of only one property is allowed. 4-Hydroxyisoleucine also complies with Pfizer's rule, i.e., $\log P > 3$ and $TPSA < 75$. Compounds which satisfy the GSK rule may have a more favorable ADMET profile. 4-HIL concurs with GSK's rule as well. The Golden Triangle rule devises the range for two parameters as follows: $200 \leq MW \leq 50$ and $-2 \leq \log D \leq 5$. Compounds satisfying this Golden Triangle rule are more likely to have favorable ADMET profiles. As per this criterion, 4-hydroxyisoleucine is rejected. However, ADMET profiles are separately described later which possess desirable pharmacokinetics.

4-HIL has zero PAINS alerts, and thus has been excluded from frequent hitters, α -screen artifacts, and reactive compounds. 4-Hydroxyisoleucine has no ALARM NMR alerts and is not a thiol reactive compound. 4-Hydroxyisoleucine has no red alert for BMS and the chelator rule and so is a desirable, non-reactive, and non-chelating compound. Medicinal chemistry parameters are summarized in Table 2.

3.3 Pharmacokinetics

3.3.1 Absorption

Caco-2 cell lines (Caucasian colon adenocarcinoma) are still widely used in absorption studies. The Caco-2 permeability score obtained was -5.99 which is lower than the minimum optimal score of -5.15 log unit. The Madin–Darby canine kidney (MDCK) permeability score was 0.01 cm/s which is in the acceptable limit. The standard scores for low permeability, medium permeability, and high passive permeability are $< 2 \times 10^{-6}$ cm/s, $2-20 \times 10^{-6}$ cm/s, and $> 20 \times 10^{-6}$ cm/s, respectively.

4-Hydroxyisoleucine has a very low probability as Pgp inhibitor as per its score of 0.01 . Therefore, it is regarded as safe from having significant drug interactions. For an inhibitor, the score is 1, while 0 score denotes a non-inhibitor. This output value exhibits the probability of being a Pgp inhibitor. The probability of being Pgp substrate is the lowest, as indicated by its score 0.05 . For a Pgp substrate, the score is 1, while for a non-substrate, the score is 0.

4-HIL is predicted to have low human intestinal absorption (HIA $< 30\%$) as indicated by its HIA score, 0.09 . The bioavailability is predicted to be greater than 20% and 30% ($F_{20\%}$ 0.004 , $F_{30\%}$ 0.002).

The oral bioavailability graph shown in Figure 1A is predicted with the SwissADMET database. This graph is based on six physicochemical

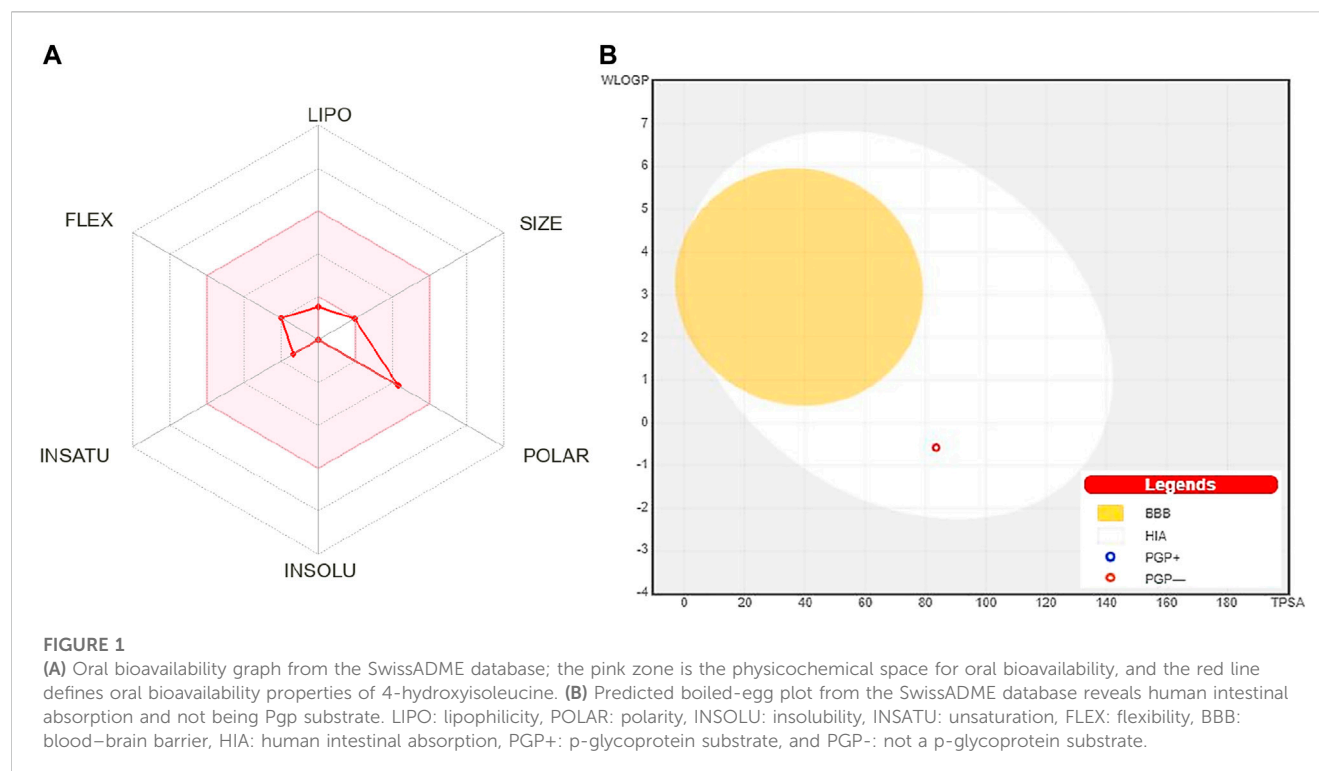


TABLE 3 Pharmacokinetics of 4-Hydroxy isoleucine calculated with ADMETLab 2.0.

Parameter	Value/probability	Parameter	Value/probability
Absorption		Metabolism	
Caco-2 permeability	–5.99	CYP1A2 inhibitor	0.01
MDCK permeability	0.01 cm/s	CYP1A2 substrate	0.06
Pgp inhibitor	0.01	CYP2C19 inhibitor	0.038
Pgp substrate	0.05	CYP2C19 substrate	0.10
HIA	0.09	CYP2C9 inhibitor	0.02
F _{20%}	0.004	CYP2C9 substrate	0.49
F _{30%}	0.002	CYP2D6 inhibitor	0.04
Distribution		CYP2D6 substrate	0.19
PPB	6.73%	CYP3A4 inhibitor	0.01
VD	0.46 L/kg	CYP3A4 substrate	0.07
BBB penetration	0.83	Excretion	
Fu	85.29%	CL	7.26 mL/min/kg
		T _{1/2}	0.68

Caco-2, Caucasian colon adenocarcinoma cell lines; MDCK, Madin–Darby canine kidney; Pgp, *para*-glycoprotein; HIA, human intestinal absorption; F, bioavailability; PPB, plasma protein binding; VD, volume of distribution; BBB, blood–brain barrier; Fu, unbound fraction in plasma; CYP, cytochrome P450; CL, clearance; T_{1/2}, half-life. *Unless otherwise specified, unit of measurement is a probable score explained in the respective text.

characteristics, i.e., lipophilicity, size, polarity, insolubility, unsaturation, and flexibility. The result of the compound 4-hydroxyisoleucine was within these limits, demonstrating a favorable physio-chemical profile, which is a prerequisite factor that must be monitored in pharmaceutical

and clinical studies. The colored zone is the suitable physicochemical space for oral bioavailability.

HIA and central nervous system (CNS) absorption needs to be checked for every biomolecule before its formulation in the

TABLE 4 Toxicity profile of 4-hydroxy isoleucine calculated with ADMETLab 2.0.

Parameter	Probability
hERG blockers	0.01
H-HT	0.11
DILI	0.02
AMES toxicity	0.01
Rat oral acute toxicity	0.11
FDAMDD	0.01
Skin sensitization	0.11
Carcinogenicity	0.03
Eye corrosion	0.02
Eye irritation	0.07
Respiratory toxicity	0.45

hERG, human ether-a-go-go-related gene; H-HT, human hepatotoxicity; DILI, drug-induced liver injury; FDAMDD, FDA maximum daily dose.

*Unless otherwise specified, the unit of measurement is a probable score explained in the respective text.

pharmaceutical or clinical trials. The blood–brain barrier penetration is essential only if the compound has a target in the CNS, but the inactive compounds on the CNS must not intersect to avoid adverse effects. The boiled-egg plot shows that 4-hydroxyisoleucine has a high gastrointestinal absorption without BBB permeability, indicating low occurrence for CNS side effects (Figure 1B).

3.3.2 Distribution

Plasma protein binding (PPB), unbound fraction in plasma (Fu), volume of distribution (VD), and blood–brain barrier (BBB) permeability were considered to study the distribution (Table 4). The predicted PPB was 10.90%. The optimal range of PPB is <90% because drugs which are high protein-bound may have a low therapeutic index. The volume of distribution (VD) was calculated as 0.39 L/kg. The optimal range for VD is 0.04–20 L/kg. 4-Hydroxyisoleucine has more tendency for blood–brain barrier penetration reflected by the score, 0.73. The output value indicates the probability of being able to cross BBB. The unbound fraction in plasma (Fu) was calculated as 87.84%. The scores for low, middle, and high unbound plasma fraction are <5%, 5–20%, and >20%, respectively. This implies that more unbound plasma fraction will be available for pharmacological action.

3.3.3 Metabolism

This parameter is important from the drug plasma concentration perspective. The database defines the ligands' probability to inhibit the enzyme or not by its placement either in category 1 (Inhibitor) or category 0 (non-inhibitor). Similarly, the probability of being substrate for the enzyme is exhibited by the score 1 or 0. Category 1 denotes the molecule is substrate, while category 0 describes it to be non-substrate of the enzyme.

4-Hydroxyisoleucine is likely a non-inhibitor of CYP1A2 as indicated by the assigned score 0.01. The CYP1A2 substrate score obtained as 0.06 implies it is a non-substrate. The probability of CYP2C19 inhibition and being CYP2C19 substrate is extremely low. 4-Hydroxyisoleucine appeared to be non-inhibitor of CYP2C9.

However, very low probability exists for being CYP2C9 substrate. Similarly, there is no incidence of CYP2D6 inhibition and being a CYP2D6 substrate. 4-Hydroxyisoleucine is not deemed a CYP3A4 inhibitor nor is it a substrate for CYP3A4.

3.3.4 Excretion

The clearance (CL) for 4-hydroxyisoleucine is calculated as 7.27 mL/min/kg which is a moderate clearance rate. If a drug has a high clearance rate, its score will be > 15 mL/min/kg; for a moderate clearance rate, it will be 5–15 mL/min/kg; and for a low clearance rate, it will be < 5 mL/min/kg. The half-life ($T_{1/2}$) value is 0.62, which reflects the probability of long half-life. The database output value may be from category 1 (long half-life, >3 h) or category 0 (short half-life, <3 h). All the pharmacokinetic parameters (absorption, distribution, metabolism, and excretion) are summarized in Table 3.

3.4 Toxicity

The toxicity parameters verified by ADMETLab include hERG blockade, human hepatotoxicity, drug-induced liver injury, AMES toxicity, rat oral acute toxicity, FDA maximum daily dose, carcinogenicity, mutagenicity, skin sensitization, eye corrosion, eye irritation, and respiratory toxicity (Table 4).

The ADMETLab 2.0 score for hERG blocker is 0.01 which shows the safety of 4-hydroxyisoleucine. The result was further in line with Pred-hERG which defined 4-hydroxyisoleucine as non-cardiotoxic with 90% confidence, as shown in Figure 2. The red dotted lines defined the area which negatively contributes to hERG blockade, while contour lines with intense green color represent a higher positive contribution of an atom to hERG blockade. 4-Hydroxyisoleucine is not human hepatotoxic (H-HT) as its score is 0.11. Molecules falling in category 1 are hepatotoxic (H-HT positive, +), while category 0 reflects non-hepatotoxic (H-HT

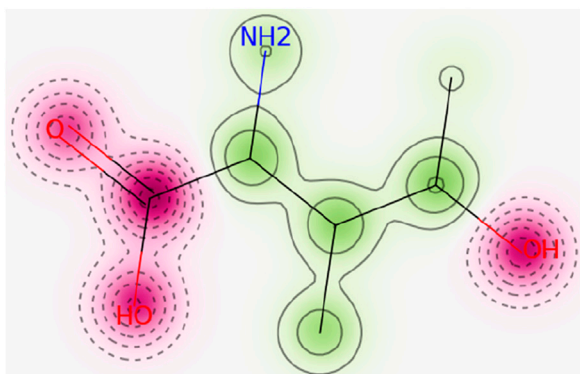


FIGURE 2

In the probability map of hERG blockage (Pred-hERG web tool), the red dotted lines reflect non-hERG blocker. Contour lines with intense green color indicate a higher positive contribution of an atom to hERG blockade.

negative, –). Similarly, there is low risk of drug-induced liver injury (DILI), 0.02. Commonly, category 1 drugs have high risk of DILI, while category 0 drugs have no risk of DILI. Furthermore, Ames toxicity score also indicated safety of 4-hydroxyisoleucine, evident from its score of 0.01. Category 1 drugs are classified as Ames positive (+, toxic), while category 0 represents Ames negative (–, non-toxic). Finally, the risk of oral toxicity is low. The rat oral acute toxicity score is 0.11, which classifies 4-hydroxyisoleucine in category 0 (low-toxicity), as category 1 indicates high-toxicity.

4-Hydroxyisoleucine has the probability of being skin non-sensitizer with a value of 0.11. Category 1 drugs are skin sensitizers, while category 0 are non-sensitizers. The risk of carcinogenicity (0.03) is lower, and the molecule is safe from adverse reactions. Category 1 drugs are carcinogens, while category 0 are non-carcinogens. A negligible risk of eye corrosion is predicted as exhibited by its score, 0.02. Category 1 drugs are corrosives, while category 0 drugs are non-corrosive to the eye. The output value indicates the probability of being non-irritant to the eye. The eye irritation score is 0.07. Category 1 drugs are irritants, while category 0 drugs are non-irritants. Finally, the respiratory toxicity score (0.45) declares borderline low probability to be free from respiratory toxicity. Category 1 indicates respiratory toxicants, and category 0 drugs are believed to be free from respiratory toxicity.

3.5 Environmental toxicity

For a drug, it is also important to be non-toxic to the environment. Therefore, ADMETLab defines some of the environmental toxicity indicators which are provided in Table 5. The bioconcentration factor, 0.17, is in favor of 4-hydroxyisoleucine non-ecotoxicity. The 50% growth inhibitory concentration (IGC₅₀) for *Tetrahymena pyriformis* is 2.43. For 96-h fathead minnow (*Pimephales promelas*), the 50% lethal concentration LC₅₀FM of 4-hydroxyisoleucine is 2.92 (Dionne et al., 2021). The score for a 48-

TABLE 5 Environmental toxicity of 4-hydroxyisoleucine calculated with ADMETLab 2.0.

Parameter	Value
Bioconcentration factors	0.167 -log ₁₀ [(mg/L)/(1,000*MW)]
IGC ₅₀ <i>Tetrahymena pyriformis</i>	2.432 -log ₁₀ [(mg/L)/(1,000*MW)]
LC ₅₀ FM (fathead minnow)	2.916 -log ₁₀ [(mg/L)/(1,000*MW)]
LC ₅₀ DM <i>daphnia magna</i>	3.199 -log ₁₀ [(mg/L)/(1,000*MW)]

IGC₅₀, 50% growth inhibitory concentration and LC₅₀, 50% lethal concentration.

h *daphnia magna* 50 percent lethal concentration (LC₅₀DM) for 4-hydroxyisoleucine is 2.36.

3.6 Tox21 pathway

There is probability, to a negligible extent, of 4-hydroxyisoleucine binding to androgen receptor (NR-AR). The output value is 0.05, indicating its safety toward genetic instability. In another approach to test its safety, it is predicted that 4-hydroxyisoleucine will not bind to androgen receptor ligand-binding domain (NR-AR-LBD). Such a reduced probability is exhibited from its lowest value 0.01.

With such a reduced probability (0.01), inactivity at aryl hydrocarbon receptor is evidenced. No activity at NR-Aromatase is predicted, as 0.01 is near to zero, which means inactive. However, 4-hydroxyisoleucine has a low probability to remain inactive at binding to the estrogen receptor (0.42), while reduced affinity for estrogen receptor ligand-binding domain exists (0.01). The peroxisome proliferator-activated receptor gamma is also unlikely to be affected by 4-hydroxyisoleucine, as the score is 0.01. 4-Hydroxyisoleucine is not predicted to be an antioxidant response element since the probability is negligible at 0.02. It is also unlikely to interfere with the activity of ATPase family AAA domain-containing protein 5 as predicted score is 0.01. Furthermore, it is unlikely to activate the heat shock factor response element as the score is 0.01. The score for non-modification of mitochondrial membrane potential is 0.01. Finally, 4-hydroxyisoleucine also has low probability (0.01) of affecting p53 (see Table 6).

3.7 Toxicophore rules

This parameter rules out the presence of a toxicophore, causing severe adverse effects. The acute toxicity rule exhibit zero alerts for acute toxicity during oral administration. The molecule is safe from genotoxic carcinogenicity or mutagenicity. Furthermore, 4-hydroxyisoleucine is clear of a non-genotoxic carcinogenic mechanism as well, with zero alerts. No toxicophore was found for skin sensitization, and 4-hydroxyisoleucine is expected to be free from skin irritation. The molecule is not toxic to aquatic life, with zero alerts for the aquatic toxicity rule (water toxicity). 4-Hydroxyisoleucine also appears biodegradable as ADMETLab revealed zero alerts for the non-biodegradable rule. 4-Hydroxyisoleucine has no enlistment in SureChEMBL database, so it has zero alerts for the

TABLE 6 Toxicity profile of 4-hydroxyisoleucine calculated with ADMETLab 2.0.

Parameter	Probability
NR-AR	0.05
NR-AR-LBD	0.01
NR-AhR	0.01
NR-Aromatase	0.01
NR-ER	0.42
NR-ER-LBD	0.01
NR-PPAR gamma	0.01
SR-ARE	0.02
SR-ATAD5	0.01
SR-HSE	0.01
SR-MMP	0.01
SR-p53	0.01

NR-AR, androgen receptor; NR-AR-LBD, androgen receptor ligand-binding domain; NR-AhR, aryl hydrocarbon receptor; NR-ER, estrogen receptor, NR-ER-LBD, estrogen receptor ligand-binding domain; NR-PPAR gamma, peroxisome proliferator-activated receptor gamma; SR-ARE, antioxidant response element; SR-ATAD5, ATPase family AAA domain-containing protein 5; SR-HSE, heat shock factor response element; SR-MMP, mitochondrial membrane potential.

TABLE 7 Toxicophore based profile of 4-hydroxy isoleucine calculated with ADMETLab 2.0.

Parameter	Value
Acute toxicity rule	0 alerts
Genotoxic carcinogenicity rule	0 alerts
Non-genotoxic carcinogenicity rule	0 alerts
Skin sensitization rule	0 alerts
Aquatic toxicity rule	0 alerts
Non-biodegradable rule	0 alerts
SureChEMBL rule	0 alerts

SureChEMBL rule and is not classified as MedChem unfriendly status (see Table 7).

3.8 DFT study results

In Figure 3, the optimized structures of seven isomers studied for 4-hydroxyisoleucine are shown, and Table 8 summarizes the energies, relative energies, HOMO/LUMO energies, and HOMO/LUMO energy gaps of these seven isomers. As can be seen from Figure 3, the isomers differ from each other by orientations of 4-hydroxy and amino groups and hydrogens of the carbonyl, hydroxyl, and amino groups. The lowest-lying isomer 5 (Figure 3) has the carbonyl oxygen of the carboxyl group and amino group located on the same molecule side and the 4-hydroxy group located in the plane approximately perpendicular to the carboxyl group, with hydrogen bonds formed between the 4-hydroxy and amino groups and between the amino group and the

carbonyl oxygen of the carboxyl group (see more detailed discussion later). Furthermore, from Table 8, it follows that all six higher-lying 4-hydroxyisoleucine isomers are quite close in energy to the global minimum isomer 5, being within 0.09 (isomer 3)–2.89 (isomer 7) kcal/mol higher in energy. We focus on the more detailed consideration of only the lowest-lying isomer 5. Further consideration of Table 4 data shows that all seven isomers have significant HOMO/LUMO gaps, 6.51–7.07 eV, with the isomer 5 having the HOMO/LUMO gap value 6.86 eV, the highest in energy isomer 7 having the smallest HOMO/LUMO gap value, 6.51 eV, and the second lowest-lying isomer 3 having the largest HOMO/LUMO gap value, 7.07 eV.

Figure 4 presents the selected bond distances, Å (Figure 4A), NBO charges on the selected atoms (Figure 4B), FMOs (Figure 4C), and MEP plot (Figure 4D) for the lowest-lying isomer 5. As aforementioned, there are several intermolecular hydrogen bonds in this isomer, between the 4-hydroxy group hydrogen and amino group nitrogen (bond distance, 2.166 Å), between the amino group hydrogens and the carbonyl oxygen of the carboxyl group (bond distances, 2.652 and 3.097 Å), and within the carboxyl group, 2.307 Å. Thus, as can be seen, intra- and also intermolecular hydrogen bonds can be formed by 4-hydroxyisoleucine, which might explain the presence of several isomers close in energy. Furthermore, NBO charge analysis (Figure 4B) shows noticeable negative charges on the oxygens and nitrogen of the molecular functional groups, −0.625 to −0.835 e, along with noticeable positive charges on the hydrogens of these functional groups, 0.372–0.503 e. These charges would explain the formation of intramolecular hydrogen bonds along with intermolecular interactions. The HOMO and LUMO plots (Figure 4C) show that essentially the whole molecule contributes to FMOs, thus being able to participate in potential processes of donating/accepting electrons. Finally, analysis of the MEP plot

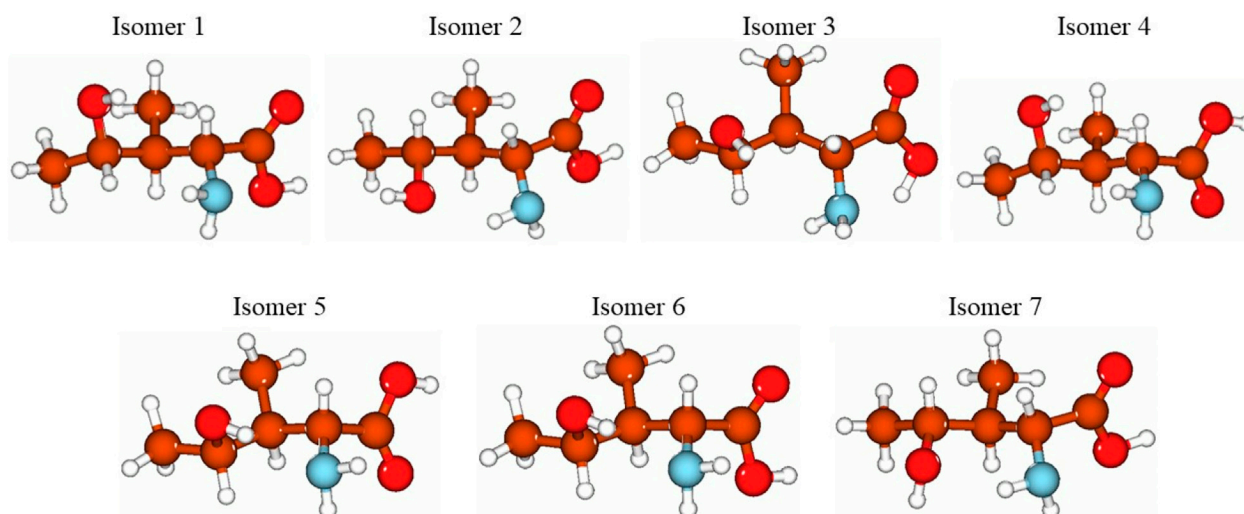


FIGURE 3

Structures of the seven studied isomers of the 4-hydroxyisoleucine optimized using the B3LYP/Def2TZVP method with the implicit effects of water. Color coding: brown for C, red for O, light blue for N, and light gray for H.

TABLE 8 Energetics of the seven optimized isomers of 4-hydroxyisoleucine calculated using the B3LYP/Def2TZVP method with the implicit effects of water.

Isomer	E_0 A.U.	$E_0 + \text{ZPEA.U.}$	DE, kcal/mol	E (HOMO/LUMO)	DE (HOMO/LUMO), eV
				A.U.	
1	-517.127629	-516.931358	2.51	-0.25518/-0.00954	6.68
2	-517.127409	-516.930878	2.65	-0.25034/-0.00949	6.55
3	-517.131486	-516.934972	0.09	-0.26491/-0.00496	7.07
4	-517.128038	-516.931992	2.25	-0.25778/-0.01037	6.73
5	-517.131629	-516.934014	0.0	-0.26509/-0.01290	6.86
6	-517.130944	-516.933490	0.43	-0.25819/-0.01357	6.66
7	-517.127028	-516.930205	2.89	-0.24850/-0.00925	6.51

(Figure 4D) shows accumulation of both negative (as indicated by red color) and positive (as indicated by blue color) electrostatic potential in the molecule. Negative MEP is accumulated on the carboxyl oxygens and on the 4-hydroxy oxygen, whereas positive MEP is accumulated on the amino group, carboxyl group hydrogen, and 4-hydroxy group hydrogen. This implies that the 4-hydroxyisoleucine molecule can behave both as a nucleophilic and electrophilic agent in chemical reactions. Also, these MEP accumulations imply possibility of intermolecular interactions for this compound.

Table 8 summarizes the values of the GRPs calculated according to Eqs 1–6 (see Materials and Methods section).

Analysis of the GRP values in Table 9 shows the following. 1) The 4-hydroxyisoleucine compound has quite a high IP value, 7.21 eV, but low EA value, 0.35 eV. This implies that this compound should be a poor electron donor and a relatively poor electron acceptor, or, in other words, should be relatively stable in redox processes, which is also supported by its significant HOMO/LUMO gap value, 6.86 eV. 2) The global

electronegativity X and global hardness η values of this compound are quite noticeable, 3.78 and 3.43 eV, respectively, which would imply that this compound might show electrophilic behavior in chemical processes but also should be relatively non-reactive and thus thermodynamically stable, which is also supported by its large HOMO/LUMO gap and noticeable value of its chemical potential, -3.78 eV. 3) Furthermore, the global softness of the compound has low value, 0.146 eV, which supports the relatively low reactivity of the compound, and global electrophilicity of the compound can be considered as relatively noticeable, 2.083 eV, which would imply that this compound might show electrophilic behavior in chemical processes.

4 Discussion

The original rule of five (Lipinski's RO5) dealt with orally active compounds and had defined four simple physicochemical

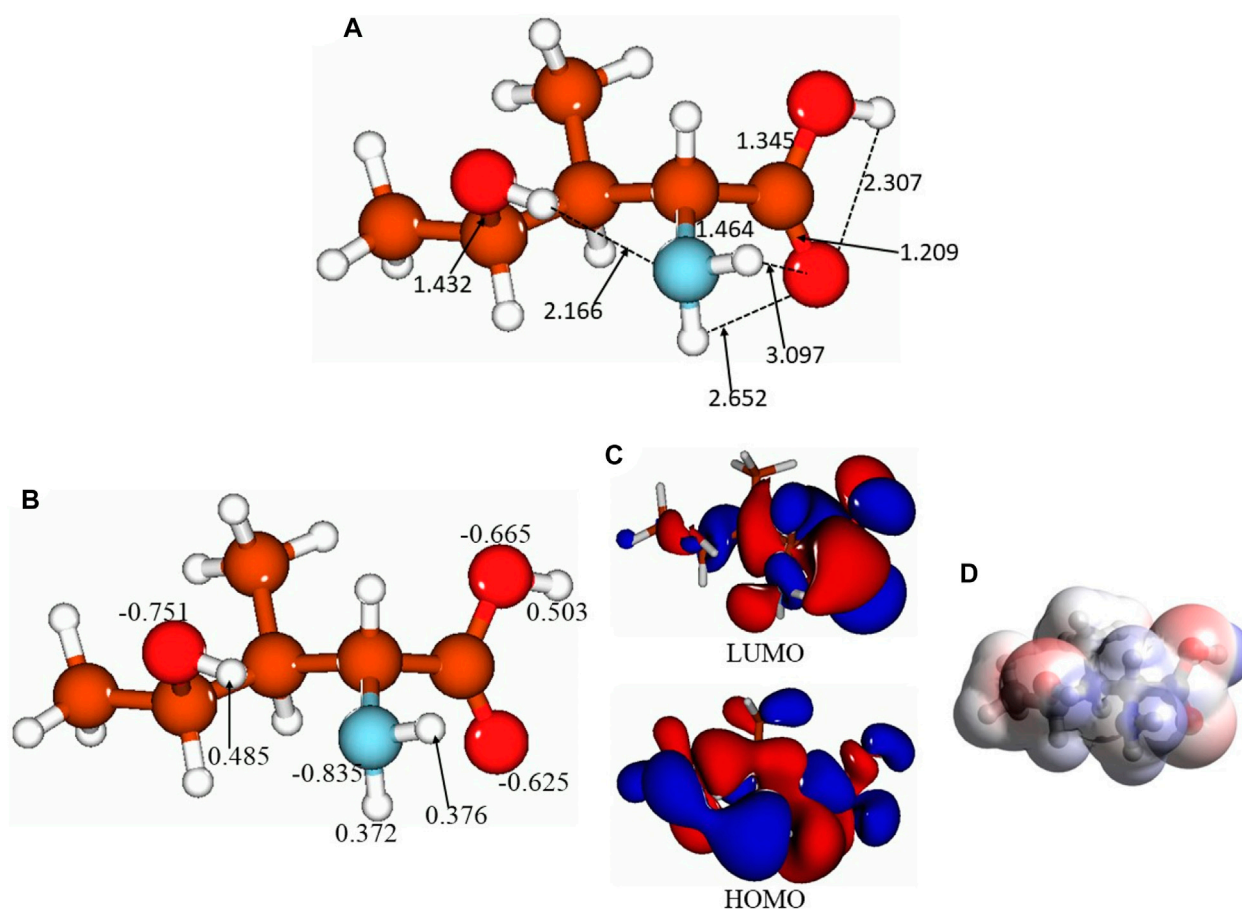


FIGURE 4
Lowest-lying isomer 5 of 4-hydroxyisoleucine: selected bond distances, Å (A), NBO charges on the selected atoms (B), FMOs (C), and MEP plot (D).

TABLE 9 Calculated GRPs for the isomer 5 (eV).

IP	EA	Gap	X	η	M	σ	ω
7.21	0.35	6.86	3.78	3.43	-3.78	0.146	2.083

parameters, i.e., molecular weight ≤ 500 , $\log p \leq 5$ representing hydrophobicity, H-bond donors ≤ 5 , and H-bond acceptors ≤ 10 . These properties have been associated with 90% of the orally active drugs that have achieved phase II clinical status (Lipinski, 2004). 4-Hydroxyisoleucine is compliant with one of the most acceptable drug-like parameters. Passing RO5 means acceptable aqueous solubility along with intestinal permeability, which are the first steps in oral bioavailability. The RO5 has found its importance where medicinal and combinatorial chemistry produced thousands of compounds with very poor physicochemical properties. However, passing RO5 never guaranteed that a compound will be drug-like. But if a compound fails, then a high probability of oral activity problems exists.

The molecular weight of 4-hydroxyisoleucine is 147.09, which is ideal for oral drug designing. The intrinsic molecular volume is considered as a molecular descriptor in modeling physicochemical properties and the biological activity. The molecular volume is one

of the most common descriptors in quantitative structure–activity relationship (QSAR) studies (Leahy, 1986; McGowan and Mellors, 1986; Consonni et al., 2002; Zhao et al., 2003a). It is related to different physicochemical properties and biological processes including intestinal absorption and BBB penetration, to name a few (Zhao et al., 2003b). The number of hydrogen bond acceptors and donors are within the required strict limits. Oral drugs usually have fewer H-bond acceptors, H-bond donors, and rotatable bonds (Bitew et al., 2021).

The parameter of rotatable bond count is used as “drug filter.” More than 10 rotatable bonds are associated with decreased rat oral bioavailability (Veber et al., 2002). The mechanistic basis for this “rotatable bond filter” is not clear, since its count does not relate to the *in vivo* clearance rate in rats. However, the filter still remains reasonable from a viewpoint of *in vitro* screening as ligand affinity decreases at an average of 0.5 kcal for each two rotatable bonds (Dashti et al., 2014). The number of aromatic and heteroaromatic rings in a molecule has a relation with aqueous solubility, lipophilicity, oral bioavailability, serum albumin binding, CYP450 inhibition, and hERG inhibition. The fewer the aromatic rings in an oral drug candidate, the greater is the developable probability of a drug candidate. However, >3 aromatic rings in a molecule are linked with poorer compound developability and an

increased risk of knock out. Even within the defined lipophilicity range, increasing aromatic rings lead to decreased aqueous solubility (Ritchie and Macdonald, 2009).

Saturated heterocycles are metabolized at the position of heteroatom or adjacent to it (Jean and Fotsch, 2012). The heteroatom is sometimes beneficial in drug metabolism like nitrogen in 4-hydroxyisoleucine. Redox reactions of drug metabolism such as heteroatom dealkylation, hydroxylation, heteroatom oxygenation, reduction, and dehydrogenation can yield active metabolites. Rarely, even conjugation reactions can give rise to an active metabolite (Obach, 2013). 4-Hydroxyisoleucine has a single rigid bond. This makes the molecule flexible enough to achieve a stable conformation at its target. The optimal range for drug-like property, TPSA, is 0–140 Å², that of 4-Hydroxyisoleucine within accurate limit (83.55 Å²). Molecules with a TPSA of ≥140 Å² would be poorly absorbed (<10% of the fraction absorbed), while a TPSA of 60 Å² favors >90% absorption (Clark, 1999).

The calculation of QED score uses eight important properties that were previously used to assess drug-likeness, i.e., molecular weight, octanol–water partition coefficient (ALOGP), number of H-bond donors, number of H-bond acceptors, molecular polar surface area, number of rotatable bonds, number of aromatic rings, and number of structural alerts (ALERTS). ALERTS is a list of undesirable chemical features such as chemical reactivity or perceived toxicity. Drugs with high QED scores achieve higher absorption and bioavailability, have lower doses, fewer drug–drug interaction warnings and P-glycoprotein interactions, and lower absorption issues due to a food-related effect. Nonetheless, it is surprising that high QED-scoring drugs have similarities to low-scoring drugs with respect to plasma free fraction, the extent of gut-wall metabolism, first-pass hepatic extraction/metabolism, volume of distribution, clearance, elimination half-life, and frequency of dosing (Ritchie and Macdonald, 2014). In this context, 4-hydroxyisoleucine lies at the borderline of attractive and unattractive molecules. Thus, this negligible amount of difference may not provide an impetus for rejection.

Fsp³ is the fraction of sp³ carbon atoms. It is the number of sp³-hybridized carbons out of the total carbon count which tells about carbon saturation of molecules and characterizes the complexity of their spatial structure (Lovering et al., 2009). Natural products generally possess a higher fraction of sp³ than synthetic compounds and so become a rich source of drugs (Jia et al., 2020). Yan and Gesteiger (2003) used it to characterize aliphatic degrees of molecules and ultimately predict solubility (Yan and Gasteiger, 2003). The optimum value of ≥0.42 for Fsp³ is considered a suitable value, and about 84% of marketed drugs meet this criterion (Kombo et al., 2013). However, sp³ content should be increased within an appropriate range; higher Fsp³ score does not always assure higher performance and can increase the difficulty of synthesis (Gerlach et al., 2019). However, critical discussions of Fsp³ scoring have helped in generating new potential descriptors, such as MCE-18, which is discussed in the following paragraphs, and calculated using the following equation (Wei et al., 2020).

$$\text{MCE-18} = (\text{AR} + \text{NAR} + \text{CHIRAL} + \text{SPIRO} + \text{sp}^3 + \text{cyc} - \text{A}_{\text{cyc}/1} + \text{sp}^3) \times \text{Q}_1.$$

A simple sp³ index (number of sp³-hybridized carbon atoms) can lead to a high rate of false-positive results, while the issue can be

solved with MCE-18 descriptor. The composition, spatial geometry, and complexity of chemical structures especially for state-of-the-art molecules are not considered in sp³ index, as these special features cannot be reflected using only the portion of sp³-hybridized carbons. They focus on the nature and quality of sp³-rich frameworks rather than sp³ counts only (Ivanenkov et al., 2019).

4-Hydroxyisoleucine is a naturally derived molecule, and therefore, MCE-18 may not be the only guide toward drug discovery. MCE-18 is applied for the assessment of pharmacological novelty and may help in designing new potential chemical entities in modern drug design. This index is also used for profiling HTS-focused libraries and prioritization of molecules. Presumably, compounds with optimal MCE-18 values have greater reliable IP positions. It is a tracer of medicinal chemistry evolution and is used to assess the “distance” between old and new chemistry. During the past 10 years, two of the top four pharmaceutical companies have primarily focused on molecules with MCE-18 values over 60, while the remaining two companies have demonstrated less drug productivity but have reached similar scores (Ivanenkov et al., 2019). The natural product score is congruent with the real-time natural source of 4-hydroxyisoleucine, namely, fenugreek seeds (Fowden et al., 1973).

4-Hydroxyisoleucine is compliant with the Lipinski rule, Pfizer rule, and GSK rule. The parameters for the Pfizer rule are logP >3 and TPSA <75. Compounds with higher log P (>3) and low TPSA (<75) are more likely to be toxic. The most significant work which describes the influence of molecular properties on toxicological outcomes has led to the “3/75 rule.” This has been devised by Pfizer from an analysis of exploratory or dose-finding toxicology studies of 245 compounds. This approach defined that compound with logP <3 and tPSA >75 Å² are expected to be non-toxic (Hughes et al., 2008; Waring et al., 2015). However, 4-hydroxyisoleucine is non-toxic and classified as safe with the Pfizer rule. The GSK rule suggests MW ≤ 400 and logP ≤ 4 (Hann, 2011) which are well qualified by 4-hydroxyisoleucine. The Golden Triangle aids in achieving metabolically stable and permeable drug candidates. It is a visualization tool developed with three parameters i.e., *in vitro* permeability, *in vitro* clearance, and computational data (Johnson et al., 2009). However, the stability and permeation of 4-hydroxyisoleucine is already demonstrated (Gowtham et al., 2022). Thus, the predicted Golden Triangle rejection may be ruled out here.

PAINS, ALARM NMR, BMS, and chelator rule are in favor of 4-hydroxyisoleucine. Pan assay interference compounds have a tendency to produce false-positive hits in high-throughput screening. The mechanism of PAINS activity is poorly understood. However, PAINS are associated with protein reactivity and non-covalent interactions (Bolz et al., 2021). ALARM NMR is critical to recognize reactive compounds that should be prioritized for lead optimization. This has emerged as a new filtering tool that can identify nuisance compounds and ultimately improve the process of hit triage. It is a La assay to detect reactive molecules by NMR by monitoring dithiothreitol-dependent ¹³C chemical shift changes brought in the human La antigen (Huth et al., 2005; Zega, 2017; Dahlin et al., 2018). Thiol reactive compounds oxidize or form covalent adducts with thiol groups of proteins. Therefore, ALARM NMR is a sensitive tool to rapidly identify compounds with increased risk of side effects in

humans, such as alcohol intolerance, reactive oxygen species generation, and drug–drug interactions (Huth et al., 2007). The BMS is an alert for mapping molecular promiscuity and identification of undesirable and reactive compounds, while the chelator rule reflects chelating compounds (Agrawal et al., 2010; Yang et al., 2021).

Caco-2 cell lines are still widely used in absorption studies. Sometime regarded as the “gold standard” technique, Caco-2 cells are used to validate other absorption assays (Artursson et al., 2001; Hidalgo, 2001; Ding et al., 2021; Murador et al., 2021). 4-Hydroxyisoleucine has a lower Caco-2 cell-predicted permeability. However, a considerable chance of MDCK permeability exists. The MDCK cell line is also used for permeability screening in drug discovery (BOKULIĆ et al., 2022). However, both ADMETLab 2.0 and SwissADME are in favor of better gastrointestinal absorption and bioavailability, indicated by the boiled-egg plot as well. However, there are no permeability results available from wet labs. 4-Hydroxyisoleucine is safe from clinically significant drug interactions. The absorption and excretion of Pgp substrate can be modified by its inducer/inhibitor, ultimately paving the way for drug interaction (Yamazaki et al., 2019). A Pgp inhibitor is more likely to have interactions (Telbisz et al., 2021; Terrier et al., 2021). ADMETLab 2.0 and SwissADME declared it as a non-Pgp substrate, diminishing the risk of potential drug interactions and drug resistance mechanisms. Metabolic enzyme inducers or inhibitors can increase or decrease the blood level of another drug which is a substrate for the target enzyme. This can lead to several pharmacokinetic drug interactions with CYP inhibitors or inducers (Bolleddula, 2021; Lang et al., 2021). Only a marginal chance of being CYP2C9 substrate is predicted for 4-hydroxyisoleucine. However, no other risk exists for a significant drug interaction. To date, there are no studies available on 4-hydroxyisoleucine–Pgp interactions. Studies on the metabolism of 4-hydroxyisoleucine are also desirable to establish a complete pharmacokinetic profile.

Some of the pharmacokinetic parameters of 4-hydroxyisoleucine (10 mg/kg) were estimated after oral administration in Wistar rats using the ultra-performance liquid chromatography–tandem mass spectrometry-based method. The V_d was estimated as $3,123.59 \pm 355.86$ mL/kg (Wadhwa et al., 2020), which is lower than that calculated by ADMETLab. The clearance of 4-hydroxyisoleucine appeared to be 204.95 ± 23.97 mL/h/kg which is less than the rate determined by ADMETLab. However, the $t_{1/2}$ 10.83 ± 1.96 h is almost equivalent to that determined by ADMETLab (Wadhwa et al., 2020). The long half-life is congruent with the low clearance rate of 4-hydroxyisoleucine.

A drug with a moderate clearance rate and a better half-life may provide dose flexibility and better patient compliance (Sleep et al., 2013). Also, 4-hydroxyisoleucine is assumed to have a desirable clearance rate and half-life. However, a risk of CNS unwanted effects can be expected as 4-hydroxyisoleucine has the probability to be BBB permeant according to ADMETLab 2.0. Contrary to this, the boiled-egg plot predicted with SwissADME showed that it is non-permeant across the BBB (Daina and Zoete, 2016).

While evaluating toxicity, not even a single risk was found for hERG blockade, H-HT, DILI, Ames toxicity, rat oral acute toxicity, skin sensitization, carcinogenicity, and ophthalmic toxicity. Cardiac

arrhythmia is one of the most frequently encountered side effects during drug discovery. This effect is related to inhibition of the hERG cardiac potassium channel. Also, this is the reason regulatory agencies such as US Food and Drug Administration (FDA) and European Medicines Agency recommend early evaluation of hERG toxicity (Garrido et al., 2020). Both ADMETLab 2.0 and Pred-hERG defined 4-hydroxyisoleucine as non-hERG blocker, being safe from cardiotoxicity. Hepatotoxicity is also highly responsible for withdrawal of an approved drug from the market. Hepatotoxicity accounts for >50% of acute liver failure cases as > 600 drugs are associated with it (Williams, 2018). Drug-induced liver injury is one of the leading factors of drug failure in clinical trials. Therefore, its assessment of drug candidates is necessary in advance as an effective strategy to decrease the drug attrition rate (He et al., 2019; Walker et al., 2020). The Ames toxicity test is an indicator for assessing mutagenic compounds' potential and use bacteria for this assay. 4-HIL is likely to be safe from these types of toxicities (Petkov et al., 2021). However, the risk of respiratory toxicity has yet to be determined.

Toxicity studies on 4-hydroxyisoleucine are scant; however, studies on standardized extracts of its natural origin exist. Singh et al. (2022) evaluated the toxicity of 4-hydroxyisoleucine using human embryonic kidney cells (HEK-293) and healthy rats. The viability of HEK-293 cells was not affected by 4-hydroxyisoleucine, and no signs of toxicity were observed in healthy rats. Another study by Swaroop et al. identified *Trigonella foenum-graecum* seed extract (source of 4-Hydroxyisoleucine) as safe and efficacious. After extensive acute oral toxicity, 28 days' sub-chronic toxicity and Ames' bacterial reverse mutation assay, no toxicity was detected (Swaroop et al., 2014).

Deshpande et al. observed IDM01 (4-hydroxyisoleucine and trigonelline-based standardized fenugreek seed extract) as safe during acute and sub-chronic preclinical toxicity in rats without mutagenicity or genotoxicity. Not a single mortality or treatment-related adverse sign was noted during acute (≤ 2000 mg/kg) and sub-chronic (90-day repeated doses of 250/500/1,000 mg per kg body weight with 28 days of recovery period). The oral median lethal dose (LD_{50}) was $>2,000$ mg/kg during the acute oral toxicity study. A non-observed adverse effect level was estimated at 500 mg/kg. The Ames test did not exhibit any mutagenicity up to 5,000 μ g/plate and did not induce any structural chromosomal aberrations at < 50 mg/culture (Deshpande et al., 2017).

A genotoxicity study of the fenugreek seed extract ($\leq 40\%$ 4-hydroxyisoleucine) was carried out *via* standard battery of tests such as reverse mutation assay, mouse lymphoma forward mutation assay, and mouse micronucleus assay. There was no genotoxicity exhibited under the tested conditions (Flammang et al., 2004).

Bioconcentration represents the accumulation of chemicals in organisms at higher concentration than its surrounding (Jambor and Weisener, 2005). Bioconcentration factors are considered to assess secondary poisoning potential and risks to human health *via* the food chain. The factor is an estimate of the residual organic chemicals used for ranking chemicals as possible hazards to the environment (Gramatica and Papa, 2003). Similarly, the toxicity of organic aromatic compounds is based on logarithm of 50% growth inhibitory concentration ($\log(IGC_{50-1})$) of *Tetrahymena pyriformis*, and fathead minnow which have extensive use in ecotoxicology and environmental safety applications (Keshavarz et al., 2021). The

environmental toxicity of 4-hydroxyisoleucine exhibited a higher LC₅₀, favoring the evidence of being non-ecotoxic. But the most acceptable evidence can be its origin as natural product, being part of nature chemistry (Avalos-Soriano et al., 2016). In the same manner, no risk of toxicity is predicted for androgen and estrogen receptors, PPAR- γ , heat shock element, mitochondrial membrane potential, and p53 (Bova et al., 2005; Ciocca et al., 2013; Khoo et al., 2014; Barna et al., 2018; Wahl and Smieško, 2018; Xi et al., 2020; Ciallella et al., 2021). Androgen-disruptors include all those chemicals which interfere with the biosynthesis, metabolism, or its physiological action. This results in an abnormal male development characteristic and poor growth and function of reproductive tract. Since androgen is the main regulator, its disruptors can harm reproductive developmental processes (Luccio-Camelo and Prins, 2011). In the same manner, estrogen disruptors can harm the female reproductive programming as the estrogen receptor is the primary determinant of female characteristics (Tsakovska et al., 2011). However, 4-hydroxyisoleucine is devoid of such toxicities and is considered non-toxic.

Mitochondrial membrane potential is generated by proton pumps and one of the essential components in the energy storage process. It is kept stable, and little variations are allowed; however, long-lasting disturbance may lead to adverse pathologies and loss of cell viability (Zorova et al., 2018). 4-Hydroxyisoleucine is free from such disruption of mitochondrial membrane potential. Similarly, genetic toxicology is important for long-term carcinogenicity during the early drug development process. This genotoxicity testing can identify the potentially hazardous drug candidates which may be knocked out at later stages of the drug development process (Custer and Sweder, 2008). However, ADMETLab recommends 4-hydroxyisoleucine as non-carcinogenic and non-genotoxic.

As all these parameters are modifiers of normal human physiology, and their disturbance leads to a red alert toxicity. Furthermore, none of the toxicophores were found in 4-hydroxyisoleucine which may have led to aquatic toxicity, skin sensitization, or created an acute toxicity response. The molecule has proven to be biodegradable and is not enlisted in the SureChEMBL database, implying medicinal chemistry friendly status (Papadatos et al., 2016; Falaguera and Mestres, 2021). The DFT study results provide certain support for the aforementioned discussion. Thus, formation of intra- and intermolecular hydrogen bonds by the 4-hydroxyisoleucine molecules, along with negative and positive MEP accumulation and significant NBO charges, would imply strong interactions with polar solvent molecules, including water, and thus acceptable aqueous solubility of the compound. Also, this would imply interaction of the 4-hydroxyisoleucine molecules with polar parts of large organic molecules. Furthermore, the shown presence of a few H-bond donors and acceptors would support the suggestion of the potential of the compound as an oral drug. Moreover, formation of several low-lying structural isomers confirms that the molecule is flexible enough to achieve a stable conformation at its target. The 4-hydroxyisoleucine stability shown by the GRP analysis supports its long half-life and stability findings achieved by other approaches.

5 Conclusion

To determine the knockout of 4-hydroxyisoleucine, a detailed ADMET profile was predicted *via* the ADMETLab 2.0 database.

This molecule has satisfactory physicochemical properties and medicinal chemistry. The pharmacokinetic profile is in the favor of oral bioavailability, efficient distribution to all tissues including the brain, and non-interruptive drug metabolism, with moderate clearance and half-life. Some of the actual PK parameters are almost comparable to *in silico* values, but further detailed studies are desirable in the lab. There is no risk of any toxicity predicted both *in silico* and inside lab, except with low probability of respiratory toxicity predicted. Even it is undesirable for this molecule to disturb some of the most important receptors/targets such as gonadal receptors, p53, heat shock factor, mitochondrial membrane potential, or PPAR- γ . The molecule is safe for the environment, with agreeable lethal concentrations for some of the most common indicators. Finally, not a single toxicophore alert was found for 4-hydroxyisoleucine. All these properties suggest the drug ability of 4-hydroxyisoleucine. DFT study results are in support of the findings obtained by other methods.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Author contributions

IA and AK have written a major part of the article, AP and KA supported in writing the draft, MD and HK supervised the overall article and finalized the submission.

Acknowledgments

This work was partially supported by Taif University Researchers Supporting Program (project number: TURSP-2020/153), Taif Univer Saudi and by the supercomputing infrastructure of the NLHPC (ECM-02).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abou El-Soud, N. H., Khalil, M. Y., Hussein, J. S., Oraby, F. S. H., Farrag, A. R. H., et al. (2007). Antidiabetic effects of fenugreek alkaloid extract in streptozotocin induced hyperglycemic rats. *J. Appl. Sci. Res.* 3 (10), 1073–1083.
- Agrawal, A., Johnson, S., Jacobsen, J., Miller, M., Chen, L. H., Pellicchia, M., et al. (2010). Chelator fragment libraries for targeting metalloproteinases. *ChemMedChem* 5 (2), 195–199. doi:10.1002/cmdc.200900516
- Artursson, P., Palm, K., and Luthman, K. (2001). Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv. Drug Deliv. Rev.* 46 (1–3), 27–43. doi:10.1016/s0169-409x(00)00128-9
- Avalos-Soriano, A., De la Cruz-Cordero, R., Rosado, J., and Garcia-Gasca, T. (2016). 4-Hydroxyisoleucine from fenugreek (*Trigonella foenum-graecum*): Effects on insulin resistance associated with obesity. *Molecules* 21 (11), 1596. doi:10.3390/molecules21111596
- Axel, D. B. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98 (7), 5648–5652. doi:10.1063/1.464913
- Banerjee, P., Eckert, A. O., Schrey, A. K., and Preissner, R. (2018). ProTox-II: A webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* 46 (1), W257–W263. doi:10.1093/nar/gky318
- Barna, J., Csermely, P., and Vellai, T. (2018). Roles of heat shock factor 1 beyond the heat shock response. *Cell. Mol. Life Sci.* 75 (16), 2897–2916. doi:10.1007/s00018-018-2836-6
- Bitew, M., Desalegn, T., Demissie, T. B., Belayneh, A., Endale, M., and Eswaramoorthy, R. (2021). Pharmacokinetics and drug-likeness of antidiabetic flavonoids: Molecular docking and DFT study. *Plos one* 16 (12), e0260853. doi:10.1371/journal.pone.0260853
- Bokulić, A., Padovan, J., Stupin-Polancec, D., and Milic, A. (2022). Isolation of MDCK cells with low expression of mdr1 gene and their use in membrane permeability screening. *Acta Pharm.* 72 (2), 275–288. doi:10.2478/acph-2022-0003
- Bolledula, J. (2021). *BPBK modeling to predict drug-drug interactions of ivosidenib as a perpetrator in cancer patients and qualification of the Simcyp platform for CYP3A4 induction*. CPT: Pharmacometrics and Systems Pharmacology.
- Bolz, S. N., Adasme, M. F., and Schroeder, M. (2021). Toward an understanding of pan-assay interference compounds and promiscuity: A structural perspective on binding modes. *J. Chem. Inf. Model.* 61 (5), 2248–2262. doi:10.1021/acs.jcim.0c01227
- Bova, M. P., Tam, D., McMahon, G., and Mattson, M. N. (2005). Troglitazone induces a rapid drop of mitochondrial membrane potential in liver HepG2 cells. *Toxicol. Lett.* 155 (1), 41–50. doi:10.1016/j.toxlet.2004.08.009
- Chakraborty, A., Pan, S., and Chattaraj, P. K. (2013). “Biological activity and toxicity: A conceptual DFT approach,” in *Applications of density functional theory to biological and bioinorganic chemistry*, 143–179.
- Ciallella, H. L., Russo, D. P., Aleksunes, L. M., Grimm, F. A., and Zhu, H. (2021). Predictive modeling of estrogen receptor agonism, antagonism, and binding activities using machine-and deep-learning approaches. *Lab. Investig.* 101 (4), 490–502. doi:10.1038/s41374-020-00477-2
- Ciocca, D. R., Arrigo, A. P., and Calderwood, S. K. (2013). Heat shock proteins and heat shock factor 1 in carcinogenesis and tumor development: An update. *Archives Toxicol.* 87 (1), 19–48. doi:10.1007/s00204-012-0918-z
- Clark, D. E. (1999). Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. *J. Pharm. Sci.* 88 (8), 807–814. doi:10.1021/js9804011
- Consonni, V., Todeschini, R., Pavan, M., and Gramatica, P. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* 42 (3), 693–705. doi:10.1021/ci0155053
- Custer, L., and Sweder, K. (2008). The role of genetic toxicology in drug discovery and optimization. *Curr. Drug Metab.* 9 (9), 978–985. doi:10.2174/138920008786485191
- Dahlin, J. L., Cuellar, M., Singh, G., Nelson, K. M., Strasser, J. M., Rappe, T., et al. (2018). ALARM NMR for HTS triage and chemical probe validation. *Curr. Protoc. Chem. Biol.* 10 (1), 91–117. doi:10.1002/cpcb.35
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7 (1), 42717. doi:10.1038/srep42717
- Daina, A., and Zoete, V. (2016). A boiled-egg to predict gastrointestinal absorption and brain penetration of small molecules. *ChemMedChem* 11 (11), 1117–1121. doi:10.1002/cmdc.201600182
- Dashti, Y., Grkovic, T., and Quinn, R. J. (2014). Predicting natural product value, an exploration of anti-TB drug space. *Nat. Product. Rep.* 31 (8), 990–998. doi:10.1039/c4np00021h
- Deshpande, P. O., Mohan, V., and Thakurdesai, P. A. (2017). Preclinical toxicological evaluation of IDM01: The botanical composition of 4-hydroxyisoleucine and trigonelline-based standardized fenugreek seed extract. *Pharmacogn. Res.* 9 (2), 138–150. doi:10.4103/0974-8490.204649
- Ding, X., Hu, X., Chen, Y., Xie, J., Ying, M., Wang, Y., et al. (2021). Differentiated Caco-2 cell models in food-intestine interaction study: Current applications and future trends. *Trends Food Sci. Technol.* 107, 455–465. doi:10.1016/j.tifs.2020.11.015
- Dionne, E., Hanson, M. L., Anderson, J. C., and Brain, R. A. (2021). Chronic toxicity of technical atrazine to the fathead minnow (*Pimephales promelas*) during a full life-cycle exposure and an evaluation of the consistency of responses. *Sci. Total Environ.* 755, 142589. doi:10.1016/j.scitotenv.2020.142589
- Falaguera, M. J., and Mestres, J. (2021). Identification of the core chemical structure in SureChEMBL patents. *J. Chem. Inf. Model.* 61 (5), 2241–2247. doi:10.1021/acs.jcim.1c00151
- Flammang, A., Cifone, M., Erexson, G., and Stankowski, L. (2004). Genotoxicity testing of a fenugreek extract. *Food Chem. Toxicol.* 42 (11), 1769–1775. doi:10.1016/j.fct.2004.07.003
- Fowden, L., Pratt, H. M., and Smith, A. (1973). 4-Hydroxyisoleucine from seed of *Trigonella foenum-graecum*. *Phytochemistry* 12 (7), 1707–1711. doi:10.1016/0031-9422(73)80391-7
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16, revision B. 01*. Wallingford, CT: Gaussian.
- Garrido, A., Lepailleur, A., Mignani, S. M., Dallemagne, P., and Rochais, C. (2020). hERG toxicity assessment: Useful guidelines for drug design. *Eur. J. Med. Chem.* 195, 112290. doi:10.1016/j.ejmech.2020.112290
- Geerlings, P., De Proft, F., and Langenaeker, W. (2003). Conceptual density functional theory. *Chem. Rev.* 103 (5), 1793–1874. doi:10.1021/cr990029p
- Gerlach, E. M., Korkmaz, M. A., Pavlinov, I., Gao, Q., and Aldrich, L. N. (2019). Systematic diversity-oriented synthesis of reduced flavones from γ -pyrones to probe biological performance diversity. *ACS Chem. Biol.* 14 (7), 1536–1545. doi:10.1021/acschembio.9b00294
- Gowtham, L., Ujjalkumar, D., Moksha, L., Roy, T. K., Halder, N., and Velpandian, T. (2022). Hydrophilic interaction LC-MS/MS method to avoid endogenous interference in the analysis of 4-hydroxy isoleucine from dietary supplementation of fenugreek. *J. Sep. Sci.* 45, 1210–1221. doi:10.1002/jssc.202100894
- Gramatica, P., and Papa, E. (2003). QSAR modeling of bioconcentration factor by theoretical molecular descriptors. *QSAR Comb. Sci.* 22 (3), 374–385. doi:10.1002/qsar.200390027
- Hann, M. M. (2011). Molecular obesity, potency and other addictions in drug discovery. *MedChemComm* 2 (5), 349–355. doi:10.1039/c1md00017a
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* 4 (1), 17. doi:10.1186/1758-2946-4-17
- He, S., Ye, T., Wang, R., Zhang, C., Zhang, X., Sun, G., et al. (2019). An *in silico* model for predicting drug-induced hepatotoxicity. *Int. J. Mol. Sci.* 20 (8), 1897. doi:10.3390/ijms20081897
- Hidalgo, I. J. (2001). Assessing the absorption of new pharmaceuticals. *Curr. Top. Med. Chem.* 1 (5), 385–401. doi:10.2174/1568026013395010
- Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., DeCrescenzo, G. A., Devraj, R. V., et al. (2008). Physicochemical drug properties associated with *in vivo* toxicological outcomes. *Bioorg. Med. Chem. Lett.* 18 (17), 4872–4875. doi:10.1016/j.bmcl.2008.07.071
- Huth, J. R., Mendoza, R., Olejniczak, E. T., Johnson, R. W., Cothron, D. A., Liu, Y., et al. (2005). Alarm NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc.* 127 (1), 217–224. doi:10.1021/ja0455547
- Huth, J. R., Song, D., Mendoza, R. R., Black-Schaefer, C. L., Mack, J. C., Dorwin, S. A., et al. (2007). Toxicological evaluation of thiol-reactive compounds identified using a la assay to detect reactive molecules by nuclear magnetic resonance. *Chem. Res. Toxicol.* 20 (12), 1752–1759. doi:10.1021/tx700319t
- Ivanenkov, Y. A., Zagibelnny, B. A., and Aladinskiy, V. A. (2019). Are we opening the door to a new era of medicinal chemistry or being collapsed to a chemical singularity? Perspective. *J. Med. Chem.* 62 (22), 10026–10043. doi:10.1021/acs.jmedchem.9b00004
- Jambor, J., and Weisener, C. (2005). The geochemistry of acid mine drainage. *Environ. Geochem.* 9, 149.
- Jean, St.D. J., Jr, and Fotsch, C. (2012). Mitigating heterocycle metabolism in drug discovery. *J. Med. Chem.* 55 (13), 6002–6020. doi:10.1021/jm300343m
- Jetté, L., Harvey, L., Eugeni, K., and Levens, N. (2000). 4-Hydroxyisoleucine: A plant-derived treatment for metabolic syndrome. *Curr. Opin. Investigational Drugs* 10(4), 353–358.
- Jia, C.-Y., Li, J. Y., Hao, G. F., and Yang, G. F. (2020). A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drug Discov. Today* 25 (1), 248–258. doi:10.1016/j.drudis.2019.10.014
- Johnson, T. W., Dress, K. R., and Edwards, M. (2009). Using the Golden Triangle to optimize clearance and oral absorption. *Bioorg. Med. Chem. Lett.* 19 (19), 5560–5564. doi:10.1016/j.bmcl.2009.08.045

- Jorio, S., Salah, M., Abou El Makarim, H., and Tabyaoui, M. (2019). Reactivity indices related to DFT theory, the electron localization function (ELF) and non-covalent interactions (NCI) calculations in the formation of the non-halogenated pyruvic esters in solution. *Mediterr. J. Chem.* 8 (6), 476–485. doi:10.13171/mjc8619072612sj
- Keshavarz, M. H., Shirazi, Z., and Sheikhabadi, P. K. (2021). Risk assessment of organic aromatic compounds to *Tetrahymena pyriformis* in environmental protection by a simple QSAR model. *Process Saf. Environ. Prot.* 150, 137–147. doi:10.1016/j.psep.2021.04.011
- Khoo, K. H., Verma, C. S., and Lane, D. P. (2014). Drugging the p53 pathway: Understanding the route to clinical efficacy. *Nat. Rev. Drug Discov.* 13 (3), 217–236. doi:10.1038/nrd4236
- Kombo, D. C., Tallapragada, K., Jain, R., Chearning, J., Mazurov, A. A., Speake, J. D., et al. (2013). 3D molecular descriptors important for clinical success. *J. Chem. Inf. Model.* 53 (2), 327–342. doi:10.1021/ci300445e
- Lagorce, D., Bouslama, L., Becot, J., Miteva, M. A., and Villoutreix, B. O. (2017). FAF-Drugs4: Free ADME-tox filtering computations for chemical biology and early stages drug discovery. *Bioinformatics* 33 (22), 3658–3660. doi:10.1093/bioinformatics/btx491
- Lang, J., Vincent, L., Chenel, M., Ogungbenro, K., and Galetin, A. (2021). Impact of hepatic CYP3A4 ontogeny functions on drug–drug interaction risk in pediatric physiologically-based pharmacokinetic/pharmacodynamic modeling: Critical literature review and ivabradine case study. *Clin. Pharmacol. Ther.* 109 (6), 1618–1630. doi:10.1002/cpt.2134
- Leahy, D. E. (1986). Intrinsic molecular volume as a measure of the cavity term in linear solvation energy relationships: Octanol–water partition coefficients and aqueous solubilities. *J. Pharm. Sci.* 75 (7), 629–636. doi:10.1002/jps.2600750702
- Lipinski, C. A. (2004). Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* 1 (4), 337–341. doi:10.1016/j.ddtec.2004.11.007
- Lovering, F., Bikker, J., and Humblet, C. (2009). Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* 52 (21), 6752–6756. doi:10.1021/jm901241e
- Luccio-Camelo, D. C., and Prins, G. S. (2011). Disruption of androgen receptor signaling in males by environmental chemicals. *J. Steroid Biochem. Mol. Biol.* 127 (1–2), 74–82. doi:10.1016/j.jsmb.2011.04.004
- McGowan, J. C., and Mellors, A. (1986). *Molecular volumes in chemistry and biology*. Chichester: E. Horwood.
- Murador, D. C., De Souza Mesquita, L. M., Neves, B. V., Braga, A. R., Martins, P. L., Zepka, L. Q., et al. (2021). Bioaccessibility and cellular uptake by caco-2 cells of carotenoids and chlorophylls from orange peels: A comparison between conventional and ionic liquid mediated extractions. *Food Chem.* 339, 127818. doi:10.1016/j.foodchem.2020.127818
- Narender, T., Puri, A., Khaliq, T., Saxena, R., Bhatia, G., Chandra, R., et al. (2006). 4-Hydroxyisoleucine an unusual amino acid as antidiabetic and antihyperglycemic agent. *Bioorg. Med. Chem. Lett.* 16 (2), 293–296. doi:10.1016/j.bmcl.2005.10.003
- Obach, R. S. (2013). Pharmacologically active drug metabolites: Impact on drug discovery and pharmacotherapy. *Pharmacol. Rev.* 65 (2), 578–640. doi:10.1124/pr.111.005439
- Papadatos, G., Davies, M., Dedman, N., Chambers, J., Gaulton, A., Siddle, J., et al. (2016). SureChEMBL: A large-scale, chemically annotated patent document database. *Nucleic Acids Res.* 44 (1), D1220–D1228. doi:10.1093/nar/gkv1253
- Petkov, P., Ivanova, H., Schultz, T., and Mekenyan, O. (2021). Criteria for assessing the reliability of toxicity predictions: I. TIMES ames mutagenicity model. *Comput. Toxicol.* 17, 100143. doi:10.1016/j.comtox.2020.100143
- Pires, D. E., Blundell, T. L., and Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58 (9), 4066–4072. doi:10.1021/acs.jmedchem.5b00104
- Reed, A. E., Curtiss, L. A., and Weinhold, F. (1988). Intermolecular interactions from a natural bond orbital, donor–acceptor viewpoint. *Chem. Rev.* 88 (6), 899–926. doi:10.1021/cr00088a005
- Ritchie, T. J., and Macdonald, S. J. (2014). How drug-like are ‘ugly’ drugs: Do drug-likeness metrics predict ADME behaviour in humans? *Drug Discov. Today* 19 (4), 489–495. doi:10.1016/j.drudis.2014.01.007
- Ritchie, T. J., and Macdonald, S. J. (2009). The impact of aromatic ring count on compound developability—are too many aromatic rings a liability in drug design? *Drug Discov. today* 14 (21–22), 1011–1020. doi:10.1016/j.drudis.2009.07.014
- Sauvaire, Y., Petit, P., Broca, C., Manteghetti, M., Baissac, Y., Fernandez-Alvarez, J., et al. (1998). 4-Hydroxyisoleucine: A novel amino acid potentiator of insulin secretion. *Diabetes* 47 (2), 206–210. doi:10.2337/diab.47.2.206
- Schaftenaar, G., and Noordik, J. H. (2000). Molden: A pre- and post-processing program for molecular and electronic structures. *J. computer-aided Mol. Des.* 14 (2), 123–134. doi:10.1023/a:1008193805436
- Singh, P., Ishteyaque, S., Prajapati, R., Yadav, K. S., Singh, R., Kumar, A., et al. (2022). Assessment of antidiabetic effect of 4-HIL in type 2 diabetic and healthy Sprague Dawley rats. *Hum. Exp. Toxicol.* 41, 096032712110618. doi:10.1177/09603271211061873
- Sleep, D., Cameron, J., and Evans, L. R. (2013). Albumin as a versatile platform for drug half-life extension. *Biochimica Biophysica Acta (BBA)—General Subj.* 1830 (12), 5526–5534. doi:10.1016/j.bbagen.2013.04.023
- Swaroop, A., Bagchi, M., Kumar, P., Preuss, H. G., Tiwari, K., Marone, P. A., et al. (2014). Safety, efficacy and toxicological evaluation of a novel, patented anti-diabetic extract of *Trigonella foenum-graecum* Seed Extract (Fenfu). *Toxicol. Mech. methods* 24 (7), 495–503. doi:10.3109/15376516.2014.943443
- Telbisz, Á., Ambrus, C., Mozner, O., Szabo, E., Varady, G., Bakos, E., et al. (2021). Interactions of potential anti-COVID-19 compounds with multispecific ABC and OATP drug transporters. *Pharmaceutics* 13 (1), 81. doi:10.3390/pharmaceutics13010081
- Terrier, J., Gaspar, F., Fontana, P., Daali, Y., Reny, J. L., Csajka, C., et al. (2021). Drug–drug interactions with direct oral anticoagulants: Practical recommendations for clinicians. *Am. J. Med.* 134 (8), 939–942. doi:10.1016/j.amjmed.2021.04.003
- Tomasi, J., Mennucci, B., and Cammi, R. (2005). Quantum mechanical continuum solvation models. *Chem. Rev.* 105 (8), 2999–3094. doi:10.1021/cr9904009
- Tsakovska, I., Pajeva, I., Alov, P., and Worth, A. (2011). Recent advances in the molecular modeling of estrogen receptor-mediated toxicity. *Adv. Protein Chem. Struct. Biol.* 85, 217–251. doi:10.1016/B978-0-12-386485-7.00006-5
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45 (12), 2615–2623. doi:10.1021/jm020017n
- Wadhwa, G., Krishna, K. V., Taliyan, R., Tandon, N., Yadav, S. S., Katiyar, C., et al. (2020). Pre-clinical pharmacokinetic and pharmacodynamic modelling study of 4-hydroxyisoleucine using validated ultra-performance liquid chromatography–tandem mass spectrometry. *RSC Adv.* 10 (10), 5525–5532. doi:10.1039/c9ra08121f
- Wahl, J., and Smieško, M. (2018). Endocrine disruption at the androgen receptor: Employing molecular dynamics and docking for improved virtual screening and toxicity prediction. *Int. J. Mol. Sci.* 19 (6), 1784. doi:10.3390/ijms19061784
- Walker, P. A., Ryder, S., Lavado, A., Dilworth, C., and Riley, R. J. (2020). The evolution of strategies to minimise the risk of human drug-induced liver injury (DILI) in drug discovery and development. *Archives Toxicol.* 94 (8), 2559–2585. doi:10.1007/s00204-020-02763-w
- Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* 14 (7), 475–486. doi:10.1038/nrd4609
- Wei, W., Cherukupalli, S., Jing, L., Liu, X., Zhan, P., et al. (2020). *Fsp3: A new parameter for drug-likeness*. Amsterdam: Drug Discovery Today.
- Weigend, F. (2006). Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* 8 (9), 1057–1065. doi:10.1039/b515623h
- Weigend, F., and Ahlrichs, R. (2005). Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* 7 (18), 3297–3305. doi:10.1039/b508541a
- Williams, D. P. (2018). Application of hepatocyte-like cells to enhance hepatic safety risk assessment in drug discovery. *Philosophical Trans. R. Soc. B Biol. Sci.* 373 (1750), 20170228. doi:10.1098/rstb.2017.0228
- Xi, Y., Zhang, Y., Zhu, S., Luo, Y., Xu, P., and Huang, Z. (2020). PPAR-mediated toxicology and applied pharmacology. *Cells* 9 (2), 352. doi:10.3390/cells9020352
- Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., et al. (2021). ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.*
- Yamazaki, S., Costales, C., Lazzaro, S., Eatemadpour, S., Kimoto, E., and Varma, M. V. (2019). Physiologically-based pharmacokinetic modeling approach to predict rifampin-mediated intestinal p-glycoprotein induction. *CPT pharmacometrics Syst. Pharmacol.* 8 (9), 634–642. doi:10.1002/psp4.12458
- Yan, A., and Gasteiger, J. (2003). Prediction of aqueous solubility of organic compounds based on a 3D structure representation. *J. Chem. Inf. Comput. Sci.* 43 (2), 429–434. doi:10.1021/ci025590u
- Yang, H., Lou, C., Sun, L., Li, J., Cai, Y., Wang, Z., et al. (2019). admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* 35 (6), 1067–1069. doi:10.1093/bioinformatics/bty707
- Yang, Z.-Y., Lu, A. P., Hou, T. J., and Cao, D. S. (2021). Scopy: An integrated negative design python library for desirable HTS/VS database design. *Briefings Bioinforma.* 22 (3), bbab194. doi:10.1093/bib/bba194
- Zega, A. (2017). NMR methods for identification of false positives in biochemical screens: Miniperspective. *J. Med. Chem.* 60 (23), 9437–9447. doi:10.1021/acs.jmedchem.6b01520
- Zhao, Y. H., Abraham, M. H., Hersey, A., and Luscombe, C. N. (2003). Quantitative relationship between rat intestinal absorption and Abraham descriptors. *Eur. J. Med. Chem.* 38 (11–12), 939–947. doi:10.1016/j.ejmech.2003.07.005
- Zhao, Y. H., Abraham, M. H., and Zissimos, A. M. (2003). Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. *J. Org. Chem.* 68 (19), 7368–7373. doi:10.1021/jo034808o
- Zorova, L. D., Popkov, V. A., Plotnikov, E. Y., Silachev, D. N., Pevzner, I. B., Jankauskas, S. S., et al. (2018). Mitochondrial membrane potential. *Anal. Biochem.* 552, 50–59. doi:10.1016/j.ab.2017.07.009



OPEN ACCESS

EDITED BY

Riccardo Nifosi,
National Research Council (CNR), Italy

REVIEWED BY

Yuqing Li,
Soochow University, China
Muslum Ilgu,
Iowa State University, United States

*CORRESPONDENCE

Ting Wang,
✉ tingwang@ncu.edu.cn
Kun-He Zhang,
✉ khzhang@ncu.edu.cn

RECEIVED 14 January 2023

ACCEPTED 26 April 2023

PUBLISHED 09 May 2023

CITATION

Zhang H-L, Lv C, Li Z-H, Jiang S, Cai D,
Liu S-S, Wang T and Zhang K-H (2023),
Analysis of aptamer-target binding and
molecular mechanisms by
thermofluorimetric analysis and
molecular dynamics simulation.
Front. Chem. 11:1144347.
doi: 10.3389/fchem.2023.1144347

COPYRIGHT

© 2023 Zhang, Lv, Li, Jiang, Cai, Liu, Wang
and Zhang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Analysis of aptamer-target binding and molecular mechanisms by thermofluorimetric analysis and molecular dynamics simulation

Hong-Li Zhang, Cong Lv, Zi-Hua Li, Song Jiang, Dan Cai,
Shao-Song Liu, Ting Wang* and Kun-He Zhang*

Department of Gastroenterology, The First Affiliated Hospital of Nanchang University, Jiangxi Institute of Gastroenterology and Hepatology, Jiangxi Clinical Research Center for Gastroenterology, Nanchang, China

Introduction: Aptamers are valuable for bioassays, but aptamer-target binding is susceptible to reaction conditions. In this study, we combined thermofluorimetric analysis (TFA) and molecular dynamics (MD) simulations to optimize aptamer-target binding, explore underlying mechanisms and select preferred aptamer.

Methods: Alpha-fetoprotein (AFP) aptamer AP273 (as the model) was incubated with AFP under various experimental conditions, and melting curves were measured in a real-time PCR system to select the optimal binding conditions. The intermolecular interactions of AP273-AFP were analysed by MD simulations with these conditions to reveal the underlying mechanisms. A comparative study between AP273 and control aptamer AP-L3-4 was performed to validate the value of combined TFA and MD simulation in selecting preferred aptamers.

Results: The optimal aptamer concentration and buffer system were easily determined from the dF/dT peak characteristics and the melting temperature (T_m) values on the melting curves of related TFA experiments, respectively. A high T_m value was found in TFA experiments performed in buffer systems with low metal ion strength. The molecular docking and MD simulation analyses revealed the underlying mechanisms of the TFA results, i.e., the binding force and stability of AP273 to AFP were affected by the number of binding sites, frequency and distance of hydrogen bonds, and binding free energies; these factors varied in different buffer and metal ion conditions. The comparative study showed that AP273 was superior to the homologous aptamer AP-L3-4.

Conclusion: Combining TFA and MD simulation is efficient for optimizing the reaction conditions, exploring underlying mechanisms, and selecting aptamers in aptamer-target bioassays.

KEYWORDS

aptamers, thermofluorimetric analysis, molecular dynamics simulation, reaction condition optimization, aptamer selection

1 Introduction

Nucleic acid aptamers, which are artificial ligands of biomolecules, are selected from synthetic random single-stranded oligonucleotide libraries through systematic evolution of ligands by exponential enrichment (SELEX) (Ellington and Szostak, 1990; Tuerk and Gold, 1990; Wu et al., 2021). Aptamers have similar binding properties to antibodies, but they are more stable and easier to prepare and modify than antibodies. Therefore, aptamers are an ideal tool for biomedical molecular detection and recognition (Zhu et al., 2015; Emrani et al., 2016; Li et al., 2019). However, the binding properties of aptamers to their targets are closely related to their structures (Davydova et al., 2020), and the structures of aptamers are sensitive to the conditions of the reaction system. Hence, the analysis of the three-dimensional structures of aptamers and their influencing factors is important for optimizing their binding properties and for selecting preferred aptamers.

Traditionally, the structural analysis of aptamers mainly depends on nuclear magnetic resonance (NMR) (Longhini et al., 2016; Cai et al., 2018) and X-ray crystallography (Ruigrok et al., 2012; Christina Schmidt et al., 2020), but universal application of these methods is not possible due to expensive instruments and complicated procedures. With the rapid development of artificial intelligence technologies, molecular structure prediction based on bioinformatics and molecular dynamics (MD) simulations is increasingly recognized and applied (Douaki et al., 2022; Poolsup et al., 2023), because structural changes in aptamers (before and after binding to their targets) can be rapidly analyzed without damaging the aptamers, the effects of metal ions in the reaction system on the interaction of aptamers and targets can be analyzed, and information on the stability, binding energy, affinity, and molecular mechanism of aptamer-target binding is provided (Hayashi et al., 2014; Hilder and Hodgkiss, 2017; Cui et al., 2020). Through MD simulations, La Penna et al. (La Penna and Chelli, 2018) found that Mg^{2+} and Na^+ could attach to the aptamer surface to enhance the interaction between the amino acid residues of protein targets and the nucleotides of aptamers; thus, Mg^{2+} and Na^+ facilitated the stability of the binding compounds. Using MD simulations, Vu et al. (Vu et al., 2018) identified the binding site of platelet-derived growth factor-B (PDGF-B) with its specific aptamer, and their binding force was mainly the electrostatic force between the positively charged amino acid of the target and the negatively charged phosphate backbone of the aptamer.

In general, it is necessary to select excellent aptamers from the dozens of aptamers generated by SELEX for subsequent applications (Darmostuk et al., 2015; Vu et al., 2018). Specificity and affinity analyses are traditional methods for selecting aptamers, but these methods are labor-intensive and often inflexible when used in complex assay settings. Thermofluorimetric analysis (TFA) is a method that can be used to evaluate aptamer-target binding and distinguish bound and free aptamers in a convenient and rapid manner. This method is based on the melting curve analysis in a polymerase chain reaction (PCR) system (Hu and Easley, 2017), and it can be used to optimize experimental conditions and select the optimal aptamer in complex reaction systems. TFA has been applied to quantify serum platelet-derived growth factor (PDGF), insulin, and prothrombin based on aptamers (Hu et al., 2015; Kim et al., 2015).

Binding between a target and an aptamer is primarily governed by the mutual adaptation of their conformations (the so-called “lock and key” relationship) (Hayashi et al., 2014), and intermolecular non-covalent bonds such as hydrogen bonds, electrostatic interaction, and van der Waals forces act as auxiliary factors in the binding between aptamer and target (Shigdar, 2019; Morozov et al., 2021), while there are also reports of aptamer and target being bound via covalent bonds (Tivon et al., 2021; Tabuchi et al., 2022). Metal ions can impact the tertiary structure of aptamers and their ability to bind to targets through their charge distribution and potential energy (Moccia et al., 2019). Therefore, it is necessary to optimize the metal ion concentrations in binding buffers to identify aptamers with a wider range of adaptability and high specificity. Conventional methods for optimization are generally based on multifactor, multilevel experiments, which are time-consuming and sometimes yield unsatisfactory results (Cheng et al., 2003; Zhou et al., 2015). Thus, a simple, fast, and effective method is needed to optimize the metal ions in the buffers.

The status of aptamer-target interactions can be simply and rapidly observed through TFA, and aptamer structures and their interactions with targets can be visually analyzed through MD simulation. Thus, we speculated that by combining TFA and MD simulations, the effects of buffers and metal ions on aptamer structures and their interactions with targets can be easily analyzed at both experimental and theoretical levels, which can help to select optimal aptamers and optimize experimental conditions. In the present study, the reported aptamer AP273 against alpha-fetoprotein (AFP) (Dong et al., 2015) and our screened aptamer AP-L3-4 against AFP-L3 were used as models for validation, by which we attempted to provide a simple and feasible method for prioritizing aptamers and optimizing experimental conditions.

2 Materials and methods

2.1 Determination of the optimal aptamer concentration by TFA

Aptamers AP273 (5'-GTG ACG CTC CTA ACG CTG ACT CAG GTG CAG TTC TCG ACT CGG TCT TGA TGT GGG TCC TGT CCG TCC GAA CCA ATC-3') and AP-L3-4 (5'-ACC GAC CGT GCT GGA CTC TGT CGA AAG GAT ACT GAG TAT TGA GGG GCG TCA GGT GGA AGA GTA TGA GCG AGC GTT GCG-3') were synthesized (Sangon Biotech (Shanghai) Co., Ltd.) and dissolved in ddH₂O to prepare a 100 nM storage solution. AP273 is an ssDNA aptamer against AFP selected using CE-SELEX by Wu's team (Dong et al., 2015); AP-L3-4 is an ssDNA aptamer against AFP-L3 selected using SELEX in our previous work.

The optimal concentration of aptamer AP273 for binding with AFP was determined by analysing the characteristics of the melting curves generated by different concentrations of aptamer reacting with a fixed concentration of AFP. Gradient concentrations of aptamer working solutions (1.25, 2.5, 5.0, 10, 20, 40, and 80 nM) were prepared by adding the typical buffer used in aptamer screening (HEPES-Na 20 mmol/L, NaCl 120 mmol/L, KCl 4 mmol/L, 2 mmol/L $MgCl_2$, 1 mmol/L $CaCl_2$, pH 7.35). Human recombinant AFP (Nearshore Protein Technology Ltd., Shanghai,

TABLE 1 Concentrations of metal ions in buffers based on orthogonal design.

Buffer	Metal ion concentration (mM)			
	Mg ²⁺	Na ⁺	K ⁺	Ca ²⁺
1	1	120	5	2.5
2	5	100	4	2.5
3	2	140	2	2.5
4	5	140	5	1
5	2	100	5	2
6	1	100	2	1
7	1	140	4	2
8	5	120	2	2
9	2	120	4	1

China) was dissolved in ddH₂O to prepare a solution of 1.45 nM (100 ng/mL). EvaGreen (20×) dye (Biotium, United States of America) was diluted to 8× working solution using ddH₂O (EvaGreen is a nucleic acid fluorescent dye commonly used in real-time PCR (Shoute and Loppnow, 2018)).

The aptamer working solution was denatured in a metal bath at 95°C for 3 min and then immediately placed in an ice bath for 3 min. Twenty microliters of the denatured solution was added to PCR tubes, followed by the addition of 5 µL of 8× EvaGreen working solution and 5 µL of human recombinant AFP solution. This mixture was incubated for 30 min at room temperature and then placed in a StepOnePlus™ real-time PCR system (Applied Biosystems Inc., United States of America) to measure melting curves (from 4°C to 80°C, with a 0.5°C rise every 10 s and fluorescence detection). These experiments were performed in three duplicate tubes. An equal volume of the buffer solution instead of the aptamer working solution was used as a blank control.

The melt region derivative data (dF/dT) of each tube was exported from the PCR system. The average dF/dT value of the three replicate tubes at each temperature point was calculated and then corrected by subtracting the average dF/dT value of the three blank control tubes. The dF/dT values were normalized (0–100). The melting curves were plotted with the temperature as the X-axis and the corrected or normalized dF/dT values as the Y-axis, and the optimal concentration of the aptamer was determined based on the melting temperature (T_m), peak height and peak area.

2.2 Determination of the optimal buffer system by TFA

To determine the optimal buffer for the experiment, the interaction of aptamer AP273 (at the optimal concentration determined above) and AFP (1.45 nM) was performed in three buffers (20 mM HEPES, 10 mM PBS and 20 mM Tris-HCl, pH = 7.35–7.45) with constant metal ion concentrations (Na⁺ 140 mM, Mg²⁺ 2 mM, K⁺ 4 mM, and Ca²⁺ 1 mM), and then the melting curves were measured via TFA. The experimental procedure and interpretation of the results were the same as described above.

2.3 Determination of optimal metal ion strength by TFA

The optimal metal ion strength for the experiment was also determined via TFA. The concentration points of metal ions were set according to the results reported in the literature as follows (Duan et al., 2012; Kang et al., 2015; Wu et al., 2017; Chen et al., 2020a): 1, 2 and 5 mM for Mg²⁺; 100, 120 and 140 mM for Na⁺; 2, 4 and 5 mM for K⁺; and 1, 2 and 2.5 mM for Ca²⁺. A 4-factor, 3-level orthogonal test (with a fixed random seed number of 300) was designed to determine the optimal concentration of metal ions. Based on the orthogonal design, nine buffers with different concentrations of metal ions were prepared using the optimal buffer system determined in the previous step (Table 1). The concentrations of aptamers and AFP and the experimental procedures were the same as described above. In subsequent MD analysis, the buffers with the optimal metal ion combination were selected according to the T_m value, peak height and peak area.

2.4 Structure prediction, docking and MD simulations of the aptamers to AFP

Structure prediction and docking of two aptamers and AFP were performed by the following software: 1) The secondary structure of the aptamers was predicted using the online software Mfold (<http://www.unafold.org/hybrid2.php>) (Zuker, 2003). Aptamer structures with the minimum energy (the lowest ΔG value) were selected. The output formats (dot-bracket notation) were used for the construction of the aptamer 3D structures (Garcia-Recio et al., 2016; Heiat et al., 2016; Subki et al., 2020). 2) The tertiary structure of aptamers was predicted by using online RNA Composer software on the basis of the secondary structure of aptamers (<https://rnacomposer.cs.put.poznan.pl>) (Popenda et al., 2012; Antczak et al., 2016). 3) The nucleotide change for the aptamer was performed by using DS (Biovia Discovery Studio) software (the uracil was changed to thymine, and the ribose backbone was changed to deoxyribose) (Chang et al., 2020). 4) All-atom (AA) energy minimization simulations of the aptamer were performed in a vacuum by using software NAnoscale Molecular Dynamics (NAMD) (Phillips et al., 2005) with the Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). 5) The 3D model of human AFP was obtained from either the AlphaFoldDB website (<https://alphafold.ebi.ac.uk/>) using code number AF-J3KMX3-F or the PDB-KB website (<https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P02771>) with the PDB code 7YIM (UniProt et al., 2023). 6) Molecular docking of the 3D structure of the aptamer and AFP was performed online using ZDOCK 3.0.2 (<https://zdock.umassmed.edu/>) (Mintseris et al., 2007; Pierce et al., 2011; Pierce et al., 2014). 7) The model with the highest ZDOCK score (Wang et al., 2019) was selected as the initial structure for MD simulations, using Amber 18 software (Lee et al., 2020), in which the aptamer was assigned OL15 nucleic acid force fields (Zgarbova et al., 2015) and the target molecule was assigned ff14SB protein force fields (Tian et al., 2020). The reaction system solvent cassette (Steinberg et al., 2019) was set up according to the metal concentrations screened by the TFA

experiment. 8) Docking, MD simulation results, hydrogen bonding, and binding sites were visualized using PyMOL software (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) (Lill and Danielson, 2011). The optimal buffer system was determined based on comprehensive analyses of the simulations.

2.5 Binding characteristics of aptamers to AFP under different ionic conditions

With optimal buffering conditions, TFA was performed for both fixed and gradient concentrations of aptamers to verify their binding characteristics to AFP. The experimental procedure and data processing were the same as above, and the binding characteristics of aptamers to AFP under various buffer systems were analyzed according to the T_m value, peak height and peak area.

2.6 Comparisons between different aptamers

In order to verify whether the combination of TFA and MD simulations can distinguish different aptamers and identify the optimal aptamer, the same experiments were performed with the aptamer AP-L3-4. Differences in the binding of aptamers

AP273 and AP-L3-4 to AFP were compared, including the binding site, the distance and frequency of hydrogen bonds, binding energy, and binding characteristics for the gradient concentrations of AFP.

3 Results

3.1 Optimal concentration of aptamers determined by TFA

The melt curves of normalized dF/dT value and the heatmaps of peak area for gradient concentrations of aptamers AP273 and AP-L3-4 binding to AFP are shown in Figure 1. The free aptamer peak (peak 1) of AP273 appeared at 2.5 nM and then rose with increasing concentration and peaked at 10 nM, while the bound aptamer peak (peak 2) almost peaked at 2.5 nM and then gradually decreased from 10 nM (Figure 1A). A similar trend was found in aptamer AP-L3-4, but the free aptamer peak topped out and the bound aptamer peak started to decrease earlier (both at 5 nM) (Figure 1B). The heatmaps of the area under the peaks showed that the bound aptamer peak of AP273 was stronger than that of AP-L3-4 at the same aptamer concentration, while the free aptamer peak was the opposite (Figures 1C, D). These findings suggest that TFA can exhibit concentration-dependent changes in aptamer-target interactions and reflect differences between aptamers.

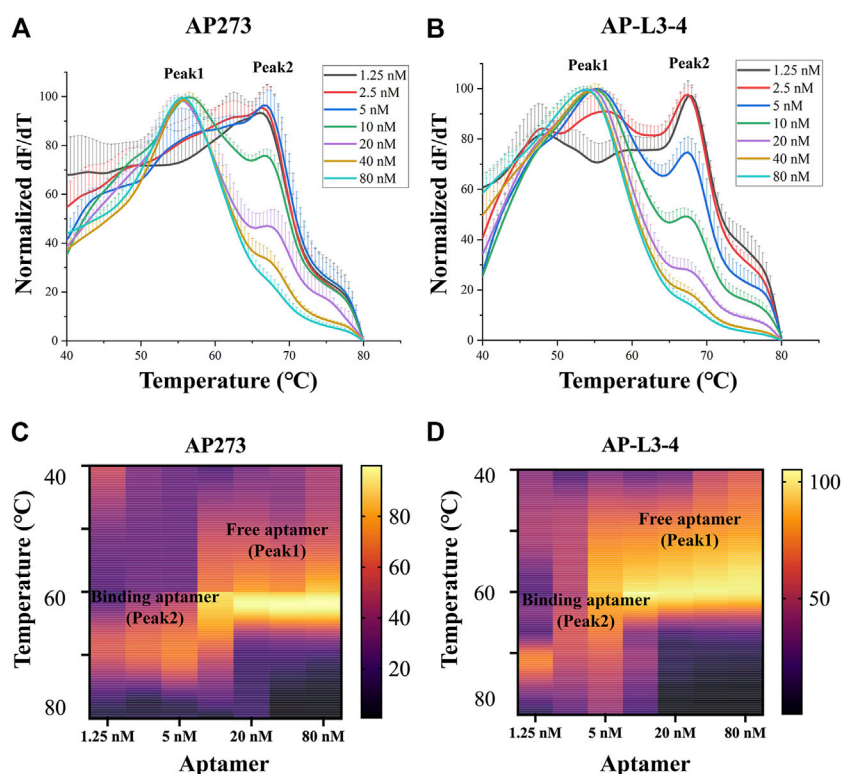


FIGURE 1

Thermofluorimetric analyses of the gradient concentrations of aptamers binding to AFP. (A) Melting curves of AP273 at 1.25–80 nM. (B) Melting curves of AP-L3-4 at 1.25–80 nM. (C) Heatmap of the area under the peaks of AP273 at 1.25–80 nM. (D) Heatmap of the area under the peaks of AP-L3-4 at 1.25–80 nM.

The results of refinement experiments for both aptamers at concentrations of 1–5 nM are shown in Figure 2. For AP273, the bound aptamer peak (peak 2) appeared at the lowest concentration (1 nM) and was maintained to the maximum concentration (5 nM), and its free aptamer peak (peak 1) appeared at 3 nM and continuously rose (Figure 2A). A similar trend was observed in AP-L3-4, although the free aptamer peak was more pronounced (Figure 2B). The heatmaps of the area under peak showed that AP273 had stronger bound aptamer peaks and weaker free aptamer peaks compared with AP-L3-4 (Figures 2C, D). Considering the need to detect different concentrations of AFP, we selected 4 nM as the optimal concentration for subsequent experiments.

3.2 Optimal buffer system and ion concentration for aptamer-target interactions determined by TFA

The melting curves of AP273 (4 nM) binding to AFP (1.45 nM) in three buffer systems (HEPES, PBS, Tris-HCl) showed different T_m values (Figure 3A), of which the HEPES buffer system had the largest T_m value and therefore was selected as the optimal buffer system for subsequent experiments. The T_m values of AP273 binding to AFP also varied in HEPES buffer solutions (buffer 1–9) with various metal ion concentrations

(according to the orthogonal design) (Figure 3B), with the lowest T_m value in buffer 4 and the highest T_m value in buffer 6. The T_m value was higher in the buffer solution with the lowest ionic strength (buffer 6) than in the buffer solution with the highest ionic strength (buffer 4) (Figure 3C). Higher T_m values indicate more stable binding between aptamers and targets, and thus buffers 5, 6, and 9, which had higher T_m values were selected as buffer solutions for the next molecular dynamics analysis, and buffer 7 with the same concentration of Mg^{2+} ions, was selected as the control buffer solution.

3.3 Secondary structure prediction of AP273 and its free energy

Prediction of the secondary structure is required before the tertiary structure of an aptamer can be determined. In the prediction of secondary structures of aptamer AP273, it was found that the amount and minimum free energy of secondary structure varied with the concentrations of Na^+ and Mg^{2+} ions (Figure 4). Increased concentrations of Mg^{2+} ions led to fewer predicted secondary structures. The concentration of Na^+ ions did not affect the amount of secondary structure, but it impacted the minimum free energy of the secondary structure. Under the same concentration of Mg^{2+} ions, higher

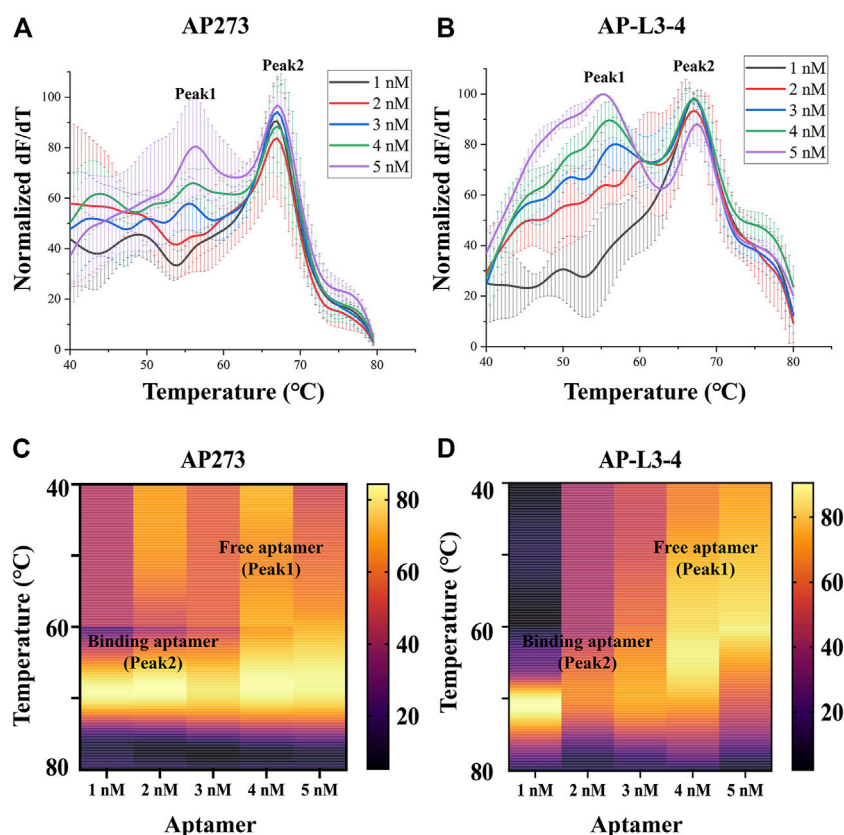


FIGURE 2

Thermofluorimetric analyses of the refined gradient concentrations of aptamers binding to AFP. (A) Melting curves of AP273 at 1–5 nM. (B) Melting curves of AP-L3-4 at 1–5 nM. (C) Heatmap of the area under the peaks of AP273 at 1–5 nM. (D) Heatmap of the area under the peaks of AP-L3-4 at 1–5 nM.

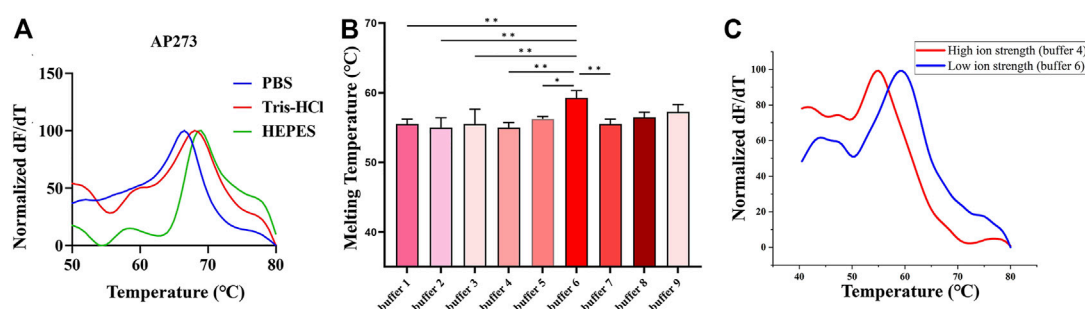


FIGURE 3

Thermofluorimetric analyses of AP273 binding to AFP in different buffer and ionic strength conditions. (A) Melting curves in different buffer systems. (B) The melting temperatures in HEPES buffer system with various metal ion concentrations. * $p < 0.05$, ** $p < 0.01$. (C) Melting curves of high and low ionic strength buffer solutions.

concentrations of Na^+ ions led to lower free energy values. The structure with the lowest value of ΔG (the most stable secondary structure) was selected for subsequent prediction of the tertiary structure.

3.4 Molecular dynamics simulation of the interaction between AP273 and AFP

The docking models of AP273 and AFP were then subjected to MD simulations in different buffer solutions (Figure 5). The results of root mean square deviation (RMSD) (Figure 5A) and root mean square fluctuation (RMSF) (Figure 5B) showed that the stability of the complex of AP273 and AFP was the highest in buffer 9 and higher than that in buffer 6. The comparisons of AP273, AFP and their complex between buffer 6 and buffer 9 (Figures 5C, D) exhibited that AFP had the highest stability and AP273 had the lowest stability, and that the AP273-AFP complex was more stable in

buffer 9 than in buffer 6. These results suggest that the metal ions in buffer systems can affect the stability of aptamer-target complexes.

The free energy of AP273 binding to AFP in various buffer solutions was calculated by the MM/GBSA method (Rastelli et al., 2010; Hou et al., 2011; Genheden and Ryde, 2015; Chen et al., 2020b) and is shown in Table 2. As shown in Table 2, buffer 6 exhibited the largest binding energy but not the best stability (Figure 5A), which is not consistent with the results of TFA experiments, indicating that the binding force and stability of aptamer-target are influenced by other complex environmental factors, such as temperature and pH, etc.

The visualization of the MD simulation results provided qualitative binding information between AP273 and AFP in different buffer solutions (Figure 6), which indicated that the binding sites and hydrogen bonds were different under various buffer solutions and that metal ions were involved in the binding.

The hydrogen bonds in AP273 binding to AFP under four buffer solutions were further quantified (Figures 7A, B). The results showed that the frequency of hydrogen bonds was the lowest in buffer 9 and

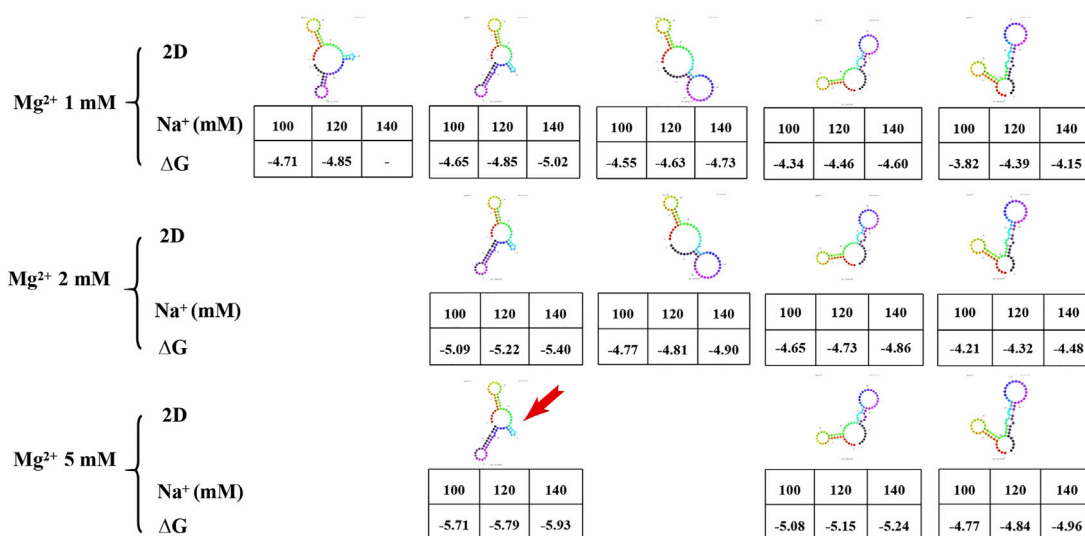


FIGURE 4

The secondary structure and free energy (ΔG) of AP273 predicted at different ion concentrations. The colour figures: secondary structures of AP273. Table: Na^+ concentration in the first row, ΔG value in the second row. Red arrow: the secondary structure selected for the prediction of tertiary structure.

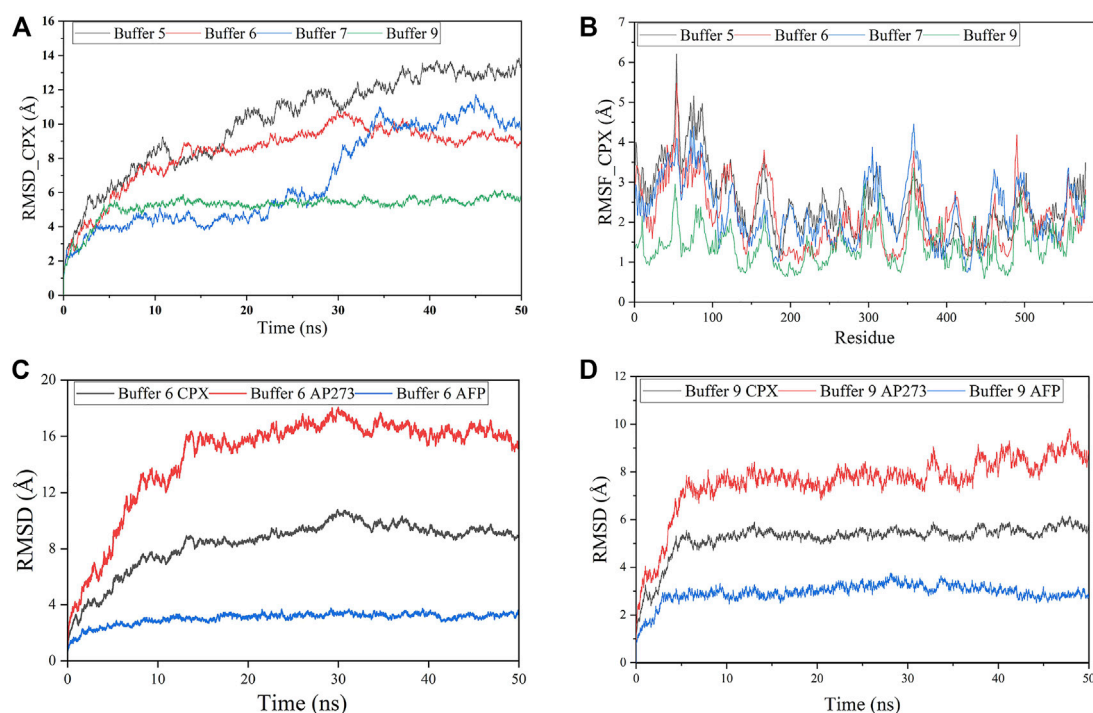


FIGURE 5

Molecular dynamics simulations of aptamer AP273 binding to AFP. (A) The root means square deviation (RMSD) of the complex of aptamer AP273 and AFP (CPX) in various buffer solutions. (B) The root mean square fluctuation (RMSF) of residues during MD simulations in various buffer solutions. (C) and (D) The RMSDs of CPX, AFP, and AP273 in buffer 6 and buffer 9.

TABLE 2 Binding free energies of AP273 and AFP predicted by MM/GBSA (kcal/mol).

	Buffer 5	Buffer 6	Buffer 7	Buffer 9
ΔE_{vdw}	-243.4	-213.4	-205.17	-223.32
ΔE_{elec}	8,949.21	8,634.37	9,121.31	10,193.54
ΔG_{GB}	-8,763.83	-8,517.34	-8,989.8	-10,041.84
ΔG_{SA}	-30.64	-27.03	-26.71	-28.68
ΔG_{bind}	-88.67	-123.4	-100.37	-100.31

AFP, alpha-fetoprotein; MM/GBSA, molecular mechanics generalized born surface area; ΔE_{vdw} , van der Waals energy; ΔE_{elec} , electrostatic energy; ΔG_{GB} , electrostatic contribution to solvation; ΔG_{SA} , nonpolar contribution to solvation; ΔG_{bind} , binding free energy.

the highest in buffer 6, and that the distance of hydrogen bonds was similar in the four buffer solutions, suggesting that the hydrogen bond frequency rather than hydrogen bond distance plays an important role in the binding between AP273 and AFP. TFA experiments were conducted under buffer 6 and buffer 9, and results showed that buffer 6 was more favorable than buffer 9 for AP273 to bind to AFP in terms of bound aptamer ratio and linear correlation (Figure 7C).

3.5 Comparison between aptamers AP273 and AP-L3-4

Comparative studies were performed between aptamers AP273 and AP-L3-4 to further verify whether the combination of TFA and MD simulation could select the aptamer with better binding properties.

Compared with AP-L3-4, AP273 binding to AFP in MD simulations had more binding sites (Figures 8A, B), a shorter hydrogen bond distance, a higher hydrogen bond frequency, and less free energy (Figure 8C). TFA results showed better linear correlation of AP273 than AP-L3-4 for detecting gradient concentrations for AFP (Figure 8D). These findings demonstrate that the binding property of AP273 to AFP is superior to that of AP-L3-4, suggesting the theoretical and experimental feasibility of combining TFA and MD simulation for preferential aptamer selection.

4 Discussion

TFA analyzes the binding of aptamers to their target molecules using melting curves. The free aptamer shows a smaller T_m value than

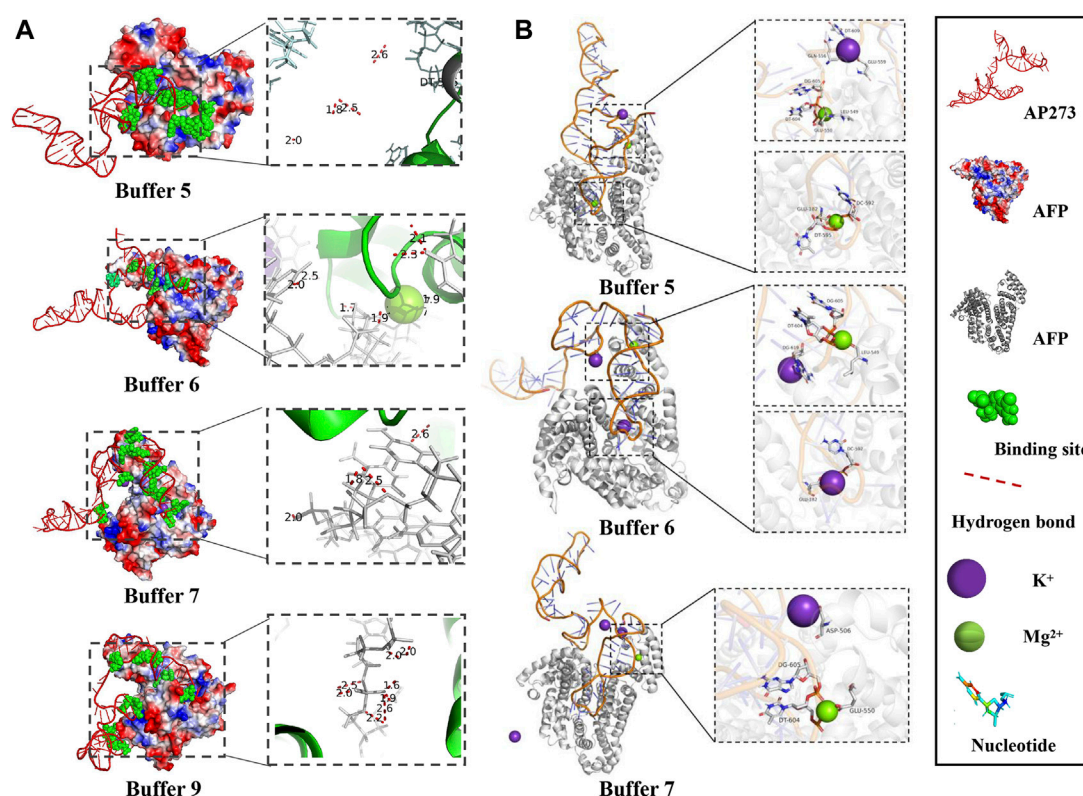


FIGURE 6

Visualization of MD simulation results in various buffer solutions. (A). Binding sites and the hydrogen bonds between AP273 and AFP in four buffer solutions. (B). The interaction of K^+ and Mg^{2+} ions with the complex of aptamer AP273 and AFP in MD simulations.

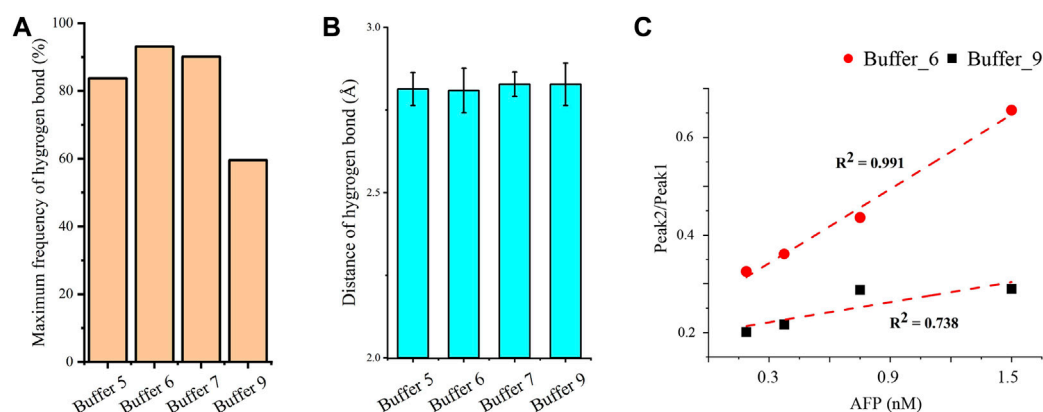


FIGURE 7

Quantitative analyses of hydrogen bonds and linear correlations in the interaction between AP273 and AFP. (A) and (B) The maximum frequency and distance of hydrogen bonds in the four buffer solutions. (C) The linear correlations between AP273 and AFP in buffer 6 and buffer 9.

that of the bound aptamer, so two peaks, peak 1 (the free aptamer peak) and peak 2 (the bound aptamer peak), can be observed on the melting curve of TFA (Kim et al., 2015; Mahmoud et al., 2019). In the present study, distinct peak 1 and peak 2 appeared on the melting curves after aptamers AP273 or AP-L3-4 were reacted with AFP, and the T_m value,

peak height and peak area were different under various experimental conditions. Using these indicators, the optimal experimental system and the best aptamer can be determined, making TFA useful to optimize the aptamer-target reaction system and to select preferred aptamers. The T_m value of a bound aptamer peak is positively correlated with the

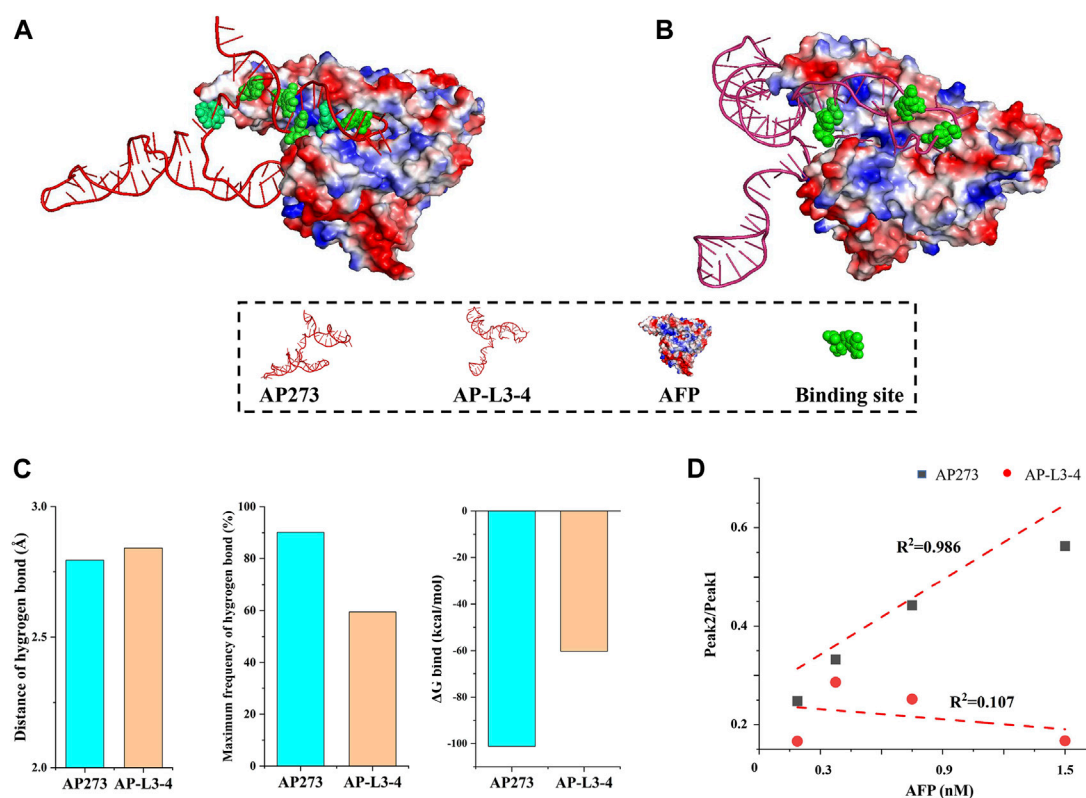


FIGURE 8

Comparison of aptamers AP273 and AP-L3-4 binding to AFP in molecular dynamics simulation and thermofluorimetric analysis. (A) and (B) Docking models of AP273 and AP-L3-4 with AFP and their binding sites (green); (C) Distances and frequencies of hydrogen bonds and binding energies of AP273 and AP-L3-4 bound to AFP; (D) The linear correlations of AP273 and AP-L3-4 bound to AFP in thermofluorimetric analysis.

stability of aptamer-target binding. In the present study, the T_m values for peak 2 differed across the three buffer systems, with the highest value in the HEPES buffer system, suggesting that the binding of AP273 to AFP is more stable in the HEPES buffer system compared to the other two buffer systems. This may be because HEPES is not prone to form complexes with metal ions in the buffer system and thus exhibits a better buffering effect (Ruzza et al., 2021). Using HEPES as the buffer salt, nine buffer solutions with various metal ion concentrations determined by orthogonal design were formulated for TFA experiments to analyze the effects of metal ions on the binding of AP273 to AFP. The T_m values of peak 2 were maximum in buffer 6 (with the lowest ionic strength) and minimum in buffer 4 (with the highest ionic strength), indicating that the high ionic strength is unfavorable for the binding of AP273 to AFP molecules. This is consistent with previous reports that a buffer solution with relatively low ionic strength is more appropriate for TFA experiments (Hayashi et al., 2014) and that immobilization of aptamers under low ionic strength conditions rather than conventional high ionic strength buffers can greatly improve the performance of the E-AB sensor (Liu et al., 2021).

The metal ions of the buffer system are important in aptamer-target binding. In the secondary structure prediction, we found that the number of secondary structures of AP273 varied with the Mg^{2+} and Na^+ concentrations and that the Mg^{2+} concentration was proportional

to the free energy (absolute value of ΔG) of the aptamer secondary structure. In MD simulations, we found that Mg^{2+} and K^+ contributed to hydrogen bonds at the binding site of the aptamer to the target molecule (Figure 6B). Metal ions are mainly embedded in the helical grooves of aptamer tertiary structures and affect the binding properties of aptamers to the target molecules in different ways (McCluskey et al., 2019). Metal ions can neutralize the negative charge of the phosphate groups of aptamers to affect the amount of charge in aptamer structures, hence affecting the structure of aptamers; in this context, Mg^{2+} affects the flexibility of aptamers and the stability of aptamer-target compounds (Zhang and Yadavalli, 2010). The aptamer structure is unstable in the binding buffer without Mg^{2+} (Zhao et al., 2020), and Mg^{2+} is important for stabilizing the two- and three-dimensional structures of aptamers (Huang et al., 2013). In addition to Mg^{2+} , the Na^+ , K^+ and pH value of a buffer system also have an effect on the binding of aptamers to targets, and the optimization of them can significantly improve the efficiency of aptamer sensors in detecting targets (Hianik et al., 2007; Belleperche and DeRosa, 2018) and significantly increase the sensitivity of aptamer sensors (Yang et al., 2022). The binding of thrombin to its aptamer also depends on pH and electrolytes (Hianik et al., 2007). The pH value may affect the ssDNA conformational changes and the electron transfer between the target protein and the aptamer, which is related to the maintenance of the three-dimensional conformation of the aptamer; and an increase in Na^+

concentration leads to a weaker binding of thrombin to the aptamer, possibly due to the shielding effect of Na^+ ions on the target. In addition, many studies have confirmed that K^+ ions can stabilize the structure of the aptamer, particularly for G-quadruplexes (Santos et al., 2022).

The stability and affinity of aptamer-target complexes involve intermolecular covalent bonds, hydrogen bonds, electrostatic interactions and van der Waals forces, in which hydrogen bonds are one of the strongest non-covalent interactions, and the number and distance of hydrogen bonds contribute to the affinity between aptamers and target molecules (Sabri et al., 2019). In the MD simulations, we found that more hydrogen bonds were formed in the binding of AP273 to AFP in buffer 6 than in buffer 9, indicating a stronger binding affinity between the two molecules in buffer 6. However, the most stable binding of AP273 to AFP was found in buffer 9 rather than buffer 6, where the hydrogen bonds were less frequent but the aptamer-target binding sites were more numerous and dispersed. This may suggest the stability is more related to the binding sites of the aptamer and the target, while the affinity is more related to the hydrogen bonds between the aptamer and the target.

In the comparative study of binding properties between aptamer AP273 and aptamer AP-L3-4, we found that the binding properties of AP273 were superior to those of AP-L3-4 in both TFA and MD simulations, suggesting that the combination of TFA and MD simulation can identify aptamers that exhibit better binding characteristics among homologous aptamers.

Additionally, we found that the stability of AP273 binding to AFP was inconsistent between TFA and MD simulations. In TFA experiments, the most stability was found in buffer 6, followed by buffer 9, buffer 5 and buffer 7, while in MD simulations, the most stability was observed in buffer 9, followed by buffer 6, buffer 7 and buffer 5. This may be due to subtle differences in experimental conditions between MD simulations and TFA. The MD simulations were performed under conditions of constant temperature (25°C), neutral reaction environment, atmospheric pressure, aqueous metal ion solution, and no buffer salt, while the TFA were performed under the HEPES buffer system and constantly increasing temperature. Furthermore, it is important to keep in mind the limitations of molecular simulation methods. The results of MD are microscopic simulations and predictions from large data sets; the prediction of aptamer 3D structure based only on the sequence is still a very unreliable task, which has by no means achieved the maturity of protein structure prediction. There is no guarantee that the prediction of an aptamer structure is close to the correct fold. Certainly, an all-atom energy minimization is not sufficient to assess the stability of the predicted aptamer structure, let alone when done in a vacuum.

5 Conclusion

In this study, the binding of aptamer AP273 to its target AFP was analyzed by TFA and underlying molecular mechanisms were analyzed through MD simulations. On the melting curve of TFA, the free and bound AP273 peaks were clearly visible, and the binding status of AP273 with AFP could be determined based on the peak-related indicators (T_m value, height or area of the peaks, and the ratio of the two peaks). These indicators varied with changes in aptamer-target

ratio, buffer system and metal ionic strength, which can be useful for optimizing experimental conditions. The analysis of MD simulation showed that the affinity and stability of AP273 binding to AFP varied at different metal ionic strengths, and the underlying mechanisms were related to the difference in the hydrogen bond frequency and binding sites. In the comparative study of AP273 and the control aptamer AP-L3-4, the two aptamers were different in the peak-related indicators and in hydrogen bonds and binding free energies, which are useful for the preferential selection of aptamers. Although the stability of AP273 binding to AFP was inconsistent between TFA and MD simulation, the combination of the two methods provided dual validation that was more intuitive than mere experimental results and more convincing than mere theoretical analysis, and it will facilitate the translational application of aptamers in bioassays.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://alphafold.ebi.ac.uk/>; <https://www.ebi.ac.uk/pdbe/pdbe-kb/proteins/P02771>.

Author contributions

H-LZ: conceptualization, methodology, software, investigation, data interpretation, data curation, writing—original draft. CL: validation, data interpretation. Z-HL: formal analysis. SJ: investigation. DC: investigation. S-SL: investigation. TW: project administration, conceptualization, data interpretation, writing—review and editing, funding acquisition, supervision. K-HZ: conceptualization, data interpretation, resources, writing—review and editing, funding acquisition, supervision.

Funding

This study is supported by the National Natural Science Foundation of China (82160444) and the Science and Technology Project of Jiangxi Province, China (20192BBG70048).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Antczak, M., Popenda, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., et al. (2016). New functionality of RNAComposer: An application to shape the axis of mR160 precursor structure. *Acta Biochim. Pol.* 63, 737–744. doi:10.18388/abp.2016_1329
- Belleperche, M., and DeRosa, M. C. (2018). pH-control in aptamer-based diagnostics, therapeutics, and analytical applications. *Pharm. (Basel)* 11, 80. doi:10.3390/ph11030080
- Cai, S., Yan, J., Xiong, H., Liu, Y., Peng, D., and Liu, Z. (2018). Investigations on the interface of nucleic acid aptamers and binding targets. *Analyst* 143, 5317–5338. doi:10.1039/c8an01467a
- Chang, Y. T., Lin, Y. C., Sun, L., Liao, W. C., Wang, C. C. N., Chou, C. Y., et al. (2020). Wilforine resensitizes multidrug resistant cancer cells via competitive inhibition of P-glycoprotein. *Phytomedicine* 71, 153239. doi:10.1016/j.phymed.2020.153239
- Chen, W., Gao, G., Jin, Y., and Deng, C. (2020). A facile biosensor for Aβ40O based on fluorescence quenching of prussian blue nanoparticles. *Talanta* 216, 120930. doi:10.1016/j.talanta.2020.120930
- Chen, Y., Zheng, Y., Fong, P., Mao, S., and Wang, Q. (2020). The application of the MM/GBSA method in the binding pose prediction of FGFR inhibitors. *Phys. Chem. Chem. Phys.* 22, 9656–9663. doi:10.1039/d0cp00831a
- Cheng, S. A., Fung, W. K., Chan, K. Y., and Shen, P. K. (2003). Optimizing electron spin resonance detection of hydroxyl radical in water. *Chemosphere* 52, 1797–1805. doi:10.1016/s0045-6535(03)00369-2
- Christina Schmidt, M. P., Klusmann, S., Betzel, C., and Betzel, C. (2020). Molecular characterization of a ghrelin-l-aptamer complex. *J. Mol. Struct.* 1204, 127510. doi:10.1016/j.molstruc.2019.127510
- Cui, X., Song, M., Liu, Y., Yuan, Y., Huang, Q., Cao, Y., et al. (2020). Identifying conformational changes of aptamer binding to theophylline: A combined biolayer interferometry, surface-enhanced Raman spectroscopy, and molecular dynamics study. *Talanta* 217, 121073. doi:10.1016/j.talanta.2020.121073
- Darmostuk, M., Rimpelova, S., Gbelcova, H., and Ruml, T. (2015). Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol. Adv.* 33, 1141–1161. doi:10.1016/j.biotechadv.2015.02.008
- Davydova, A. S., Timoshenko, V. V., Lomzov, A. A., Pyshnyi, D. V., Venyaminova, A. G., and Vorobyeva, M. A. (2020). G-quadruplex 2'-F-modified RNA aptamers targeting hemoglobin: Structure studies and colorimetric assays. *Anal. Biochem.* 611, 113886. doi:10.1016/j.ab.2020.113886
- Dong, L., Tan, Q., Ye, W., Liu, D., Chen, H., Hu, H., et al. (2015). Screening and identifying a novel ssDNA aptamer against alpha-fetoprotein using CE-SELEX. *Sci. Rep.* 5, 15552. doi:10.1038/srep15552
- Douaki, A., Garoli, D., Inam, A., Angeli, M. A. C., Cantarella, G., Rocchia, W., et al. (2022). Smart approach for the design of highly selective aptamer-based biosensors. *Biosens. (Basel)* 12, 574. doi:10.3390/bios12080574
- Duan, N., Wu, S., Chen, X., Huang, Y., and Wang, Z. (2012). Selection and identification of a DNA aptamer targeted to *Vibrio parahaemolyticus*. *J. Agric. Food Chem.* 60, 4034–4038. doi:10.1021/jf300395z
- Ellington, A. D., and Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 346, 818–822. doi:10.1038/346818a0
- Emrani, A. S., Danesh, N. M., Ramezani, M., Taghdisi, S. M., and Abnous, K. (2016). A novel fluorescent aptasensor based on hairpin structure of complementary strand of aptamer and nanoparticles as a signal amplification approach for ultrasensitive detection of cocaine. *Biosens. Bioelectron.* 79, 288–293. doi:10.1016/j.bios.2015.12.025
- Garcia-Recio, E. M., Pinto-Diez, C., Perez-Morgado, M. I., Garcia-Hernandez, M., Fernandez, G., Martin, M. E., et al. (2016). Characterization of MNK1b DNA aptamers that inhibit proliferation in MDA-mb231 breast cancer cells. *Mol. Ther. Nucleic Acids* 5, e275. doi:10.1038/mtna.2015.50
- Genheden, S., and Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* 10, 449–461. doi:10.1517/17460441.2015.1032936
- Hayashi, T., Oshima, H., Mashima, T., Nagata, T., Katahira, M., and Kinoshita, M. (2014). Binding of an RNA aptamer and a partial peptide of a prion protein: Crucial importance of water entropy in molecular recognition. *Nucleic Acids Res.* 42, 6861–6875. doi:10.1093/nar/gku382
- Heiat, M., Najafi, A., Ranjbar, R., Latifi, A. M., and Rasaei, M. J. (2016). Computational approach to analyze isolated ssDNA aptamers against angiotensin II. *J. Biotechnol.* 230, 34–39. doi:10.1016/j.jbiotec.2016.05.021
- Hianik, T., Ostatna, V., Sonlajterova, M., and Grman, I. (2007). Influence of ionic strength, pH and aptamer configuration for binding affinity to thrombin. *Bioelectrochemistry* 70, 127–133. doi:10.1016/j.bioelechem.2006.03.012
- Hilder, T. A., and Hodgkiss, J. M. (2017). The bound structures of 17β-estradiol-binding aptamers. *Chemphyschem* 18, 1881–1887. doi:10.1002/cphc.201700363
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model* 51, 69–82. doi:10.1021/ci100275a
- Hu, J., and Easley, C. J. (2017). Homogeneous assays of second messenger signaling and hormone secretion using thermofluorimetric methods that minimize calibration burden. *Anal. Chem.* 89, 8517–8523. doi:10.1021/acs.analchem.7b02229
- Hu, J., Kim, J., and Easley, C. J. (2015). Quantifying aptamer-protein binding via thermofluorimetric analysis. *Anal. Methods* 7, 7358–7362. doi:10.1039/c5ay00837a
- Huang, R., Xiong, W., Wang, D., Guo, L., Lin, Z., Yu, L., et al. (2013). Label-free aptamer-based partial filling technique for enantioseparation and determination of DL-tryptophan with micellar electrokinetic chromatography. *Electrophoresis* 34, 254–259. doi:10.1002/elps.201200464
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph* 14 (33–8), 33–38. doi:10.1016/0263-7855(96)00018-5
- Kang, L., Yang, B., Zhang, X., Cui, L., Meng, H., Mei, L., et al. (2015). Enzymatic cleavage and mass amplification strategy for small molecule detection using aptamer-based fluorescence polarization biosensor. *Anal. Chim. Acta* 879, 91–96. doi:10.1016/j.aca.2015.03.030
- Kim, J., Hu, J., Bezerra, A. B., Holtan, M. D., Brooks, J. C., and Easley, C. J. (2015). Protein quantification using controlled DNA melting transitions in bivalent probe assemblies. *Anal. Chem.* 87, 9576–9579. doi:10.1021/acs.analchem.5b03432
- La Penna, G., and Chelli, R. (2018). Structural insights into the osteopontin-aptamer complex by molecular dynamics simulations. *Front. Chem.* 6, 2. doi:10.3389/fchem.2018.00002
- Lee, T. S., Allen, B. K., Giese, T. J., Guo, Z., Li, P., Lin, C., et al. (2020). Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug Discovery. *J. Chem. Inf. Model* 60, 5595–5623. doi:10.1021/acs.jcim.0c00613
- Li, G., Zeng, J., Liu, H., Ding, P., Liang, J., Nie, X., et al. (2019). A fluorometric aptamer nanoprobe for alpha-fetoprotein by exploiting the FRET between 5-carboxyfluorescein and palladium nanoparticles. *Mikrochim. Acta* 186, 314. doi:10.1007/s00604-019-3403-z
- Lill, M. A., and Danielson, M. L. (2011). Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des.* 25, 13–19. doi:10.1007/s10822-010-9395-8
- Liu, Y., Canoura, J., Alkhamis, O., and Xiao, Y. (2021). Immobilization strategies for enhancing sensitivity of electrochemical aptamer-based sensors. *ACS Appl. Mater. Interfaces* 13, 9491–9499. doi:10.1021/acsami.0c20707
- Longhini, A. P., LeBlanc, R. M., Becette, O., Salguero, C., Wunderlich, C. H., Johnson, B. A., et al. (2016). Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res.* 44, e52. doi:10.1093/nar/gkv1333
- Mahmoud, M., Laufer, S., and Deigner, H. P. (2019). An aptamer based thermofluorimetric assay for ethanolamine. *Biochimie* 158, 233–237. doi:10.1016/j.biochi.2019.01.014
- McCluskey, K., Boudreault, J., St-Pierre, P., Perez-Gonzalez, C., Chauvier, A., Rizzi, A., et al. (2019). Unprecedented tunability of riboswitch structure and regulatory function by sub-millimolar variations in physiological Mg²⁺. *Nucleic Acids Res.* 47, 6478–6487. doi:10.1093/nar/gkz316
- Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R., and Weng, Z. (2007). Integrating statistical pair potentials into protein complex prediction. *Proteins* 69, 511–520. doi:10.1002/prot.21502
- Moccia, F., Platella, C., Musumeci, D., Batool, S., Zumrut, H., Bradshaw, J., et al. (2019). The role of G-quadruplex structures of LIGS-generated aptamers R1.2 and R1.3 in IgM specific recognition. *Int. J. Biol. Macromol.* 133, 839–849. doi:10.1016/j.ijbiomac.2019.04.141
- Morozov, D., Mironov, V., Moryachkov, R. V., Shchugoreva, I. A., Artyushenko, P. V., Zamay, G. S., et al. (2021). The role of SAXS and molecular simulations in 3D structure elucidation of a DNA aptamer against lung cancer. *Mol. Ther. Nucleic Acids* 25, 316–327. doi:10.1016/j.omtn.2021.07.015
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi:10.1002/jcc.20289
- Pierce, B. G., Hourai, Y., and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* 6, e24657. doi:10.1371/journal.pone.0024657
- Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30, 1771–1773. doi:10.1093/bioinformatics/btu097
- Poolsup, S., Zaripov, E., Huttman, N., Minic, Z., Artyushenko, P. V., Shchugoreva, I. A., et al. (2023). Discovery of DNA aptamers targeting SARS-CoV-2 nucleocapsid protein and protein-binding epitopes for label-free COVID-19 diagnostics. *Mol. Ther. Nucleic Acids* 31, 731–743. doi:10.1016/j.omtn.2023.02.010
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., et al. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* 40, e112. doi:10.1093/nar/gks339

- Rastelli, G., Del Rio, A., Degliesposti, G., and Sgobba, M. (2010). Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* 31, 797–810. doi:10.1002/jcc.21372
- Ruigrok, V. J. B., Levisson, M., Hekelaar, J., Smidt, H., Dijkstra, B. W., and Van der Oost, J. (2012). Characterization of aptamer-protein complexes by X-ray crystallography and alternative approaches. *Int. J. Mol. Sci.* 13, 10537–10552. doi:10.3390/ijms130810537
- Ruzza, P., Honisch, C., Hussain, R., and Siligardi, G. (2021). Free radical generation in far-UV synchrotron radiation circular dichroism assays-protein and buffer composition contribution. *Int. J. Mol. Sci.* 22, 11325. doi:10.3390/ijms222111325
- Sabri, M. Z., Abdul Hamid, A. A., Sayed Hitam, S. M., and Abdul Rahim, M. Z. (2019). *In silico* screening of aptamers configuration against hepatitis B surface antigen. *Adv. Bioinforma.* 2019, 6912914. doi:10.1155/2019/6912914
- Santos, T., Lopes-Nunes, J., Alexandre, D., Miranda, A., Figueiredo, J., Silva, M. S., et al. (2022). Stabilization of a DNA aptamer by ligand binding. *Biochimie* 200, 8–18. doi:10.1016/j.biochi.2022.05.002
- Shigdar, S. (2019). Aptamer-based diagnostics and therapeutics. *Pharm. (Basel)* 12, 6. doi:10.3390/ph12010006
- Shoute, L. C. T., and Loppnow, G. R. (2018). Characterization of the binding interactions between EvaGreen dye and dsDNA. *Phys. Chem. Chem. Phys.* 20, 4772–4780. doi:10.1039/c7cp06058k
- Steinberg, L., Russo, J., and Frey, J. (2019). A new topological descriptor for water network structure. *J. Cheminform* 11, 48. doi:10.1186/s13321-019-0369-0
- Subki, A., Ho, C. L., Ismail, N. F. N., Zainal Abidin, A. A., and Balia Yusof, Z. N. (2020). Identification and characterisation of thiamine pyrophosphate (TPP) riboswitch in *Elaeis guineensis*. *PLoS One* 15, e0235431. doi:10.1371/journal.pone.0235431
- Tabuchi, Y., Yang, J., and Taki, M. (2022). Relative nuclease resistance of a DNA aptamer covalently conjugated to a target protein. *Int. J. Mol. Sci.* 23, 7778. doi:10.3390/ijms23147778
- Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Miguez, A. N., et al. (2020). ff19SB: Amino-Acid-Specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16, 528–552. doi:10.1021/acs.jctc.9b00591
- Tivon, Y., Falcone, G., and Deiters, A. (2021). Protein labeling and crosslinking by covalent aptamers. *Angew. Chem. Int. Ed. Engl.* 60, 16035–16040. doi:10.1002/ange.202101174
- Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510. doi:10.1126/science.2200121
- UniProt, C., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., et al. (2023). UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi:10.1093/nar/gkac1052
- Vu, C. Q., Rotkrua, P., Soontornworajit, B., and Tantirungrotechai, Y. (2018). Effect of PDGF-B aptamer on PDGFR β /PDGF-B interaction: Molecular dynamics study. *J. Mol. Graph. Model* 82, 145–156. doi:10.1016/j.jmglm.2018.04.012
- Wang, Q. L., Cui, H. F., Du, J. F., Lv, Q. Y., and Song, X. (2019). *In silico* post-SELEX screening and experimental characterizations for acquisition of high affinity DNA aptamers against carcinoembryonic antigen. *RSC Adv.* 9, 6328–6334. doi:10.1039/c8ra10163a
- Wu, H., Liu, R., Kang, X., Liang, C., Lv, L., and Guo, Z. (2017). Fluorometric aptamer assay for ochratoxin A based on the use of single walled carbon nanohorns and exonuclease III-aided amplification. *Mikrochim. Acta* 185, 27. doi:10.1007/s00604-017-2592-6
- Wu, L., Wang, Y., Xu, X., Liu, Y., Lin, B., Zhang, M., et al. (2021). Aptamer-based detection of circulating targets for precision medicine. *Chem. Rev.* 121, 12035–12105. doi:10.1021/acs.chemrev.0c01140
- Yang, F., Li, J., Dong, H., Wang, G., Han, J., Xu, R., et al. (2022). A novel label-free electrochemiluminescence aptasensor using a tetrahedral DNA nanostructure as a scaffold for ultrasensitive detection of organophosphorus pesticides in a luminol-H(2)O(2) system. *Analyst* 147, 712–721. doi:10.1039/d1an02060a
- Zgarbova, M., Sponer, J., Otyepka, M., Cheatham, T. E., 3rd, Galindo-Murillo, R., and Jurecka, P. (2015). Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-dna. *J. Chem. Theory Comput.* 11, 5723–5736. doi:10.1021/acs.jctc.5b00716
- Zhang, X., and Yadavalli, V. K. (2010). Molecular interaction studies of vascular endothelial growth factor with RNA aptamers. *Analyst* 135, 2014–2021. doi:10.1039/c0an00200c
- Zhao, Q., Bai, Y., and Wang, H. (2020). Directing a rational design of aptamer-based fluorescence anisotropy assay for sensitive detection of immunoglobulin E by site-specific binding study. *Talanta* 217, 121018. doi:10.1016/j.talanta.2020.121018
- Zhou, C., Hu, J., Ma, H., Yagoub, A. E., Yu, X., Owusu, J., et al. (2015). Antioxidant peptides from corn gluten meal: Orthogonal design evaluation. *Food Chem.* 187, 270–278. doi:10.1016/j.foodchem.2015.04.092
- Zhu, Y., Cai, Y., Xu, L., Zheng, L., Wang, L., Qi, B., et al. (2015). Building an aptamer/graphene oxide FRET biosensor for one-step detection of bisphenol A. *ACS Appl. Mater. Interfaces* 7, 7492–7496. doi:10.1021/acsami.5b00199
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi:10.1093/nar/gkg595



OPEN ACCESS

EDITED BY

Jose Luis Cabellos,
Polytechnic University of Tapachula,
Mexico

REVIEWED BY

Christof Holzer,
Karlsruhe Institute of Technology (KIT),
Germany
Gerardo Martinez-Guajardo,
Universidad Autónoma de Zacatecas,
Mexico

*CORRESPONDENCE

Zikuan Wang,
✉ zwang@kofo.mpg.de
Wenjian Liu,
✉ liuwj@sdu.edu.cn

RECEIVED 14 July 2023

ACCEPTED 23 October 2023

PUBLISHED 10 November 2023

CITATION

Wang X, Wu C, Wang Z and Liu W (2023),
When do tripdoublet states fluoresce? A
theoretical study of copper(II) porphyrin.
Front. Chem. 11:1259016.
doi: 10.3389/fchem.2023.1259016

COPYRIGHT

© 2023 Wang, Wu, Wang and Liu. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

When do tripdoublet states fluoresce? A theoretical study of copper(II) porphyrin

Xingwen Wang¹, Chenyu Wu¹, Zikuan Wang^{1,2*} and Wenjian Liu^{1*}

¹Qingdao Institute for Theoretical and Computational Sciences, Shandong University, Qingdao, China,

²Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany

Open-shell molecules rarely fluoresce, due to their typically faster non-radiative relaxation rates compared to closed-shell ones. Even rarer is the fluorescence from states that have two more unpaired electrons than the open-shell ground state, since they involve excitations from closed-shell orbitals to vacant-shell orbitals, which are typically higher in energy compared to excitations from or out of open-shell orbitals. States that are dominated by the former type of excitations are known as tripdoublet states when they can be described as a triplet excitation antiferromagnetically coupled to a doublet state, and their description by unrestricted single-reference methods (e.g., U-TDDFT) is notoriously inaccurate due to large spin contamination. In this work, we applied our spin-adapted TDDFT method, X-TDDFT, and the efficient and accurate static-dynamic-second order perturbation theory (SDSPT2), to the study of the excited states as well as their relaxation pathways of copper(II) porphyrin; previous experimental works suggested that the photoluminescence of some substituted copper(II) porphyrins originate from a tripdoublet state, formed by a triplet ligand $\pi \rightarrow \pi^*$ excitation antiferromagnetically coupled with the unpaired d electron. Our results demonstrated favorable agreement between the X-TDDFT, SDSPT2 and experimental excitation energies, and revealed noticeable improvements of X-TDDFT compared to U-TDDFT, not only for vertical excitation energies but also for adiabatic energy differences. These suggest that X-TDDFT is a reliable tool for the study of tripdoublet state fluorescence. Intriguingly, we showed that the aforementioned tripdoublet state is only slightly above the lowest doublet excited state and lies only slightly higher than the lowest quartet state, which suggests that the tripdoublet of copper(II) porphyrin is long-lived enough to fluoresce due to a lack of efficient non-radiative relaxation pathways; an explanation for this unusual state ordering is given. Indeed, thermal vibration correlation function (TVCF)-based calculations of internal conversion, intersystem crossing, and radiative transition rates confirm that copper(II) porphyrin emits thermally activated delayed fluorescence (TADF) and a small amount of phosphorescence at low temperature (83 K), in accordance with experiment. The present contribution is concluded by a few possible approaches of designing new molecules that fluoresce from tripdoublet states.

KEYWORDS

tripdoublet, fluorescence, photophysics, spin-adapted TDDFT, SDSPT2

1 Introduction

Fluorescence, while ubiquitous in organic and organometallic molecules, is in most cases observed in closed-shell systems. It is well-known that introducing an open-shell impurity, such as dioxygen (Schmidt, 2006), a stable organic radical (Green et al., 1990) or a transition metal ion (Varnes et al., 1972), frequently quenches the fluorescence of a closed-shell molecule (Evans, 1957). One reason of this phenomenon is that the addition of an unpaired electron to a system typically introduces additional low-lying states, in particular charge transfer states that involve an electron exciting from or out of the new open-shell orbital (O). Moreover, while spin-conserving single excitations of a singlet reference determinant from closed-shell (C) to vacant-shell (V) orbitals, hereafter termed CV excitations following our previous works (Li and Liu, 2010; Li et al., 2011; Li and Liu, 2011; Wang et al., 2020), give rise to $n_C n_V$ singlet excited states and $n_C n_V$ triplet excited states (where n_C and n_V are the number of closed-shell and vacant-shell orbitals, respectively), with an $M_S = 1/2$ doublet determinant one obtains $2n_C n_V$ excitations that are mixtures of doublets and quartets (the Ψ_i^a and $\Psi_i^{\bar{a}}$ determinants in Figure 1; here orbitals without overbars denote α orbitals, and those with overbars denote β ones). They can be linearly combined to make $n_C n_V$ pure doublet states, but the other linear combination remains a mixture of doublet and quartet:

$$\Psi_{\text{singdoublet}} = \frac{1}{\sqrt{2}} (\Psi_i^a + \Psi_i^{\bar{a}}), \quad (1)$$

$$\Psi_{\text{mixed}} = \frac{1}{\sqrt{2}} (\Psi_i^a - \Psi_i^{\bar{a}}). \quad (2)$$

In spin-adapted TDDFT methods, the latter are spin-adapted to give $2n_C n_V$ pure doublet states and $n_C n_V$ quartet states, by mixing with the $n_C n_V$ spin flip-up excitations from the $M_S = -1/2$ component of the reference determinant, i.e., the $\Psi_{it}^{\bar{a}}$ determinants in Figure 1 (Li and Liu, 2010; Li et al., 2011; Li and Liu, 2011):

$$\Psi_{\text{tripdoublet}} = \frac{1}{\sqrt{6}} (-\Psi_i^a + \Psi_i^{\bar{a}} + 2\Psi_{it}^{\bar{a}}), \quad (3)$$

$$\Psi_{\text{quartet}} = \frac{1}{\sqrt{3}} (\Psi_i^a - \Psi_i^{\bar{a}} + \Psi_{it}^{\bar{a}}). \quad (4)$$

Note that both the “singdoublets” and “tripdoublets” are pure doublet states. While the singdoublets Eq. 1 (which we called the CV(0) states in our previous works (Li and Liu, 2010; Li et al., 2011; Li and Liu, 2011)) are direct analogs of singlet excited states out of a singlet reference, the tripdoublets Eq. 3 (CV(1) states) do not have analogs in closed-shell systems, and create extra spin-allowed non-radiative relaxation pathways compared to when the reference determinant is singlet. This further contributes to the short excited state lifetimes of doublet systems. As a consequence, doublet molecules (and open-shell molecules in general) are rarely fluorescent.

Still, there exist open-shell molecules that do fluoresce, which have found applications in, e.g., organic light-emitting diodes (OLEDs) (He et al., 2019; Gao et al., 2023). However, their fluorescence usually originates from an excited state that has only one unpaired electron, i.e., a CO or OV excited state (where CO stands for a single excitation from a closed-shell orbital to an open-shell one; similar for OV), instead of a CV excited state. This can be partly rationalized by approximating the excitation energies of the system by orbital energy differences. Under this approximation, there is at least one CO state and one OV state below any given CV state, since the lowest CV excitation energy is the sum of the excitation energies of a CO state and an OV state (Figure 1). Therefore, the lowest CV state tends to be a rather high excited state of the system, and thus usually has more energetically accessible non-radiative relaxation pathways than the low-lying CO and OV states do, rendering fluorescence from CV states especially hard to achieve. To counter this, one may try to inhibit the non-radiative relaxation of the CV state to lower excited states. However, the sheer number of non-radiative relaxation pathways that one would have to inhibit poses a great challenge for designing an open-shell molecule that fluoresces from a CV state. Alternatively, one may design a system where the orbital energy difference approximation fails dramatically, allowing the lowest CV state to become the first excited state, or only

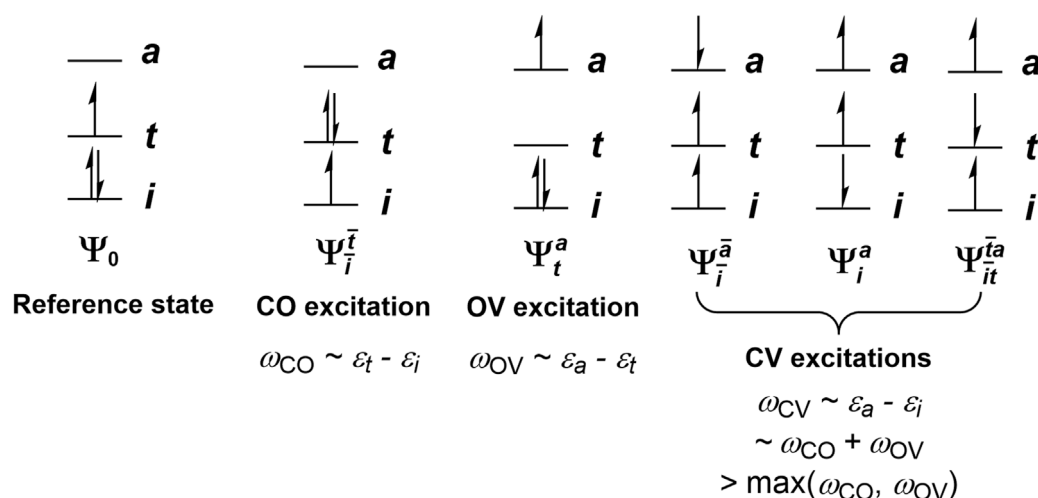


FIGURE 1

Schematic depictions of closed-open (CO), open-vacant (OV), and closed-vacant (CV) excitations, and their approximate excitation energies as predicted from restricted open-shell Kohn-Sham (ROKS) orbital energy differences.

slightly higher than the first excited state. In this case, the fluorescence from the CV state only needs to compete with the intersystem crossings (ISCs) to the lowest quartet state(s) and the internal conversion (IC) to the ground state and lower doublet excited state(s), which are the only energy downhill non-radiative relaxation pathways available to the CV state. In particular, note that when the CV excitations shown in Figure 1 linearly combine to give singdoublets, tripdoublets and quartets via Eqs 3, 4, there is an energy splitting that usually places the quartet below the tripdoublet, and the tripdoublet below the singdoublet; while the former is a consequence of Hund's rule, the latter can be rationalized by applying Hund's rule after neglecting the coupling of the open-shell orbital to the closed-shell and vacant-shell ones. This gives tripdoublets a much greater chance than singdoublets for emitting fluorescence with an appreciable quantum yield. Nevertheless, the singdoublet-tripdoublet splitting appears to be small in general, compared to the orbital energy difference that one would have to overcome, which can amount to several eVs. Hence, even the fluorescence from tripdoublets proves to be scarce.

The present paper represents a preliminary attempt to unveil some of the factors that enable an open-shell molecule to fluoresce from a tripdoublet state, via a case study of copper(II) porphyrin complexes. Copper(II) porphyrin complexes (Figure 2), like most porphyrin complexes, show two intense visible absorption bands near 390–420 nm and 520–580 nm (Gouterman, 1959; Eastwood and Gouterman, 1969); they are conventionally termed the B and Q bands, respectively. Eastwood and Gouterman (1969) studied the luminescence of copper(II) porphyrin molecules in the solid state by exciting their Q bands, suggesting that the emission may originate from one of the two low-lying $\pi \rightarrow \pi^*$ states, 2T or 4T (here the 2, 4 represent the overall spin multiplicity of the complex, and T denotes that the "local" spin multiplicity of the porphyrin ring is triplet). They speculated that a rapid equilibrium may exist between the 2T and 4T states. The equilibrium ratio of these two states is largely dependent on the energy gap (ΔE_{DQ}) between them and the temperature, via the Boltzmann distribution. The radiative transition from the 2T state to the ground state is spin-allowed, making it much faster than the phosphorescence from the 4T state. Thus, when ΔE_{DQ} is small and the temperature is high, the experimentally observed rapid emission is predominantly from the 2T

state. Conversely, when ΔE_{DQ} is large and the temperature is low, a slow emission attributed to the phosphorescence of the 4T state was observed instead, due to the concentration of the 4T state largely overwhelming that of the 2T state. Thus, molecules such as copper 2,3,7,8,12,13,17,18-octaalkylporphyrin (CuOAP), which possess small ΔE_{DQ} values, exhibit luminescence primarily in the form of fluorescence from the 2T state at liquid nitrogen temperature, whereas copper 5,10,15,20-tetraphenylporphyrin (CuTPP) with a larger ΔE_{DQ} mainly undergoes phosphorescence from the 4T state at the same temperatures. The unsubstituted copper porphyrin (CuP) is the most interesting of all, as pure phosphorescence was observed at low temperatures (35 K), which gradually gives way to fluorescence when the temperature was elevated, eventually giving pure fluorescence at 143 K (Bohandy and Kim, 1980). Similar results have been obtained by following works with different techniques and/or solvents (Magde et al., 1974; Kobayashi et al., 2008).

The simple and intuitive picture has since been supplemented by subsequent works, which also excited the B band, and proposed that charge transfer (CT) states may play an important role in the relaxation of the initial bright state to the essentially dark 2T state. Yan and Holten (1988) investigated the excited state relaxation processes of CuTPP and CuOEP at different temperatures and in different solvents, proposing possible pathways involving intermediate states that are probably ligand-to-metal CT (LMCT) states. This is supported by the gas-phase mass spectrometry experiments by Ha-Thi et al. (2013), although the precise composition of the CT state remains uncertain. Understanding the excited-state relaxation pathways of copper porphyrins is crucial for gaining insights into their photophysical processes and controlling their optical properties. In particular, whether any other excited state(s) lie below the 2T state may have a profound influence on whether the 2T state fluoresces or not, as follows from Kasha's rule. Meanwhile, the energy gap of the 2T and 4T states is important for the relative concentration of the two states, and therefore the relative intensities of fluorescence from the 2T state and the phosphorescence from the 4T state, i.e., whether the experimentally observed luminescence should be attributed to fluorescence or phosphorescence, or both.

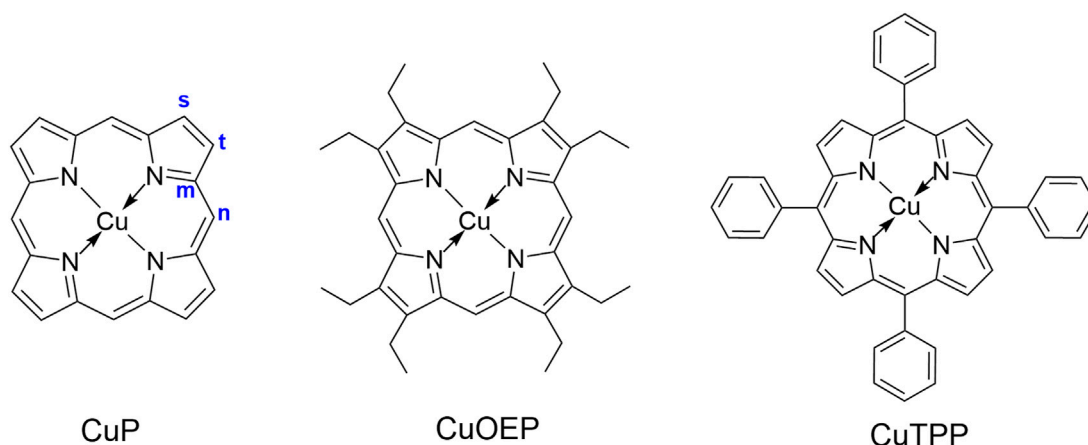


FIGURE 2
Molecular structures of CuP, CuOEP and CuTPP.

Despite the importance of tripdoublet fluorescence and the long history of experimental studies of copper porphyrins, accurate computational studies of this system prove to be difficult, as traditional unrestricted single-reference methods like U-TDDFT suffer from severe spin contamination issues, leading to systematically underestimated excitation energies. In particular, tripdoublet states are the worst scenario for U-TDDFT, as when the reference state itself is not spin-contaminated, the errors of the U-TDDFT $\langle S^2 \rangle$ values reach the theoretical maximum of singly excited states, i.e., 2, for tripdoublet states (Li and Liu, 2010; Li et al., 2011; Li and Liu, 2011; Li and Liu, 2016a; Wang et al., 2020). While multireference methods trivially solve the spin contamination problems, it is notoriously difficult to obtain an accurate multireference description of the electronic structure of metalloporphyrins, due to the complex interplay between static and dynamic correlation. In this study, we employed the methods developed by our group, namely, X-TDDFT (Li and Liu, 2011; Wang et al., 2020) and SDSPT2 (Liu and Hoffmann, 2014; Lei et al., 2017) (static-dynamic-static second-order perturbation theory), to address these challenges and provide a rational description of the photophysical processes in CuP. As the first rigorous spin-adapted TDDFT method (Li and Liu, 2011), X-TDDFT gives spin-adapted excited states even when the reference state is open-shell, thereby generally giving better excitation energies, as well as better transition matrix elements involving the excited states. The recent development of the analytic gradient of X-TDDFT (Wang et al., 2020) allowed us to use X-TDDFT for excited state geometry optimization and seminumerical Hessian calculations as well. For vertical excitation calculations, we could afford to use SDSPT2, which also served as a reference for benchmarking X-TDDFT and U-TDDFT.

2 Computational details

All DFT, TDDFT, and SDSPT2 calculations were performed using a development version of the Beijing Density Functional (BDF) package (Liu et al., 1997; Liu et al., 2003; Liu et al., 2004a; Liu et al., 2004b; Zhang et al., 2020). Geometry optimizations were conducted using the PBE0 (Adamo and Barone, 1999; Ernzerhof and Scuseria, 1999) functional and x2c-SVPall (Pollak and Weigend, 2017) basis set in the gas phase, including Grimme's D3 dispersion correction (Grimme et al., 2010; Grimme et al., 2011), as implemented in the BDF software; relativistic effects were considered at the spin-free exact two component (sf-X2C) level (Liu et al., 2003; Li et al., 2014b; Liu et al., 1995; Liu et al., 2004b; Liu et al., 2004a). For transition metal complexes (especially when excited states are considered), the choice of the optimum functional may not be obvious. Herein, TDDFT calculations were performed using four different functionals [BP86 (Perdew, 1986a; Perdew, 1986b; Becke, 1988), B3LYP (Becke, 1993; Stephens et al., 1994), PBE0 and ω B97X (Chai and Head-Gordon, 2008)] in conjunction with the x2c-TZVPall (Pollak and Weigend, 2017) basis set, followed by benchmarking the results against SDSPT2 and experimental results, and the PBE0 functional was chosen based on its satisfactory and uniform accuracy (see Section 3.1 for details). The TDDFT/x2c-TZVPall energies differ from the TDDFT/x2c-SVPall ones by within 0.04 eV for ligand excited states

(2T_1 , 2T_2 , 2S_1 , 2S_2), within 0.14 eV for d-d excited states, and up to 0.29 eV for CT states. Based on this, we used TD-PBE0/x2c-SVPall in all rate constant calculations that involve only the ligand excited states; for rate constants that involve the d-d states (where even TDDFT/x2c-TZVPall predicts wrong state orderings), we calculated the adiabatic excitation energies of the corresponding states using SDSPT2 vertical absorption energies plus the adiabatic/vertical excitation energy differences calculated at the TD-PBE0/x2c-SVPall level, and calculated all the other requisite quantities with TD-PBE0/x2c-SVPall. The orbital diagrams were drawn and visualized with VMD v.1.9.4 (Humphrey et al., 1996), using cube files generated with the help of Multiwfn v.3.8(dev) (Lu and Chen, 2012).

The calculations of ISC rate constants were conducted by the ESD module of the ORCA program, version 5.0.4 (Neese, 2012; Neese, 2018; Neese et al., 2020; Neese, 2022), using the thermal vibration correlation function (TVCF) method based on a multimode harmonic oscillator model. Other rate constants involved in the excited state relaxation process were calculated by the MOMAP package, version 2022A (Peng et al., 2007; Niu et al., 2008; Niu et al., 2020), again using the TVCF method and a harmonic approximation of the potential energy surfaces. The default parameters of the two programs were used in all TVCF calculations, except for the "tmax" parameter in the MOMAP calculations (which controls the propagation time of the TVCF), which was set to 3,000 fs. For the ${}^2T_1 \rightarrow {}^2dd_3$ IC rate constant, the tmax = 3,000 fs calculation suffered from numerical instabilities, so we chose to use tmax = 1,500 fs instead. All necessary transition matrix elements, including the transition dipole moments, non-adiabatic coupling matrix elements (NACMEs) (Li et al., 2014a; Li and Liu, 2014; Wang et al., 2021), spin-orbit coupling matrix elements (SOCMEs) (Li et al., 2013; Li et al., 2014a; Li et al., 2014b), as well as the seminumerical Hessians necessary for the TVCF calculations, were calculated by BDF. Note however that all NACMEs were computed by U-TDDFT instead of X-TDDFT, since the theory of X-TDDFT NACMEs has not been developed yet; similarly, geometry optimization and frequency calculations of the 4T_1 state were performed at the unrestricted Kohn-Sham (UKS) level, which is justified by the small spin contamination ($\langle S^2 \rangle$ deviation < 0.1) of this state. The ALDA0 noncollinear exchange-correlation (XC) kernel (Li and Liu, 2012) was used in all spin flip-up Tamm-Dancoff approximation (TDA) calculations (i.e., calculation of quartet states from a doublet reference), which has proven essential for obtaining correct spin state splittings (Li and Liu, 2016b). Duschinsky rotation was considered whenever applicable. The Herzberg-Teller effect was only considered while calculating the radiative relaxation rates, but not the ISC rates, due to program limitations; however this should not change the qualitative conclusions of this paper, since all ISC processes whose Franck-Condon contributions are negligible or zero are expected to contribute negligibly to the photophysics of CuP. Although we have implemented the interface for calculating the Herzberg-Teller effect of phosphorescence by BDF and MOMAP, the computation of the geometric derivatives of the doublet-quartet transition dipole moments by finite differences proved to be numerically ill-behaved, as the $M_S = \pm 1/2$ and $M_S = \pm 3/2$ microstates of the 4T state mix strongly when the geometry is perturbed; note that this phenomenon seems to be related to the

involvement of quartet states, since we have never observed similar behavior in triplet phosphorescence rate calculations. We thus estimated the total phosphorescence rate by assuming that the ratios of the Franck-Condon and Herzberg-Teller rates are the same for fluorescence and phosphorescence. This treatment is justified by the observation that the geometries and vibrational frequencies of the 2T_1 and 4T_1 states are very similar. For the SDSPT2 calculations, we employed the x2c-TZVPall basis set for the Cu and N atoms, and x2c-SVPall for the remaining atoms. The active space of the SDSPT2 calculations was selected through the iCAS (imposed automatic selection and localization of complete active spaces) method (Lei et al., 2021) using x2c-SVPall for all atoms, followed by projecting onto the aforementioned mixed basis, and the orbitals were optimized using the iCISCF [iterative configuration interaction (iCI)-based multiconfigurational self-consistent field (SCF) theory] method (Guo et al., 2021), which provided a reference wavefunction for the SDSPT2 calculation. An active space of CAS (13, 14) was used in this study. The B-band, Q-band and CT states involved in the excited state relaxation process mainly involve the Cu 3d and 4d orbitals, plus the four porphyrin π orbitals of the Gouterman four-orbital model (Gouterman, 1959), making a minimal active space of CAS (13, 14). The chosen active space thus properly describes the primary excited states of interest for investigation. Expanding the active space further would result in unnecessary computational overhead without providing additional insights. All SDSPT2 calculations reported herein include the Pople correction.

3 Results and discussion

3.1 Absorption process

As is well-known, density functionals generally have difficulties with simultaneously describing local excitation (LE) and CT states with good accuracy. Since we could only afford to do the geometry optimizations and frequency calculations under the DFT and TDDFT levels, a suitable functional that qualitatively reproduces the SDSPT2 excitation energies has to be chosen by comparing the TDDFT vertical absorption energies of a few common functionals with SDSPT2 data. B3LYP and PBE0 are generally common choices for the excited states of metalloporphyrins, and BP86 is often used to optimize their ground-state structures. Pure functionals usually tend to underestimate excitation energies, but empirically, their description of the Q band (an LE state) is better than many hybrid functionals, as will be confirmed by our calculation results. As CT states are involved in the relaxation process of the excited states of copper porphyrin, range-separated hybrid functionals (which provide good descriptions of CT states in general) may prove to be suitable as well. These considerations gave a list of four representative functionals, BP86, B3LYP, PBE0 and ω B97X, that were subjected to benchmark calculations.

Different functionals display distinct behaviors for the excitation energies of CuP compared to the results obtained from SDSPT2, as shown in Figure 3. The two characteristic absorption bands of the porphyrin molecule correspond to the 2S_1 (Q band) and 2S_2 (B band) states, which are the only bright states of most porphyrin complexes

in the visible region. They are also the only excited states for which accurate experimental vertical absorption energies are available: in benzene they have been measured as 2.25 and 3.15 eV, respectively (Eastwood and Gouterman, 1969). Moreover, the absorption energy of the 2T_1 state has been measured by fluorescence excitation spectra experiments, but only for certain substituted porphyrins: for example, the 2T_1 absorption energy of CuEtiO (EtiO = etioporphyrin I) was measured in *n*-octane as 1.81 eV, while the emission energy from the same state in the same solvent was 1.79 eV (Eastwood and Gouterman, 1969). Assuming that the Stokes shift of the 2T_1 state is independent of the porphyrin substituents, and combined with the experimental emission energy of the 2T_1 state of CuP in the same solvent (1.88 eV) (Eastwood and Gouterman, 1969), we obtain an estimate of the experimental 2T_1 absorption energy of CuP as 1.90 eV. Gratifyingly, the SDSPT2 excitation energies of all three states agree with the experimental values to within 0.3 eV, which is typical of the accuracy of SDSPT2 (Song et al., 2022) and confirms the suitability of SDSPT2 as a benchmark reference for CuP. The B3LYP functional performs better for the two bright states 2S_1 and 2S_2 than the other functionals (Figure 3), with results closer to the SDSPT2 calculations, suggesting its suitability for localized excitations in the porphyrin system. However, it (as well as the pure functional BP86) performs poorly in describing the dark charge transfer (CT) states, significantly underestimating their energies, as expected. In contrast, the range-separated functional ω B97X shows better agreement with the CT states compared to SDSPT2 results, but its description of the 2S_2 state is rather poor, with energies notably higher than the SDSPT2 results. The PBE0 functional represents a compromise between the two classes of functionals and provides more accurate overall descriptions of the LE and CT states, giving results closer to the SDSPT2 calculations. In particular, PBE0 gives the best predictions of all the d-d excitation energies, although a significant overestimation is still seen for the lowest d-d state, 2dd_1 . Considering the overall performance in describing different states, we chose to use PBE0 for the remaining part of the present study.

The 2S_1 and 2S_2 states are almost spin-adapted states with minimal spin contamination, even at the U-TDDFT level (Table 1), since they are dominated by singdoublet excitations. As shown in Figure 3, both X-TDDFT and U-TDDFT provide similar descriptions for these two states; note however that functionals with large amounts of HF exchange generally overestimate the excitation energies of these two states, especially 2S_2 . At the TDDFT levels, the CT states are dominated by CO-type excitations (from π to $3d_{x^2-y^2}$), which are also spin-adapted. Both U-TDDFT and X-TDDFT show comparable performance in describing the CT states. However, both methods display large errors compared to SDSPT2 for the CT states. Table 1 presents the excitation energies and the corresponding dominant excited state compositions, computed at the ground state structure of CuP. It can be observed that the CT states are predominantly composed of double excitations, which cannot be described by the present TDDFT calculations. Despite this, functionals with large amounts of HF exchange still perform notably better, as is generally expected for CT states. The 2T_1 and 2T_2 states correspond to tripdoublet excitations (from π to π^*), and they suffer from significant spin contamination at the U-TDDFT level, since instead of pure doublets,

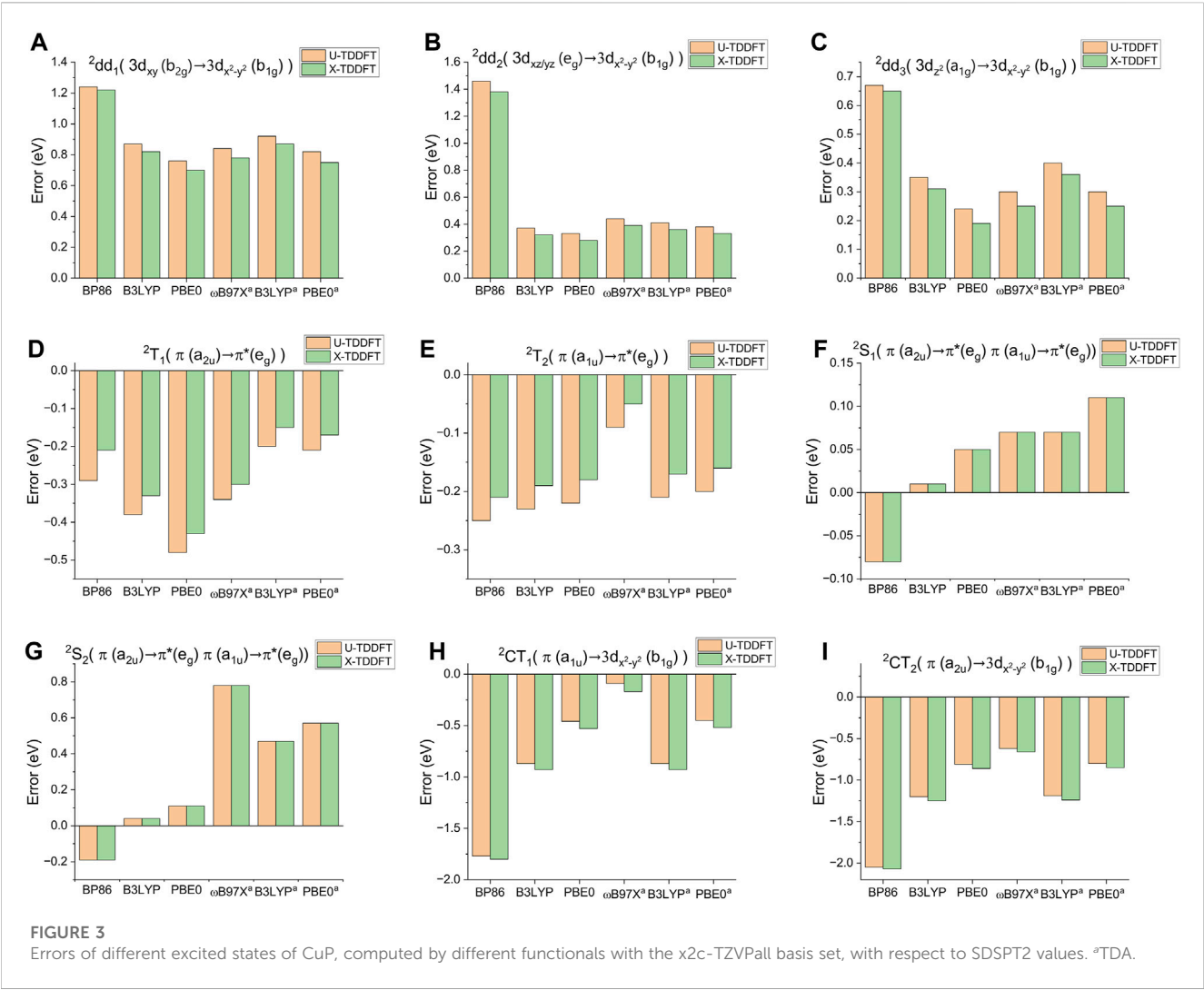


TABLE 1 The SDSPT2/x2c-TZVPall(Cu,N)/x2c-SVPall(C,H) excitation energies (in eV) computed at the sf-X2C-PBE0/x2c-SVPall ground state structure of CuP, along with the corresponding excited state compositions.

State	ΔE	$\Delta \langle S^2 \rangle$	Dominant transitions
${}^2\text{dd}_1$	1.85	0.0019	$3d_{xy} \rightarrow 3d_{x^2-y^2}$ 88.0%
${}^2\text{dd}_2$	2.02	0.0022	$3d_{xz}/3d_{yz} \rightarrow 3d_{x^2-y^2}$ 87.8%
${}^2\text{dd}_3$	2.13	0.0016	$3d_{z^2} \rightarrow 3d_{x^2-y^2}$ 88.0%
${}^2\text{T}_1$	2.19	1.9994	$\pi(a_{2u}) \rightarrow \pi^*(e_g)$ 87.1%
${}^2\text{T}_2$	2.32	1.9968	$\pi(a_{1u}) \rightarrow \pi^*(e_g)$ 86.7%
${}^2\text{S}_1$	2.43	0.0031	$\pi(a_{1u}) \rightarrow \pi^*(e_g)$ 56.9%, $\pi(a_{2u}) \rightarrow \pi^*(e_g)$ 36.1%
${}^2\text{S}_2$	3.41	0.0115	$\pi(a_{2u}) \rightarrow \pi^*(e_g)$ 52.5%, $\pi(a_{1u}) \rightarrow \pi^*(e_g)$ 31.8%
${}^2\text{CT}_1$	3.31	0.0066	$[\pi(a_{1u}) \rightarrow \text{Cu } 3d_{x^2-y^2}(b_{1g}) + \text{Cu } 3d_{xz}/3d_{yz}(e_g) \rightarrow \pi^*(e_g)]$ 65.6%
			$\pi(a_{1u}) \rightarrow \text{Cu } 3d_{x^2-y^2}(b_{1g})$ 25.8%
${}^2\text{CT}_2$	3.51	0.0086	$[\pi(a_{2u})/\pi(a_{2u}) \rightarrow \text{Cu } 3d_{x^2-y^2}(b_{1g}) + \text{Cu } 3d_{xz}/3d_{yz}(e_g) \rightarrow \pi^*(e_g)]$ 61.7%
			$\pi(a_{2u}) \rightarrow \text{Cu } 3d_{x^2-y^2}(b_{1g})$ 21.6%

$\Delta \langle S^2 \rangle$: difference of the excited state's $\langle S^2 \rangle$ value with the ground state $\langle S^2 \rangle$, computed at the U-TD-PBE0/x2c-TZVPall level. Transitions in square brackets represent double excitations.

U-TDDFT can only describe these tripdoublet states as a heavy mixture of doublets and quartets, e.g.:

$$\Psi(^2T_1)^{U-TDDFT} \approx -\sqrt{\frac{1}{3}}\Psi(^2T_1)^{X-TDDFT} + \sqrt{\frac{2}{3}}\Psi(^4T_1, M_S = 1/2)^{X-TDDFT}, \quad (5)$$

As follows from Eqs 2–4, U-TDDFT thus systematically underestimates the excitation energies of the 2T_1 and 2T_2 states, since the energies of quartets are in general lower than the corresponding tripdoublets, as discussed in the Introduction. In Section 3.3 we will also see that part of the underestimation is due to the failure of U-TDDFT to reproduce the energy degeneracy of $\Psi(^4T_1, M_S = 1/2)$ and $\Psi(^4T_1, M_S = 3/2)$. On the other hand, X-TDDFT avoids spin contamination through implicitly incorporating extra double excitations necessary for spin-adapting the tripdoublet states (Eq. 3), and therefore performs systematically better than U-TDDFT for all the functionals studied herein. The improvements of the excitation energies (~ 0.05 eV) may seem small, but have profound influences on the magnitude and even the sign of the 2T_1 – 4T_1 gap, and therefore on the ratio of fluorescence and phosphorescence emission, as will be detailed in Section 3.3. Finally, X-TDDFT also improves upon the U-TDDFT excitation energies of the d-d excited states, despite that the latter are almost free from spin contamination. Already from the calculated absorption energies, one can draw some conclusions about the photophysical processes of CuP. The vertical absorption energy differences between the 2T_1 and 2S_1 states, as well as between the 2S_2 , 2CT_1 and 2CT_2 states, are very small (0.1–0.2 eV). Therefore, once CuP is excited to the bright 2S_1 state by visible light, the molecule is expected to undergo a cascade of ultrafast IC processes, all the way till the doublet state, 2T_1 . While the fate of the 2S_2 state is not immediately obvious from the excitation energies, it is well-known that the S_2 states of porphyrin complexes typically undergo ultrafast IC to the S_1 state on the hundred fs to 1 ps timescale (Bräm et al., 2019), suggesting that the 2S_2 state of CuP will also quickly relax to the 2T_1 state via 2S_1 , even without the presence of 2CT_1 ; the presence of 2CT_1 creates a further intermediate energy level between 2S_1 and 2S_2 and is expected to accelerate the IC further. The availability of an ultrafast IC cascade also means the ISC from these high-lying excited states are probably unimportant, especially considering that copper is a relatively light element. These findings are in qualitative agreement with the experimental observation that the 2S_2 states of substituted copper(II) porphyrins relax to the 2T_1 states in gas phase through a two-step process via the intermediacy of a CT state, with time constants 65 fs and 350–2,000 fs, respectively, depending on the substituents (Ha-Thi et al., 2013); the fact that the CT $\rightarrow ^2T_1$ IC is slower than the $^2S_2 \rightarrow$ CT IC is consistent with the trend of our calculated energy gaps. In solution, the 2S_1 state of Cu(II) protoporphyrin IX dimethyl ester was known to relax to 2T_1 within 8 ps (Kobayashi et al., 2008), and for CuTPP as well as CuOEP the same relaxation was also found to occur within the picosecond timescale (Magde et al., 1974). Recently, the decay rates of the 2S_1 state were measured as 50 fs and 80 fs for CuTPP and CuOEP, respectively, in cyclohexane (Bräm et al., 2019). The 2S_1 state lifetime of CuP itself was also estimated, although

indirectly from the natural width of the 0–0 peak of the Q band, as 30 fs (Noort et al., 1976). Quantitative computation of these IC rates is however beyond the scope of the paper, as the narrow energy gaps and possible involvement of conical intersections probably necessitate nonadiabatic molecular dynamics simulations. Nevertheless, a $^2S_2 \rightarrow ^2CT_1 \rightarrow ^2S_1/^2T_2 \rightarrow ^2T_1$ IC pathway can still be tentatively proposed based on the energy ordering alone. Besides, there are also three d-d excited states with energies slightly lower (within 0.34 eV) than the 2T_1 state. There is thus the possibility that the 2T_1 state may further relax to either one of these three d-d states, which we will return to later. Finally, it is worth noting that the use of the accurate SDSPT2 method, as opposed to TDDFT, is crucial for obtaining a reliable estimate of the qualitative trend of the excited state energies. BP86 predicts that the CT states lie below the 2T states, leading to a qualitatively wrong IC pathway; ω B97X, on the other hand, grossly overestimates the energy of 2S_2 and would underestimate its tendency to undergo IC to the CT states (Figure 3). While B3LYP and PBE0 predict reasonable excited state orderings, they still underestimate the energies of the CT states due to the presence of double excitation contributions, which cannot be correctly described under the adiabatic TDDFT framework.

3.2 Analysis of the equilibrium geometry of the 2T_1 state

Since all higher lying excited states are predicted to convert to 2T_1 over a short timescale, to study the luminescence of CuP [and probably also other Cu(II) porphyrin complexes bearing alkyl or aryl substituents, given that these substituents do not change excitation energies drastically (Eastwood and Gouterman, 1969)], it should suffice to study the radiative and non-radiative processes starting from the 2T_1 state. Unlike the three 2dd states, whose Franck-Condon emissions are symmetry forbidden, the emission from the 2T_1 state is symmetry allowed, and existing experimental results unequivocally point to a ligand state luminescence, rather than luminescence from a d-d excited state. Therefore, accurately predicting the equilibrium geometry of the 2T_1 state is crucial for subsequent studies.

Some selected bond lengths for the optimized ground state and excited state structures are provided in Table 2. The difference in ground state bond lengths between the UKS and ROKS methods is extremely small (< 0.0001 Å), as can be seen from their root mean square deviation (RMSD), which can be attributed to the extremely small UKS spin contamination of the ground state of CuP ($\langle S^2 \rangle_{PBE0} = 0.7532$). The doubly degenerate 2T_1 state, which belongs to the doubly degenerate E_u irreducible representation (irrep) under the D_{4h} group, undergoes Jahn-Teller distortion to give a D_{2h} structure, where two of the opposing Cu-N bonds are elongated but the corresponding pyrrole rings remain almost intact, while the other two Cu-N bonds are almost unchanged but the corresponding pyrrole rings exhibit noticeable deformation. The U-TDDFT and X-TDDFT bond lengths of the 2T_1 state show larger deviations than the UKS and ROKS ground state ones, with the largest deviation exceeding 0.001 Å [the C-C (mn) bond], which is also reflected in the RMSD values. However, the structure

TABLE 2 The equilibrium bond lengths (in Å) of the ground state (2S_0) and the first doublet excited state (2T_1) of the CuP molecule.

State	Cu-N	C-N	C-C (mn) ^a	C-C (mt) ^a	C-C (st) ^a
UKS					
2S_0	2.0148	1.3652	1.3881	1.4410	1.3611
U-TDDFT					
2T_1	2.0190	1.3683	1.4155	1.4145	1.3885
	2.0416	1.3674	1.3842	1.4447	1.3584
ROKS					
2S_0	2.0148	1.3652	1.3881	1.4410	1.3612
X-TDDFT					
2T_1	2.0198	1.3681	1.4152	1.4149	1.3886
	2.0413	1.3668	1.3853	1.4442	1.3590
RMSD ^b	0.00002				
RMSD ^c	0.00102				

^aSee Figure 2 for the labeling of atoms.

^bThe RMSD (Å) between the optimized 2S_0 state structures obtained using UKS and ROKS.

^cThe RMSD (Å) between the optimized 2T_1 state structures obtained using U-TDDFT and X-TDDFT.

differences are still small on an absolute scale. This suggests that the coupling of the unpaired Cu(II) d electron and the porphyrin triplet is weak, so that a reasonable tripdoublet state geometry is obtained even if this coupling is described qualitatively incorrectly (as in U-TDDFT). By contrast, our previous benchmark studies on small molecules (where the coupling between unpaired electrons is much larger) revealed that X-TDDFT improves the U-TDDFT bond lengths by 0.01–0.05 Å on average, depending on the functional and the molecule (Li and Liu, 2011; Wang et al., 2020).

3.3 Relaxation processes of the 2T_1 state

As revealed by the above analyses, the relaxation process from high-lying excited states to the 2T_1 state is rapid, and the only energetically accessible relaxation pathways are the radiative (fluorescence) and non-radiative (IC) relaxations from 2T_1 to the ground state 2S_0 , the IC from 2T_1 to one of the lower 2dd states, as well as the ISC from 2T_1 to 4T_1 . The 4T_1 state can furthermore convert back to the 2T_1 state through reverse ISC (RISC), or relax to the ground state via radiative (phosphorescence) or non-radiative (ISC) pathways (Figure 4). While the RISC from 4T_1 to the 2dd states is also possible, they are expected to be much slower than the spin-allowed $^2T_1 \rightarrow ^2dd$ IC processes, and are therefore neglected in the present study.

Before we discuss the quantitative values of transition rates, we first analyze the relevant electronic states from the viewpoint of point group symmetry. The equilibrium structures of the 2T_1 and 4T_1 states are both distorted owing to the Jahn-Teller effect, and possess only D_{2h} symmetry, compared to the D_{4h} symmetry of the ground state equilibrium structure. The implications are two-fold: the double degeneracy of the nT_1 ($n = 2$ or 4) state at the D_{4h} geometry (where they both belong to the E_u irrep) is lifted to

give two adiabatic states, hereafter termed the $^nT_1(1)$ and $^nT_1(2)$ states, respectively, where $^nT_1(1)$ is the state with the lower energy; and the potential energy surface of the $^nT_1(1)$ state has two chemically equivalent D_{2h} minima, $^nT_1(1)(X)$ and $^nT_1(1)(Y)$, where different pairs of Cu-N bonds are lengthened and shortened (see the schematic depictions in Figure 4). Although $^nT_1(1)(X)$ and $^nT_1(1)(Y)$ are on the same adiabatic potential energy surface, their electronic wavefunctions represent different diabatic states, as they belong to the B_{3u} and B_{2u} irreps, respectively. The $^nT_1(1)(X)$ structure is diabatically connected to $^nT_1(2)(Y)$ (i.e., the $^nT_1(2)$ state at the equilibrium structure of $^nT_1(1)(Y)$) via a D_{4h} conical intersection, while $^nT_1(1)(Y)$ is diabatically connected to $^nT_1(2)(X)$ via the same conical intersection. Thus, the $^nT_1(2)(X)$ and $^nT_1(2)(Y)$ states are expected to undergo ultrafast IC from the D_{4h} conical intersection, to give the $^nT_1(1)(Y)$ and $^nT_1(1)(X)$ states as the main products, respectively. The direct transition from $^nT_1(2)$ to states other than $^nT_1(1)$ can therefore be neglected. From the irreps of the electronic states, we conclude that certain ISC transitions are forbidden by spatial symmetry. These include the transitions between $^2T_1(1)(X)$ and $^4T_1(1)(X)$, between $^2T_1(1)(Y)$ and $^4T_1(1)(Y)$, and between any one of the $^4T_1(1)$ structures and 2S_0 . All IC and radiative transitions, plus the ISC transitions between $^2T_1(1)(X)$ and $^4T_1(1)(Y)$ as well as between $^2T_1(1)(Y)$ and $^4T_1(1)(X)$, are symmetry allowed. While symmetry forbidden ISC processes can still gain non-zero rates from the Herzberg-Teller effect, we deem that the rates are not large enough to have any noticeable consequences. On one hand, the two symmetry forbidden ISC pathways between the $^2T_1(1)$ and $^4T_1(1)$ states are overshadowed by the two symmetry allowed ones, so that the total ISC rate between $^2T_1(1)$ and $^4T_1(1)$ is undoubtedly determined by the latter alone. The ISC from $^4T_1(1)$ to 2S_0 , on the other hand, has to compete with the IC process from $^2T_1(1)$ to 2S_0 in order to affect the quantum yield or the dominant relaxation pathway of the system noticeably, but the latter process is both spin-

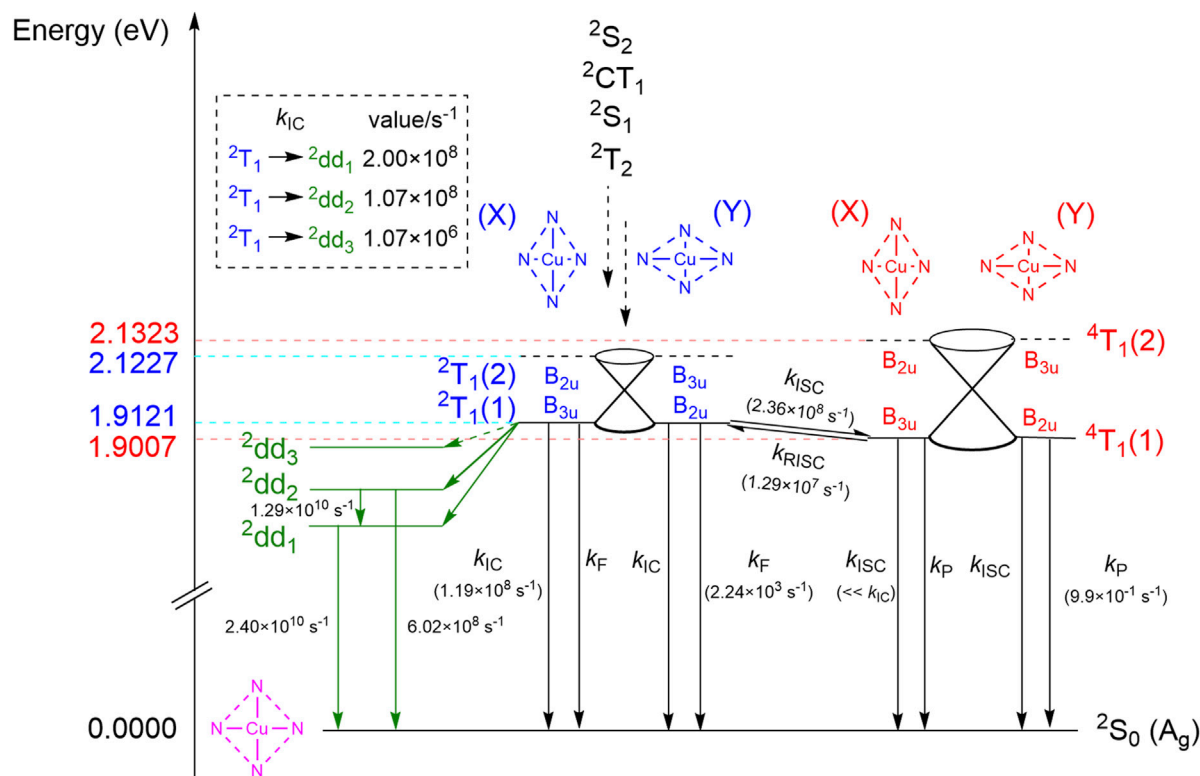


FIGURE 4

Radiative and non-radiative relaxation pathways of the 2T_1 state. Both the 2T_1 and 4T_1 states are splitted by the Jahn-Teller effect to give two adiabatic states, labeled (1) and (2). Each of the (1) states have two equivalent D_{2h} equilibrium structures, labeled (X) and (Y). The (2) states do not have equilibrium structures and are connected with the corresponding (1) states via conical intersections. The X-TDA(doublet)/U-TDA(quartet) adiabatic excitation energies of the (1) states, as well as the energies of the (2) states at the equilibrium geometries of their corresponding (1) states, are shown on the left. The transition rates are calculated at 83 K in the gas phase. The forward and reverse ISC rates between $^2T_1(1)(X)$ and $^4T_1(1)(Y)$ are equal to those between $^2T_1(1)(Y)$ and $^4T_1(1)(X)$ by symmetry, but the former ISC processes are omitted for clarity. Transition rates that are obviously equal by symmetry reasons are shown only once. Transitions out of the 2dd_3 state are not shown since the quantum yield of the 2dd_3 state is expected to be negligible based on the small $^2T_1 \rightarrow ^2dd_3$ IC rate.

allowed and spatial symmetry-allowed, while the former is forbidden in both aspects. We therefore neglect all ISC rates whose Franck-Condon contributions are zero by spatial symmetry.

We then calculated the rate constants for all transitions between 2T_1 , 4T_1 and 2S_0 whose rates are non-negligible, by the TVCF method. The rates (Figure 4) were calculated at 83 K, the temperature used in the quantum yield studies of Eastwood and Gouterman (1969); the latter studies gave a luminescence quantum yield of 0.09, in a solvent mixture of diethyl ether, isopentane, dimethylformamide and ethanol. The accurate treatment of solvation effects is however complicated and beyond the scope of the paper, so that all transition rates were computed in the gas phase. Our calculated k_{ISC} from 2T_1 to 4T_1 ($2.36 \times 10^8 \text{ s}^{-1}$) is slightly smaller than the total IC rate from 2T_1 to 2S_0 as well as to the three 2dd states ($4.27 \times 10^8 \text{ s}^{-1}$), suggesting that treating the 2T_1 and 4T_1 states as a rapid equilibrium [as in, e.g., Ake and Gouterman (1969) and Bohandy and Kim (1980)] is not justified at least in the gas phase. The IC from 2T_1 to 2dd_3 is two orders of magnitudes slower than other non-radiative relaxation pathways of 2T_1 , and is therefore not considered viable in the remaining discussions. At 83 K, the RISC from 4T_1 to 2T_1 is 11% of the forward ISC rate. Both rates are in favorable agreement with the experimental values of Cu(II) protoporphyrin IX dimethyl ester in benzene, $k_{ISC} = 1.6 \times$

10^9 s^{-1} and $k_{RISC} = 5.6 \times 10^8 \text{ s}^{-1}$, at room temperature (Kobayashi et al., 2008). Our computed ISC and RISC rates give a 2T_1 -to- 4T_1 equilibrium concentration ratio of 1:9.0 when all IC processes are neglected, but our kinetic simulation shows that the steady state concentration ratio is 1:50.7 when the latter are considered, further illustrating that treating the 2T_1 - 4T_1 interconversion as a fast equilibrium can lead to noticeable error. Nevertheless, the fluorescence rate of 2T_1 still exceeds the phosphorescence rate of 4T_1 by three orders of magnitude, which more than compensates for the low steady state concentration of 2T_1 . Similar conclusions could be derived from the rates reported in Ake and Gouterman (1969) ($3.6 \times 10^3 \text{ s}^{-1}$ and $8.3 \times 10^{-1} \text{ s}^{-1}$, respectively), calculated from semiempirical exchange and SOC integrals and experimental absorption oscillator strengths, which agree surprisingly well with the rates that we obtained here. Kinetic simulation suggests that 99.2% of the total luminescence at this temperature is contributed by fluorescence, and only 0.8% is due to phosphorescence. This can be compared with the experimental finding by Bohandy and Kim (1980) that the phosphorescence of CuP at 86 K is observable as a minor 0-0 peak besides the 0-0 fluorescence peak, with a fluorescence to phosphorescence ratio of about 5:1 to 10:1 [as estimated from Figure 5 of Bohandy and Kim (1980)]; however note that this study was performed in a triphenylene solid matrix.

The total luminescence quantum yield is predicted by our kinetic simulations to be 5.3×10^{-6} , four orders of magnitude smaller than the experimental quantum yield (0.09) in solution. We believe one possible reason is that the 2T_1 – 4T_1 gap of CuP is larger in solution than in the gas phase. This can already be seen from the experimental 2T_1 – 4T_1 0–0 gaps of CuP in solid matrices with different polarities: the 0–0 gap was measured in polymethylmethacrylate as 500 cm^{-1} (Smith and Gouterman, 1968), but 310 – 320 cm^{-1} in *n*-octane (Noort et al., 1976) and 267 cm^{-1} in triphenylene (Bohandy and Kim, 1980). Therefore, the 0–0 gap in the gas phase is probably smaller than 267 cm^{-1} , and indeed, our X-TDA calculations predict an adiabatic 2T_1 – 4T_1 gap of 92 cm^{-1} in the gas phase. The larger 2T_1 – 4T_1 gap in solution compared to the gas phase is expected to introduce a Boltzmann factor of $\exp(-(E_{\text{sol}} - E_{\text{gas}})/RT)$ to k_{RISC} , while changing the other rates negligibly. Setting $E_{\text{sol}} = 267\text{ cm}^{-1}$ and $E_{\text{gas}} = 92\text{ cm}^{-1}$, we obtain a solution phase k_{RISC} of $9.02 \times 10^5\text{ s}^{-1}$, from which kinetic simulations give a fluorescence-phosphorescence ratio of 8.6:1, in quantitative agreement with experiment (Bohandy and Kim, 1980). Setting $E_{\text{sol}} = 500\text{ cm}^{-1}$ [as appropriate for the polar solvent used in Eastwood and Gouterman (1969)] gives $k_{\text{RISC}} = 1.60 \times 10^4\text{ s}^{-1}$, and a total luminescence quantum yield of 4.0×10^{-5} , with 13% contribution from fluorescence and 87% from phosphorescence. The remaining discrepancy ($\sim \times 2200$) of the experimental and calculated quantum yields can be attributed to the restriction of the molecular vibrations of CuP by the low temperature (and thus viscous) solvent, which is expected to suppress the IC process significantly.

Interestingly, U-TDA completely fails to reproduce the qualitative picture of Figure 4 and predicts a 2T_1 – 4T_1 adiabatic gap of the wrong sign (-276 cm^{-1}), violating Hund's rule. At first sight, this may seem surprising: since the U-TDA “tripdoublet state” is a mixture of the true tripdoublet state and the quartet state, the U-TDA 2T_1 energy should lie in between the energies of the true 2T_1 state and the 4T_1 state, which means that the U-TDA 2T_1 – 4T_1 gap should be smaller than the X-TDA gap but still have the correct sign. However, the U-TDA 2T_1 state is contaminated by the $M_S = 1/2$ component of the 4T_1 state (Eq. 5), while a spin flip-up U-TDA calculation of the 4T_1 state gives its $M_S = 3/2$ component. The two spin components obviously have the same energy in the exact non-relativistic theory and in all rigorous spin-adapted methods, but not in U-TDA, even when the ground state is not spin-contaminated (Li and Liu, 2012; Li and Liu, 2016b). This shows that the restoration of the degeneracy of spin multiplets by the random phase approximation (RPA) correction in X-TDA (Li and Liu, 2011) indeed leads to qualitative improvement of the excitation energies, instead of being merely a solution to a conceptual problem. It also shows that estimating the tripdoublet energy by extrapolating from the energies of the spin-contaminated tripdoublet and the quartet by, e.g., the Yamaguchi method (Soda et al., 2000) does not necessarily give a qualitatively correct estimate of the spin-pure tripdoublet energy. The inverted doublet-quartet gap introduces qualitative defects to the computed photophysics of CuP. Already when the doublet-quartet gap is zero, the Boltzmann factor is expected to raise the k_{RISC} to $9.24 \times 10^7\text{ s}^{-1}$, reducing the ratio of phosphorescence in the total luminescence to 0.11%. Further raising the quartet to reproduce the U-TDA doublet-quartet gap will reduce the k_{ISC} to $1.42 \times 10^6\text{ s}^{-1}$, which reduces the ratio of phosphorescence to 0.0007%. These values are obviously in much worse agreement with the experiments (Bohandy and Kim, 1980).

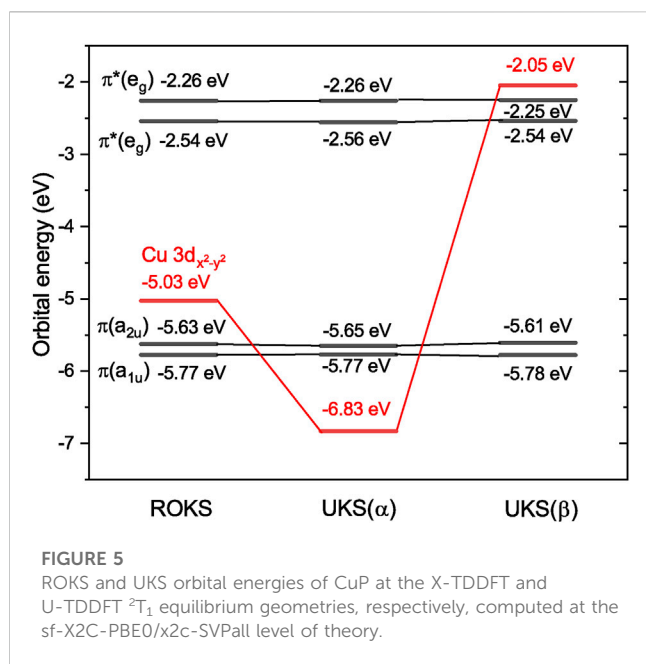
Finally, we briefly comment on the luminescence lifetimes. The luminescence of CuP is known to decay non-exponentially (Smith and Gouterman, 1968), so its luminescence lifetime can only be approximately determined. The luminescence lifetime of CuP has been determined as $400\text{ }\mu\text{s}$ (Smith and Gouterman, 1968) at 80 K in polymethylmethacrylate, and a biexponential decay with lifetimes 155 and $750\text{ }\mu\text{s}$ was reported (Eastwood and Gouterman, 1969) at 78 K in methylphthalylethylglycolate. The same references also reported that the luminescence lifetimes of CuOEP and CuTPP are also within the 50 – $800\text{ }\mu\text{s}$ range. However, in a room temperature toluene solution the luminescence lifetimes of CuOEP and CuTPP were reported to be 115 and 30 ns, respectively (Liu et al., 1995), and a few nanoseconds in the gas phase (Ha-Thi et al., 2013). If we define the luminescence lifetime as the time needed for $1 - 1/e \approx 63.2\%$ of the luminescence to be emitted, then kinetic simulations from our X-TDDFT rate constants give a gas-phase luminescence lifetime of 5.7 ns at 83 K, which is much shorter than the low-temperature condensed phase results, but in reasonable agreement with the room-temperature solution phase and especially the gas-phase experimental results. However, using the ISC and RISC rate constants consistent with the U-TDA doublet-quartet gap, one obtains a lifetime of 2.4 ns, differing noticeably from the X-TDDFT result. Thus, our results suggest that X-TDDFT/X-TDA gives non-negligible corrections upon the luminescence lifetime of U-TDDFT/U-TDA, and also confirm that the discrepancy of the experimental and calculated quantum yields is probably due to suppression of the IC of 2T_1 by the low temperature solvent.

3.4 Discussions

As mentioned in the Introduction, the simple orbital energy difference model based on a restricted open-shell determinant (Figure 1) predicts that the excitation energy of the lowest tripdoublet of any doublet molecule is at least the sum of the excitation energies of the first two excited states (as long as the ROKS ground state satisfies the *aufbau* rule). It therefore comes as a surprise that the lowest tripdoublet state of CuP (2T_1) is only barely higher than the lowest excited state (2dd_1) by 0.34 eV (Table 1), even though the ROKS ground state of CuP is indeed an *aufbau* state (Figure 5). This suggests a failure of the ROKS orbital energy difference model.

To understand why the ROKS orbital energies fail qualitatively for describing the excited state ordering of CuP, despite that the X-TDDFT method (which uses the ROKS determinant as the reference state) still gives reasonable excitation energies as compared to SDSPT2, we note that the α and β Fock matrices of an ROKS calculation are in general not diagonal under the canonical molecular orbital (CMO) basis. Only the unified coupling operator \mathbf{R} , assembled from blocks of the CMO Fock matrices,

$$\mathbf{R} = \begin{pmatrix} \frac{1}{2}(\mathbf{F}_{\text{CC}\alpha} + \mathbf{F}_{\text{CC}\beta}) & \mathbf{F}_{\text{CO}\beta} & \frac{1}{2}(\mathbf{F}_{\text{CV}\alpha} + \mathbf{F}_{\text{CV}\beta}) \\ \mathbf{F}_{\text{OC}\beta} & \frac{1}{2}(\mathbf{F}_{\text{OO}\alpha} + \mathbf{F}_{\text{OO}\beta}) & \mathbf{F}_{\text{OV}\alpha} \\ \frac{1}{2}(\mathbf{F}_{\text{VC}\alpha} + \mathbf{F}_{\text{VC}\beta}) & \mathbf{F}_{\text{VO}\alpha} & \frac{1}{2}(\mathbf{F}_{\text{VV}\alpha} + \mathbf{F}_{\text{VV}\beta}) \end{pmatrix}, \quad (6)$$



Is diagonal (Hirao and Nakatsuji, 1973). Note that herein we have used the Guest-Saunders parameterization (Guest and Saunders, 1974) of the diagonal blocks of \mathbf{R} , which is the default choice of the BDF program, although our qualitative conclusions are unaffected by choosing other parameterizations. However, the leading term of the X-TDDFT calculation is not simply given by the eigenvalue differences of \mathbf{R} ,

$$\Delta'_{i\alpha\sigma,j\beta\tau} = \delta_{\sigma\tau}\delta_{ij}\delta_{ab}(R_{aa} - R_{ii}), \quad (7)$$

but rather from the α and β Fock matrices themselves via

$$\Delta_{i\alpha\sigma,j\beta\tau} = \delta_{\sigma\tau}(\delta_{ij}F_{ab\sigma} - \delta_{ab}F_{ji\sigma}). \quad (8)$$

Here i, a represent occupied CMOs, j, b virtual CMOs, and σ, τ spin indices. For the diagonal matrix element of an arbitrary single excitation, Eqs 7, 8 differ by the following term:

$$\Delta_{i\alpha\sigma,i\alpha\sigma} - \Delta'_{i\alpha\sigma,i\alpha\sigma} = \frac{1}{2}((F_{aaa} - F_{aaa'}) - (F_{iis} - F_{iis'})), \quad (9)$$

where σ' is the opposite spin of σ . For a general hybrid functional, the Fock matrix element differences in Eq. 9 are given by (where p is an arbitrary CMO, c_x is the proportion of HF exchange, and v^{xc} is the XC potential)

$$F_{pp\beta} - F_{pp\alpha} = c_x(pt|pt) + (v_{pp\beta}^{xc} - v_{pp\alpha}^{xc}), \quad (10)$$

Assuming, for the sake of simplicity, that there is only one open-shell orbital t in the reference state. Assuming that the XC potential behaves similarly as the exact exchange potential, the difference Eq. 10 is positive, and should usually be the largest when $p = t$, while being small when p is spatially far from t . The corollary is that the orbital energy difference approximation Eq. 7 should agree well with the X-TDDFT leading term Eq. 8 for CV excitations (where the difference is proportional to the small exchange integral $(pt|pt)$), but underestimate the excitation energies of CO and OV excitations by a correction proportional to the large $(tt|tt)$ integral.

The underestimation of CO and OV excitation energies by ROKS orbital energy differences opens up the possibility of engineering a system to break the $\omega_{ia} = \omega_{it} + \omega_{ta}$ constraint inherent in the ROKS orbital energy difference model, and make the tripdoublet state the lowest excited state or only slightly higher than the lowest excited state. Possible approaches include:

1. Increase the difference Eq. 9 for the CO and OV states, while keeping it small for the lowest CV state, so that all CO and OV states are pushed above the lowest CV state. This is most easily done by making the open-shell orbital t very compact, which naturally leads to a larger $F_{tt\beta} - F_{tt\alpha}$ (due to a larger $(tt|tt)$) but a smaller $F_{pp\beta} - F_{pp\alpha}$ $p \in \{i, a\}$ (due to a small absolute overlap between the p and t orbitals).
2. Reduce the orbital energy gap between the highest doubly occupied orbital and the lowest unoccupied orbital, which also helps to reduce the excitation energy of the lowest CV state. However, a too small orbital energy gap will favor the IC of the tripdoublet to the ground state, which may quench the fluorescence of the tripdoublet. As already mentioned in Section 3.3, the IC rate of CuP is already large enough to make CuP only barely fluorescent (quantum yield $< 10^{-5}$) in the gas phase, and a viscous solvent seems to be required to suppress the IC contribution and make the fluorescence stronger.

Now, it becomes evident that CuP fits the above design principles very well. The unpaired electron in the ground state of CuP is on the Cu $3d_{x^2-y^2}$ orbital (Figure 6), which is spatially localized. Moreover, the Cu $3d_{x^2-y^2}$ orbital occupies a different part of the molecule than the ligand π and π^* orbitals, which results in a small absolute overlap between the orbitals and helps to reduce the effect of Eq. 9 on the CV excitation energies. To quantitatively assess the effect of Eq. 9 on the CO and OV excitation energies, we note that the X-TDDFT leading term Eq. 8 is nothing but the UKS orbital energy difference, if the shape differences of the UKS and ROKS orbitals are neglected. Therefore, we have plotted the UKS orbital energies of CuP in Figure 5 as well. Intriguingly, the α Cu $3d_{x^2-y^2}$ orbital now lies below the porphyrin $\pi(a_{1u})$ and $\pi(a_{2u})$ orbitals, while the β Cu $3d_{x^2-y^2}$ orbital lies above the porphyrin $\pi^*(e_g)$ orbitals. Therefore, the differences of UKS orbital energies predict that the lowest excited states of CuP are the CV states obtained from exciting an electron from $\pi(a_{1u})$ and $\pi(a_{2u})$ to $\pi^*(e_g)$. This is not only consistent with our U-TD-PBE0 excitation energies, but also the X-TD-PBE0 results (save for the 2S_2 state, which is higher than 2CT_1 computed at the respective levels of theory), despite that the latter method is spin-adapted. Note also that although the $(tt|tt)$ integral leads to a huge splitting between the α and β Cu $3d_{x^2-y^2}$ orbitals, the splitting is only barely enough for the UKS orbital energy differences to predict a tripdoublet first excited state: if the β Cu $3d_{x^2-y^2}$ orbital were just 0.2 eV lower, one would predict that the CO-type CT excitation $\pi(a_{2u}) \rightarrow$ Cu $3d_{x^2-y^2}$ is lower than the lowest tripdoublet $\pi(a_{2u}) \rightarrow \pi^*(e_g)$. Alternatively, one may say that the HOMO-LUMO gap of the porphyrin ligand is barely narrow enough to fit within the energy window between the α and β Cu $3d_{x^2-y^2}$ orbitals, which clearly illustrates the importance of using a narrow-gap ligand for designing systems with a low-lying tripdoublet excited state.

Besides, we note that our SDSPT2 vertical absorption energy results suggest that three d-d excited states, 2dd_1 , 2dd_2 and 2dd_3 , lie

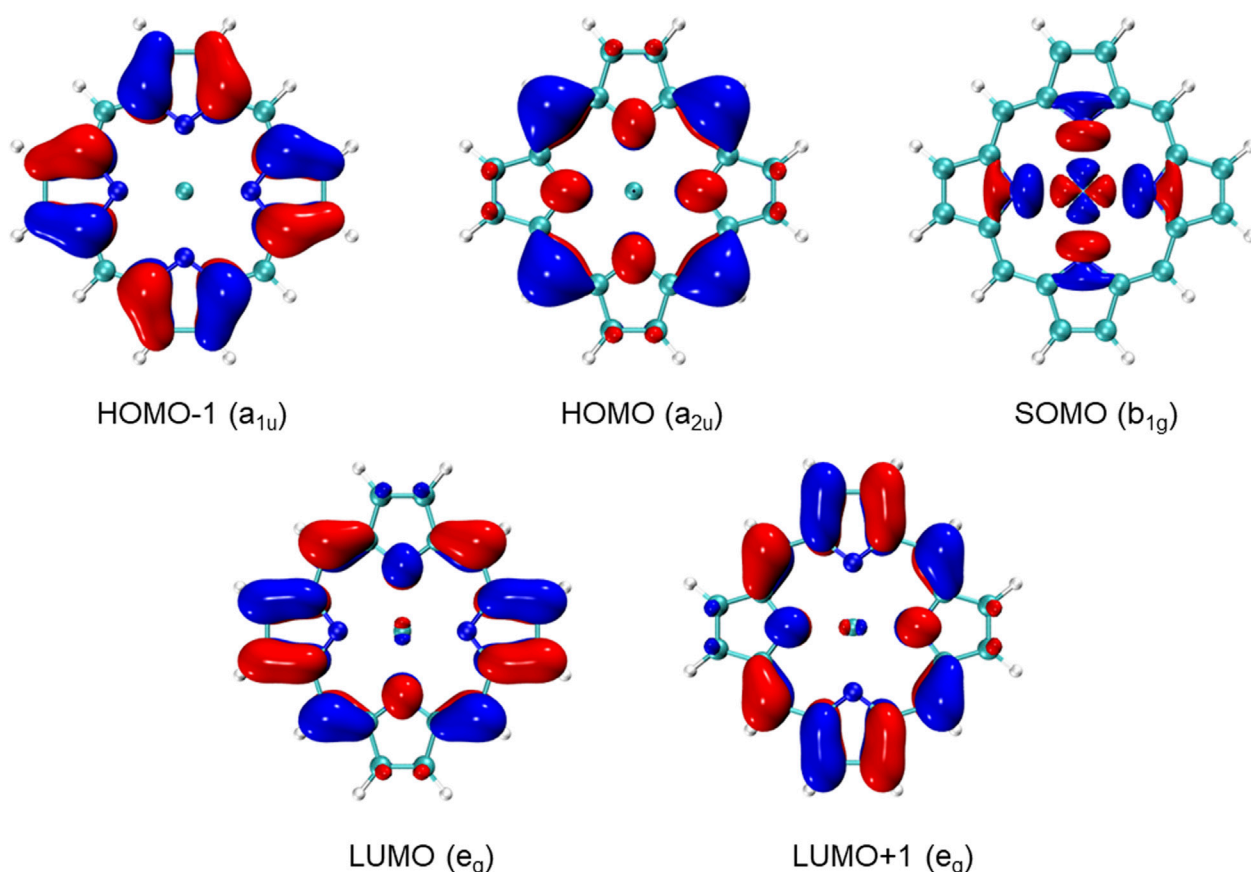


FIGURE 6

ROKS frontier molecular orbitals of CuP, computed at the sf-X2C-PBE0/x2c-SVPall level of theory.

barely below the 2T_1 state, although the actual energy ordering of 2dd_3 and 2T_1 cannot be said with full certainty. This means that CuP undergoes anti-Kasha fluorescence, which (as mentioned in the Introduction) is difficult to achieve especially for open-shell molecules. The 2dd states are very close in energy to the 2T_1 state, strongly favoring an ultrafast IC process, while the 2dd states themselves relax radiationlessly to the ground state on the hundred picosecond timescale, mainly through the cascade $^2dd_2 \rightarrow ^2dd_1 \rightarrow ^2S_0$ (Figure 4). In reality, however, our calculations suggest a $^2T_1 \rightarrow ^2dd_1$ IC rate of only $2.00 \times 10^8 s^{-1}$ (with other $^2T_1 \rightarrow ^2dd$ IC processes even slower), which although competitive with the $^2T_1 \rightarrow ^2S_0$ IC rate, is very small for an IC process with an adiabatic gap of only 0.42 eV (the estimated SDSPT2 adiabatic excitation energies of the 2dd_1 , 2dd_2 , 2dd_3 and 2T_1 being 1.66, 1.90, 2.06 and 2.08 eV, respectively). The reason is that the 2T_1 and 2dd states differ by a double excitation, while the nuclear derivative operator in the NACMEs is a one-electron operator, making the IC forbidden to first order. Moreover, to transform the 2T_1 state to any of the 2dd states, one has to perform an LMCT excitation as well as an MLCT excitation (from $\alpha \pi^*(e_g)$ to α Cu $3d_{x^2-y^2}$, and from β Cu $3d_{xy}/3d_{xz}/3d_{yz}/3d_{z^2}$ to $\beta \pi(a_{2u})$), both of which involve occupied/virtual orbital pairs with poor overlap. Therefore, the drastically different natures of the ligand-centered 2T_1 state and the metal-centered 2dd states effectively prevented the quenching of the luminescence by the 2dd

states, despite that the 2T_1 and 2dd states are extremely close in energy.

To conclude this section, we briefly note that even making the first doublet excited state a tripdoublet state still does not guarantee the realization of tripdoublet fluorescence. Two remaining potential obstacles are (1) the IC of the tripdoublet state to the ground state and (2) the ISC of the tripdoublet to the lowest quartet state (which is almost always lower than the lowest tripdoublet state owing to Hund's rule). Both can be inhibited by making the molecule rigid, which is indeed satisfied by the porphyrin ligand in CuP. Alternatively, if the ISC from the first quartet state to the ground state is slow (as is the case of CuP, thanks to the spatial symmetry selection rules), and the gap between the first doublet and the first quartet is comparable to the thermal energy kT at the current temperature, then the quartet state can undergo RISC to regenerate the tripdoublet state, which can then fluoresce. This is well-known as the thermally activated delayed fluorescence (TADF) mechanism (Parker and Hatchard, 1961; Endo et al., 2011; Yang et al., 2017), although existing TADF molecules typically fluoresce from singlets and use a triplet "reservoir state" to achieve delayed fluorescence. In order for the TADF mechanism to outcompete the phosphorescence from the first quartet state, both the phosphorescence rate and the doublet-quartet gap have to be small. While the low phosphorescence rate of CuP can be explained by the fact that copper is a relatively light element, the small 2T_1 – 4T_1 gap of CuP

can be attributed to the distributions of the frontier orbitals of CuP. Recall that the X-TDDFT gap between a tripdoublet excitation Eq. 3 and the associated quartet excitation Eq. 4 is exactly given by the X-RPA gap (Li and Liu, 2011), which is equal to $\frac{3}{2}((it|it) + (ta|ta))$. However, both of the two integrals are small for the 2T_1 and 4T_1 states of CuP, since the orbitals i and a reside on the ligand while t is localized near the metal atom (Figure 6). Such a clean spatial separation of the metal and ligand CMOs (despite the close proximity of the metal and the ligand) can further be attributed to the fact that the Cu $3d_{x^2-y^2}$ orbital has a different irrep than those of the ligand π and π^* orbitals, preventing the delocalization of the open-shell orbital to the π system of the porphyrin ligand; while the Cu $3d_{x^2-y^2}$ orbital can still delocalize through the σ bonds of the ligand, the delocalization is of limited extent due to the rather local electronic structures of typical σ bonds (Figure 6). Incidentally, the only other class of tripdoublet-fluorescing metalloporphyrins that we are aware of, i.e., vanadium(IV) oxo porphyrin complexes (Ake and Gouterman, 1969; Gouterman et al., 1970), are characterized by a single unpaired electron in the $3d_{xy}$ orbital, whose mixing with the ligand π and π^* orbitals is also hindered by symmetry mismatches. Whether this can be extended to a general strategy of designing molecules that fluoresce from tripdoublet states (or more generally, molecules that possess small doublet-quartet gaps) will be explored in the future. Finally, we briefly note that the design of doublet molecules with TADF and/or phosphorescence is also an interesting subject and deserves attention in its own right.

4 Conclusion

Fluorescence of open-shell molecules from tripdoublet states is a rare and underexplored phenomenon, for which traditional excited state methods such as U-TDDFT are unreliable due to severe spin contamination. In this work, we employed the high-precision method SDSPT2 to obtain accurate excitation energies of the CuP molecule, which suggests that the bright states obtained by light absorption relax to a low-lying doublet state, 2T_1 , via a cascade of ultrafast IC processes, in agreement with experiments. 2T_1 is a tripdoublet state composed of a triplet ligand state antiferromagnetically coupled with the unpaired electron of Cu(II), and contrary to predictions from ROKS orbital energy differences, the only excited states lower than 2T_1 are the d-d excited states 2dd_1 , 2dd_2 and 2dd_3 , and the nonradiative relaxations from 2T_1 to the 2dd states are slower than expected from their energy gaps, due to the large differences in excited state compositions. Using the SDSPT2 results as a benchmark, we found that the X-TDDFT method provides a more accurate description of the 2T_1 state (which exhibits considerable spin contamination) compared to U-TDDFT, while for the CO excitations, U-TDDFT and X-TDDFT show similar performance.

In addition to vertical absorption calculations and structural analyses, we conducted a detailed analysis of the relaxation rate constants of the excited states of CuP. Our results suggest that, in the gas phase and at low temperature (83 K), CuP emits fluorescence from the lowest tripdoublet state 2T_1 with a very small quantum yield ($\sim 10^{-5}$), and the contribution of phosphorescence is negligible. These results complement the experimental results in solution phase and solid matrix, which gave a lower but still greater than unity fluorescence-to-phosphorescence ratio and a much higher luminescence

quantum yield. The discrepancies are nicely explained by a solvent-dependent 2T_1 - 4T_1 gap and the viscosity of the solvent. Furthermore, we confirm the presence of an equilibrium between the first doublet state 2T_1 and the first quartet state 4T_1 , the latter of which functions as a reservoir of the 2T_1 state, although the steady state concentration ratio of these two states deviates noticeably from their equilibrium constant. CuP therefore represents an interesting example of a TADF molecule that emits fluorescence through a doublet-doublet transition, instead of the much more common singlet-singlet pathway. Notably, U-TDA predicts a doublet-quartet gap of the wrong sign, due to the spin contamination of the doublet state as well as the breaking of the spin multiplet degeneracy of the quartet state. Although the error is small (< 0.05 eV), it translates to a large error in the luminescence lifetime and (even more) the contribution of phosphorescence to the total luminescence. This again highlights the importance of using spin-adapted approaches in the study of open-shell systems, even when the excitation energy errors of unrestricted methods are small.

Based on the computational results, we proposed a few possible approaches that can be used to design new doublet molecules that fluoresce from tripdoublets: 1) keep the open-shell orbital of the molecule spatially compact, to open up a gap between the α and β UKS orbital energies of the open-shell orbital; 2) make the gap between the highest doubly occupied orbital and the lowest vacant orbital small enough so that both orbitals fit into the gap between the α and β open-shell orbitals, but not overly small as to encourage IC of the lowest tripdoublet state to the ground state; 3) make the molecule rigid to minimize unwanted non-radiative relaxation processes; 4) avoid introducing heavy elements in order to suppress unwanted ISC and phosphorescence processes; 5) localize the open-shell orbital and the frontier π/π^* orbitals onto different molecular fragments, and (if possible) make them belong to different irreps, to minimize the doublet-quartet gap; and 6) when the presence of doublet excited states below the tripdoublet is unavoidable, make those states differ significantly in composition from the tripdoublet state, to avoid the quenching of tripdoublet fluorescence by excited state-excited state IC. We hope that the present work will facilitate the discovery of novel molecules that fluoresce from tripdoublet states. Moreover, we expect that the success of the X-TDDFT and SDSPT2 methods will encourage the use of these two methods in the excited state studies of other systems.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Author contributions

XW: Data curation, Investigation, Writing—original draft, Software, Visualization. CW: Software, Funding acquisition, Methodology, Validation, Writing—review and editing. ZW: Methodology, Validation, Writing—review and editing, Conceptualization, Data curation, Investigation, Supervision,

Writing—original draft. WL: Conceptualization, Funding acquisition, Resources, Supervision, Writing—review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Grant Nos 21833001, 21973054, 22101155), Mountain Tai Climbing Program of Shandong Province, and Key-Area Research and Development Program of Guangdong Province (Grant No. 2020B0101350001). ZW gratefully acknowledges generous financial support by the Max Planck society.

Acknowledgments

The authors acknowledge the computational software provided by the Institute of Scientific Computing Software in Shandong

University and Qingdao BDF Software Technology Co., Ltd. The authors also acknowledge that an earlier version of this manuscript was submitted as a preprint on arXiv.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adamo, C., and Barone, V. (1999). Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* 110, 6158–6170. doi:10.1063/1.478522
- Ake, R. L., and Gouterman, M. (1969). Porphyrins XIV. Theory for the luminescent state in VO, Co, Cu complexes. *Theor. Chim. Acta* 15, 20–42. doi:10.1007/BF00526463
- Becke, A. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A At., Mol. Opt. Phys.* 38, 3098–3100. doi:10.1103/PhysRevA.38.3098
- Becke, A. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi:10.1063/1.464913
- Bohandy, J., and Kim, B. F. (1980). Temperature dependence of Mg porphyrin, Cu porphyrin, and Pd porphyrin luminescence. *J. Chem. Phys.* 73, 5477–5481. doi:10.1063/1.440093
- Bräm, O., Cannizzo, A., and Chergui, M. (2019). Ultrafast broadband fluorescence up-conversion study of the electronic relaxation of metalloporphyrins. *J. Phys. Chem. A* 123, 1461–1468. doi:10.1021/acs.jpca.9b00007
- Chai, J.-D., and Head-Gordon, M. (2008). Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* 10, 6615–6620. doi:10.1039/B810189B
- Eastwood, D., and Gouterman, M. (1969). Porphyrins: XII. luminescence of copper complexes at liquid nitrogen temperature. *J. Mol. Spectrosc.* 30, 437–458. doi:10.1016/0022-2852(69)90276-8
- Endo, A., Sato, K., Yoshimura, K., Kai, T., Kawada, A., Miyazaki, H., et al. (2011). Efficient up-conversion of triplet excitons into a singlet state and its application for organic light emitting diodes. *Appl. Phys. Lett.* 98, 083302. doi:10.1063/1.3558906
- Ernzerhof, M., and Scuseria, G. E. (1999). Assessment of the perdue–burke–ernzerhof exchange–correlation functional. *J. Chem. Phys.* 110, 5029–5036. doi:10.1063/1.478401
- Evans, D. F. (1957). 772. Magnetic perturbation of singlet–triplet transitions. Part II. *J. Chem. Soc.*, 3885–3888. doi:10.1039/JR9570003885
- Gao, S., Cui, Z., and Li, F. (2023). Doublet-emissive materials for organic light-emitting diodes: exciton formation and emission processes. *Chem. Soc. Rev.* 52, 2875–2885. doi:10.1039/D2CS00772J
- Gouterman, M. (1959). Study of the effects of substitution on the absorption spectra of porphyrin. *J. Chem. Phys.* 30, 1139–1161. doi:10.1063/1.1730148
- Gouterman, M., Mathies, R. A., Smith, B. E., and Caughey, W. S. (1970). Porphyrins. XIX. Triplet-doublet and quartet luminescence in Cu and VO complexes. *J. Chem. Phys.* 52, 3795–3802. doi:10.1063/1.1673560
- Green, S. A., Simpson, D. J., Zhou, G., Ho, P. S., and Blough, N. V. (1990). Intramolecular quenching of excited singlet states by stable nitroxyl radicals. *J. Am. Chem. Soc.* 112, 7337–7346. doi:10.1021/ja00176a038
- Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* 132, 154104. doi:10.1063/1.3382344
- Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. doi:10.1002/jcc.21759
- Guest, M., and Saunders, V. R. (1974). On methods for converging open-shell Hartree-Fock wave-functions. *Mol. Phys.* 28, 819–828. doi:10.1080/00268977400102171
- Guo, Y., Zhang, N., Lei, Y., and Liu, W. (2021). iCISCF: an iterative configuration interaction-based multiconfigurational self-consistent field theory for large active spaces. *J. Chem. Theory Comput.* 17, 7545–7561. doi:10.1021/acs.jctc.1c00781
- Ha-Thi, M.-H., Shafizadeh, N., Poisson, L., and Soep, B. (2013). An efficient indirect mechanism for the ultrafast intersystem crossing in copper porphyrins. *J. Phys. Chem. A* 117, 8111–8118. doi:10.1021/jp4008015
- He, C., Li, Z., Lei, Y., Zou, W., and Suo, B. (2019). Unraveling the emission mechanism of radical-based organic light-emitting diodes. *J. Phys. Chem. Lett.* 10, 574–580. doi:10.1021/acs.jpclett.8b03864
- Hirao, K., and Nakatsuji, H. (1973). General scf operator satisfying correct variational condition. *J. Chem. Phys.* 59, 1457–1462. doi:10.1063/1.1680203
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- Kobayashi, T., Huppert, D., Straub, K. D., and Rentzepis, P. M. (2008). Picosecond kinetics of copper and silver protoporphyrins. *J. Chem. Phys.* 70, 1720–1726. doi:10.1063/1.437689
- Lei, Y., Liu, W., and Hoffmann, M. R. (2017). Further development of SDSPT2 for strongly correlated electrons. *Mol. Phys.* 115, 2696–2707. doi:10.1080/00268976.2017.1308029
- Lei, Y., Suo, B., and Liu, W. (2021). iCAS: imposed automatic selection and localization of complete active spaces. *J. Chem. Theory Comput.* 17, 4846–4859. doi:10.1021/acs.jctc.1c00456
- Li, Z., and Liu, W. (2010). Spin-adapted open-shell random phase approximation and time-dependent density functional theory. I. Theory. *J. Chem. Phys.* 133, 064106. doi:10.1063/1.3463799
- Li, Z., and Liu, W. (2011). Spin-adapted open-shell time-dependent density functional theory. III. An even better and simpler formulation. *J. Chem. Phys.* 135, 194106. doi:10.1063/1.3660688
- Li, Z., and Liu, W. (2012). Theoretical and numerical assessments of spin-flip time-dependent density functional theory. *J. Chem. Phys.* 136, 024107. doi:10.1063/1.3676736
- Li, Z., and Liu, W. (2014). First-order nonadiabatic coupling matrix elements between excited states: a Lagrangian formulation at the CIS, RPA, TD-HF, and TD-DFT levels. *J. Chem. Phys.* 141, 014110. doi:10.1063/1.4885817
- Li, Z., and Liu, W. (2016a). Critical assessment of td-dft for excited states of open-shell systems: I. doublet-doublet transitions. *J. Chem. Theory Comput.* 12, 238–260. doi:10.1021/acs.jctc.5b01158
- Li, Z., and Liu, W. (2016b). Critical assessment of time-dependent density functional theory for excited states of open-shell systems: II. doublet-quartet transitions. *J. Chem. Theory Comput.* 12, 2517–2527. doi:10.1021/acs.jctc.5b01219

- Li, Z., Liu, W., Zhang, Y., and Suo, B. (2011). Spin-adapted open-shell time-dependent density functional theory. II. Theory and pilot application. *J. Chem. Phys.* 134, 134101. doi:10.1063/1.3573374
- Li, Z., Suo, B., and Liu, W. (2014a). First order nonadiabatic coupling matrix elements between excited states: implementation and application at the TD-DFT and pp-TDA levels. *J. Chem. Phys.* 141, 244105. doi:10.1063/1.4903986
- Li, Z., Suo, B., Zhang, Y., Xiao, Y., and Liu, W. (2013). Combining spin-adapted open-shell TD-DFT with spin-orbit coupling. *Mol. Phys.* 111, 3741–3755. doi:10.1080/00268976.2013.785611
- Li, Z., Xiao, Y., and Liu, W. (2014b). On the spin separation of algebraic two-component relativistic Hamiltonians: molecular properties. *J. Chem. Phys.* 141, 054111. doi:10.1063/1.4891567
- Liu, F., Cunningham, K. L., Uphues, W., Fink, G. W., Schmolt, J., and McMillin, D. R. (1995). Luminescence quenching of copper (ii) porphyrins with Lewis bases. *Inorg. Chem.* 34, 2015–2018. doi:10.1021/ic00112a011
- Liu, W., and Hoffmann, M. R. (2014). SDS: the 'static-dynamic-static' framework for strongly correlated electrons. *Theor. Chem. Acc.* 1, 12–1512. doi:10.1007/s00214-014-1481-x
- Liu, W., Hong, G., Dai, D., Li, L., and Dolg, M. (1997). The Beijing four-component density functional program package (BDF) and its application to EuO, EuS, YbO and YbS. *Theor. Chem. Acc.* 96, 75–83. doi:10.1007/s002140050207
- Liu, W., Wang, F., and Li, L. (2003). The Beijing density functional (BDF) program package: methodologies and applications. *J. Theor. Comput. Chem.* 2, 257–272. doi:10.1142/S0219633603000471
- Liu, W., Wang, F., and Li, L. (2004a). "Recent advances in relativistic density functional methods," in *Encyclopedia of computational chemistry*. Editors P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, and H. F. Schaefer (Chichester, UK: Wiley).
- Liu, W., Wang, F., and Li, L. (2004b). "Relativistic density functional theory: the BDF program package," in *Recent advances in relativistic molecular theory*. Editors K. Hirao and Y. Ishikawa (Singapore: World Scientific), 257–282.
- Lu, T., and Chen, F. (2012). Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592. doi:10.1002/jcc.22885
- Magde, D., Windsor, M. W., Holten, D., and Gouterman, M. (1974). Picosecond flash photolysis: transient absorption in Sn(IV), Pd(II), and Cu(II) porphyrins. *Chem. Phys. Lett.* 29, 183–188. doi:10.1016/0009-2614(74)85008-6
- Neese, F. (2012). The ORCA program system. *WIREs Comput. Mol. Sci.* 2, 73–78. doi:10.1002/wcms.81
- Neese, F. (2018). Software update: the ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.* 8, e1327. doi:10.1002/wcms.1327
- Neese, F. (2022). Software update: the ORCA program system—version 5.0. *WIREs Comput. Mol. Sci.* 12, e1606. doi:10.1002/wcms.1606
- Neese, F., Wennmohs, F., Becker, U., and Riplinger, C. (2020). The ORCA quantum chemistry program package. *J. Chem. Phys.* 152, 224108. doi:10.1063/5.0004608
- Niu, Y., Li, W., Peng, Q., Geng, H., Yi, Y., Wang, L., et al. (2020). Molecular materials property prediction package (MOMAP) 1.0: a software package for predicting the luminescent properties and mobility of organic functional materials. *Mol. Phys.* 116, 1078–1090. doi:10.1080/00268976.2017.1402966
- Niu, Y., Peng, Q., and Shuai, Z. (2008). Promoting-mode free formalism for excited state radiationless decay process with Duschinsky rotation effect. *Sci. China Chem. B* 51, 1153–1158. doi:10.1007/s11426-008-0130-4
- Noort, M., Jansen, G., Canters, G., and van der Waals, J. (1976). High resolution spectra of palladium, platinum and copper porphyrins in n-octane crystals. *Spectrochim. Acta A Mol. Spectrosc.* 32, 1371–1375. doi:10.1016/0584-8539(76)80183-3
- Parker, C. A., and Hatchard, C. G. (1961). Triplet-singlet emission in fluid solutions. phosphorescence of eosin. *Trans. Faraday Soc.* 57, 1894–1904. doi:10.1039/TF9615701894
- Peng, Q., Yi, Y., Shuai, Z., and Shao, J. (2007). Excited state radiationless decay process with Duschinsky rotation effect: formalism and implementation. *J. Chem. Phys.* 126, 114302. doi:10.1063/1.2710274
- Perdew, J. (1986a). Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B Condens. Matter Mater. Phys.* 33, 8822–8824. doi:10.1103/physrevb.33.8822
- Perdew, J. (1986b). Erratum: density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B Condens. Matter Mater. Phys.* 34, 7406. doi:10.1103/PhysRevB.34.7406
- Pollak, P., and Weigend, F. (2017). Segmented contracted error-consistent basis sets of double- and triple- ζ valence quality for one- and two-component relativistic all-electron calculations. *J. Chem. Theory Comput.* 13, 3696–3705. doi:10.1021/acs.jctc.7b00593
- Schmidt, R. (2006). Photosensitized generation of singlet oxygen. *Photochem. Photobiol.* 82, 1161–1177. doi:10.1562/2006-03-03-IR-833
- Smith, B., and Gouterman, M. (1968). Quartet luminescence from copper porphyrins. *Chem. Phys. Lett.* 2, 517–519. doi:10.1016/0009-2614(63)80001-9
- Soda, T., Kitagawa, Y., Onishi, T., Takano, Y., Shigeta, Y., Nagao, H., et al. (2000). *Ab initio* computations of effective exchange integrals for H–H, H–He–H and Mn2O2 complex: comparison of broken-symmetry approaches. *Chem. Phys. Lett.* 319, 223–230. doi:10.1016/S0009-2614(00)00166-4
- Song, Y., Guo, Y., Lei, Y., Zhang, N., Liu, W., and Qi, J. (2022). The static-dynamic-static family of methods for strongly correlated electrons: methodology and benchmarking. *Top. Curr. Chem.* 19, 181–187. doi:10.2174/1567202619666220614153209
- Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). *Ab initio* calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* 98, 11623–11627. doi:10.1021/j100096a001
- Varnes, A. W., Dodson, R. B., and Wehry, E. (1972). Interactions of transition-metal ions with photoexcited states of flavines. fluorescence quenching studies. *J. Am. Chem. Soc.* 94, 946–950. doi:10.1021/ja00758a037
- Wang, Z., Li, Z., Zhang, Y., and Liu, W. (2020). Analytic energy gradients of spin-adapted open-shell time-dependent density functional theory. *J. Chem. Phys.* 153, 164109. doi:10.1063/5.0025428
- Wang, Z., Wu, C., and Liu, W. (2021). NAC-TDDFT: time-dependent density functional theory for nonadiabatic couplings. *Acc. Chem. Res.* 54, 3288–3297. doi:10.1021/acs.accounts.1c00312
- Yan, X., and Holten, D. (1988). Effects of temperature and solvent on excited-state deactivation of copper (II) octaethyl- and tetraphenylporphyrin: relaxation via a ring-to-metal charge-transfer excited state. *J. Phys. Chem.* 92, 5982–5986. doi:10.1021/j100332a029
- Yang, Z., Mao, Z., Xie, Z., Zhang, Y., Liu, S., Zhao, J., et al. (2017). Recent advances in organic thermally activated delayed fluorescence materials. *Chem. Soc. Rev.* 46, 915–1016. doi:10.1039/C6CS00368K
- Zhang, Y., Suo, B., Wang, Z., Zhang, N., Li, Z., Lei, Y., et al. (2020). BDF: a relativistic electronic structure program package. *J. Chem. Phys.* 152, 064113. doi:10.1063/1.5143173



OPEN ACCESS

EDITED BY

Nino Russo,
University of Calabria, Italy

REVIEWED BY

Jonathan E. Cisterna,
Universidad Católica del Norte, Chile
Sajjad Hussain Sumrra,
University of Gujrat, Pakistan

*CORRESPONDENCE

M. Alcolea Palafox,
✉ alcolea@ucm.es
Irena P. Kostova,
✉ irenakostova@yahoo.com
Lozan Todorov,
✉ ltodorov@pharmfac.mu-sofia.bg

RECEIVED 16 June 2024

ACCEPTED 04 October 2024

PUBLISHED 06 November 2024

CITATION

Palafox MA, Belskaya NP, Todorov L,
Hristova-Avakumova N and Kostova IP (2024)
Molecular properties of a triazole–Ce(III)
complex with antioxidant activity: structure,
spectroscopy, and relationships with related
derivatives. Influence of the ligands in
the complex.
Front. Chem. 12:1450106.
doi: 10.3389/fchem.2024.1450106

COPYRIGHT

© 2024 Palafox, Belskaya, Todorov, Hristova-Avakumova and Kostova. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Molecular properties of a triazole–Ce(III) complex with antioxidant activity: structure, spectroscopy, and relationships with related derivatives. Influence of the ligands in the complex

M. Alcolea Palafox^{1*}, Nataliya P. Belskaya², Lozan Todorov^{3*},
Nadya Hristova-Avakumova⁴ and Irena P. Kostova^{3*}

¹Departamento de Química Física, Facultad de Ciencias Químicas, Universidad Complutense, Madrid, Spain, ²Department of Technology for Organic Synthesis, Ural Federal University, Yekaterinburg, Russia, ³Department of Chemistry, Faculty of Pharmacy, Medical University – Sofia, Sofia, Bulgaria, ⁴Department of Medical Physics and Biophysics, Faculty of Medicine, Medical University of Sofia, Sofia, Bulgaria

A novel Ce(III) complex with the triazole ligand 2b, which presents four H-bonded sites with amino acids of the MMP-2 receptor, was synthesized. The experimental IR and Raman spectra of this Ce(III) complex were well-interpreted based on their comparison to the theoretical scaled spectra using the scaling equations determined by two procedures and four density functional theory (DFT) levels. Therefore, the structure predicted for the synthesized Ce(III) complex was clearly characterized and confirmed. The potential antioxidant action of this complex was compared with the analogous La(III) complex, and it was found that the coordination of ligand 2b with Ce(III) improves the ligand's ability to participate in single-electron transfer (SET), as observed in the ABTS⁺ assay, and this complex seems to scavenge the stable radical much more actively compared to its La(III) counterpart. Additionally, interactions with potassium superoxide and sodium hypochlorite indicate a high pro-oxidant behavior of the complex. The effects of different ligands on the geometric parameters, atomic charges, and molecular properties of the Ce(III) complex were analyzed at four DFT levels, and several relationships were clearly established. These relationships can facilitate the selection of new ligands with improved properties in the design of novel lanthanide–triazole carboxylate complexes with promising biological activity. The ligand size increase in the complexes facilitates the electronic transfer of negative charge, and the low HOMO (highest occupied molecular orbital)–LUMO (lowest unoccupied molecular orbital) energy gap indicates a large reactivity and low energy for their excitation.

KEYWORDS

cerium(III) complex, 1,2,3-triazoles, antioxidant activity, structural relationships, infrared, Raman

1 Introduction

Coordination compounds with rare earth elements are being extensively developed for medical use (Goswami and Kostova, 2022; Lewandowski et al., 2005), especially as promising candidates for anticancer therapeutics (Fricker, 2006; Musib et al., 2023; Zhang et al., 2023). Lanthanide (Ln) complexes with a number of drugs (hymecromone, umbellipherone, mendiaxon, warfarin, coumachlor, and nifedipine) have demonstrated cytotoxicity against the HL-60 myeloid cell line in preclinical studies (Chundawat et al., 2021). The lanthanide–tetraphenylporphyrin complexes have advanced into clinical trials (Evens, 2004; Mody et al., 2001).

Special interest appears in the cerium complexes, which have been reported to have important pharmaceutical properties. For example, the cerium–humic acid complex has bacteriostatic potency, inhibiting the growth of several dangerous bacterial strains (Zhang et al., 2000), cerium–curcumin and cerium–quercetin complexes exhibit toxic effects against both breast and melanoma cancer cells used in photodynamic therapy (PDT) (Hosseinzadeh et al., 2021), cerium–ofloxacin and 2,2'-bipyridine complexes show antimicrobial and anticancer activities against breast and colon cell lines (Abd El-Hamid et al., 2019), and azamacrocyclic–cerium complexes promote the hydrolysis of the phosphodiester bond of supercoiled DNA (Feng et al., 2013). In addition, series of cerium(III) bipyridyl, phenanthroline, and related complexes have been reported with *in vitro* activity against cancer cell lines (Fricker, 2006). Despite the important properties of cerium complexes, they have received far less attention than lanthanum complexes, which is one of the reasons for their investigations in the present study.

The coordination chemistry of the lanthanoid metals is of interest in a variety of fields, such as molecular magnetic materials, catalysts, luminescent thermometers, or MRI contrast agents in bioimaging, nonlinear optics, and up-conversion materials (Bünzli, 2015; Li and Yan, 2020; Omodara et al., 2019; Zhang et al., 2021; Zhao et al., 2021), and the selection of a specific ligand for this coordination is an important and difficult task.

Ln(III) ions tend to form stable chemical bonds with organic compounds containing oxygen, nitrogen, or sulfur atoms (O, N, and S). Among the *N*-donor ligands, preferred heterocycles for Ln complexation are pyridine, pyrazole, tetrazole, and oxazole. There are well-known Ln(III) complexes with terpyridine, coumarin, acridine, porphyrin, quercetin, 1,10-phenanthroline, and benzotriazole (Fricker, 2006). Neutral *N*-1-substituted 1,2,3-triazoles have not been applied extensively as coordination units for Ln-complex formation because their binding affinities are not strong enough to overcome affinities with anions such as NO₃[−], F[−], Cl[−], and Br[−] (Juwarker et al., 2009) or the hydration enthalpy of Ln(III) ions (Guo et al., 2023). 2-Substituted-1,2,3-triazoles have a more symmetrical electronic distribution that could not stabilize such constructions. However, the triazole core is quite attractive as it is biocompatible and tends to endow molecules with a broad profile of biological activities and diverse pharmacophore properties (Agalave et al., 2011; Bonandi et al., 2017; Bozorov et al., 2019; Giroud et al., 2018; Keri et al., 2015; Li et al., 2018; Zhang et al., 2009).

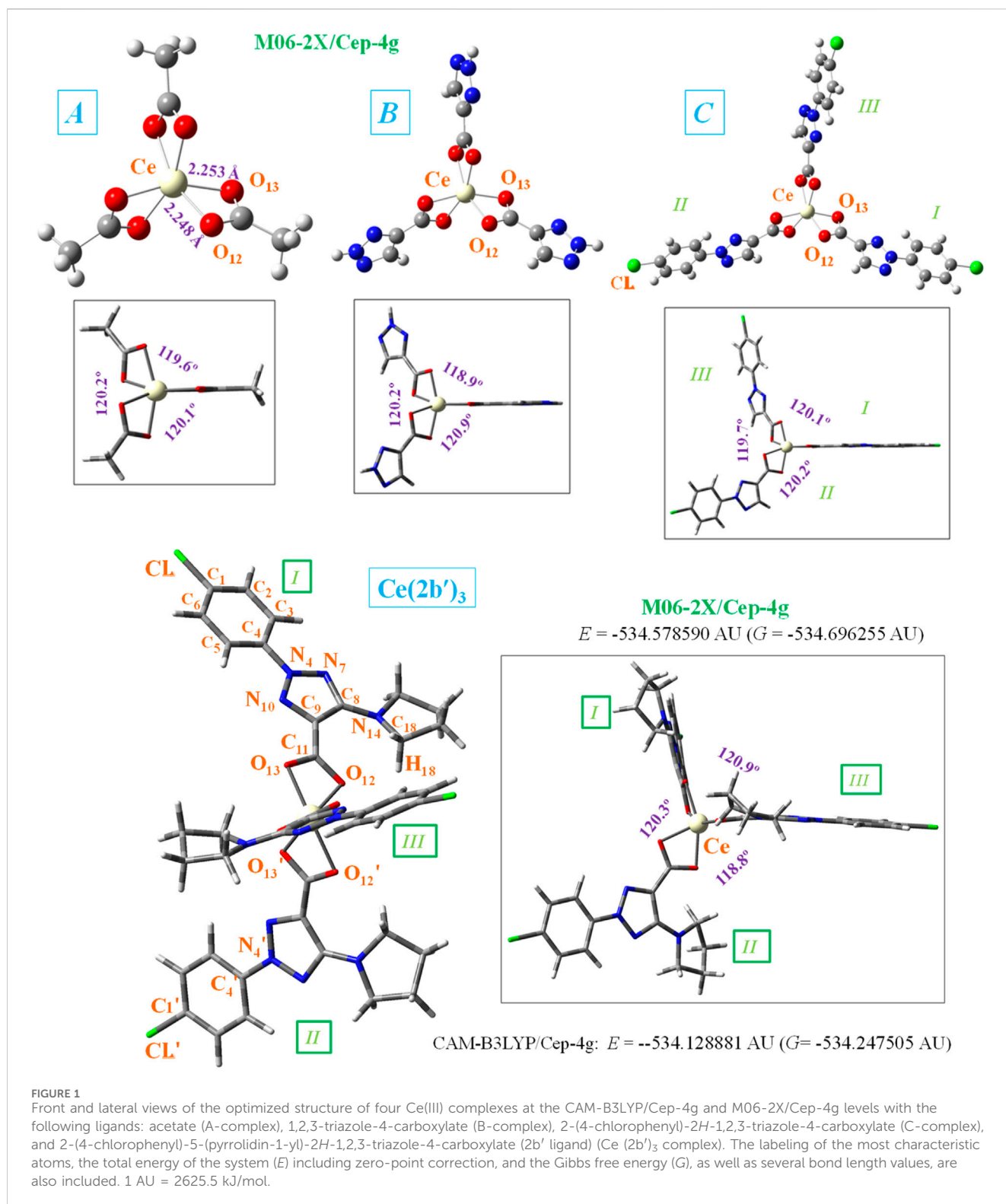
Triazoles are of particular interest to modern medicine for several reasons. In addition to having significant dipole moments,

they are characterized by high stability in terms of redox reactions and hydrolysis. Nitrogen atoms in the structure can also participate in hydrogen bond formation, thus enhancing solubility (Dheer et al., 2017). The facility of the triazole ring to form these H-bonds is considered the main reason for its ability to enhance binding to target molecules (Raić-Malić and Mescić, 2015; Kumar et al., 2013; Saini and Dwivedi, 2013). Moreover, in a variety of studies (Slavova et al., 2020), triazoles have been utilized as linkers. Particularly, the triazole ring is not just a passive linker, and it contributes to biological activity by binding to target molecules via H-bonding and dipole interactions (Kolb and Sharpless, 2003). When combined with another compound, it can produce new biological effects, increasing the activity of molecules (Slavova et al., 2020). In terms of mechanisms of action, triazole-containing molecules can act as plant gibberellin inhibitors, carbonic anhydrase inhibitors, competitive inhibitors of *para*-aminobenzoic acid, cellular microtubule stabilizers, nucleoside analogs, etc.

Since there are many examples that have demonstrated the synergetic biological effect due to coordination with transition metals or lanthanides (Slavova et al., 2020; Gil-Moles and Concepción Gimeno, 2024; Podolski-Renić et al., 2024; Adhikari et al., 2024; Alcolea Palafox et al., 2023a), the 1,2,3-isomer has been selected because it has been shown to have the best properties (Agalave et al., 2011; Bonandi et al., 2017; Bozorov et al., 2019; Giroud et al., 2018; Slavova et al., 2020; Hrimla et al., 2021) as a ligand and is suitable to coordinate with the Ce(III) ion. In order to overcome the possible problem of instability of the desired complex, we employed a well-known approach based on attaching the side substituents and functionalities to heterocycles with the additional nucleophilic centers (Zafar et al., 2023; Hussain Sumrra et al., 2020).

To facilitate the coordination of the triazole ring with the Ce(III) ion, this was carried out through the carboxylate group at C4 (C9 in Figure 1) of the triazole ring. In addition, aryl and pyrrolidine groups were bonded to the triazole ring to provide liposolubility to facilitate cell penetration. A chlorine atom was also attached to the aryl group, constituting the 2-(4-chlorophenyl)-5-(pyrrolidin-1-yl)-2*H*-1,2,3-triazole-4-carboxylate ligand. This triazole ligand has been labeled as 2b (sodium salt) and 2b' (its anionic form) (Alcolea Palafox et al., 2023b), and such notation will be used here. The ligand 2b has been previously studied from the structural, spectroscopic, and molecular docking points of view (Alcolea Palafox et al., 2023b; Saffronov et al., 2023; Palafox et al., 2023). The cerium cation is now coordinated with this ligand via the carboxylate groups, linked in a three-dimensional coordination complex, as shown in Figure 1. This arrangement is similar to that corresponding to the La(III) complex first obtained by us (Alcolea Palafox et al., 2023a), as well as in complexes of other carboxylic acid derivatives (Zhao et al., 2021). The *in vitro* behavior of this new complex toward a wide range of RS-generating model systems revealed potential therapeutic applications as hydroxyl scavengers and potential pro-oxidants (Alcolea Palafox et al., 2023a; Alcolea Palafox et al., 2023b).

Because the most accurate of the quantum chemical methods are still too expensive to apply in routine research, density functional theory (DFT) has emerged during the past decades as a powerful methodology for the simulation of chemical systems. DFT methods are less computationally demanding than other theoretical methods with a similar accuracy, or even better in some cases, such as the theoretical prediction of vibrational spectra (Palafox, 2018). They



are being able to include electron correlation in the calculations at a fraction of time of post-Hartree-Fock methodologies. Therefore, these DFT methods have a widespread application and are becoming increasingly popular.

Among the multiple applications of DFT methods, one based on the molecular docking approach has been used against SARS-CoV-2

(Subbiah et al., 2022) and promising drug targets for cancer (Noreen et al., 2023; Sert et al., 2020). However, our interest in the present manuscript is related to the interpretation of the spectra, which have been developed specifically over the last decade. Although they have been used for the assignment of the UV/vis (Sert et al., 2020; Hassan et al., 2023; Barim and Akman, 2022) and ¹H and ¹³C NMR spectra

(Sert et al., 2020; Rajaraman et al., 2022), in the present work, we focus our attention only on the IR and Raman vibrational spectra (Dodoo et al., 2023; Evangalin et al., 2018; Singh et al., 2023). Vibrational spectroscopy is one of the most powerful techniques for the characterization of medium-sized molecules. An accurate interpretation of vibrational experimental spectra using DFT methods is essential in many fields of chemistry and has become an important part of spectrochemical and quantum chemical investigations (Palafox, 2018; Alcolea Palafox, 2017).

However, the proper assignment of spectra is often not straightforward, and a large overestimation of the calculated vibrational wavenumbers is expected. This overestimation (which may be due to many different factors that are usually not even considered in the theory, such as anharmonicity, errors in the computed geometry, Fermi resonance, and solvent effects) can be significantly reduced with the use of transferable empirical parameters for the calculated wavenumbers. The use of suitable DFT quantum-chemical methods and scaling procedures remarkably reduces the risk in the assignment and can accurately determine the contribution of different modes in an observed band. Now, these procedures appear to be extensively used in the journals of vibrational spectroscopy, and we applied them in this study.

Now, the two main aims of the current investigation are as follows: 1) the first aim was to synthesize and characterize a newly synthesized cerium(III) complex of the 2-(4-chlorophenyl)-5-(pyrrolidin-1-yl)-2H-1,2,3-triazole-4-carboxylate (**2b'**) ligand and study its potential antioxidant activity. This ligand appears to be of great interest as it presents H-bonded sites with the matrix metalloproteinase-2 (MMP-2) receptor (Palafox et al., 2023), which is one of the main models for angiogenesis and tumor development (Jezierska and Motyl, 2009). The product develops as an amorphous precipitate and does not provide crystals suitable for X-ray molecular structure determination. Since the crystal structure data on this newly synthesized Ce (**2b'**)₃ coordination complex are not available, in such cases, theoretical calculations are very informative and useful. Moreover, theoretical approaches for the determination of geometrical parameters, vibrational frequencies with accurate scaling procedures, hydrogen bond strengths, model metal(III)-ligand bonding, and conformation of the molecular structure proposed at a high level of theory are very helpful for extracting reliable structural information. Therefore, these theoretical approaches were carried out in this study, and they helped predict correctly the metal coordination polyhedron of the studied complex.

The potential antioxidant activity of the Ce(III) complex was assessed utilizing multiple reactive species (RS)-generating model systems. Therefore, the impact of the Ce(III) complex on the physiologically important hydroxyl radicals (two different experiments), superoxide radical ions, and hypochlorite ions was investigated. Potential mechanisms of action—single-electron transfer (SET) and hydrogen atom transfer (HAT)—will be elucidated with the aid of the 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic) free radical (ABTS^{•+}) and 2,2-diphenyl-1-picrylhydrazyl radical (DPPH) assays. The complex's behavior in these model systems at the highest concentration tested was compared to a well-established positive control: 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox) (Goswami and Kostova, 2022). 2) Another aim was to reveal the impact of structural fragments on the features of the geometry and electronic

structure of the triazole complex. For this purpose, three simpler complexes, namely, A- (acetate), B- (1,2,3-triazole-4-carboxylate), and C- (2-(4-chlorophenyl)-2H-1,2,3-triazole-4-carboxylate), were visualized and used as models for quantum mechanical chemical calculations (Figure 1). Therefore, new relationships with different ligands were established to find better properties, which will facilitate the selection of new ligands for further design of new complexes with improved properties.

2 Materials and methods

2.1 Materials

All compounds used were of pro analysis grade from Sigma-Aldrich. Ethanol (95%) and bi-distilled water were utilized for the preparation of the solution. Trichloroacetic acid (TCA), thiobarbituric acid (TBA), and K-Na-phosphate buffered saline (PBS) solutions were required for the deoxyribose degradation assay, according to that reported in Alcolea Palafox et al. (2023a). 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was required, as well as other reagents for the Fenton reaction MTT assay. Participation in hydrogen atom transfer (HAT) reactions was elucidated following the protocols (Kedare and Singh, 2011; Molyneux, 2004). The participation of the Ce(III) complex in single-electron transfer (SET) reactions was determined according to that reported in Erel (2004a) and Erel (2004b). Luminol-dependent chemiluminescence (LDCL) was utilized to assess scavenging of (O₂^{•-}) radicals.

2.2 Experiment

The Ce(III) complex was synthesized from Merck's Ce(NO₃)₃·6H₂O salt (99%) and the **2b** ligand (Safronov et al., 2023). The procedure is similar to that followed in the synthesis of the La (**2b'**)₃ complex (Alcolea Palafox et al., 2023a). The chemical composition of the newly obtained Ce(III) complex was characterized via elemental analysis. The binding mode in the Ce(III) complex was confirmed via vibrational spectroscopy. The synthesized Ce (**2b'**)₃ complex was investigated as KBr pellet at room temperature for the IR spectrum, which was plotted in the 3,900–400 cm⁻¹ region utilizing a Bruker IFS25 FTIR spectrophotometer. The Raman spectrum was registered in the 4,000–0-cm⁻¹ range using a Horiba Jobin Yvon's spectrometer.

UV-vis spectra were recorded using a Shimadzu UV-1601 spectrophotometer. For LDCL measurements, an LKB 1251 luminometer (Bio-Orbit, Turku, Finland) was set at 37°C and connected to an IT-type computer via a serial interface. Three samples have been tested in each concentration. Averages and standard deviations were calculated. Relative changes within the limits of experimental errors were not discussed.

2.3 Methods

Stock solutions of the investigated ligand **2b** (1·10⁻³ M) and its complex Ce (**2b'**)₃ (1·10⁻⁴ M) were prepared using bi-distilled water. In order to test the molarities of 1·10⁻⁶ M or lower, these solutions were

further diluted as necessary. A $1 \cdot 10^{-3}$ M stock solution of Trolox in PBS was prepared. The activity of the cerium complex was compared to that of Trolox at the highest concentration tested ($3 \cdot 10^{-5}$ M).

For each tested molarity, three replicates were run, each representing an individual datapoint. Average values and standard deviations were calculated. The impact of the tested compounds on the RS-generating model systems was statistically verified via one-way ANOVA, followed by the Bonferroni post-test. Differences with $p < 0.05$ were considered statistically significant. Relative changes within the limits of the experimental error were not discussed.

2.3.1 Deoxyribose degradation assay

Hydroxyl radicals (OH^\cdot) were generated via UV-induced water radiolysis (Burns and Sims, 1981) through a modified protocol (Halliwell et al., 1987). The TBA assay was applied to assess the degradation of 2-deoxyribose (2-DR) to malondialdehyde (MDA). The 2-DR degradation degree is calculated as radical-scavenging activity (RSA) by Equation 1:

$$\text{RSA, \%} = \frac{A_{\text{control}} - (A_{\text{sample}} - A_{\text{blank}})}{A_{\text{control}}} \cdot 100, \quad (1)$$

where “control” sample excludes the Ce(III) complex. “Sample” is composed of 2-DR (0.5 mL) + Ce(III) complex solution + PBS (up to 5 mL). This 5-mL sample was irradiated using UV, and TBA and TCA were added following the procedure described in Alcolea Palafox et al. (2023a).

The *in vitro* scavenging activity of the sodium salt of the ligand (2b') and its La(III) complex in the presence of several free radical-generating model systems was discussed in Alcolea Palafox et al. (2023a). Herein, the behavior of 2b and its Ce(III) complex on RS, generated by six model systems, is revealed and compared to that of the respective La(III) complex.

2.3.2 ABTS, DPPH, and Fenton reaction MTT assays

Three types of samples, namely, “blank,” “control,” and “sample,” were tested using the kinetic function of the apparatus with the lag time of 10 s and measuring time of 600 s. In each sample composition, 1.0 mL is the total volume in ABTS, whereas it is 2.0 mL in DPPH and MTT assays. Results are presented as RSA, which is similar to the 2-deoxyribose degradation assay, as shown in Table 1.

ABTS: Prior to experimentation, two reagents (R1 and R2) are prepared (Erel, 2004a), where R1 is Na-acetate buffer with pH = 5.8, while R2 is composed of ABTS dissolved in Na-acetate buffer with pH = 3.8, along with the addition of H_2O_2 to form ABTS^+ radical. Absorbance was determined at 660 nm.

DPPH: It was performed according to literature data (Kedare and Singh, 2011; Molyneux, 2004; Chrzczanowicz et al., 2008), and where DPPH' absorbance changes were measured at $\lambda = 517$ nm.

MTT: The reduction in MTT to formazan by Fenton-generated OH^\cdot was investigated. Ascorbic acid increases the formation of OH^\cdot . Formazan production causes an increase in the absorbance at $\lambda = 578$ nm.

2.3.3 LDCL in the presence of KO_2 or NaClO

With KO_2 , the “control” composition (1.0 mL) contains luminol (0.05 mL) + KO_2 (0.05 mL) + PBS. The “sample” composition (1.0 mL) contains the Ce(III) complex + luminol (0.05 mL) + KO_2

TABLE 1 Fenton reaction with ABTS, DPPH, and MTT assay sample compositions. Values are in mL.

Assay	Reagent	Blank	Control	Sample
ABTS	Tested compound	0.1	-	0.1
	R1	0.86	0.86	0.86
	R2	-	0.04	0.04
	Bi-distilled water	0.04	0.1	-
DPPH	Tested compound	0.2	-	0.2
	DPPH	-	1.8	1.8
	Ethanol	1.8	-	-
	Bi-distilled water	-	0.2	-
MTT	Tested compound	0.2	-	0.2
	MTT	0.2	0.2	0.2
	$\text{Fe}^{2+}/\text{H}_2\text{O}_2/\text{Na}_2\text{-EDTA}$	-	0.1	0.1
	Ascorbic acid	-	0.1	0.1
	Bi-distilled water	up to 2.0	to 2.0	to 2.0

(0.05 mL) + PBS. The kinetic mode of the apparatus was utilized. Measurements are taken with 2-s delay from the total of 10 s. Data are presented as the CL-SI index by Equation 2:

$$\text{CL} - \text{SI, \%} = \frac{I_{\text{sample}}}{I_{\text{control}}} \cdot 100. \quad (2)$$

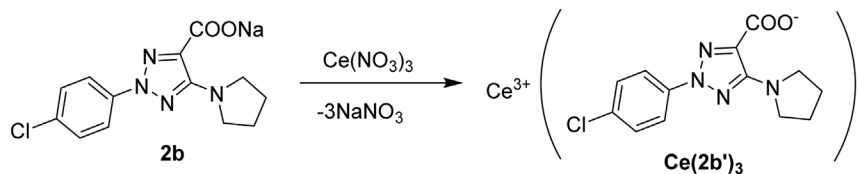
In the absence of KO_2 , the background signal has been withdrawn from all data recorded.

With NaClO, the “control” composition (1.0 mL) contains NaClO (0.15 mL) + luminol (0.05 mL) + PBS. The “sample” composition (1.0 mL) contains Ce(III) complex + luminol (0.05 mL) + NaClO (0.15 mL) + PBS. The background signal is recorded in the absence of hypochlorite and is withdrawn from all data recorded. Data are also shown as the CL-SI index.

2.4 Computational details

Four theoretical levels were used for the optimization of all Ce(III) complexes. Density functional theory (DFT) methods (Seminario and Politzer, 1995) were chosen for this purpose since they have provided accurate vibrational wavenumber values in biomolecules that are in good agreement with the experimental values and in better agreement compared to those determined by the MP2 method (Palafox, 2018). Among these DFT methods, the Minnesota M06-2X functional was the most preferred because it is one of the best meta-generalized gradient functionals to analyze dispersion bound in large systems (Riley and Hobza, 2011; Riley et al., 2010), especially in biomolecules with non-covalent weak interactions, like those included in the present work. Moreover, this method has also shown a large applicability in chemistry (Zhao and Truhlar, 2011).

The B3LYP functional has been chosen second because it has yielded excellent results in the computation of the IR and Raman wavenumbers of biomolecules and is better than other DFT methods



SCHEME 1
Synthesis of the Ce(III) complex.

(Palafox, 2018; Alcolea Palafox, 2000). In particular, with the Ce(III) ion, it performs slightly better than other hybrid functionals (Kullgren et al., 2010). However, the B3LYP functional alone is not appropriate for reproducing systems with non-covalent weak interactions, like those observed with our complexes; therefore, the spectra obtained differ from the experimental spectra. For this reason, the D3-B3LYP and the CAM-B3LYP methods (Yanai et al., 2004), which combine the hybrid qualities of B3LYP and the long-range correction (Tawada et al., 2004), were also used to improve the calculated structural parameters. The Cep-4g basis set is the only set that appears available in the Gaussian-16 program package for cerium (Frisch et al., 2016). Although it is a very small basis set, the vibration spectra obtained with this basis set can be well correlated with the experimental spectra, and therefore, a good characterization of the synthesized Ce (2b')₃ complex was carried out. Even with the smaller Ce(III) complexes labeled as A, B, and C as shown in Figure 1, the three DFT methods required a charge +1 on the whole system. With the neutral charge 0 on the system, the DFT methods used in the Gaussian-16 program package does not run, indicating an inconsistency in the ground state calculations (multiplicity = 1). The authors consider this unusual since the three positive charges on the Ce ion are compensated by three 2b' anions. Notably, this could be due to the default of the Cep-4g basis for the Ce ion. Despite the fact that the effect on the organic ligands is very slight and their calculated spectra are in accordance with the experimental spectra, various Ce(III) and Ce(IV) complexes with organic molecules have been reported (Chundawat et al., 2021; Evens, 2004; Das et al., 2014; Levin et al., 2016). The UNIX version of the Gaussian-16 program with standard parameters was operated in the Brigit super-computer of the Complutense University of Madrid.

All optimized complexes showed positive wavenumbers, indicating a minimum in the potential energy surface. For this task, harmonic wavenumber computations were performed at the same level of the corresponding optimization process.

2.4.1 Scaling the wavenumbers

The calculated wavenumbers by the theoretical DFT methods appear overestimated due to different reasons (Alcolea Palafox, 2019). To correct this overestimation, different scaling procedures have been reported for each specific method and basis set (Alcolea Palafox, 2000; Alcolea Palafox, 2019), obtaining a noticeable improvement in the wavenumbers. To get an accurate assignment of the experimental bands, the scaling is the normal procedure followed by the authors. Among the different procedures available, the linear scaling equation (LSE) (Equations 3–6) and

polynomic scaling equation (PSE) (Equations 7–10) procedures appear as the most appropriate, and therefore, they were used in the present study. The specific equations utilized here for the four levels of computation are as follows:

$$\text{LSE procedure: } \nu^{\text{scal}} = 195.7 + 0.8706 \nu^{\text{cal}} \text{ at B3LYP/Cep-4g level,} \quad (3)$$

$$\nu^{\text{scal}} = 194.0 + 0.8714 \nu^{\text{cal}} \text{ at D3-B3LYP/Cep-4g level,} \quad (4)$$

$$\nu^{\text{scal}} = 184.0 + 0.8646 \nu^{\text{cal}} \text{ at CAM-B3LYP/Cep-4g level,} \quad (5)$$

$$\nu^{\text{scal}} = 185.0 + 0.8613 \nu^{\text{cal}} \text{ at M06-2X/Cep-4g level,} \quad (6)$$

$$\text{PSE procedure: } \nu^{\text{scal}} = 90.6 + 1.0305 \nu^{\text{cal}} - 0.0000403 \cdot (\nu^{\text{cal}})^2 \text{ at B3LYP/Cep-4g level,} \quad (7)$$

$$\nu^{\text{scal}} = 92.7 + 1.0253 \nu^{\text{cal}} - 0.0000388 \cdot (\nu^{\text{cal}})^2 \text{ at D3-B3LYP/Cep-4g level,} \quad (8)$$

$$\nu^{\text{scal}} = 78.7 + 1.0216 \nu^{\text{cal}} - 0.0000390 \cdot (\nu^{\text{cal}})^2 \text{ at CAM-B3LYP/Cep-4g level,} \quad (9)$$

$$\nu^{\text{scal}} = 78.0 + 1.0204 \nu^{\text{cal}} - 0.0000394 \cdot (\nu^{\text{cal}})^2 \text{ at M06-2X/Cep-4g level.} \quad (10)$$

3 Results and discussion

The *N*-2-aryl-triazole ligand 2b was prepared according to the procedures reported in the literature (Alcolea Palafox et al., 2023a; Safronov et al., 2023; Palafox et al., 2023). The complex Ce (2b')₃ was synthesized by the interaction of triazole sodium salt 2b with Ce(NO₃)₃·6H₂O at a molar ratio of 3:1 in the water solution (Scheme 1) with a yield of 85% and characterized via elemental analysis and vibrational spectroscopy.

The elemental analysis of the Ce(III) complex of 2-(4-chlorophenyl)-5-(pyrrolidin-1-yl)-2H-1,2,3-triazole-4-carboxylic acid: (% calculated/found): Ce (2b')₃ · 2H₂O: C: 44.61/45.02; H: 4.00/3.78; N: 16.01/15.68; and H₂O: 3.43/3.98; Ce: 13.35/12.95, where 2b' = C₁₃H₁₂N₄O₂Cl[−].

The binding mode of the ligand to Ce(III) ions was elucidated by recording the IR and Raman spectra of the complex, as compared with those of the free ligand and the theoretical predictions. The vibrational fundamentals from the IR and Raman spectra were analyzed by comparing these modes with literature sources in combination with the results derived from DFT calculations (i.e., harmonic vibrational wavenumbers and their Raman scattering activities) for the ligand and Ce(III) complex.

TABLE 2 Several selected optimized geometrical parameters calculated with different DFT methods and the Cep-4g basis set in the Ce(III) complexes shown in Figure 1. Bond lengths (r) are in Å, and bond angles and dihedral angles (\angle) are in degrees. The data were from ligand I.

	A complex			B complex			C complex			Ce(2b') ₃			
Parameter	B3LYP	CAM-B3LYP	M06-2X	B3LYP	CAM-B3LYP	M06-2X	B3LYP	CAM-B3LYP	M06-2X	B3LYP	D3-B3LYP	CAM-B3LYP	M06-2X
r (C ₉ –C ₁₁)	1.657	1.648	1.638	1.587	1.581	1.575	1.583	1.577	1.571	1.576	1.570	1.563	1.556
r (C=O ₁₂)	1.434	1.423	1.413	1.434	1.422	1.410	1.436	1.423	1.413	1.437	1.437	1.425	1.415
r (C=O ₁₃)	1.434	1.423	1.412	1.435	1.423	1.414	1.436	1.423	1.413	1.442	1.441	1.433	1.425
r (Ce–O ₁₂)	2.287	2.256	2.248	2.290	2.260	2.261	2.288	2.257	2.253	2.304	2.292	2.264	2.260
r (Ce–O ₁₃)	2.286	2.255	2.253	2.282	2.251	2.240	2.281	2.251	2.245	2.275	2.272	2.231	2.224
\angle (C ₉ –C ₁₁ = O ₁₂)	123.7	123.9	123.4	122.2	122.4	122.0	122.1	122.4	121.7	125.9	125.4	126.6	125.5
\angle (O=C=O)	112.5	112.1	113.2	114.2	113.8	114.8	114.2	113.7	114.8	113.7	113.9	112.7	113.8
\angle (C=O ₁₂ –Ce)	92.3	92.3	91.9	90.9	91.0	90.3	90.9	91.1	90.5	90.9	90.8	91.1	90.4
\angle (C=O ₁₃ –Ce)	92.3	92.4	91.7	91.2	91.4	91.0	91.2	91.3	90.8	91.9	91.5	92.2	91.6
\angle (O ₁₂ –Ce–O' ₁₃)	153.9	155.3	100.5	154.3	155.5	100.5	153.8	155.3	101.7	152.4	153.3	154.8	99.7
\angle (O ₁₃ –Ce–O' ₁₂)	102.2	100.8	102.1	102.7	101.2	100.1	103.1	101.3	100.4	104.5	103.4	101.6	99.9
\angle (C ₁₁ –O ₁₂ ...O' ₁₂ –C' ₁₁)	57.6	60.6	–98.1	56.9	60.5	–98.8	55.9	59.8	–99.1	56.4	59.6	61.7	–103.5
\angle (C ₉ –C ₁₁ ... C' ₁₁ –C' ₉)	–0.1	–0.2	0.5	0.2	0.0	1.8	0.0	0.0	2.7	–0.2	–1.1	0.0	–18.0
\angle (C' ₉ –C' ₁₁ ... C'' ₁₁ –C'' ₉)	0.1	0.3	–0.5	2.1	2.6	0.2	0.0	0.1	1.0	–0.1	–1.1	–0.1	16.7
\angle (C ₁₁ ...Ce... C' ₁₁)	120.1	120.2	119.6	120.2	120.2	118.9	119.9	120.1	120.1	120.0	119.9	119.8	120.3
\angle (C' ₁₁ ...Ce... C'' ₁₁)	120.0	119.6	120.2	119.0	119.0	120.9	120.1	120.0	119.7	119.8	119.8	119.9	118.8

TABLE 3 Atomic polar tensor charges calculated with different DFT methods and the Cep-4g basis set in the Ce(III) complexes shown in Figure 1. The ligand APT values are only shown.

	A complex			B complex			C complex			Ce(2b') ₃			
Atom	B3LYP	CAM-B3LYP	M06-2X	B3LYP	CAM-B3LYP	M06-2X	B3LYP	CAM-B3LYP	M06-2X	B3LYP	D3-B3LYP	CAM-B3LYP	M06-2X
Ce	2.366	2.814	2.752	2.982	3.464	3.436	3.657	4.147	4.117	2.844	3.056	4.457	4.494
N ₄	-	-	-	-0.596	-0.608	-0.625	-0.210	-0.440	-0.383	-0.058	-0.109	-0.466	-0.430
N ₇	-	-	-	-0.013	-0.018	-0.031	-0.111	-0.076	-0.105	-0.099	-0.109	-0.124	-0.133
C ₈	-	-	-	-0.071	-0.090	-0.079	0.021	-0.093	-0.072	0.444	0.453	0.596	0.610
C ₉	-0.450	-0.461	-0.512	-0.682	-0.664	-0.695	-1.065	-1.042	-1.100	-1.048	-1.096	-1.342	-1.392
N ₁₀	-	-	-	0.253	0.262	0.276	0.335	0.443	0.472	0.105	0.157	0.411	0.426
C ₁₁	0.830	0.831	0.810	1.459	1.446	1.441	1.954	1.871	1.900	1.458	1.591	2.071	2.134
O ₁₂	-0.733	-0.819	-0.794	-0.921	-0.995	-0.989	-1.205	-1.241	-1.260	-0.909	-0.974	-1.323	-1.363
O ₁₃	-0.747	-0.823	-0.797	-0.937	-1.011	-0.992	-1.167	-1.224	-1.201	-0.925	-1.001	-1.380	-1.413

3.1 Molecular structure of the cerium complex

The cerium(III) ion, similar to other lanthanide ions, appears to coordinate well with oxygen atoms rather than nitrogen atoms (Peica et al., 2006). This could be due to the large flexibility of the carboxylic oxygens, which facilitates bonding with Ce³⁺. Therefore, the starting geometry to be optimized was that with the Ce(III) ion coordinated through the COO⁻ group with three ligands. For simplicity, as shown at the bottom of Figure 1, the total energy (*E*), which includes the ZPE (zero-point vibrational energy) correction, and the Gibbs energy (*G*) value were only shown for the Ce (2b')₃ complex and with the CAM-B3LYP and M06-2X methods. Because the optimized structure by the B3LYP and D3-B3LYP methods appears noticeably distorted, they are included in Supplementary Figures S1, S2. This distortion is due to the lack of long-range interactions of B3LYP for stabilizing the complex to a symmetric arrangement.

In the complex formation with the cerium ion, the CO bonds of the ligands are lengthened, as compared to the free form or in a dimer form (Alcolea Palafox et al., 2023b), which causes a slight shortening of the C₉–C₁₁ bond length. This feature in the triazole ring leads to a decrease in the N₇ = C₈ and C₉ = N₁₀ double-bond character (an increase in their bond lengths) and shortening of the neighbor N₄–N₁₀ and C₈–N₁₄ bonds. It is noted that the lengthening of these bond lengths leads to more rotated triazole substituents, and thus, they can interact more easily with other ligands, especially through the pyrrolidine ring. Therefore, with cerium binding, the triazole ring bonds and angles are slightly modified, which consequently results in the modification of their molecular properties.

Different ligands have little impact on the almost symmetric arrangement with the Ce(III) ion, as shown in Figure 1. Therefore, all methods and complexes place these ligands at angles (C₁₁...Ce...C'₁₁ and C'₁₁...Ce...C''₁₁) very close to 120.0° and with very little rotation, with C₉–C₁₁...C'₁₁–C'₉ and C'₉–C'₁₁...C''₁₁–C''₉ torsional angles close to 0°, with the exception of the calculated values in the Ce (2b')₃ complex by the M06-2X method, -18.0° and 16.7°, respectively. As expected, the main differences appear in the

coordination distances between the carboxylate oxygens and Ce(III) ion, which significantly affects the neighboring O–C₁₁ (slightly, for example, 1.434, 1.434, 1.436, and 1.437 Å) and (more prominently: 1.657, 1.587, 1.583, and 1.576 Å) C₁₁–C₉ bond lengths. These differences are presented in Table 2, which includes several selected optimized geometrical parameters in one of the ligands (labeled as *I*) at three DFT levels. The notation used for labeling the atoms is from that reported in the 2b' ligand (Safronov et al., 2023). Large differences also appear among the three DFT levels used, with the values by CAM-B3LYP closer to M06-2X than to B3LYP.

The optimized structure of the aryl ring in complexes C- and Ce (2b')₃ is full planar at all three DFT levels used. It is almost coplanar with the triazole ring, with the C₅–C₄–N₄–N₁₀ torsional angle of ca. -0.4° in the C complex and slightly larger due to the effect of the pyrrolidine ring in the Ce (2b')₃ complex, -4.7° by B3LYP and -2.3° by M06-2X. Similar values are also observed in the isolated ligands of C and Ce (2b')₃, which indicates a very small impact of the Ce(III) ion in this coplanarity.

The triazole ring is also full planar in B and C complexes at all three DFT levels and with torsional angle values lower than 0.5° in Ce (2b')₃. By contrast, the pyrrolidine substituent appears out-of-plane, as expected, and out-of-coplanarity with the triazole ring plane, C₉–C₈–N₁₄–C₁₈ = -20.2° by M06-2X and 16.6° by CAM-B3LYP. However, this value is noticeably lower than that calculated in the 2b ligand alone by the M06-2X method, -40.3°. This can be explained by the strong intramolecular H-bond O₁₂...H₁₈ between the carboxylate oxygen O₁₂ and the pyrrolidine hydrogen H₁₈, 1.711 Å vs. 2.605 Å in the Ce (2b')₃ complex. This H-bond forces the rotation of the pyrrolidine ring.

In the Ce (2b')₃ complex, the C₁₁ carbon atom appears slightly rotated related to the triazole ring plane, with this rotation being slightly lesser with O₁₂ than with O₁₃. Therefore, the torsional angle C₈–C₉–C=O₁₂ with O₁₂ has a smaller value of -1.7° by M06-2X (4.8° by CAM-B3LYP) vs that with O₁₃, N₁₀–C₉–C=O₁₃ of -8.2° (8.9° by CAM-B3LYP). This difference is due to a weak intramolecular H-bond O₁₂...H₁₈, 2.605 Å by M06-2X (2.332 Å by CAM-B3LYP), but it has not been observed in complexes B and C which lack the pyrrolidine ring. Therefore, in complexes B and

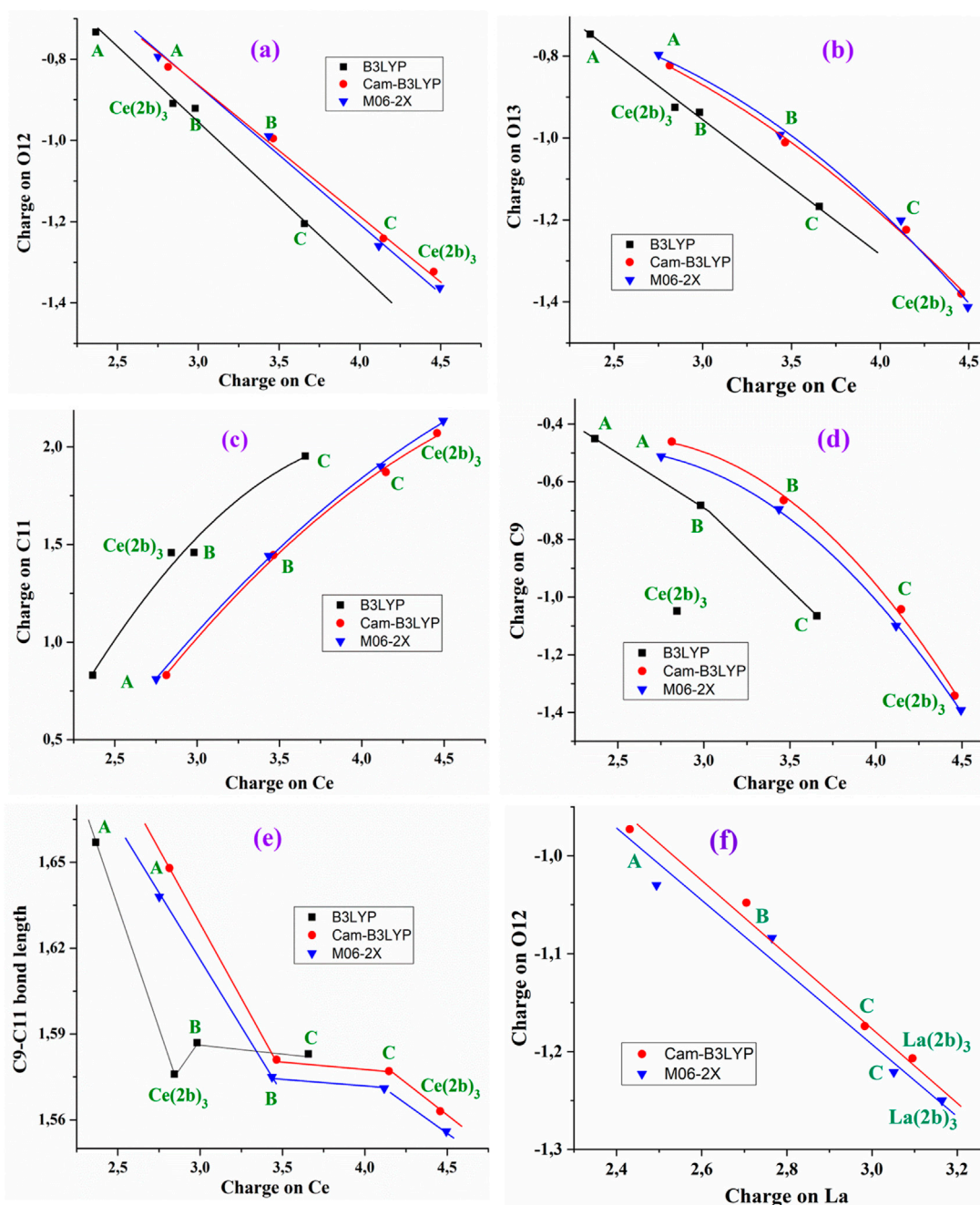


FIGURE 2

Relationships established at the B3LYP/Cep-4g, CAM-B3LYP/Cep-4g, and M06-2X/Cep-4g levels between the positive calculated atomic APT charge on the cerium ion in the A, B, C, and $\text{Ce}(2b)_3$ complexes with (A) the atomic charge on O_{12} , (B) the atomic charge on O_{13} , (C) the atomic charge on C11, (D) the atomic charge on C9, (E) the C9–C11 bond length, and (F) relationship established at the CAM-B3LYP/Lan12dz and M06-2X/Lan12dz levels between the atomic APT charge on the lanthanum ion and the atomic charge on O_{12} .

C, the carboxylate group appears coplanar to the triazole ring plane and with the same $\text{N}_{10}\text{--C}_9\text{--C}=\text{O}_{13}$ and $\text{C}_8\text{--C}_9\text{--C}=\text{O}_{12}$ torsional angle values of -0.2° . This intramolecular H-bond in the $\text{Ce}(2b')_3$ complex is also the main reason for the difference in Ce--O_{12} and Ce--O_{13} coordination distance values, especially in the C--O_{12} and C--O_{13} bond lengths. They are different in the $\text{Ce}(2b')_3$ complex but same in A, B, and C complexes.

Different ligands also have little influence on the calculated bond angles, such as $\text{C}=\text{O}_{12}\text{--Ce}$ (90.3° by M06-2X in the B complex and

90.4° in the $\text{Ce}(2b')_3$ complex), the $\text{C}=\text{O}_{13}\text{--Ce}$ angle (91.0° in the B complex and 91.6° in the $\text{Ce}(2b')_3$ complex), as well as the OCO , $\text{O}_{12}\text{--Ce--O}'_{13}$, and $\text{O}_{13}\text{--Ce--O}'_{12}$ angles, with differences lower than 2° , as shown in Table 2.

Significant differences are observed if we compare complexes $\text{Ce}(2b')_3$ and $\text{La}(2b')_3$, published previously by Alcolea Palafox et al. (2023a). For example, the lengths of the C–O and C–C bonds surrounding the metal cation in the cerium complex are noticeably shorter (approximately 7%–13% by M06-2X), whereas

TABLE 4 Molecular properties and global chemical reactivity descriptors (eV) calculated at different DFT levels in the Ce(III) complexes.

Molecular property	A complex		B complex		C complex		Ce(2b') ₃			
	CAM-B3LYP	M06-2X	CAM-B3LYP	M06-2X	CAM-B3LYP	M06-2X	B3LYP	D3-B3LYP	CAM-B3LYP	M06-2X
Rotational constants (GHz): A	0.575	0.582	0.133	0.135	0.023	0.028	0.018	0.018	0.018	0.019
B	0.572	0.582	0.129	0.131	0.023	0.020	0.018	0.018	0.018	0.017
C	0.328	0.331	0.070	0.071	0.013	0.012	0.011	0.011	0.011	0.011
C _v (cal/mol-K)	51.7	53.6	86.8	83.8	163.8	162.4	231.5	231.0	225.9	224.1
S (cal/mol-K)	134.3	142.9	188.4	181.6	297.1	296.2	387.8	388.5	383.0	379.4
Dipole moment (debye)	0.098	0.130	0.040	0.036	5.416	2.571	11.523	10.916	10.604	5.193
HOMO	−0.553	−0.565	−0.512	−0.513	−0.423	−0.421	−0.342	−0.344	−0.390	−0.390
LUMO	−0.290	−0.350	−0.288	−0.346	−0.279	−0.337	−0.306	−0.305	−0.239	−0.296
E _g	0.263	0.214	0.224	0.167	0.144	0.083	0.107	0.039	0.151	0.093
IP	0.553	0.565	0.512	0.513	0.423	0.421	0.342	0.344	0.390	0.390
EA	0.290	0.350	0.288	0.346	0.279	0.337	0.306	0.305	0.239	0.296
χ	0.421	0.457	0.400	0.430	0.351	0.379	0.324	0.324	0.315	0.343
η	0.131	0.107	0.112	0.083	0.072	0.042	0.018	0.019	0.076	0.047
S	0.066	0.054	0.056	0.042	0.036	0.021	0.009	0.009	0.038	0.023

the La³⁺–O bonds are noticeably shorter (approximately 10%–11%). These differences in the spatial structures of the complexes may affect their behavior in biological systems and the magnitude or selectivity of their biological effects.

3.2 APT atomic charges and relationships established

In the complexes under the theoretical study, the oxygen atoms have the highest negative charge, as expected, making them appear to be the most reactive. Due to their high reactivity, it is also expected that they would play a key role in the H-bonding of these complexes to amino acids of the cancer cell proteins. N₄ and N₁₄ nitrogen atoms also have a large negative charge, although lower than that of the oxygen atoms, and therefore, these atoms are also expected to participate in the biological activity of the synthesized complexes under the investigation. N₇ nitrogen has a very small negative charge, and N₁₀ has a small positive charge. Thus, these nitrogen atoms should be less active in potential interactions with biological targets.

C₈ and C₁₁ carbon atoms have a high positive charge due to the fact that they are bonded with highly negatively charged atoms. In the calculations performed, the Ce(III) ion appears positively charged around 3–4e depending on the DFT method used and complex studied, as shown in Table 3. It is smaller by B3LYP (between 2.3e and 3.7e) and larger by CAM-B3LYP and M06-2X methods (between 2.4e and 4.5e). CAM-B3LYP and M06-2X methods yield similar results. From A to Ce (2b')₃ complexes, the cerium charge is increased. This could be explained by a

better charge distribution on the cerium ion in larger ligands than in smaller ligands, such as in A complex.

The charge of the Ce(III) ion has a large influence on the bond lengths and atomic charges in the A-, B-, C-, and Ce (2b')₃ complexes, and therefore, several relationships can be well-established by the B3LYP, CAM-B3LYP, and M06-2X methods, as shown in Figure 2. For instance, an increase in the positive calculated atomic APT charge on the cerium atom shows good linear relationship to an incremental increase in the negative charge on O₁₂ and O₁₃ atoms, as shown in Figures 2A,B. The magnitude of the electronic charge lost by the Ce(III) ion is almost the same as that transferred to both O₁₂ and O₁₃. Increasing the ligand size appears to facilitate this negative electron transfer. Calculations show a large change in the A complex when the triazole ring is inserted (B-complex) and a notable increase with the additional insertion of the aryl ring (C complex). The addition of the pyrrolidine ring increases that electron transfer only a little. B3LYP differs in this point. The values by CAM-B3LYP and M06-2X methods are very close, deviating somewhat from those calculated by the B3LYP method. These increased deviations are presented in Figures 2C–E. Moreover, B3LYP shows large discrepancies in the calculated values for the Ce (2b')₃ complex, and for this reason, the results obtained by B3LYP were not discussed in the present study.

The increase in the negative charge on the oxygen atoms leads to a similar increase in the positive charge on the C₁₁ atom, as shown in Figure 2C. Consequently, the atomic charge on C₉ is negative and increased, as shown in Figure 2D. These features show the facility for electron transfer in the presented complexes.

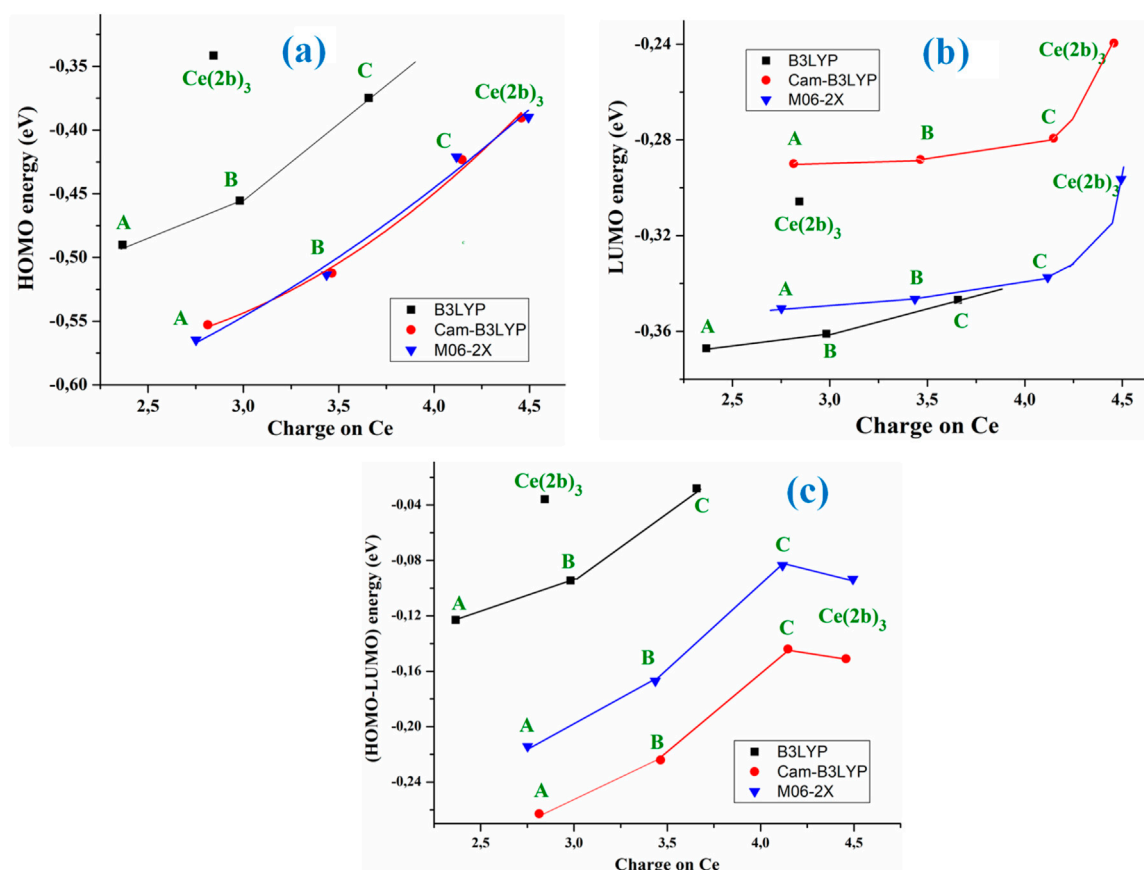


FIGURE 3 Relationships between the positive calculated atomic APT on the cerium ion in the A, B, C, and Ce (2b')₃ complexes using the B3LYP, CAM-B3LYP, and M06-2X methods with (A) the HOMO energy orbital, (B) the LUMO energy orbital, and (C) the (HOMO-LUMO) energy difference.

The charge variation in the cerium ion was also related to the C₉-C₁₁ bond length, as shown in Figure 2E. Although this relation is not linear, there is a relationship between both parameters. In general, the charge variation in the ligand atoms leads to changes in the geometrical parameters. Supplementary Figure S3 shows several relationships with the increase in the cerium charge, leading to an increase in the Ce-O₁₂ and C=O₁₂ bond lengths and a decrease in Ce-O₁₃. The relationships plotted in Figure 2 have also been observed with the lanthanum ion (Alcolea Palafox et al., 2023a). For example, the relationship calculated with the better Lanl2dz basis set between the atomic APT charge on the lanthanum ion and the atomic charge on O₁₂ is shown in Figure 2F. A larger increase is observed in the positive charge value on the central cerium cation and atom C₁₁ with the increase of the negative charge in C₉, O₁₂, and O₁₃ in comparison with La cations.

3.3 Molecular properties

Several parameters, such as rotational constants, heat capacity at constant volume, dipole moments, molecular orbitals, and other global chemical descriptors in the A, B, C, and Ce (2b')₃ complexes, have been determined and are presented in Table 4. The large symmetry obtained in the optimized complexes leads to values of the

rotational constant similar in A, B, and C. Very small differences are observed in its calculation by the four DFT methods used. Its value decreases with the increase in the systems' complexity, i.e., from A to Ce (2b')₃.

The constant volume heat capacity (C_v) value differs slightly between the DFT methods used, being slightly higher with CAM-B3LYP compared to M06-2X, with the exception of the A complex. As expected, its value markedly increases with system complexity. Entropy (S) values also increase with the system complexity, although to a lesser extent.

Dipole moment values indicate an almost null water solubility in A and B complexes but a remarkable increase when the aryl and pyrrolidine rings are inserted in the ligands, complexes C and Ce (2b')₃. In these C and Ce (2b')₃ complexes, the calculated value by M06-2X is almost half that by the three other DFT methods.

HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) values have also been determined. Their values slightly decrease as system complexity increases, and they appear in a linear relationship with the atomic charge on the Ce ion, with the exception of the Ce (2b')₃ complex values, as shown in Figures 3A,B.

The M06-2X and CAM-B3LYP methods calculate almost the same HOMO energy orbital value but differ largely in the LUMO value. With these computed energies, the global chemical reactivity

TABLE 5 Calculated, scaled, and experimental wavenumbers (ν , cm^{-1}) in the $\text{Ce}(\text{2b}')_3$ complex by CAM-B3LYP and M06-2X methods. Relative infrared intensity (A) in %, relative Raman intensity (S) in %, and Raman depolarization ratios for plane (DP) and unpolarized incident light (DU). For each vibration of the tetramer, the wavenumber with the highest IR intensity is indicated in bold style and that with the highest Raman intensity is indicated in italic style. The relative IR and Raman intensities were only shown for these wavenumbers. DP and DU values were from the most intense Raman line. The number of the ring mode corresponds to Wilson's notation (Varsányi, 1974).

Calculated by CAM-B3LYP					scaled		Experiment		Characterization by CAM-B3LYP
ν	A	S	DP	DU	LSE	PSE	IR	Raman	
3324, 3323, 3323	47	1	0.70	0.82	3057	3043	2968.1 s		20b, $\nu(\text{C5-H})$ in aryl (100)
3312, 3312, 3312	4	3	0.64	0.78	3048	3034			7b, $\nu(\text{C6-H})$ in aryl (100)
3303, 3303, 3303	13	0	0.75	0.86	3040	3028			20a, $\nu(\text{C2-H})$ in aryl (100)
3211, 3211, 3211	5	3	0.01	0.02	2960	2957	2872.1 m		$\nu_s(\text{C-H})$ in C15H_2 in pyrrolidine (100)
1664 , 1660, 1660	56	35	0.09	0.16	1623	1671	1577.2 br, vs		$\nu(\text{C}_8\text{-N}_{14})$ (45) + $\nu_s(\text{CC})$ (34)
1644, 1644, 1644	0	0	0.22	0.36	1605	1653			8b, $\nu(\text{C=C})$ in aryl (89)
1635, 1635, 1635	0	92	0.18	0.31	1598	1645		1595.0 vs	8a, $\nu(\text{C=C})$ in aryl (82)
1558, 1525 , <i>1525</i>	97	100	0.75	0.86	1503	1546	1500.1 s	1504.5 s	$\nu_{\text{as}}(\text{COO})$ + $\nu(\text{C}_9\text{-C}_{11})$ + $\nu(\text{C}_4\text{N})$ + 19a, $\nu(\text{CC,CH})$
<i>1475</i> , 1464 , 1464	100	60	0.04	0.08	1450	1491	1484.4 vs		19a, $\nu(\text{CC,CH})$ + $\nu(\text{C}_9\text{-C}_{11})$ + $\nu(\text{triazole})$
<i>1385</i> , 1383, 1383	7	7	0.04	0.07	1380	1417	1418.0 w		$\nu_{\text{as}}(\text{NNN, CN})$ + 19a, $\nu(\text{CC})(35)$ + $\nu_s(\text{COO})$
1362, 1362, 1362	2	1	0.75	0.86	1362	1398	1398.3 m		19b, $\nu(\text{CC})(85)$
<i>1347</i> , 1339 , 1339	26	12	0.10	0.18	1342	1377	1372.2 vs	1376.8 vs	$\nu_s(\text{COO})$ + $\nu(\text{triazole})$ + 3, $\delta(\text{CH})$ + $\delta(\text{pyrrolidine})$
1337, 1336, 1336	7	16	0.05	0.09	1339	1374			$\nu_s(\text{COO})$ (49) + $\delta(\text{pyrrolidine})$ + $\nu(\text{triazole})$
1306, 1305, 1305	6	1	0.75	0.86	1312	1345	1343.1 s		$\delta_s(\text{C-H})$ in pyrrolidine + $\nu_{\text{as}}(\text{COO})$
1303, 1301, <i>1301</i>	3	19	0.75	0.86	1309	1342			$\nu_{\text{as}}(\text{CO}_{12})$ + 3, $\delta(\text{C-H})$ aryl + $\delta(\text{triazole})$
<i>1259</i> , 1248 , 1248	40	3	0.35	0.52	1263	1293	1301.8 vs		$\nu(\text{triazole})$ + $\nu_{\text{as}}(\text{CO}_{13})$ + $\Gamma(\text{pyrrolidine})$
1238, <i>1234</i> , 1233	41	7	0.75	0.86	1250	1279	1285.1 s		$\nu_{\text{as}}(\text{CO}_{13})$ + $\nu(\text{C-N})$ triazole + $\Gamma(\text{pyrrolidine})$
1216, 1215, 1215	3	3	0.75	0.86	1234	1262			$\nu(\text{CO}_{12})$ + $\nu(\text{C-N})$ triazole + $\delta(\text{C-H})$ in aryl
1189, 1188, 1188	1	0	0.75	0.86	1211	1237	1246.8 m		$\delta_s(\text{C-H})$ in pyrrolidine
1163, 1163, 1163	1	0	0.75	0.86	1190	1214	1218.1 m		$\gamma_{\text{as}}(\text{C-H})$ in pyrrolidine
1131, 1130, 1130	3	1	0.75	0.86	1161	1183	1178.3 m		$\gamma_{\text{as}}(\text{C-H})$ in pyrrolidine + $\nu_s(\text{NNN})$
1112, 1109, 1109	9	2	0.75	0.86	1143	1164		1171.5 w	$\nu_s(\text{COO})$ + $\nu(\text{NCCN})$ + $\gamma_s(\text{CC,CH})$
1011, 1010, 1010	4	1	0.75	0.86	1057	1071	1091.2 vs	1089.9 m	$\nu(\text{triazole})$ + $\nu(\text{C-CL})$ + $\nu_s(\text{COO})$ + 18a, $\delta(\text{CC,CH})$
956, 956, 956	0	0	0.73	0.84	1011	1020	1011.9 m	1012.8 w	$\gamma_{\text{as}}(\text{CC,CH})$ in pyrrolidine

(Continued on following page)

TABLE 5 (Continued) Calculated, scaled, and experimental wavenumbers (ν , cm^{-1}) in the $\text{Ce}(\text{2b}')_3$ complex by CAM-B3LYP and M06-2X methods. Relative infrared intensity (A) in %, relative Raman intensity (S) in %, and Raman depolarization ratios for plane (DP) and unpolarized incident light (DU). For each vibration of the tetramer, the wavenumber with the highest IR intensity is indicated in bold style and that with the highest Raman intensity is indicated in italic style. The relative IR and Raman intensities were only shown for these wavenumbers. DP and DU values were from the most intense Raman line. The number of the ring mode corresponds to Wilson's notation (Varsányi, 1974).

Calculated by CAM-B3LYP					scaled		Experiment		Characterization by CAM-B3LYP
ν	A	S	DP	DU	LSE	PSE	IR	Raman	
932, 931, 931	1	0	0.09	0.16	989	996	969.1 vs		ν_{as} (triazole) + ν_{s} (COO) + $\gamma(\text{CC},\text{CH})$
887, 887, 887	0	4	0.02	0.04	951	954			$\gamma(\text{CC})$ pyrrolidine + ν_{as} (NNN)
840, 839, 839	4	22	0.04	0.08	909	908	914.3 w	970.2 s	ν_{as} (NNN) + 12, $\delta(\text{CCC})$ in aryl
742, 742, 742	9	0	0.75	0.86	826	815	857 w-m		17b, $\gamma(\text{C-H})$ in aryl
713, 709 , 709	4	11	0.07	0.13	797	783	829.9 vs		γ_{s} (COO) + $\nu(\text{triazole})$ + $\gamma(\text{CC})$ pyrrolidine
567, 557 , 557	12	2	0.75	0.86	666	636	654 m		$\gamma(\text{triazole})$ + δ_{as} (COO) + 4, $\gamma(\text{CCC})$
555, 554, 554	4	1	0.75	0.86	663	633	647.1 m		$\Gamma(\text{triazole})$ + δ_{as} (COO) + 4, $\gamma(\text{CCC})$
474, 474, 473	16	2	0.05	0.10	594	554	509.0 m		δ_{as} (COO) + $\delta(\text{triazole})$ + $\delta(\text{CC})$ in aryl
371, 371 , 369	4	0	0.75	0.85	505	452	466.9 m		$\nu(\text{aryl}, \text{C-CL})$ + $\delta(\text{COO})$ + $\delta(\text{triazole})$
Calculated by M06-2X			Scaled		Experiment		Characterization by M06-2X		
ν	A	S	LSE	PSE	IR	Raman			
3326, 3326, 3325	37	1	3050	3036	2968.1 s		7b, $\nu(\text{C}_5\text{-H})$ in aryl (100)		
3314, 3314, 3313	3	1	3039	3027			20b, $\nu(\text{C}_6\text{-H})$ in aryl (100)		
3308, 3308, 3307	15	0	3034	3022			20a, $\nu(\text{C}_2\text{-H})$ in aryl (100)		
3223, 3223, 3223	6	5	2961	2957	2872.1 m		ν_{s} (C-H) in C_{15}H_2 pyrrolidine (100)		
1702, 1700 , 1698	46	20	1649	1699	1577.2 br, vs		$\nu(\text{C}_8\text{-N}_{14})$ + ν_{s} (NNN, NC) in triazole		
1652, 1651, 1651	0	1	1607	1655			8b, $\nu(\text{C}=\text{C})$ in aryl (93)		
1646, 1646, 1645	1	100	1603	1651		1595.0 vs	8a, $\nu(\text{C}=\text{C})$ in aryl (93)		
1584, 1550 , 1549	100	73	1520	1565	1500.1 s	1504.5 s	ν_{as} (COO) + $\nu(\text{C}_9\text{-C}_{11})$ + ν_{s} (CCN) triazole + $\nu(\text{C}_4\text{-N}_4)$		
1493, 1482, 1481	94	98	1461	1503	1484.4 vs		19a, $\nu(\text{CC}, \text{CH})$ + $\nu(\text{C}_4\text{-N}_4)$ + $\nu(\text{C}_9\text{-C}_{11})$		
1407, 1405 , 1405	5	10	1395	1434	1418.0 w		ν_{as} (NNN) in triazole + 19a, $\nu(\text{CC})$ in aryl		
1378, 1378, 1378	2	0	1372	1409	1398.3 m		19b, $\nu(\text{CC},\text{CH})$ in aryl		
1358, 1357, 1352	28	9	1349	1386	1372.2 vs	1376.8 vs	ν_{s} (COO) + ν_{s} ($\text{C}_9\text{-N}_{10}$) triazole + 19b, $\nu(\text{CC},\text{CH})$ in aryl		
1345, 1344, 1344	8	19	1343	1378			ν_{as} (NNN) + ν_{as} (COO) + 19a, $\nu(\text{CC})$ + $\delta(\text{CC})$ pyrrolidine		

(Continued on following page)

TABLE 5 (Continued) Calculated, scaled, and experimental wavenumbers (ν , cm^{-1}) in the $\text{Ce}(\text{2b}')_3$ complex by CAM-B3LYP and M06-2X methods. Relative infrared intensity (A) in %, relative Raman intensity (S) in %, and Raman depolarization ratios for plane (DP) and unpolarized incident light (DU). For each vibration of the tetramer, the wavenumber with the highest IR intensity is indicated in bold style and that with the highest Raman intensity is indicated in italic style. The relative IR and Raman intensities were only shown for these wavenumbers. DP and DU values were from the most intense Raman line. The number of the ring mode corresponds to Wilson's notation (Varsányi, 1974).

Calculated by M06-2X			Scaled		Experiment		Characterization by M06-2X
ν	A	S	LSE	PSE	IR	Raman	
1330, 1325 , 1325	2	14	1326	1361			$\nu_{\text{as}}(\text{COO}) + \nu_{\text{as}}(\text{triazole}) + \nu_{\text{as}}(\text{CNC})$ pyrrolidine + 19a, $\nu(\text{CC})$
1309, 1308, 1308	0	4	1312	1345	1343.1 s		$\delta_{\text{s}}(\text{C-H})$ in pyrrolidine + $\nu_{\text{s}}(\text{NNN})$
1283, 1274, 1269	81	11	1278	1309	1301.8 vs		$\nu(\text{triazole})$ (45) + $\nu(\text{CO}_{13})$ (40) + $\delta_{\text{s}}(\text{CC,CH})$ pyrrolidine (12)
1242, 1242, 1241	1	4	1255	1285	1285.1 s		$\nu_{\text{as}}(\text{CO}_{12}) + \nu_{\text{s}}(\text{C}_9\text{NN})$ triazole + $\delta_{\text{s}}(\text{CC,CH})$ in pyrrolidine
1193, 1192, 1192	1	1	1212	1238	1246.8 m		$\delta_{\text{s}}(\text{C-H})$ in pyrrolidine
1168, 1166 , 1166	5	1	1189	1214	1218.1 m		$\gamma_{\text{as}}(\text{C-H})$ in pyrrolidine + 3, $\delta(\text{CH}) + \nu_{\text{s}}(\text{NNN})$
1134, 1132, 1131	8	1	1159	1182			$\nu_{\text{s}}(\text{COO}) + \nu(\text{triazole}) + \delta_{\text{as}}(\text{C-H})$ in pyrrolidine
1119, 1119, 1118	1	1	1149	1170	1178.3 m	1171.5 w	$\gamma_{\text{as}}(\text{C-H})$ in pyrrolidine + $\nu_{\text{s}}(\text{triazole})$
1027, 1025, 1025	4	1	1068	1083	1091.2 vs	1089.9 m	$\nu_{\text{s}}(\text{NNN}) + \nu(\text{C-CL}) + 18\text{a}, \delta(\text{C-H})$ in aryl
959, 958, 958	0	1	1010	1019	1011.9 m	1012.8 w	$\gamma_{\text{as}}(\text{CC, CH})$ in pyrrolidine
945, 944, 944	3	1	998	1006			$\nu_{\text{as}}(\text{NC}_9\text{-C}_8) + \delta_{\text{as}}(\text{CC, CH})$ in pyrrolidine
921, 920, 920	3	1	977	983	969.1 vs		$\nu_{\text{as}}(\text{NNN}) + 18\text{a}, \delta(\text{C-H})$ in aryl
851, 851, 851	2	19	918	918		970.2 s	$\nu_{\text{as}}(\text{NNN})(36) + 12, \delta(\text{CCC})$ in aryl
751, 750, 750	7	0	831	821	857 w-m		17b, $\gamma(\text{C-H})$ in aryl
725, 724, 723	3	10	808	795	829.9 vs		$\gamma_{\text{s}}(\text{COO}) + \gamma(\text{CC, CH})$ pyrrolidine + $\delta(\text{triazole})$
575 , 573, 573	0	1	680	652	654 m		$\gamma(\text{triazole}) + \gamma_{\text{as}}(\text{COO}) + 8\text{b}, \delta(\text{CCC})$ in aryl
572, 565, 564	11	2	671	641	647.1 m		$\Gamma(\text{triazole}) + \gamma_{\text{as}}(\text{COO})$
478, 476, 475	16	2	594	554	509.0 m		$\delta_{\text{as}}(\text{COO}) + \delta(\text{triazole}) + \delta(\text{CC})$ in aryl
372, 370, 363	4	1	498	443	466.9 m		$\nu(\text{aryl, C-CL}) + \delta_{\text{as}}(\text{COO}) + \delta(\text{triazole})$

descriptors were determined, which facilitates a better understanding of the stability and reactivity of the four Ce(III) complexes studied in the present work. The energy gap (Eg) between HOMO and LUMO frontier orbitals (HOMO–LUMO gap) appears as one of the meaningful characteristics of molecules, and it facilitates the characterization of chemical reactivity and kinetic stability. By all methods, the Eg value was linearly related to the Ce atomic charge (Figure 3C) and it decreases as the complexity of the system increases, with the exception of the Ce (2b')₃ complex that has a similar value to the C complex. A high Eg value shows that the molecule (or complex system) is less polarizable, which is generally related to low chemical reactivity and hence - high kinetic stability. Since the Eg values decrease with the increase in system complexity and the lowest value appears in the C complex, it means that this C complex is more reactive than Ce (2b')₃, with the A complex being the most stable one. The pyrrolidine ligand appears to slightly reduce the reactivity of the Ce (2b')₃ complex.

The low Eg values obtained in the Ce (2b')₃ and C complexes with the four DFT methods used reveal their noticeable chemical reactivity, and, therefore, small energies are required for excitation.

With the HOMO and LUMO energies, several global chemical reactivity descriptors were calculated according to the following well-known formulas (Equations 11–15):

$$IP = -E_{HOMO}, \quad (11)$$

$$EA = -E_{LUMO}, \quad (12)$$

$$\chi = -(E_{HOMO} + E_{LUMO})/2, \quad (13)$$

$$\eta = (E_{LUMO} - E_{HOMO})/2, \quad (14)$$

$$S = \frac{1}{2} \eta. \quad (15)$$

Due to the large reactivity of the Ce (2b')₃ complex, its computed ionization potential (IP) value appears slightly lowered with the four DFT methods used. It is noted that the calculations by CAM-B3LYP and M06-2X methods lead to the same value, 0.390 eV, remaining very close to those determined by B3LYP and D3-B3LYP methods. Compared to the other Ce(III) complexes (A, B and C), these values are the lowest, increasing as complexity decreases (the Ce atomic charge decreases).

Electron affinity (EA) appears to be noticeably lower than IP, and it slightly increases in the complexes as complexity decreases, as expected. The M06-2X method predicts higher values than the CAM-B3LYP method. The low electronegativity (χ) observed is in agreement with neutral complexes, and it slightly increases as complexity decreases. The chemical hardness (η) and global softness (S) values indicate the opposition of a system to an alteration in its number of electrons. These values are low in the investigated complexes, and they slightly increase as complexity decreases. When the values of η are low, the system (or the complex system in our case) is named soft; when they are high, the system is identified as hard. Therefore, according to the low values obtained in our complexes, they are soft systems, with small gaps and an easily modified electron density.

In addition, the significant difference should be noted in the global chemical reactivity descriptors of the complex Ce (2b')₃ and the same complex with La³⁺, the La (2b')₃. For instance, all level results showed an increase in the dipole moment of the cerium complex. Such a deviation in properties can lead to different

biological behavior and both qualitative and quantitative differences in their interaction with biological targets.

3.4 Vibrational analysis

The confirmation of the proposed molecular structure of the synthesized complex Ce (2b')₃ was achieved through a detailed analysis of the calculated and experimental IR and Raman spectra. For this purpose, a first theoretical–experimental comparison of the full IR spectra in the 3,750–400 cm^{−1} range has been carried out, and it is shown in [Supplementary Figure S4](#) with the scaled values obtained by the four DFT levels used. The same comparison but with the Raman spectra and in the 3,750–0 cm^{−1} range is included in [Supplementary Figure S5](#). The identification and characterization of all vibrations in the 3,500–400 cm^{−1} range is collected in [Supplementary Table S1](#) for the calculations with the CAM-B3LYP method, while the results with the M06-2X method are included in [Supplementary Table S2](#). All values correspond to the most stable conformer, with the arrangement of the ligands shown in [Figure 1](#). Because these tables are too long, a resume is included in [Table 5](#), including only the frequencies with high IR or Raman intensities, as well as those characteristics of the complex. [Table 5](#) is divided into two parts corresponding to the results by the CAM-B3LYP and M06-2X methods.

3.4.1 General comparison of the IR and Raman spectra

In the first comparison of the scaled IR spectra with the experimental spectra plotted in [Supplementary Figure S4](#), the following observations are pointed:

- (i) A good agreement between the scaled theoretical spectra with the experimental spectra has been noted. In particular, the scaled strongest vibrations have their corresponding vibrations in the experimental spectrum. This feature confirms the scaling carried out on the calculated wavenumbers; therefore, the applied theoretical methods appear appropriate. Thus, in general, the assignments proposed could be considered true, identifying most of the computed modes in their normal ranges.
- (ii) A very broad and strong band centered at 3400.8 cm^{−1} has been observed in the experimental spectrum. Judging by its position and broad intensity, it can be only assigned to the O–H stretching ν mode corresponding to water molecules strongly H-bonded to the nitrogen and oxygen atoms of the three 2b' ligands in the Ce (2b')₃ complex. These water molecules were not included in our optimized theoretical complex due to the fact that in previous studies with La(III) complexes, the water molecules only slightly affected the carboxylate group, the other groups remaining unaffected ([Peica et al., 2006](#)). This hydration appears due to the spatial arrangement of the ligands in the complex, which leaves cavities that can be occupied by water molecules. As expected, this band is not observed in the Raman spectrum.
- (iii) Another broad but very strong experimental IR band is observed at 1577.2 cm^{−1}. Its large broadening can be interpreted as a result of the additional contribution of

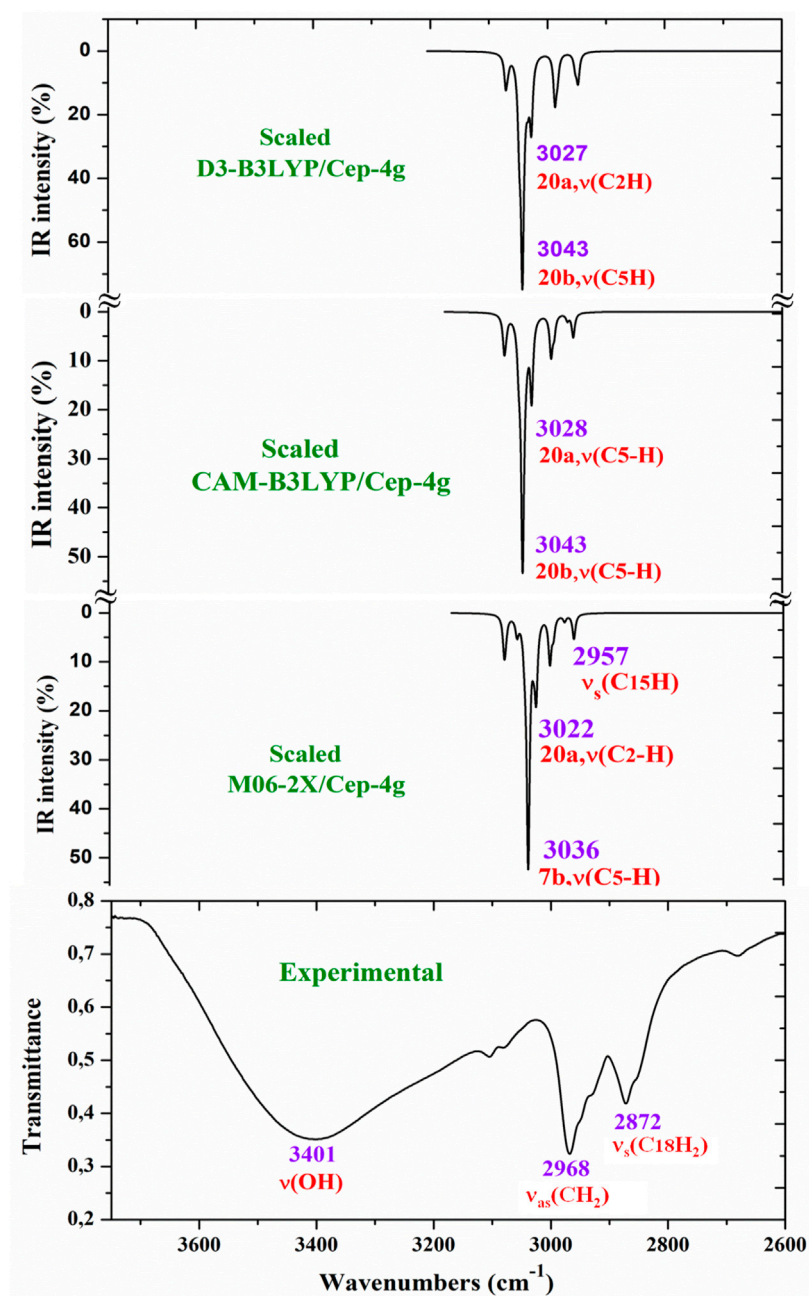


FIGURE 4 Comparison of the scaled IR spectra with the B3LYP, CAM-B3LYP, and M06-2X methods by the PSE procedure with the experimental spectrum in the 3750–2600 cm^{-1} range.

the in-plane bending $\delta(\text{OH})$ mode of these hydrated water molecules to the main assignment of this band corresponding to the $\text{C}_8\text{-N}_{14}$ and C-C stretching, as shown in Table 5.

- (iv) A noticeable resemblance between the scaled spectra obtained by the D3-B3LYP, CAM-B3LYP, and M06-2X methods that include long-range correction has been observed, while that by B3LYP differs noticeably. As compared to the experimental spectrum, the two best methods are CAM-B3LYP and M06-2X, and for this reason, their spectra were analyzed in detail and are

included in Table 5. In this analysis, the scaled wavenumbers by CAM-B3LYP appear slightly more accurate than by M06-2X. This is why CAM-B3LYP has been mainly utilized in the experimental spectra assignment.

- (v) The coordination of the 2b' ligands to the Ce(III) ion remarkably modified the IR and Raman spectra. The spectra appear to differ from those obtained with the 2b' ligand anion alone (Alcolea Palafox et al., 2023a).

The same analysis has been carried out using the Raman spectra, as shown in Supplementary Figure S5. In this case, the scaled spectra

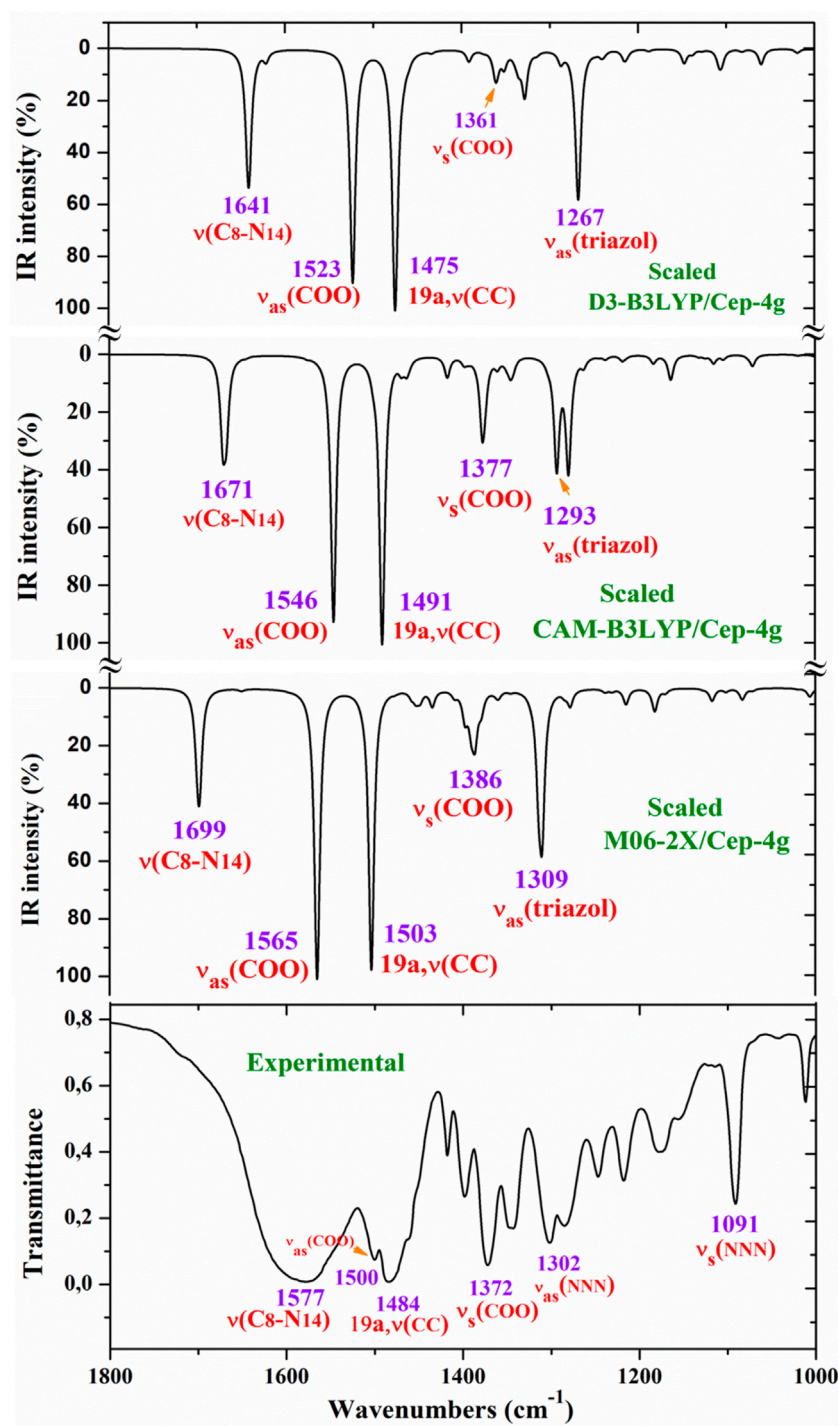


FIGURE 5
Comparison of the scaled IR spectra at three different levels by the PSE procedure with the experimental spectrum in the 1800–1000- cm^{-1} range.

show large differences among the DFT methods. CAM-B3LYP and M06-2X methods again seem to be more accurate with bands showing good agreement with the experimental bands. Unfortunately, the experimental spectrum appears with noticeable background noise, which complicates the detection and further analysis of all weak lines, making this comparison difficult. The appearance of a broad Raman line with medium intensity at 73.1 cm^{-1} was observed in the experimental

spectrum, which has not been detected in the theoretical spectra. Due to this feature, it was not shown in Table 5.

3.4.2 Specific comparison of the IR and Raman spectra

For a comprehensive and specific comparison of the distinct scaled and experimental frequencies, the spectra were divided into three ranges. In the IR spectra, these ranges were 3750–2600 cm^{-1}

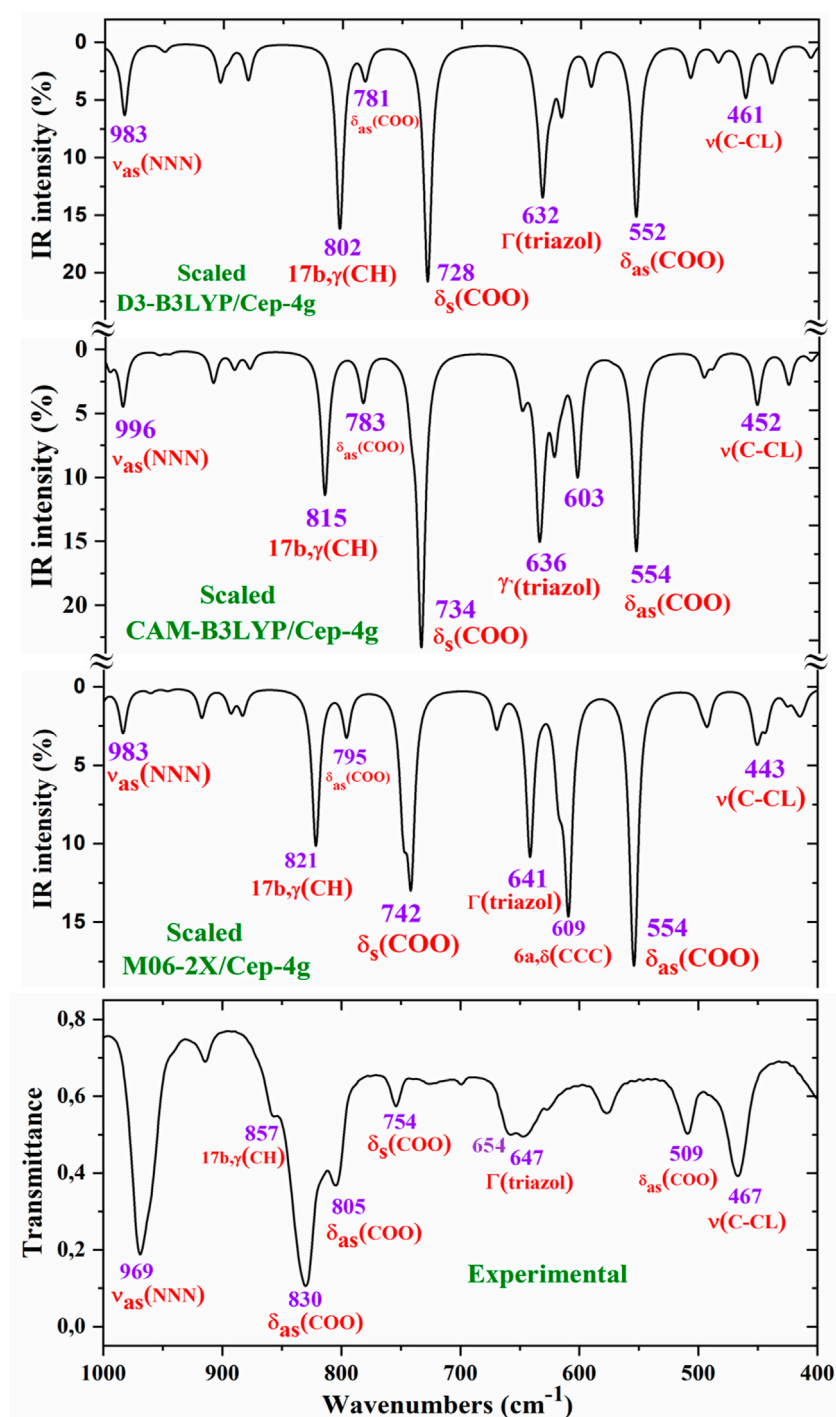


FIGURE 6 Comparison of the scaled IR spectra at three different levels by the PSE procedure with the experimental spectrum in the 1000–400-cm⁻¹ range.

(Figure 4), 1800–1000 cm⁻¹ (Figure 5), and 1000–400 cm⁻¹ (Figure 6). In the Raman spectrum, the comparison was only carried out in the 1800–800-cm⁻¹ range, as shown in Figure 7, in which the Raman lines can be clearly identified, because of the noticeable background noise of the experimental spectrum. In these figures, for simplicity, only the assignment of strong and characteristic vibrational modes was included. For the final assignment of the experimental bands, characterization obtained

by the different DFT methods was considered, as well as the assignment reported in related molecules.

For this purpose, a resume with the most characteristic frequencies determined in the Ce (2b')₃ complex with the CAM-B3LYP and M06-2X methods is shown in Table 5. The first column presents the three computed frequencies for each mode related to the three 2b' ligands of the complex. Of these three frequencies, the frequency with higher computed IR intensity was typed in bold style,

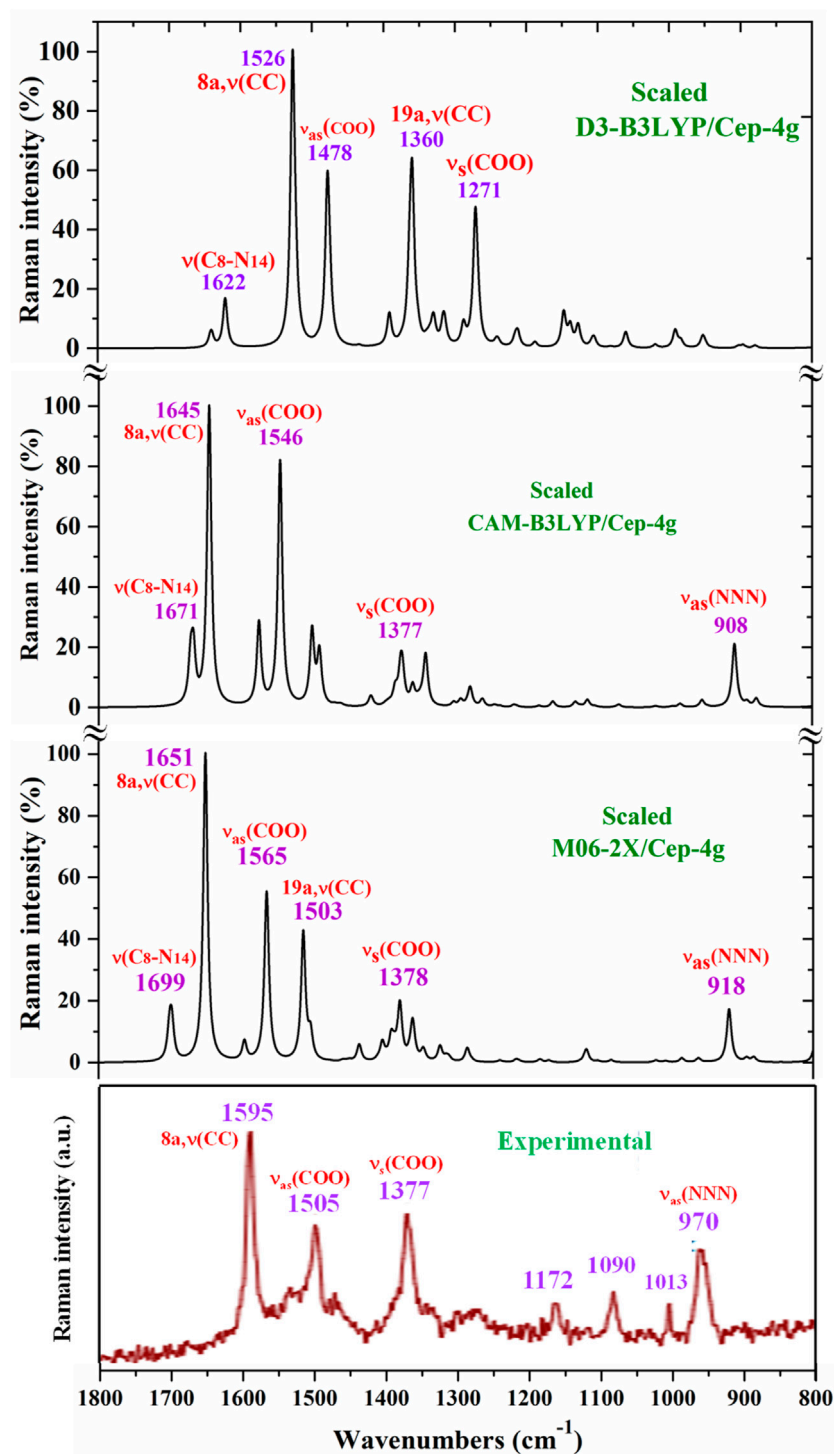


FIGURE 7
Comparison in the 1800–800- cm^{-1} range of the scaled Raman spectrum by the PSE procedure and at three different levels with the experimental spectrum.

whereas the frequency with higher Raman intensity was typed in italic style. This notation was not used if the intensity of these three wavenumbers is similar or very weak. The second and third columns list the relative IR and Raman intensities (%) corresponding to the frequencies of the first column typed in bold style and in italic style, respectively. These relative intensities were

determined by normalizing each computed data to the strongest intensity of the spectrum. In the CAM-B3LYP method was included two new columns, the fourth and fifth, with the Raman depolarization ratios for plane (DP) and unpolarized incident light (DU), respectively.

The values included correspond to the wavenumbers typed in italic style in the first column. The scaled wavenumbers by the LSE

or PSE procedures were listed in the next two columns. For simplicity, only the wavenumbers typed in bold style are included as scaled values in these two columns. The observed experimental IR and Raman bands with their corresponding intensities were collected in the next two columns, respectively. Finally, the last column shows the principal characterization of the computed wavenumbers determined by the CAM-B3LYP (first part of the Table) and M06-2X (second part of the Table) methods. The % contribution (PEDs) of the distinct modes to a calculated wavenumber was added only in few cases.

The main purpose of this vibrational study is the identification and characterization of the synthesized Ce (2b')₃ coordination complex. For this reason, the main focus was on the most characteristic modes and the strongest IR and Raman bands, in order to validate the structure that was planned and optimized in Figure 1. Since the spectra with the CAM-B3LYP method seem to be the most precise, their frequencies have been mainly used for discussion. In specific cases, the scaled values obtained by the M06-2X method were also discussed. For simplicity, the scaled values obtained through the PSE scaling procedure were mainly used in the discussion due to their slightly better agreement with experimental results than those with the LSE procedure.

The analysis and assignment of different vibrational modes have been carried out under the following sections: 1) the COO[−] group modes, 2) the triazole ring modes, and 3) the aryl ring modes. Due to the low quality of the Raman spectra, the vibrational modes with the Ce(III) ion were not detected in the experimental spectrum, and therefore, they were not discussed in the present work.

3.4.2.1 The carboxylate COO[−] group modes

The vibrations of this group appear to be of great interest because it has been found (Palafox et al., 2023) that the ligand is fixed to the amino acid chain of the receptor through this group. In solid state samples (Sokrates, 2004), the asymmetric ν_{as} (COO[−]) mode of the carboxylate group appears as a strong IR band in the 1600–1560 cm^{−1} range, while the ν_s (COO[−]) mode appears at lower wavenumbers, although the spectrum of this region is complicated due to intermolecular H-bonding. In the 2b ligand alone, ν_{as} (COO[−]) stretching has been scaled at 1713 cm^{−1} with very strong IR intensity (Palafox et al., 2023). Nevertheless, in the Ce (2b')₃ complex, a noticeable red shift to lower values is expected due to the remarkably longer C–O bonds used to form the six O–Ce coordination bonds. In the CAM-B3LYP method, a large contribution of the asymmetric ν_{as} (COO[−]) mode was identified in the calculated wavenumber at 1525 cm^{−1} (scaled at 1503 cm^{−1} via the LSE procedure) with very strong IR intensity, the second strongest of the spectrum, in very good agreement with the experimentally strong IR band at 1500.1 cm^{−1} and with the strong Raman line at 1504.5 cm^{−1}. This feature was also observed with the M06-2X method, appearing scaled at 1520 cm^{−1} by LSE, which is in good agreement with the experimental spectrum. Large contributions of the asymmetric stretching vibrations ν_{as} (CO₁₂) and ν_{as} (CO₁₃) were also identified in the scaled wavenumbers by CAM-B3LYP at 1342, 1293, 1279, and 1262 cm^{−1}, which are in good agreement with the experimentally very strong IR band at 1301.8 cm^{−1} and with the strong band at 1285.1 cm^{−1}. The M06-2X method confirms these assignments in the scaled values at 1309 and 1285 cm^{−1}, respectively.

The largest contribution of the symmetric ν_s (COO[−]) mode was found in the calculated wavenumber at 1336 cm^{−1} (scaled at 1374 cm^{−1} via the PSE procedure), which is consistent with that found in the 1420–1400-cm^{−1} range in solid-state samples of related compounds (Sokrates, 2004). A large contribution of the asymmetric stretching vibration ν_s (COO[−]) was clearly identified by CAM-B3LYP in the scaled wavenumber by PSE at 1377 cm^{−1} with strong IR intensity and medium Raman intensity, which is in excellent agreement with the very strong bands at 1372.3 (IR) and 1376.8 cm^{−1} (Raman). The M06-2X method confirms this assignment with scaled wavenumbers at 1386 and 1378 cm^{−1}.

3.4.2.2 Triazole ring modes

The triazole group vibrations are also of great interest as their nitrogen atoms appear to be weakly H-bonded to the amino acids of the receptor (Palafox et al., 2023). The geometry and vibrational frequencies of the 1,2,3-triazole ring in different triazoles have already been studied (Aziz et al., 2014; El-Azhary et al., 1998; Törnkvist et al., 1991), and our computations and experimental results are in good agreement with them. Therefore, for simplicity, the analysis of different vibrations of this group has only been focused in the assignment of the strongest bands.

NNN modes: The ν_{as} (NNN) stretching was characterized by CAM-B3LYP as strongly coupled with the 19a aryl mode in the scaled frequency with weak IR intensity at 1417 cm^{−1}, which is in good agreement with the weak experimental IR band at 1418.0 cm^{−1}. The results by M06-2X are in good agreement with this assignment with a scaled wavenumber at 1434 cm^{−1} and also with weak IR intensity. Due to the large background noise of the experimental Raman spectrum, the weak lines predicted for this mode were not clearly identified.

Other triazole ring stretching modes with asymmetric character and strongly coupled with other modes were identified in the scaled wavenumbers at 996 and 908 cm^{−1} and are related to the very strong experimental IR band at 969.1 cm^{−1} and weak band at 914.3 cm^{−1}, respectively. In the experimental Raman spectrum, this mode appears as a strong line at 970.2 cm^{−1} in accordance with the medium intensity vibration predicted by M06-2X and CAM-B3LYP methods. D3-B3LYP dramatically fails in its prediction, as well as in other bands, and for this reason, their values were not considered. The Raman line detected at 1012.8 cm^{−1} was also assigned to this mode.

The symmetric stretching ν_s (NNN) mode was identified at lower wavenumbers and strongly coupled with the symmetric ν_s (COO[−]) mode, as well as with other ring modes. It was predicted by CAM-B3LYP with medium–strong IR intensity at 1377 cm^{−1} and corresponds to the very strong IR band at 1372.2 cm^{−1}. The very strong experimental Raman line observed at 1376.8 cm^{−1} can be assigned to this mode or to the neighbor scaled wavenumber at 1374 cm^{−1}. A similar scaled wavenumber at 1378 cm^{−1} was calculated by the M06-2X method. In the 2b ligand alone (isolated state), it has been scaled at 1360 cm^{−1} and assigned to the experimental IR band at 1340.5 cm^{−1} (Palafox and Rastogi, 2013).

C₈–N₁₄ modes: The stretching mode has been scaled via the LSE procedure with CAM-B3LYP at 1623 cm^{−1} with strong IR intensity, and it corresponds to the very strong and broad band observed in the experimental IR spectrum at 1577.2 cm^{−1}. This difference in the wavenumbers and the broadening of the experimental IR band can

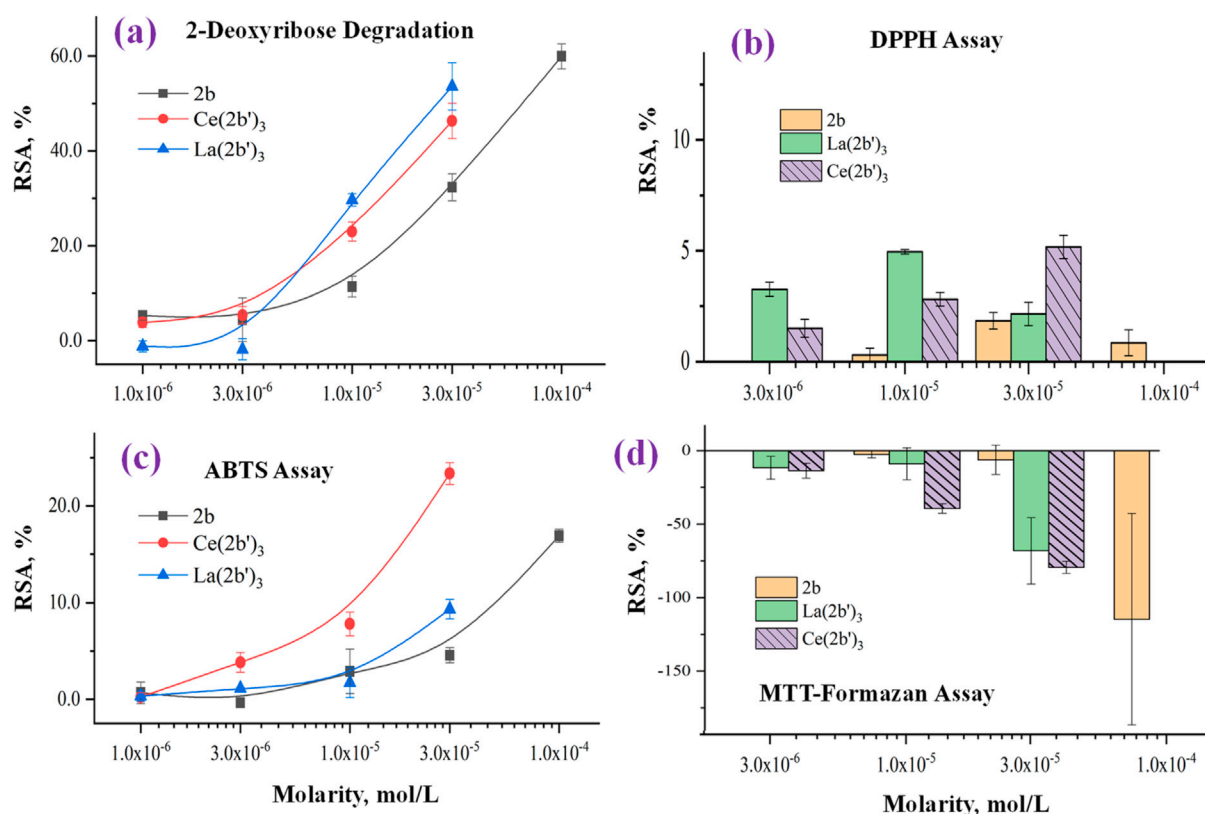


FIGURE 8
Impact of 2b, Ce (2b')₃, and La (2b')₃ on (A) 2-deoxyribose degradation, (B) DPPH[•], (C) DPPH[•], and (D) MTT-formazan transformation by Fenton-generated OH[•]. Data = mean ± StDev, *p* < 0.05, *N* = 3. A higher result means higher scavenging activity.

be due to the contribution of the δ (O-H) mode corresponding to the hydrated water molecules. This assignment was in agreement to that found by M06-2X in the scaled wavenumber at 1649 cm⁻¹ by the LSE procedure. This stretching mode was identified in the isolated 2b ligand in the scaled wavenumber at 1556 cm⁻¹ (Palafox et al., 2023), which is in good agreement with the experimental bands with medium intensities at 1543.9 cm⁻¹ (IR) and 1550.6 cm⁻¹ (Raman).

3.4.2.3 Aryl ring modes

The aryl and the pyrrolidine groups appear to provide liposolubility to the complex, facilitating its membrane cell crossing, and in addition, the aryl group establishes π - π and π -alkyl interactions, stabilizing the ligand binding to the amino acids of the receptor (Palafox et al., 2023). For this reason, the vibrations of this aryl group are analyzed in the present study. For the assignments of the ring modes, the Varsanyi notation (Varsanyi, 1974) for a 1,4-disubstituted benzene derivative was followed. Therefore, the substituent modes of the aryl ring for stretching vibrations correspond to 7a and 13 modes, for in-plane vibrations to 9b and 15 modes, and for out-of-plane vibrations to 10b and 11 modes. In particular, the C4-N4 bond is represented by 13, 15, and 10b modes, while the C-CL bond is represented by 7a, 9b, and 11 modes. In a chloro-substituted benzene derivative (Palafox and Rastogi, 2013), mode 7a ν (C-CL) was found at 359 cm⁻¹, mode 9b δ (C-CL) at 310 cm⁻¹, and mode 11 γ (C-CL) at 93 cm⁻¹.

The assignments of many of the remaining ring modes are shown in Table 1S and do not need further analysis; thus, interest was directed only to the strongest vibrations in order to confirm the structure of the synthesized complex.

The aromatic C-H stretching modes appear mainly in the 3200–2950 cm⁻¹ region theoretically as almost pure modes (100% PED) with weak or very weak IR and Raman intensities. Thus, only mode 20b with a scaled value by the PSE procedure at 3043 cm⁻¹ correlated with the experimental IR band with strong intensity at 2968.1 cm⁻¹. This wavenumber is exactly the same as the wavenumber found in an isolated 2b ligand (Palafox et al., 2023), which can be explained by the negligible effect of the Ce(III) ion on the far aryl ring.

The aromatic C-C stretching vibrations, modes 8a and 8b, appear as nearly pure modes but with %PED of approximately 90%. Through the CAM-B3LYP method, mode 8a was scaled by the LSE procedure at 1598 cm⁻¹ with practically null IR intensity, but being the second vibration with the highest Raman intensity, which is in excellent agreement with the very strong Raman line at 1595.0 cm⁻¹. A similar scaled wavenumber at 1603 cm⁻¹ is predicted by M06-2X. Mode 8b was calculated by both DFT methods with practically null IR and Raman intensities, and therefore, it has not been detected in the experimental spectra.

Mode 19a was calculated by both DFT methods with very strong IR and Raman intensities at 1491 cm⁻¹, which is in excellent agreement with the experimental strong IR band at 1484.4 cm⁻¹.

Mode **19b** was scaled at 1398 cm^{-1} with weak IR and very weak Raman intensities, which is in agreement with the experimental IR band with medium intensity detected at 1398.3 cm^{-1} .

3.5 Free radical-scavenging activity

Oxidative stress (OS) is a physiological state associated with an imbalance between the production and elimination of reactive species in a living organism. In terms of human health, OS is generally associated with the development of a number of pathologies (Hajam et al., 2022). However, there are some cases in which OS is beneficial, particularly in cancer treatment (Van Loenhout et al., 2020), immunological response (Lauridsen, 2019), and/or pharmacological treatment (Albesa et al., 2004) of microbial diseases. Lanthanide ions and their coordination compounds are known for their antimicrobial (Cota et al., 2019; Wakabayashi et al., 2016) and anticancer properties (Patyal et al., 2023), which are associated with their luminescent properties (Zhang et al., 2023) and ionic mimicry, particularly toward iron (Todorov et al., 2019). Herein, the interaction between the novel complex $\text{Ce}(\text{2b}')_3$ and two hydroxyl radical generating model systems that involve UV-induced water radiolysis (Burns and Sims, 1981) and the physiologically significant Fenton reaction (Levin et al., 2016) is reported. Hydroxyl radicals are highly aggressive reactive species that tend to “attack” double bonds in biomolecules, causing lipid peroxidation and radical chain reactions and, therefore, can ultimately inflict a variety of pathologies in living organisms (Negre-Salvayre et al., 2010). 2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic) free radical (ABTS^+) and 2,2-diphenyl-1-picrylhydrazyl radical (DPPH $^\cdot$) assays were additionally performed to evaluate the ability of $\text{Ce}(\text{2b}')_3$ to scavenge radicals by single-electron transfer (SET) and hydrogen atom transfer (HAT) reactions, respectively. The radical-scavenging activities of the sodium salt **2b** and its La(III) complex, $\text{La}(\text{2b}')_3$, have previously been reported (Alcolea Palafox et al., 2023a) and are presented together with the new results for $\text{Ce}(\text{2b}')_3$ in order to observe the impact (if any) of the type of coordination center on radical-scavenging activity.

3.5.1 Impact of $\text{Ce}(\text{2b}')_3$ on 2-deoxyribose degradation

The impact of the Ce(III) complex on 2-deoxyribose degradation as a result of UV-induced water radiolysis is presented in Figure 8A. The observed effect with $\text{Ce}(\text{2b}')_3$ is concentration-dependent, as is the case with the previously reported **2b** and $\text{La}(\text{2b}')_3$. At concentrations below $1 \cdot 10^{-5}\text{ M}$, the activity of the complex is very mild ($\text{RSA} < 5\%$). At $1 \cdot 10^{-5}\text{ M}$ and higher, the cerium complex is more active than the ligand in the same concentration. At $3 \cdot 10^{-5}\text{ M}$, $\text{Ce}(\text{2b}')_3$ and $\text{La}(\text{2b}')_3$ have similar activity ($\text{RSA} = 46 \pm 4\%$ and $54 \pm 5\%$, respectively). The observed impact of the coordination center in this model system appeared to be very mild, with both complexes exhibiting similar activity within the tested range of molarities.

3.5.2 Impact of $\text{Ce}(\text{2b}')_3$ on a model system containing the stable radical DPPH $^\cdot$

The ability of $\text{Ce}(\text{2b}')_3$ to exchange hydrogen with DPPH $^\cdot$ is presented in Figure 8B. Previous experiments have shown that

the ligand **2b** and the complex $\text{La}(\text{2b}')_3$ manifest very mild activity in this model system - $\text{RSA} < 5\%$ at even the highest concentration with the effect decreasing to 0 at $3 \cdot 10^{-6}\text{ M}$ and lower. The case of $\text{Ce}(\text{2b}')_3$ is very similar. At $3 \cdot 10^{-5}\text{ M}$, this complex has $\text{RSA} = 5.2 \pm 0.5\%$. The effect decreases in a concentration-dependent manner to $1.5\% \pm 0.4\%$ at $3 \cdot 10^{-6}\text{ M}$. DPPH $^\cdot$ is scavenged by HAT, which is similar to OH $^\cdot$ (Galano, 2015). However, its relatively large size may be implicated in the observed low activity in this model system when compared to the 2-deoxyribose degradation model. $\text{Ce}(\text{2b}')_3$ is more active than **2b** at three times higher molarity, a sign that Ce(III) possibly influences the distribution of electron density within the coordinated ligands, increasing the activity of hydrogen atoms that, in turn, could interact with DPPH $^\cdot$.

3.5.3 Impact of $\text{Ce}(\text{2b}')_3$ on a model system containing the stable radical ABTS^+

The ability of the novel complex to participate in electron-exchange reactions with ABTS^+ is shown in Figure 8C. A concentration-dependent impact is observed with RSA increasing from $0.24 \pm 0.44\%$ at $1 \cdot 10^{-6}\text{ M}$ to $23 \pm 1\%$ at $3 \cdot 10^{-5}\text{ M}$. At the same molarities, the activity of $\text{Ce}(\text{2b}')_3$ is higher than that of **2b** and even the lanthanum(III) counterpart at $3 \cdot 10^{-6}\text{ M}$ or more. The activity of the Ce(III) complex tends to be higher than that of **2b** at three times higher concentration. At the highest tested molarity ($3 \cdot 10^{-5}\text{ M}$) $\text{Ce}(\text{2b}')_3$ has $\text{RSA} = 23 \pm 1\%$, which is higher compared to $9 \pm 1\%$ for $\text{La}(\text{2b}')_3$ ($3 \cdot 10^{-5}\text{ M}$) and $16.9 \pm 0.7\%$ for the ligand **2b** ($1 \cdot 10^{-4}\text{ M}$). Unlike La(III), the complexation of **2b** with Ce(III) seems to promote electron-exchange with ABTS^+ rather than suppress it.

3.5.4 Impact of $\text{Ce}(\text{2b}')_3$ on MTT-formazan transformation triggered by Fenton reaction-derived hydroxyl radicals

The Fenton reaction is a clinically significant chemical process (Kell, 2009) that involves the transition metal-catalyzed production of hydroxyl radicals from H_2O_2 . OH $^\cdot$ are highly reactive species that tend to attack molecular sites with conjugated double bonds, causing molecular fragmentation, lipid peroxidation, and generation of malondialdehyde-like products. Since both the Fenton reaction and water radiolysis produce OH $^\cdot$, similar activity in both model systems would also be expected. This, however, is not the case for $\text{Ce}(\text{2b}')_3$, which is consistent with previously published observations with **2b** and $\text{La}(\text{2b}')_3$. Results are presented in Figure 8D.

Previous observations demonstrate that in this model system, the ligand **2b** is inert at $3 \cdot 10^{-5}\text{ M}$ or less. At $1 \cdot 10^{-4}\text{ M}$, it actually increases MTT-formazan formation to a significant degree ($\text{RSA} = -115 \pm 72\%$), a sign of pro-oxidant action. At $3 \cdot 10^{-5}\text{ M}$, $\text{Ce}(\text{2b}')_3$ also seems to act as a pro-oxidant ($\text{RSA} = -79 \pm 4\%$), which is similar to the previously reported $\text{La}(\text{2b}')_3$ ($\text{RSA} = -68 \pm 23\%$). This effect decreases in a concentration-dependent manner. At $1 \cdot 10^{-5}\text{ M}$, $\text{Ce}(\text{2b}')_3$ seems to behave as a slightly more potent prooxidant compared to $\text{La}(\text{2b}')_3$ ($\text{RSA} = -39 \pm 3\%$ and $9.0 \pm 10.9\%$ respectively). $\text{Ce}(\text{2b}')_3$ seems to act as a more potent prooxidant than **2b** at concentrations $3 \cdot 10^{-5}\text{ M}$ and lower. At $3 \cdot 10^{-6}\text{ M}$ and $1 \cdot 10^{-5}\text{ M}$, it increases MTT-formazan

transformation to a greater extent than the ligand 2b at three times greater molarity.

3.5.5 Chemiluminometric assays: Impact of Ce (2b')₃ on LDCL in the presence of KO₂ and NaOCl

In previous work [Alcolea Palafox et al. \(2023a\)](#), the effects of 2b and La (2b')₃ on LDCL are reported in the presence of NaOCl and KO₂. The ligand and its La (III) complex behaved as weak pro-oxidants in the presence of superoxide. In the NaOCl model system, the ligand seems to show a tendency to behave as a mild concentration-dependent antioxidant (CL-SI = 64 ± 40%), while the La(III) complex behaved as a pro-oxidant (CL-SI = 186 ± 6%, 3·10⁻⁵ M). The methodologies involved dissolving the compounds in 100–200 µL of DMSO and further diluting to 10 mL of distilled water. In order to maintain experimental conditions as close to physiological conditions as possible, the experiments were repeated, this time without the application of DMSO, a compound known for its antioxidant activity. The results of the 2b and La (2b')₃ experiments (without DMSO),

as well as the research on Ce (2b')₃, using the updated protocol are shown below.

The superoxide ion is an oxygen radical that is produced in the human body by one-electron reduction of oxygen. It also participates in the body's defense against pathogens and in a multitude of cell signaling pathways ([Che et al., 2016](#); [Trist et al., 2021](#)). The ability of the ligand 2b, La (2b')₃, and Ce (2b')₃ to scavenge KO₂-derived superoxide is presented in [Figure 9A](#).

All three compounds are almost inert at 3·10⁻⁶ M or lower. As molarities increase, LDCL in the presence of 2b and La (2b')₃ increases up to CL-SI = 205 ± 6% for 2b at 1·10⁻⁴ M and CL-SI = 136 ± 8% for La (2b')₃ at 3·10⁻⁵ M, a sign of a the pro-oxidant effect. Conversely, Ce (2b')₃ at 3·10⁻⁵ M scavenges superoxide with CL-SI = 69 ± 3%.

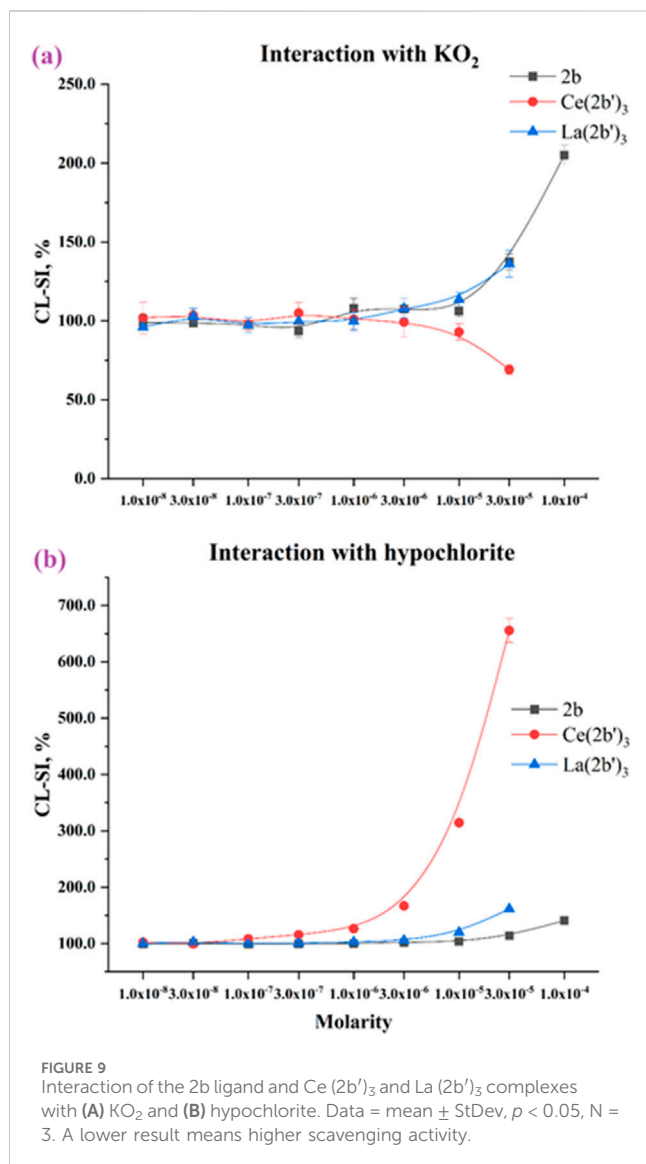
Hypochlorous acid is part of the human body's immune defense ([Pattison and Davies, 2006](#)), a product of the enzyme myeloperoxidase. It acts as a strong, oxidizing, non-specific bactericide, associated with multiple pathologies ([Strzepa et al., 2017](#)). The impact of 2b, La (2b')₃, and Ce (2b')₃ on LDCL in the presence of sodium hypochlorite is presented in [Figure 9B](#). All three compounds behave as significant pro-oxidants. At 1·10⁻⁴ M, 2b has CL-SI = 140 ± 2% and at 3·10⁻⁵ M La (2b')₃ has CL-SI = 161 ± 2%. At the highest tested molarities, the effect of the La(III) complex (at 3·10⁻⁵ M) is higher than that of the ligand at three times the concentration (1·10⁻⁴ M). Considering that each complex bears three ligands, it seems that the La(III) ion potentiates the pro-oxidant behavior of 2b in this model system. Most striking are the results of the cerium complex. Pro-oxidant behavior is observed at molarities as low as 1·10⁻⁷ M (CL-SI = 115 ± 1%), increasing in a concentration-dependent manner to as high as CL-SI = 656 ± 21% at 3·10⁻⁵ M.

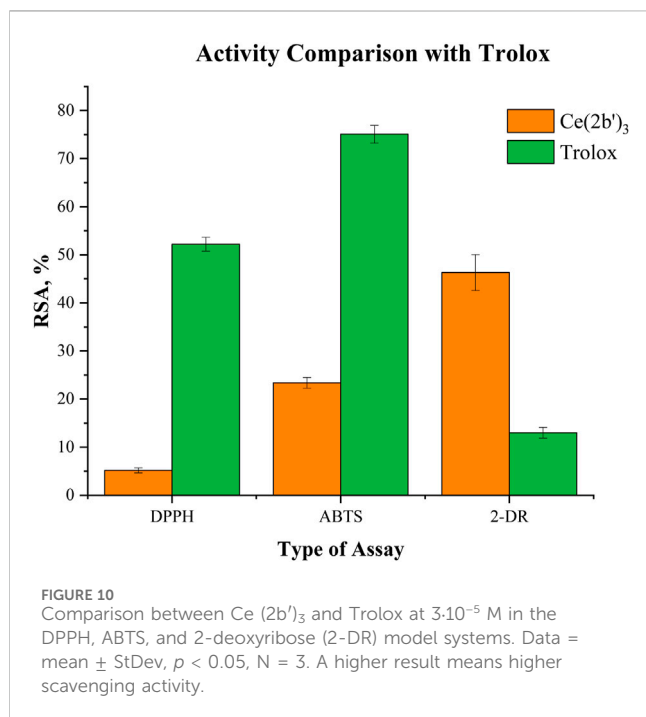
3.5.6 Comparison between the activity of Ce (2b')₃ and Trolox

A well-established positive control (Trolox) was tested at 3·10⁻⁵ M in order to compare its activity with the activity of the cerium complex at the same molarity. As Trolox does not dissolve well in bi-distilled water, PBS (pH = 7.45) was utilized to produce a clear 1·10⁻³ M stock solution. Since ferric ions form a precipitate in PBS, Trolox could not be tested in the Fenton reaction model system. Results obtained from the spectrophotometric assays are presented in [Figure 10](#).

Compared to Trolox, the cerium complex manifests significantly lower DPPH scavenging activity (RSA = 52.2 ± 1.4% versus 5.2 ± 0.2%). A similar trend is observed in the ABTS assay: RSA = 75.1 ± 1.8% for Trolox and RSA = 23.4 ± 1.1% for Ce (2b')₃. Those results contrast with observations from the 2-deoxyribose degradation assay. OH· scavenging activity of the complex in this model system is much higher than that of the positive control RSA = 46.3 ± 3.7% (complex) versus RSA = 13.0 ± 1.1% (Trolox).

At 3·10⁻⁵ M, the cerium complex performs as a moderate scavenger of superoxide in the KO₂ model system (CL-SI = 69 ± 3%). At the same molarity, Trolox behaves as a much stronger scavenger (CL-SI = 25 ± 2%). When it comes to interaction with NaOCl, Ce (2b')₃ behaves as a very potent pro-oxidant CL-SI = 656 ± 21%, while Trolox almost totally extinguishes LDCL CL-SI = 1.5 ± 0.1%.





4 Summary and conclusion

A novel rare-earth metal ion complex with a biologically active ligand was synthesized, and its structure, electronic properties, and antioxidant activity were carefully studied by the experimental methods and quantum-mechanical calculations. The composition and structure of the tri-coordinated complex of Ce³⁺ and 2-(4-chlorophenyl)-5-(pyrrolidin-1-yl)-2*H*-1,2,3-triazole-4-carboxylate were determined using elemental analysis and spectral data. The structural features of Ce (2b')₃ and global chemical reactivity descriptors were revealed based on quantum chemical calculations using four DFT levels in comparison with three other models of cerium complexes.

The influence of different ligands on the charge and molecular properties of the Ce(III) complex was analyzed, and established and global reactivity descriptors were determined and demonstrated: 1) the better charge distribution on the cerium ion in larger systems; 2) the HOMO–LUMO gap value linearly decreases as system complexity increases with the exception of C complex that has a similar value to Ce (2b')₃. The low energy value indicates a large reactivity and low energy for excitation. 3) The analysis of the experimental and calculated IR and Raman spectra supported the proposed structure of the obtained Ce(III) complex. For the analysis of these spectra, detailed comparisons were carried out using mainly the M06-2X, CAM-B3LYP, and D3-B3LYP methods, obtaining the best results with the CAM-B3LYP method.

The complex Ce (2b')₃ scavenges hydroxyl radicals, generated by UV-induced water radiolysis to a greater extent than the positive control Trolox at 3·10⁻⁵ M. In this particular model system the type of coordination center, La(III) or Ce(III), has a very mild impact on activity. In line with previous observations on 2b and La (2b')₃, the ability of Ce (2b')₃ to participate in HAT with DPPH is very limited. These results, starkly contrasting with the results derived from the 2-

deoxyribose degradation assay, suggest that the low DPPH-scavenging activity may be due to steric hindrance rather than the low hydrogen-donating capacity of the complex. Coordination with Ce(III) seems to improve the ligand's ability to participate in SET, as observed in the ABTS⁺ assay. Ce (2b')₃ seems to scavenge the stable radical much more actively compared to its La(III) counterpart (almost threefold higher RSA values). In this case, we can comfortably deduce that the electron-exchanging capacity is significantly impacted by the type of metal ion, which is coordinated with 2b. Ce (2b')₃ reaffirms previous observations that 2b and its lanthanide complexes tend to behave as pro-oxidants in the presence of the clinically significant Fenton reaction. A concentration-dependent increase in MTT-formazan formation was observed with Ce (2b')₃, which is similar to La (2b')₃. Ionic mimicry allows lanthanide ions to competitively replace iron from biomolecules (e.g., iron-dependent enzymes), impairing their functions and therefore, exhibiting chemotherapeutic effects. This would also cause an increase in “free” iron that can yield toxic OH[•] via the Fenton reaction. On top of that, the abovementioned reaction seems to be enhanced in the presence of the ligand 2b both in free form and as a component of a complex. Unlike the pro-oxidant 2b and its La(III) complex, Ce (2b')₃ acts as an antioxidant in the presence of KO₂. A very strong pro-oxidant effect of the Ce(III) complex is observed in the presence of sodium hypochlorite, an RS that serves as a component of the body's immune defense against xenobiotics. The present research introduces to the reader a second biologically active lanthanide complex with 2b. Both La(III) and Ce(III) complexes manifest interesting biological behavior, serving as antioxidants in some model systems and as antioxidants in others. Additionally, the type of the lanthanide coordination center (in this case Ce³⁺ instead of La³⁺) was associated with a change in biological behavior, which is observed prominently in the ABTS, KO₂, and NaOCl assays. The most logical next steps were left for further research, which involve investigations on other lanthanide complexes of 2b, selecting those with most prominent activities and consequently testing them against bacterial/cancer cell lines. The results presented herein could lay a solid foundation for the synthesis of additional, novel lanthanide–triazole carboxylate complexes with promising biological activity. For this task, the relationships established could facilitate the selection of new ligands with improved properties in this design of new complexes.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#); further inquiries can be directed to the corresponding authors.

Author contributions

MP: writing–review and editing, writing–original draft, validation, supervision, software, methodology, investigation, formal analysis, and conceptualization. NB: writing–review and editing and supervision. LT: writing–review and editing, supervision, investigation, and conceptualization. NH-A: writing–review and editing and investigation. IK: writing–review and editing, validation, supervision, and funding acquisition.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the European Union - NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria (BG-RRP-2.004-0004-C01).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abd El-Hamid, S. M., Sadeek, S. A., Zordok, W. A., and El-Shwiniy, W. H. (2019). Synthesis, spectroscopic studies, DFT calculations, cytotoxicity and antimicrobial activity of some metal complexes with ofloxacin and 2,2'-bipyridine. *J. Molec. Struct.* 1176, 422–433. doi:10.1016/j.molstruc.2018.08.082
- Adhikari, S., Nath, P., Das, A., Datta, A., Baidya, N., Duttaroy, A. K., et al. (2024). A review on metal complexes and its anti-cancer activities: recent updates from *in vivo* studies. *Biomed. and Pharmacother.* 171, 116211. doi:10.1016/j.biopha.2024.116211
- Agalave, S. G., Maujan, S. R., and Pore, V. S. (2011). Click chemistry: 1, 2, 3-triazoles as pharmacophores. *Chem.-An Asian J.* 6 (10), 2696–2718. doi:10.1002/asia.201100432
- Albesa, I., Becerra, M. C., Battán, P. C., and Páez, P. L. (2004). Oxidative stress involved in the antibacterial action of different antibiotics. *Biochem. Biophys. Res. Comm.* 317 (2), 605–609. doi:10.1016/j.bbrc.2004.03.085
- Alcolea Palafox, M. (2000). Scaling factors for the prediction of vibrational spectra. I. Benzene molecule. *Int. J. Quantum Chem.* 77, 661–684. doi:10.1002/(SICI)1097-461X(2000)77:3<661::AID-QUA7>3.0.CO;2-J
- Alcolea Palafox, M. (2017). Computational chemistry applied to vibrational spectroscopy: a tool for characterization of nucleic acid bases and some of their 5-substituted derivatives. *Phys. Sci. Rev.* 2 (8), 1–21. doi:10.1515/psr-2016-0132
- Alcolea Palafox, M. (2019). “DFT computations on vibrational spectra: scaling procedures to improve the wavenumbers,” in *Density functional theory: advances in applications (book)* (Boston, USA: Walter de Gruyter, Inc.), 147–191. chapter 10.
- Alcolea Palafox, M., Belskaya, N. P., and Kostova, I. P. (2023b). Peculiarities of the spatial and electronic structure of 2-aryl-1, 2, 3-triazol-5-carboxylic acids and their salts on the basis of spectral studies and DFT calculations. *Int. J. Molec. Sci.* 24 (18), 14001. doi:10.3390/ijms241814001
- Alcolea Palafox, M., Belskaya, N. P., Todorov, L. T., and Kostova, I. P. (2023a). Structural study of a La (III) complex of a 1, 2, 3-triazole ligand with antioxidant activity. *Antioxidants* 12, 1872. doi:10.3390/antiox12101872
- Aziz, S. G., Elroby, S. A., Alyoubi, A., Osman, O. I., and Hilal, R. (2014). Experimental and theoretical assignment of the vibrational spectra of triazoles and benzotriazoles. Identification of IR marker bands and electric response properties. *J. Mol. Model.* 20 (3), 2078. doi:10.1007/s00894-014-2078-y
- Barim, E., and Akman, F. (2022). Study on vibrational spectroscopy, molecular property, UV-VIS, HOMO-LUMO energies and MEP analysis of N-[2-(4-Bromo-benzoyl)-benzofuran-3-yl]-acrylamide monomer by DFT method. *Pigment Resin Technol.* 51 (1), 69–79. doi:10.1108/PRT-05-2020-0051
- Bonandi, E., Christodoulou, M. S., Fumagalli, G., Perdicchia, D., Rastelli, G., and Passarella, D. (2017). The 1, 2, 3-triazole ring as a bioisostere in medicinal chemistry. *Drug Discov. Today* 22 (10), 1572–1581. doi:10.1016/j.drudis.2017.05.014
- Bozorov, K., Zhao, J., and Aisa, H. A. (2019). 1, 2, 3-Triazole-containing hybrids as leads in medicinal chemistry: a recent overview. *Bioorg. Med. Chem.* 27 (16), 3511–3531. doi:10.1016/j.bmc.2019.07.005
- Bünzli, J.-C. G. (2015). On the design of highly luminescent lanthanide complexes. *Coord. Chem. Rev.* 293, 19–47. doi:10.1016/j.ccr.2014.10.013
- Burns, W. G., and Sims, H. E. (1981). Effect of radiation type in water radiolysis. *J. Chem. Soc., Faraday Trans. 1 Phys. Chem. Cond. Phases.* 77 (11), 2803–2813. doi:10.1039/f19817702803
- Che, M., Wang, R., Li, X., Wang, H.-Y., and Zheng, X. S. (2016). Expanding roles of superoxide dismutases in cell regulation and cancer. *Drug Discov. Today* 21 (1), 143–149. doi:10.1016/j.drudis.2015.10.001
- Chrzczanowicz, J., Gawron, A., Zwolinska, A., de Graft-Johnson, J., Krajewski, W., Krol, M., et al. (2008). Simple method for determining human serum 2, 2-diphenyl-1-picryl-hydrazyl (DPPH) radical scavenging activity—possible application in clinical studies on dietary antioxidants. *Clin. Chem. Lab. Med.* 46 (3), 342–349. doi:10.1515/cclm.2008.062
- Chundawat, N. S., Jadoun, S., Zarrintaj, P., and Chauhan, N. P. S. (2021). Lanthanide complexes as anticancer agents: a review. *Polyhedron* 207, 115387. doi:10.1016/j.poly.2021.115387
- Cota, I., Marturano, V., and Tylkowski, B. (2019). Ln complexes as double faced agents: study of antibacterial and antifungal activity. *Coord. Chem. Rev.* 396, 49–71. doi:10.1016/j.ccr.2019.05.019
- Das, B., Ghosh, K., and Baruah, J. B. (2014). Tris-dipicolinate cerium complexes bearing dications of arginine, histidine, and ornithine. *Syn. React. Inorg. Metal-Org. Nano-Metal Chem.* 44 (2), 251–257. doi:10.1080/15533174.2013.770754
- Dheer, D., Singh, V., and Shankar, R. (2017). Medicinal attributes of 1,2,3-triazoles: current developments. *Bioorg. Chem.* 71, 30–54. doi:10.1016/j.bioorg.2017.01.010
- Dodoo, D., Snitka, V., and Vaitiekunaite, D. (2023). Traceability of bilberries (*Vaccinium myrtillus* L.) of the Baltic-Nordic region using surface-enhanced Raman spectroscopy (SERS): DFT simulation-based DNA analysis. *Spectrochim. Acta, Part A Molec. Biomol. Spectros.* 288, 122192. doi:10.1016/j.saa.2022.122192
- El-Azhary, A., Suter, H., and Kubelka, J. (1998). Experimental and theoretical investigation of the geometry and vibrational frequencies of 1, 2, 3-triazole, 1, 2, 4-triazole, and tetrazole anions. *J. Phys. Chem. A* 102 (3), 620–629. doi:10.1021/jp9719568
- Erel, O. (2004a). A novel automated direct measurement method for total antioxidant capacity using a new generation, more stable ABTS radical cation. *Clin. Biochem.* 37 (4), 277–285. doi:10.1016/j.clinbiochem.2003.11.015
- Erel, O. (2004b). A novel automated method to measure total antioxidant response against potent free radical reactions. *Clin. Biochem.* 37 (2), 112–119. doi:10.1016/j.clinbiochem.2003.10.014
- Evangelin, J. H., Bharanidharan, S., and Dhandapani, A. (2018). Molecular spectroscopic investigations of (E)-1-(4-methylbenzylidene) urea using DFT method. *Arch. Chem. Res.* 2 (2), 1–10. doi:10.21767/2572-4657.100018
- Evans, A. M. (2004). Motexafin gadolinium: a redox-active tumor selective agent for the treatment of cancer. *Curr. Opin. Oncol.* 16 (6), 576–580. doi:10.1097/01.cco.0000142073.29850.98
- Feng, F. M., Cai, S., Liu, F. A., and Xie, J. Q. (2013). Studies of DNA-binding and DNA-cutting mechanism of an azamacrocyclic cerium complex with carboxyl branch. *Progr. React. Kinet. Mech.* 38 (3), 283–294. doi:10.3184/146867813X13738207456695
- Fricker, S. P. (2006). The therapeutic application of lanthanides. *Chem. Soc. Rev.* 35 (6), 524–533. doi:10.1039/b509608c
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16, revision C. 01*. Wallingford CT: Gaussian, Inc.
- Galano, A. (2015). Free radicals induced oxidative stress at a molecular level: the current status, challenges and perspectives of computational chemistry based protocols. *J. Mex. Chem. Soc.* 59 (4), 231–262. doi:10.29356/jmcs.v59i4.81
- Gil-Moles, M., and Concepción Gimeno, M. (2024). The therapeutic potential in cancer of terpyridine-based metal complexes featuring group 11 elements. *ChemMedChem* 19 (10), e202300645. doi:10.1002/cmdc.202300645
- Giroud, M., Kuhn, B., Saint-Auret, S., Kuratli, C., Martin, R. E., Schuler, F., et al. (2018). 2-H-1, 2, 3-Triazole-based dipeptidyl nitriles: potent, selective, and

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2024.1450106/full#supplementary-material>

- trypanocidal rhodesain inhibitors by structure-based design. *J. Med. Chem.* 61 (8), 3370–3388. doi:10.1021/acs.jmedchem.7b01870
- Goswami, A. K., and Kostova, I. (2022). *Medicinal and biological inorganic chemistry*. Berlin: Walther de Gruyter GmbH and Co KG.
- Guo, C., Su, F., Su, P., Yu, X., and Li, X. (2023). Luminescent terpyridine-based metallo-supramolecular systems: from design to applications. *Sci. China Chem.* 1–23, 1869–1870. doi:10.1007/s11426-023-1601-0
- Hajam, Y. A., Rani, R., Ganie, S. Y., Sheikh, T. A., Javaid, D., Qadri, S. S., et al. (2022). Oxidative stress in human pathology and aging: molecular mechanisms and perspectives. *Cells* 11 (3), 552. doi:10.3390/cells11030552
- Halliwell, B., Gutteridge, J. M., and Aruoma, O. I. (1987). The deoxyribose method: a simple “test-tube” assay for determination of rate constants for reactions of hydroxyl radicals. *Anal. Biochem.* 165 (1), 215–219. doi:10.1016/0003-2697(87)90222-3
- Hassan, A. U., Sumrra, S. H., Zafar, W., Imran, M., Sadaf, N., and Irfan, M. (2023). Enriching the compositional tailoring of NLO responsive dyes with diversity oriented electron acceptors as visible light harvesters: a DFT/TD-DFT approach. *Molec. Phys.* 121 (1), e2148585. doi:10.1080/00268976.2022.2148585
- Hosseinizadeh, R., Khorsandi, K., Sadat Esfahani, H., Habibi, M., and Hosseinizadeh, G. (2021). Preparation of cerium-curcumin and cerium-quercetin complexes and their LEDs irradiation assisted anticancer effects on MDA-MB-231 and A375 cancer cell lines. *Photodiagnosis Photodyn. Ther.* 34, 102326. doi:10.1016/j.pdpdt.2021.102326
- Hrimla, M., Oubella, A., Laamari, Y., Lahoucine, B., Ghaleb, A., Itto, M. Y. A., et al. (2021). Click synthesis, anticancer activity, and molecular docking investigation of some functional 1,2,3-triazole derivatives. *Biointerface Res. Appl. Chem.* 12 (6), 7633–7667. doi:10.33263/BRIAC126.76337667
- Hussain Sumrra, S., Anees, M., Asif, A., Nadeem Zafar, M., Mahmood, K., Faizan Nazar, M., et al. (2020). Synthesis, structural, spectral and biological evaluation of metals endowed 1,2,4-triazole. *Bull. Chem. Soc. Ethiop.* 34 (2), 335–351. doi:10.4314/bcse.v34i2.11
- Jezińska, A., and Motyl, T. (2009). Matrix metalloproteinase-2 involvement in breast cancer progression: a mini-review. *Med. Sci. Monit.* 15 (2), RA32–40. doi:10.2210/pdb1CK7/pdb
- Juwarke, H., Lenhardt, J. M., Castillo, J. C., Zhao, E., Krishnamurthy, S., Jamiolkowski, R. M., et al. (2009). Anion binding of short, flexible aryl triazole oligomers. *J. Org. Chem.* 74 (23), 8924–8934. doi:10.1021/jo901966f
- Kedare, S. B., and Singh, R. (2011). Genesis and development of DPPH method of antioxidant assay. *J. Food Sci. Technol.* 48 (4), 412–422. doi:10.1007/s13197-011-0251-1
- Kell, D. B. (2009). Iron behaving badly: inappropriate iron chelation as a major contributor to the aetiology of vascular and other progressive inflammatory and degenerative diseases. *BMC Med. Genom.* 2 (1), 2–79. doi:10.1186/1755-8794-2-2
- Keri, R. S., Patil, S. A., Budagumpi, S., and Triazole, B. M. N. (2015). Triazole: a promising antitubercular agent. *Chem. Biol. Drug Des.* 86 (4), 410–423. doi:10.1111/cbdd.12527
- Kolb, H. C., and Sharpless, K. B. (2003). The growing impact of click chemistry on drug discovery. *Drug Discov. Today* 8 (24), 1128–1137. doi:10.1016/S1359-6446(03)02933-7
- Kullgren, J., Castleton, C. W., Müller, C., Ramo, D. M., and Hermansson, K. (2010). B3LYP calculations of cerium oxides. *J. Chem. Phys.* 132, 054110. doi:10.1063/1.3253795
- Kumar, R., Yar, M. S., Chaturvedi, S., and Srivastava, A. (2013). Triazole as pharmaceuticals potentials. *Int. J. Pharm. Tech. Res.* 5 (4), 1844–1869.
- Lauridsen, C. (2019). From oxidative stress to inflammation: redox balance and immune system. *Poult. Sci.* 98 (10), 4240–4246. doi:10.3382/ps/pey407
- Levin, J. R., Dorfner, W. L., Dai, A. X., Carroll, P. J., and Schelter, E. J. (2016). Density functional theory as a predictive tool for cerium redox properties in nonaqueous solvents. *Inorg. Chem.* 55 (24), 12651–12659. doi:10.1021/acs.inorgchem.6b01779
- Lewandowski, W., Kalinowska, B., and Lewandowska, H. (2005). The influence of metals on the electronic system of biologically important ligands. Spectroscopic study of benzoates, salicylates, nicotines and isoorotates. Review. *J. Inorg. Biochem.* 99 (7), 1407–1423. doi:10.1016/j.jinorgbio.2005.04.010
- Li, L., and Yan, M. (2020). Recent progresses in exploring the rare earth based intermetallic compounds for cryogenic magnetic refrigeration. *J. Alloys. Comp.* 823, 153810. doi:10.1016/j.jallcom.2020.153810
- Li, Z., Wang, Y.-F., Zeng, C., Hu, L., and Liang, X.-J. (2018). Ultrasensitive tyrosinase-activated turn-on near-infrared fluorescent probe with a rationally designed urea bond for selective imaging and photodamage to melanoma cells. *Anal. Chem.* 90 (6), 3666–3669. doi:10.1021/acs.analchem.7b05369
- Mody, T. D., Fu, L., and Sessler, J. L. (2001). Texaphyrins: synthesis and development of a novel class of therapeutic agents. *Progr. Inorg. Chem.*, 551–598. doi:10.1002/9780470166512.ch5
- Molyneux, P. (2004). The use of the stable free radical diphenylpicrylhydrazyl (DPPH) for estimating antioxidant activity. *Songklanakarin J. Sci. Technol.* 26 (2), 211–219.
- Musib, D., Mukherjee, M., and Roy, M. (2023). Emerging trends of La(III)-based compounds as the strategic tools for photodynamic therapy. *Inorg. Chim. Acta.* 558, 121751. doi:10.1016/j.ica.2023.121751
- Negre-Salvayre, A., Auge, N., Ayala, V., Basaga, H., Boada, J., Brenke, R., et al. (2010). Pathological aspects of lipid peroxidation. *Free Radic. Res.* 44 (10), 1125–1171. doi:10.3109/10715762.2010.498478
- Noreen, S., Sumrra, S. H., Chohan, Z. H., Mustafa, G., and Imran, M. (2023). Synthesis, characterization, molecular docking and network pharmacology of bioactive metallic sulfonamide-isatin ligands against promising drug targets. *J. Molec. Struct.* 1277, 134780. doi:10.1016/j.molstruc.2022.134780
- Omodara, L., Pitkäaho, S., Turpeinen, E.-M., Saavalainen, P., Oravissjärvi, K., and Keiski, R. L. (2019). Recycling and substitution of light rare earth elements, cerium, lanthanum, neodymium, and praseodymium from end-of-life applications-A review. *J. Clean. Prod.* 236, 117573. doi:10.1016/j.jclepro.2019.07.048
- Palafox, M. A. (2018). DFT computations on vibrational spectra: scaling procedures to improve the wavenumbers. *Phys. Sci. Rev.* 3 (6), 20170184. doi:10.1515/psr-2017-0184
- Palafox, M. A., Belskaya, N. P., and Kostova, I. P. (2023). Study of the molecular architectures of 2-(4-Chlorophenyl)-5-(pyrrolidin-1-yl)-2H-1,2,3-triazole-4-carboxylic acid using their vibrational spectra, quantum chemical calculations and molecular docking with MMP-2 receptor. *Pharmaceutics* 15, 2686. doi:10.3390/pharmaceutics15122686
- Palafox, M. A., and Rastogi, V. (2013). Spectra and structure of benzonitriles and some of its simple derivatives. *Asian J. Phys.* 22 (3), 1–30.
- Pattison, D., and Davies, M. (2006). Reactions of myeloperoxidase-derived oxidants with biological substrates: gaining chemical insight into human inflammatory diseases. *Curr. Med. Chem.* 13 (27), 3271–3290. doi:10.2174/092986706778773095
- Patyal, M., Kaur, K., Bala, N., Gupta, N., and Malik, A. K. (2023). Innovative lanthanide complexes: shaping the future of cancer/tumor chemotherapy. *J. Trace Elem. Med. Biol.* 80, 127277. doi:10.1016/j.jtemb.2023.127277
- Peica, N., Kostova, I., and Kiefer, W. (2006). Theoretical and experimental studies on binding mode of 3, 5-pyrazoledicarboxylic acid in its new La (III) complex. *Chem. Phys.* 325, 411–421. doi:10.1016/j.chemphys.2006.01.017
- Podolski-Renić, A., Čipak Gašparović, A., Valente, A., López, Ó., Bormio Nunes, J. H., Kowol, C. R., et al. (2024). Schiff bases and their metal complexes to target and overcome (multidrug) resistance in cancer. *Eur. J. Med. Chem.* 270, 116363. doi:10.1016/j.ejmech.2024.116363
- Raica-Malic, S., and Mescic, A. (2015). Recent trends in 1,2,3-Triazole-nucleosides as promising anti-infective and anticancer agents. *Curr. Med. Chem.* 22 (12), 1462–1499. doi:10.2174/0929867322666150227150127
- Rajaraman, D., Athishu Anthony, L., Sundararajan, G., Shanmugam, M., and Arunkumar, A. (2022). Synthesis, NMR, anti-oxidant, anti-cancer activity, Molecular docking, DFT Calculations and *in silico* ADME analysis of 3'-benzoyl-4'-phenyl-5'-(piperazin-1-ylmethyl)spiro[indoline-3,2'-pyrrolidin]-2-one derivatives. *J. Molec. Struct.* 1267, 133551. doi:10.1016/j.molstruc.2022.133551
- Riley, K. E., and Hobza, P. (2011). Noncovalent interactions in biochemistry. *Comput. Molec. Sci.* 1 (1), 3–17. doi:10.1002/wcms.8
- Riley, K. E., Pitonák, M., Jurecka, P., and Hobza, P. (2010). Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.* 110 (9), 5023–5063. doi:10.1021/cr1000173
- Safronov, N. E., Kostova, I. P., Palafox, M. A., and Belskaya, N. P. (2023). Combined NMR spectroscopy and quantum-chemical calculations in fluorescent 1, 2, 3-Triazole-4-carboxylic acids fine structures analysis. *Int. J. Molec. Sci.* 24, 8947–8966. doi:10.3390/ijms24108947
- Saini, M. S., and Dwivedi, J. (2013). Synthesis and biological significances of 1, 2, 4-triazole and its derivatives: a review. *Int. J. Pharm. Sci. Res.* 4 (8), 2866. doi:10.13040/IJPSR.0975-8232.4(8).2866-79
- Seminario, J. M., and Politzer, P. (1995). “Modern density functional theory: a tool for chemistry,” 2. Amsterdam: Elsevier.
- Sert, Y., El-Hiti, G. A., Gökce, H., Uzun, F., Abdel-Wahab, B. F., and Kariuki, B. M. (2020). DFT, molecular docking and experimental FT-IR, laser-Raman, NMR and UV investigations on a potential anticancer agent containing triazole ring system. *J. Molec. Struct.* 1211, 128077. doi:10.1016/j.molstruc.2020.128077
- Singh, J. S., Shahid Khan, M., and Uddin, S. (2023). A DFT study of vibrational spectra of 5-chlorouracil with molecular structure, HOMO-LUMO, MEPs/ESPs and thermodynamic properties. *Polym. Bull.* 80, 3055–3083. doi:10.1007/s00289-022-04181-7
- Slavova, K. I., Todorov, L. T., Belskaya, N. P., Alcolea Palafox, M., and Kostova, I. P. (2020). Developments in the application of 1,2,3-triazoles in cancer treatment. *Recent Pat. Anti-cancer Drug Dis.* 15, 92–112. doi:10.2174/1574892815666200717164457
- Sokrates, T. M. G. (2004). *Infrared and Raman characteristic group frequencies: tables and charts*. Wiley.
- Strzepa, A., Pritchard, K. A., and Dittel, B. N. (2017). Myeloperoxidase: a new player in autoimmunity. *Cell. Immunol.* 317, 1–8. doi:10.1016/j.cellimm.2017.05.002
- Subbiah, J., Ramuthai, M., and Murugan, P. (2022). DFT calculations of 2,5-diphenyl furan against SARS-CoV-2 mpro based on molecular docking approach. *Asian J. Chem.* 34 (7), 1665–1674. doi:10.14233/ajchem.2022.23622
- Tawada, Y., Tsuneda, T., Yanagisawa, S., Yanai, T., and Hirao, K. (2004). A long-range-corrected time-dependent density functional theory. *J. Chem. Phys.* 120 (18), 8425–8433. doi:10.1063/1.1688752

- Todorov, L., Kostova, I., and Traykova, M. (2019). Lanthanum, Gallium and their impact on oxidative stress. *Curr. Med. Chem.* 26 (22), 4280–4295. doi:10.2174/0929867326666190104165311
- Törnkvist, C., Bergman, J., and Liedberg, B. (1991). Geometry and vibrations of the 1, 2, 3-triazole anion. A theoretical and experimental study. *J. Phys. Chem.* 95 (8), 3119–3123. doi:10.1021/j100161a030
- Trist, B. G., Hilton, J. B., Hare, D. J., and Crouch, P. J. (2021). Superoxide dismutase 1 in health and disease: how a Frontline antioxidant becomes neurotoxic. *Angew. Chem. Int. Ed.* 60 (17), 9215–9246. doi:10.1002/anie.202000451
- Van Loenhout, J., Peeters, M., Bogaerts, A., Smits, E., and Deben, C. (2020). Oxidative stress-inducing anticancer therapies: taking a closer look at their immunomodulating effects. *Antioxidants* 9 (12), 1188. doi:10.3390/antiox9121188
- Varsányi, G. (1974). Assignment for vibrational spectra of seven hundred benzene derivatives. *Hilder* 1.
- Wakabayashi, T., Ymamoto, A., Kazaana, A., Nakano, Y., Nojiri, Y., and Kashiwazaki, M. (2016). Antibacterial, antifungal and nematocidal activities of rare earth ions. *Biol. Trace Elem. Res.* 174, 464–470. doi:10.1007/s12011-016-0727-y
- Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Letts.* 393, 51–57. doi:10.1016/j.cplett.2004.06.011
- Zafar, W., Ashfaq, M., and Hussain Sumrra, S. (2023). A review on the antimicrobial assessment of triazole-azomethine functionalized frameworks incorporating transition metals. *J. Molec. Struct.* 1288, 135744. doi:10.1016/j.molstruc.2023.135744
- Zhang, H., Feng, J., Zhu, W., Liu, C., and Gu, J. (2000). Bacteriostatic effects of cerium-humic acid complex - an experimental study. *Biol. Trace Elem. Res.* 73 (1), 29–36. doi:10.1385/BTER:73:1:29
- Zhang, H., Ryono, D. E., Devasthale, P., Wang, W., O'Malley, K., Farrelly, D., et al. (2009). Design, synthesis and structure–activity relationships of azole acids as novel, potent dual PPAR α/γ agonists. *Bioorg. Med. Chem. Letts.* 19 (5), 1451–1456. doi:10.1016/j.bmcl.2009.01.030
- Zhang, S., Saji, S. E., Yin, Z., Zhang, H., Du, Y., and Yan, C. H. (2021). Rare-earth incorporated alloy catalysts: synthesis, properties, and applications. *Adv. Mater.* 33 (16), 2005988. doi:10.1002/adma.202005988
- Zhang, T., Zheng, H., Zhang, F., Chu, C., Liao, T., Xie, L., et al. (2023). Rare-earth scintillating nanoparticles for X-ray induced photodynamic therapy. *J. Luminescence* 261, 119862. doi:10.1016/j.jlumin.2023.119862
- Zhao, J., Mei, D., Wang, W., Wu, Y., and Xue, D. (2021). Recent advances in nonlinear optical rare earth structures. *J. Rare Earths.* 39 (12), 1455–1466. doi:10.1016/j.jre.2021.07.005
- Zhao, Y., and Truhlar, D. G. (2011). Applications and validations of the Minnesota density functionals. *Chem. Phys. Lett.* 502 (1–3), 1–13. doi:10.1016/j.cplett.2010.11.060

Frontiers in Chemistry

Explores all fields of chemical science across the periodic table

Advances our understanding of how atoms, ions, and molecules come together and come apart. It explores the role of chemistry in our everyday lives - from electronic devices to health and wellbeing.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

