



Automated Fragmentation QM/MM Calculation of NMR Chemical Shifts for Protein-Ligand Complexes

Xinsheng Jin¹, Tong Zhu^{1,2}, John Z. H. Zhang^{1,2,3} and Xiao He^{1,2,4*}

¹ State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, East China Normal University, Shanghai, China,

² NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, China, ³ Department of Chemistry, New York University, New York, NY, United States, ⁴ National Engineering Research Centre for Nanotechnology, Shanghai, China

OPEN ACCESS

Edited by:

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

Reviewed by:

Ahmet Altun,
Max-Planck-Institut für
Kohlenforschung, Germany
Giovanni La Penna,
Consiglio Nazionale Delle Ricerche
(CNR), Italy

*Correspondence:

Xiao He
xiaoh@phy.ecnu.edu.cn

Specialty section:

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

Received: 30 January 2018

Accepted: 16 April 2018

Published: 08 May 2018

Citation:

Jin X, Zhu T, Zhang JZH and He X
(2018) Automated Fragmentation
QM/MM Calculation of NMR Chemical
Shifts for Protein-Ligand Complexes.
Front. Chem. 6:150.
doi: 10.3389/fchem.2018.00150

In this study, the automated fragmentation quantum mechanics/molecular mechanics (AF-QM/MM) method was applied for NMR chemical shift calculations of protein-ligand complexes. In the AF-QM/MM approach, the protein binding pocket is automatically divided into capped fragments (within ~200 atoms) for density functional theory (DFT) calculations of NMR chemical shifts. Meanwhile, the solvent effect was also included using the Poisson-Boltzmann (PB) model, which properly accounts for the electrostatic polarization effect from the solvent for protein-ligand complexes. The NMR chemical shifts of neocarzinostatin (NCS)-chromophore binding complex calculated by AF-QM/MM accurately reproduce the large-sized system results. The ¹H chemical shift perturbations (CSP) between apo-NCS and holo-NCS predicted by AF-QM/MM are also in excellent agreement with experimental results. Furthermore, the DFT calculated chemical shifts of the chromophore and residues in the NCS binding pocket can be utilized as molecular probes to identify the correct ligand binding conformation. By combining the CSP of the atoms in the binding pocket with the Glide scoring function, the new scoring function can accurately distinguish the native ligand pose from decoy structures. Therefore, the AF-QM/MM approach provides an accurate and efficient platform for protein-ligand binding structure prediction based on NMR derived information.

Keywords: AF-QMMM, NMR chemical shift, protein-ligand binding, scoring function, structure prediction

INTRODUCTION

Structure-based computational methods are useful tools for analyzing the binding modes and affinities for protein-ligand complexes (Grinter and Zou, 2014). With the development of X-ray crystallography and nuclear magnetic resonance (NMR) technology, more than 100,000 high resolution three-dimensional protein structures have been determined, which is helpful for finding lead compounds and therapeutic targets (Ferreira et al., 2015). As compared to experimental methods, computational approaches, such as molecular docking, are the fast and efficient ways for drug discovery. In molecular docking programs, the scoring functions are used to approximate the binding free energy and hence to rank the simulated decoy poses (Sliwoski et al., 2014). Most scoring functions are roughly derived from force-field-based (Morris et al., 1998; Englebienne and Moitessier, 2009), empirical (Eldridge et al., 1997; Murray et al., 1998) or knowledge-based

potentials (Gohlke et al., 2000; Huang and Zou, 2006). However, based on parameterized functions, these scoring methods are usually not accurate enough to differentiate the experimental structure from the docked decoy structures, and sometimes the rankings from different software suites may be inconsistent (Śledź and Cafilisch, 2018). For solving this problem, much effort has been devoted to the development of docking methods by introducing the experimental structural information or quantum mechanical (QM) calculations for scoring the native and predicted poses (Mohan et al., 2005; Grinter and Zou, 2014; Adeniyi and Soliman, 2017).

The chemical shift is one of the most effective and precise NMR parameters in reflecting the local chemical environment around the atom, which plays an important role in structure determination and refinement (Zhu et al., 2014; Bratholm and Jensen, 2017). For NMR chemical shift calculations, the empirical chemical shift prediction softwares include ShiftS (Xu and Case, 2001; Moon and Case, 2007), ShiftX (Neal et al., 2003), ShiftX2 (Han et al., 2011), CamShift (Robustelli et al., 2010), PROSHIFT (Meiler, 2003; Meiler and Baker, 2003), SHIFTCALC (Williamson and Craven, 2009), ProCS (Christensen et al., 2014; Bratholm and Jensen, 2017), CheShift (Vila et al., 2009; Garay et al., 2014) and Sparta+ (Shen and Bax, 2010). These empirical methods are fast in computational speed, and have been successful in predicting backbone chemical shifts. As the empirical formulas for these models are derived from fitting the experimental or QM calculated chemical shift database and the high-quality structures, these models are not well suited for accurate prediction of NMR chemical shifts for some complex systems such as protein-ligand complexes, non-standard protein residues or non-canonical base pairs in nucleic acid systems (Swails et al., 2015). The quantum mechanical chemical shift calculations are in principle able to predict the NMR chemical shifts for any complex systems (Lodewyk et al., 2012; Hartman and Beran, 2014; Merz, 2014). For protein NMR chemical shift calculations, Cui and Karplus had proposed a very effective QM/MM approach (Cui and Karplus, 2000), Gao et al. developed fragment molecular orbital (FMO) method (Gao et al., 2007, 2010), Exner and coworkers utilized the adjustable density matrix assembler (ADMA) approach (Frank et al., 2012; Victora et al., 2014), Tan and Bettens developed the combined fragmentation method (CFM) (Tan and Bettens, 2013), and He and coworkers developed the automated fragmentation quantum mechanics/molecular mechanics (AF-QM/MM) method (He et al., 2009, 2014; Zhu et al., 2012, 2013, 2014, 2015; Swails et al., 2015; Jin et al., 2016). These fragment-based QM methods have been successfully applied for NMR chemical shift calculation of proteins and nucleic acids.

On the basis that chemical shifts or chemical shift perturbations (CSP) are sensitive to the variations of chemical environment, these parameters are quite suitable for structure determination (Case, 1998; Cavalli et al., 2007; Shen and Bax, 2015). Many NMR-based methods have been developed for prediction of protein-ligand binding modes (Medek et al., 2000; Cioffi et al., 2008; Riedinger et al., 2008; Aguirre et al., 2014). McCoy and Wyss utilized proton CSP data, induced by aromatic ring current effect in the ligands, to locate the ring position

of docking structures (McCoy and Wyss, 2002). Recently, Ten Brink et al. compared the experimental and simulated CSPs to verify protein conformational changes and developed the CSP-based docking method (Ten Brink et al., 2015). Merz et al. developed CSP-based scoring functions to determine the binding poses for protein-ligand complexes (Wang et al., 2007; Yu et al., 2017). However, most of these scoring functions only calculate the proton chemical shift on proteins. The scoring functions could be more efficient by taking ligand ^1H chemical shifts into consideration.

In this work, we applied the AF-QM/MM approach for NMR chemical shift calculations on protein-ligand binding complexes. Based on DFT calculations, the ^1H chemical shifts on both protein and ligand are available for structure determination and improving the scoring functions. In the framework of the AF-QM/MM approach, the ligand is also divided into smaller fragments, and hence it significantly reduced the computational cost for ^1H NMR chemical shift calculation on the large ligand. In this study, the neocarzinostatin (NCS) protein is selected as the test case for the AF-QM/MM method because of its importance in cancer therapy. NCS has experimental chemical shifts for both apo and holo forms (Myers et al., 1988; Mohanty et al., 1994; Schaus et al., 2001; Takashima et al., 2005; Wang and Merz, 2010). Furthermore, by comparing the AF-QM/MM calculated chemical shifts with experiment data, a chemical shift based scoring function was developed to rank the native and predicted protein-ligand binding poses.

This paper is organized as follows: first, a benchmark test was performed using the AF-QM/MM method for NMR chemical shift calculations of protein-ligand complex. The computed results are compared to the large-sized system NMR chemical shift calculations. Subsequently, AF-QM/MM calculated chemical shifts are compared with the experimental results for both apo and holo NCS structures. Next, the performance of chemical shift based and conventional energy based scoring functions on the rankings of predicted protein-ligand binding poses is discussed. Finally, the hybrid scoring function, that combines the calculated NMR chemical shifts and binding energy, is applied to rank the experimental structure and other docked binding poses.

COMPUTATIONAL APPROACHES

Structure Preparation

The X-ray structures of apo and holo NCS were download from the Protein Data Bank (PDB ids: 1NOA and 1NCO, respectively). The experimental chemical shift data of apo protein and chromophore were obtained from previous studies (Myers et al., 1988; Mohanty et al., 1994). The holo NCS experiment chemical shifts were downloaded from Biological Magnetic Resonance Data Bank (BMRB entry: 5969). The structure minimization of the protein X-ray structures was performed using the AMBER12 program (Case et al., 2012) with the ff99SB force field. The apo and holo NCS structures were solvated in a truncated octahedral periodic box of TIP3P water molecules with each side at least 10 Å from the nearest solute atom (Jorgensen and Jenson, 1998). After the entire system was neutralized with counter ions, 1,000

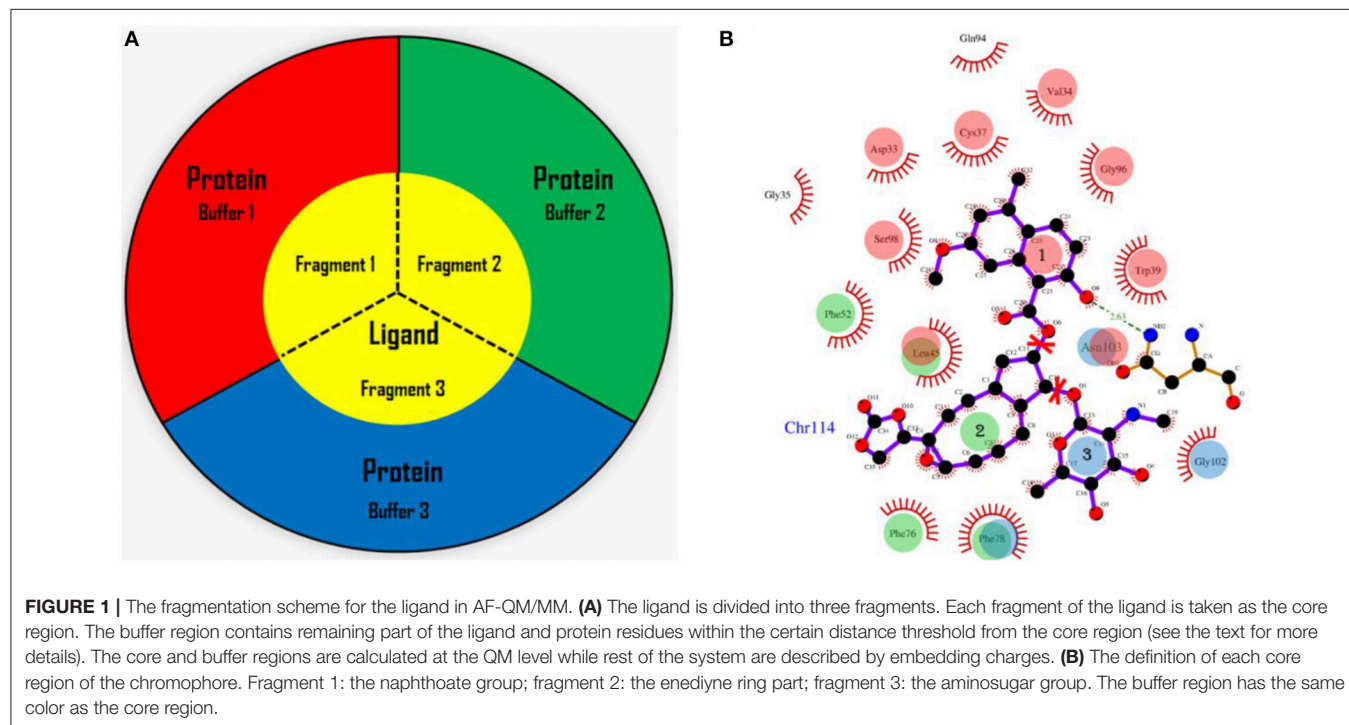
steps of steepest descent algorithm following with 4,000 steps of conjugate gradient method were used to remove the improper contacts of the system. For obtaining the force field parameters of the ligand, the general AMBER force field (GAFF) (Wang et al., 2004) and AM1-bond charge correlations (AM1-BCC) charge model were utilized for the ligand (Jakalian et al., 2002). The molecular docking was performed using the Glide module in the Schrödinger program (Friesner et al., 2004; Halgren et al., 2004). The scoring function used in this study was Glide XP. In this study, the protein structure was fixed when the ligand was docked into the binding site. Therefore, we did not include the flexibility of the protein during molecular docking. Based on the optimized experimental structure using the molecular force field, 38 docking poses of the ligand predicted by Glide, whose RMSDs range from 1.5 to 10.5 Å with reference to the native position, were selected for subsequent chemical shift calculations at the QM level.

The AF-QM/MM Method for NMR Chemical Shift Calculation of the Protein-Ligand Complex

In the AF-QM/MM approach (He et al., 2009, 2014; Zhu et al., 2012, 2013; Swails et al., 2015; Jin et al., 2016), the apo protein is divided into individual residue by cutting through the peptide bonds. The number of fragments is the same as the number of residues in the protein. Each fragment contains a core region (each amino acid) in the protein, and the buffer region which contains the nearby residues surrounding the core region. Both the core and buffer regions are treated with quantum mechanics (QM) while the residues outside the buffer region are described

by molecular mechanics (MM). The details for the definition of the buffer regions are described in our previous work (He et al., 2009; Zhu et al., 2012, 2013). For the holo protein studied in this work, we developed the fragmentation scheme for the ligand and its surrounding protein residues. As shown in **Figure 1**, we also divided the chromophore into three parts by cutting the C-O single bond. Fragment 1 contains the naphthoate group, fragment 2 includes the enediyne ring and fragment 3 has the aminosugar group, respectively. For each fragment of the ligand (taken as each core region), the rest part of the ligand and the protein residues surrounding the core region (each fragment of the ligand) are taken as the buffer region (see **Figure 1**). The distance criteria for selecting the buffer region for each fragment of the ligand is the same as that for each residue in the protein.

In this study, we adopt the following distance-dependent criteria to include residues within the buffer region of each core region for ligand: (1) if a heavy atom of the residue is less than 3.5 Å away from any atom in the core region, (2) if the distance of a hydrogen atom of the residue is less than 3.0 Å away from any atom in the core region. The cutoff enables a sufficient size of the buffer region for the convergence of chemical shift calculations on the core region (Flaig et al., 2012). The remaining residues beyond the buffer region are described by embedding charges to account for the electrostatic field outside the QM region. The protein charges were obtained directly from the ff99SB force field. For assigning the buffer region for each residue (where each residue is defined as the core region), the ligand is treated as a whole molecule (non-fragmented), and the definition of the buffer region follows the same criteria as the apo protein. The dangling bonds are capped with hydrogen atoms for constructing the closed-shell fragment.



The fragment QM calculations were carried out in parallel at the B3LYP/6-31G** level. All QM calculations were performed using the Gaussian 09 package (Frisch et al., 2009). Only the NMR isotropic shieldings of the core region atoms were collected from each fragment QM calculation. The ^1H chemical shifts are obtained by referencing to that of the tetramethylsilane (TMS) at the same computational level, which is 31.66 ppm. The implicit solvation model was applied to approximate the solvent effect. The protein charge distribution polarizes the dielectric solution and creates a reaction field to act back on the solute until equilibrium is reached. The reaction field acting on the solute can be effectively represented by the induced charges on the cavity surface. In this work, the surface charges are calculated by the Poisson-Boltzmann (PB) model using the Delphi program (Rocchia et al., 2001). The set of point charges of the MM environment and on the molecular surface, which represents the reaction field, are used as the background charges in the QM calculation. Because the computational cost of QM chemical shift calculations will be dramatically increased on multiple configurations when the conformational sampling effect is taken into account. In this study, the optimized X-ray structure using molecular force field was taken as a representative configuration for the ensemble averaging structure.

Scoring Functions

To differentiate the native protein-ligand binding structure from decoy poses, here we propose a scoring function based on NMR chemical shifts (CS_{score}), which is simply the root-mean-square deviation (RMSD) of computed

chemical shifts with reference to the experimental values,

$$\text{CS}_{\text{score}} = \sqrt{\frac{\sum_{i=1}^N (\delta_{\text{H}}^i - \delta_{\text{exp}}^i)^2}{N}} \quad (1)$$

where δ_{H}^i is the chemical shift of i th hydrogen atom on the ligand and nearby residues, and δ_{exp}^i is the experimental chemical shift of the corresponding atom in the native complex (holo NCS). N is the number of atoms whose chemical shifts were selected as molecular probe to characterize the NCS-chromophore binding structure. In this study, N was set to 31 for holo NCS, 21 of which are non-amide protons on the chromophore, and the other 10 hydrogen atoms are those with experimental chemical shift perturbations (CSP, between the bound and unbound complexes) greater than 0.5 ppm from residues in the binding site of the protein (see **Figure 2** and Table S1 of the Supplementary Materials).

It is worth noting that, in Equation (1), we could also add the chemical shifts of the hydrogen atoms that experimentally do not change upon ligand binding. A false docking pose may cause significant deviations in CSPs for those hydrogen atoms. However, there are many of such protons on the residues in the binding pocket, which will average out the final score to make the scoring function incapable of distinguishing the native structure from the decoy sets. Protons with experimental perturbations greater than 0.5 ppm are more sensitive to the binding pose, therefore we took those atoms into account in the scoring function. Furthermore, although NMR chemical shifts of amide

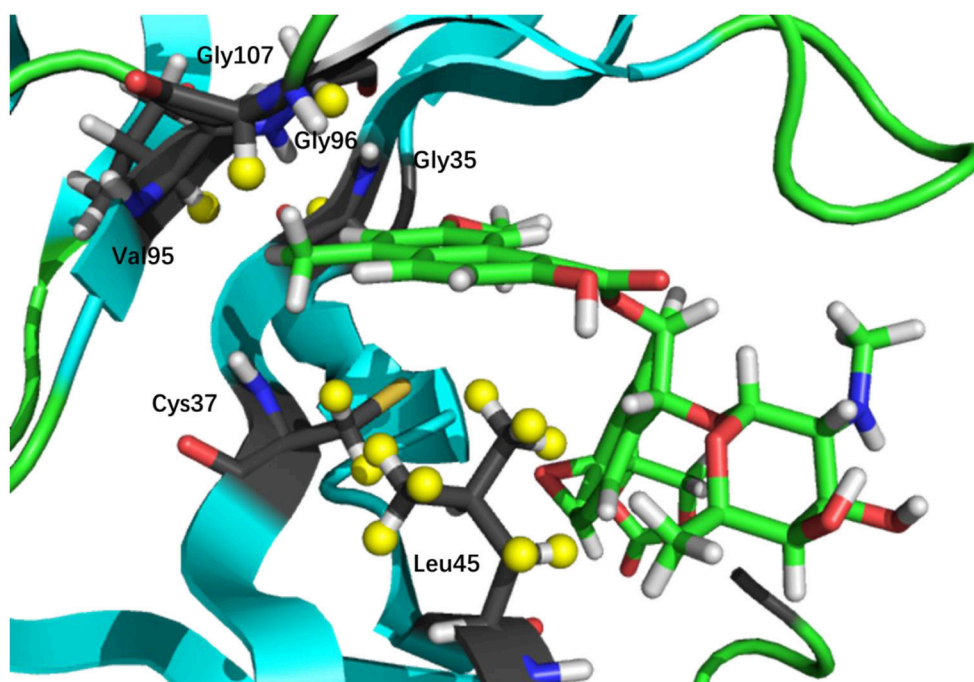


FIGURE 2 | 3D structure of the binding pocket. The nonpolar hydrogen atoms on the ligand, and protein protons whose CSP values (between bound and unbound complexes) are greater than 0.5 ppm (denoted in yellow) are chosen in the chemical shift based scoring functions (namely, CS_{score} and $\text{CS}_{\text{Gscore}}$).

protons are also very sensitive to the local chemical environment of the binding pocket, these atoms were excluded owing to the lack of experimental data.

The second scoring function (CS_{Gscore}) we propose here, is a linear combination of CS_{score} and Glide score,

$$CS_{Gscore} = CS_{score} + \alpha \text{Glide Score} \quad (2)$$

where α is a weighting factor. In this study, the ranges of CS_{score} and Glide Score are 0.42~2.90 and -10.96~10.98 (see Table S2 of the Supplementary Materials), respectively, and thus we choose $\alpha = (2.90 - 0.42) / (10.98 - (-10.96)) \approx 0.1$. By adding the Glide score to CS_{score} , the unphysical structure with an unfavorable binding energy will be avoided. Since both the CS_{score} and Glide score will be smaller as the docking pose gets closer to the experimental structure, the native docking pose will give the lowest CS_{Gscore} value.

RESULTS AND DISCUSSION

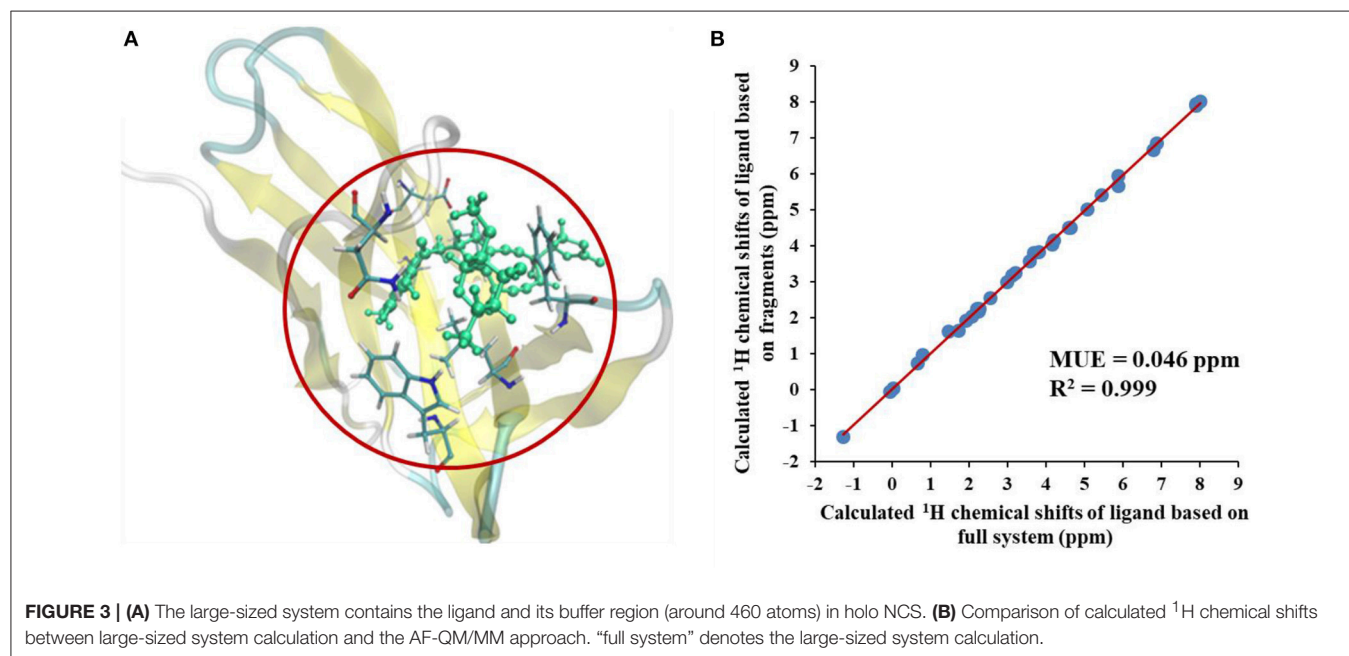
Benchmark Test of AF-QM/MM on the Native NCS-Chromophore Binding Complex

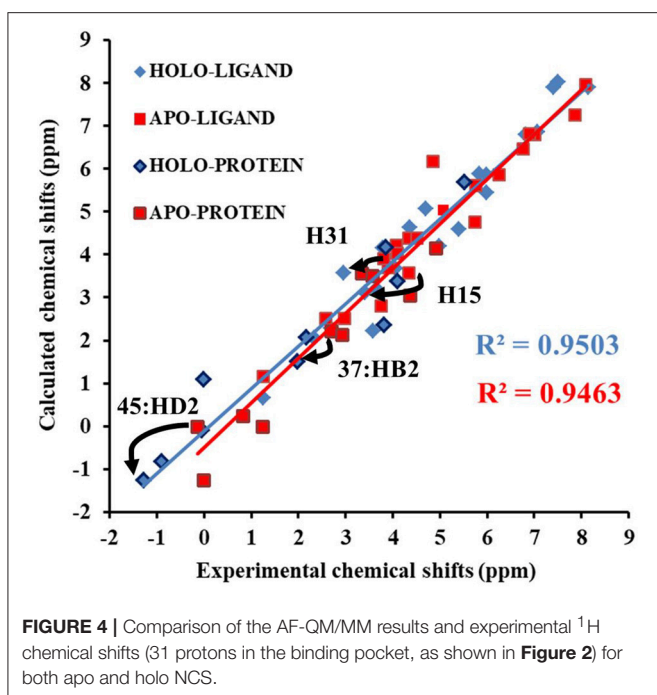
We first compared the calculated chemical shifts on the chromophore between AF-QM/MM and large-sized system calculations. Because the holo NCS contains more than 1500 atoms and was too large to perform full system QM calculations, we alternatively used the entire ligand and its buffer region for large-sized QM calculation. The other atoms beyond the buffer region are taken as background charges, and the PB surface charges for the entire complex are also placed to approximate the implicit solvent. The computed chemical shifts from such a model system (around 460 atoms in the QM region) are taken as the reference values. Here, we only compare the

chemical shifts on the ligand between AF-QM/MM calculation and large-sized system calculation. As shown in **Figure 3**, the ^1H chemical shifts on the ligand calculated by the AF-QM/MM method (where the ligand is divided into three parts) are in good agreement with large-sized system calculation. The mean unsigned error (MUE) between AF-QM/MM results and chemical shifts from the large-sized system calculation is 0.046 ppm, and the RMSD between them is 0.051 ppm. The results demonstrate that the AF-QM/MM approach can accurately reproduce the large-sized system calculation. Furthermore, at the DFT level, the total computational cost was reduced by 36%, from 5,601 min (CPU time) by the large-sized system calculation to 3,585 min by dividing the ligand into 3 fragments. In addition, the 3 fragment-based QM/MM calculations were carried out in parallel. Therefore, the computational wall time could be further reduced by approximately 2/3.

Next, we compare the calculated chemical shifts for apo and holo NCS (31 protons in the binding pocket, as shown in **Figure 2**) with the experimental values. In this benchmark test, the AF-QM/MM results correlate well with the experiment (see **Figure 4**). For the bound complex, the MUE between the calculated and experimental chemical shifts is 0.44 ppm, and the RMSD is 0.57 ppm. For the unbound protein and ligand, the MUE and RMSD between calculated and experimental chemical shifts are 0.45 and 0.62 ppm, respectively.

The prediction of chemical shift perturbations (CSP) between apo and holo NCS could further validate the accuracy of the AF-QM/MM approach. Among the hydrogen atoms in the binding pocket, H15, H31 of the ligand, HD2 of Leu45 and HB2 of Cys37 are significantly influenced by the ring current effect, where they are close to the aromatic rings in the native holo structure (see **Figure 5**). As a result, large upfield shifts upon ligand binding were observed for those atoms. The experimental CSPs of those four atoms (namely, H15, H31, Leu45:HD2, and Cys37:HB2) are





−0.93, −0.86, −1.15, and −0.72 ppm, respectively, while the AF-QM/MM results of them are −0.45, −0.34, −1.25, and −0.71 ppm. The results show that large chemical shift perturbations between apo and holo protein-ligand systems could be accurately predicted by the AF-QM/MM approach.

Performance of the CS_{score} Scoring Function

The Glide scores of the 38 docked decoys and the native holo NCS are shown in **Figure 6**. The energy based scoring function is capable of distinguishing the experimental structure from the docked poses whose structural RMSDs are larger than 4 Å with reference to the native pose. However, for the docked poses whose RMSDs are between 2 and 4 Å, the Glide scores of them are sometimes very close to the experimental structure (see Pose 7 in **Figure 6A**). In contrast, the CS_{score} is easier to discriminate the experimental structure from the decoy sets whose structural RMSDs are around 2 Å. This is mainly due to that chemical shifts are quite sensitive to the local chemical environment at the binding site. In CS_{score} , protons in ligand and the selected hydrogen atoms in protein residues serve as molecular probes to detect the binding environment. When the protons have different close contacts between the native and decoy structures, such as the interactions with aromatic rings or hydrogen bonding, the calculated chemical shift of certain protons may have substantial deviations between different binding modes. Therefore, the change of NMR chemical shifts of protons could clearly reflect the corresponding binding interactions between the protein and ligand.

The comparison of calculated chemical shifts between the native and docking poses could probe the structural changes of ligand binding poses among them. **Figure 7** shows that, for

the positions of the naphthoate group and enediyne ring in Pose 7, the chromophore is very close to the native binding structure. However, the aminosugar group is pointing to a different direction, which results in the large chemical shifts deviation for H13 and H15 on the ligand (see **Table 1**).

The weakness of CS_{score} is that for decoys with larger structural RMSDs, the chemical shift based scoring function might be not as efficient as those with low structural RMSDs. In high structural RMSD range, some ligand poses might be close to the apo state (fewer interactions with the protein), and the CS_{score} score for the apo state of chromophore is 0.60. Therefore, the rankings of those poses are not very sensitive to the structural RMSDs using CS_{score} . The example cases for poses 10 and 18 will be discussed in Section Improvement of the hybrid $\text{CS}_{\text{Gscore}}$ scoring function.

Improvement of the Hybrid $\text{CS}_{\text{Gscore}}$ Scoring Function

Figure 8 shows that the $\text{CS}_{\text{Gscore}}$ score is capable of differentiating the experimental structure from the decoys for the protein-ligand complex. In this work, the weighting factor α in Equation (2) was set to 0.1 to make the CS_{score} and Glide scores on the same scale. As shown in **Figure 8**, for poses whose structural RMSDs are around 2–4 Å, the CS_{score} score from NMR chemical shifts dominates the scoring function. Even though the Glide score for the decoy structures are close to the experimental structure (see **Figure 6B**), the CS_{score} could discriminate the experimental structure from the decoys, resulting that the $\text{CS}_{\text{Gscore}}$ (combination of CS_{score} and Glide scores) ranked the native binding pose clearly as the most favorable structure. On the other hand, for decoy poses whose structural RMSDs are larger than 4 Å, the energy based scoring function (Glide score) has the major impact on the $\text{CS}_{\text{Gscore}}$, which makes the decoys with large structural RMSDs deviates more from the experimental structure as compared to the CS_{score} score (**Figure 6B**).

The pose 10 was previously scored low in CS_{score} , whose chemical shift RMSD is 0.70 ppm between the calculated and experimental data. The naphthoate group of the ligand was flipped in the docking structure, while locations of the other regions of the chromophore are similar to the native binding pose (see **Figure 9**). On the naphthoate group, H23 has the largest chemical shift deviation between the native binding structure and pose 10. In the native binding pose, H23 is beside the aromatic ring, which causes downfield chemical shift on H23. While in Pose 10, the H23 atom moves away from the indole ring of Trp39, resulting that its calculated chemical shift is much lower than that of the experimental structure (see **Table 2**). Furthermore, **Figure 9a** shows that in Pose 10, H31 moves away from the phenyl ring of Phe52, and the corresponding chemical shielding decreased, resulting in higher NMR chemical shift. Furthermore, because the aromatic ring position of the naphthoate group moved in the docking structure, ^1H chemical shift on surrounding protein residues, such as Cys37:HB3, also changed significantly. Meanwhile, since the naphthoate ring strongly influenced the ^1H chemical shifts from the unbound

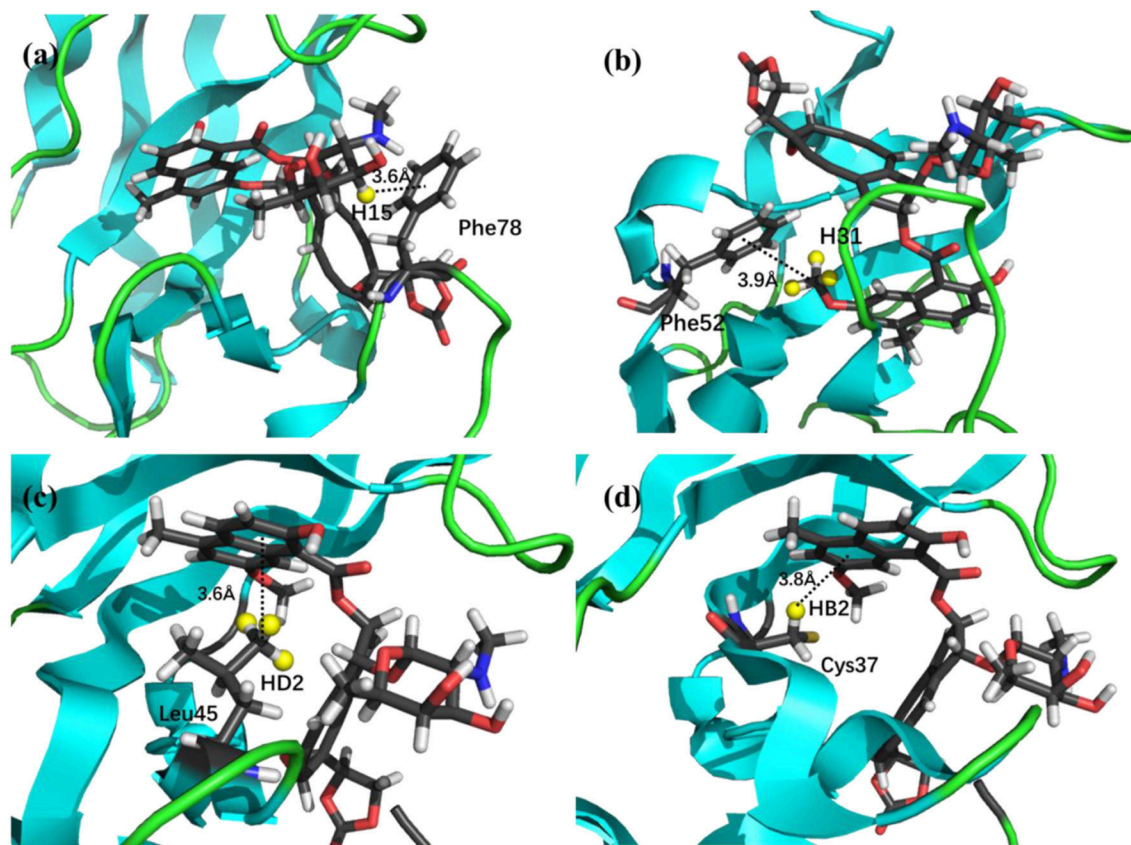


FIGURE 5 | X-ray structure of the holo NCS. The CSPs for protons of H15 (a) and H31 (b) in the chromophore, Leu45:HD2 (c) and Cys37:HB2 (d) in the NCS, are significantly influenced by the ring current effect.

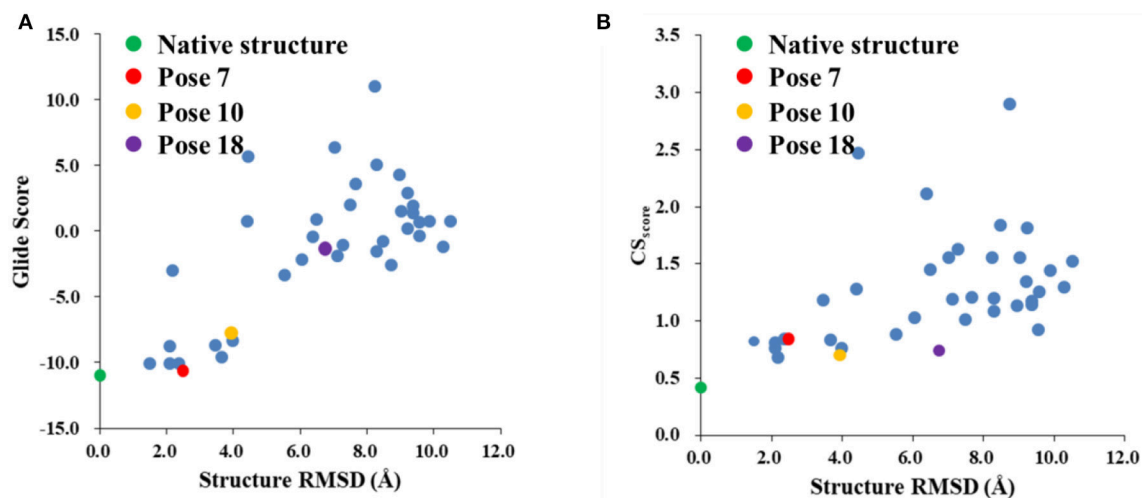


FIGURE 6 | The Glide (A) and CSscore (B) scores with different structural RMSDs of the chromophore in the holo NCS. The green dot represents the score of the binding pose in the X-ray structure of the holo NCS. The red, yellow and violet dots denote the docking poses 7, 10 and 18, respectively.

to bound states, straying away of the naphthoate group in Pose 10 caused that the chemical shifts are less affected upon ligand binding, which results in slightly higher CS_{score} score

(see Figure 6B, but other groups are close to the native state). Considering the physical non-bonded interactions between NCS and chromophore, as the naphthoate group stays deep inside the

binding pocket in the native state, which contributes most to the NCS-chromophore binding energy. Therefore, the structural deformation in pose 10 caused that the interaction energy between them became weaker, and the corresponding Glide score gave the lower rank. In the hybrid scoring function CS_{Gscore} , the ranking of pose 10 has a clear separation from the experimental structure, because it incorporates both the NMR chemical shift deviations from the experimental data and the physical interaction energy between the protein and ligand.

The docking pose 18 is also the case that the energy function is more important than the CS_{score} in ranking the docking poses. **Figure 10** shows that the ligand position of pose 18 almost translated to the direction away from the binding pocket. As the naphthoate ring moved away, the chemical shieldings of

Cys37:HB2, Cys37:HB3 and Leu45:HD2 decreased as compared to the native state (see **Table 3**). The ligand pose 18 is closer

TABLE 1 | The comparison between the experimental and calculated chemical shifts (in ppm).

	Experiment	Native	Pose 7
LIG:H13	5.98	5.87	4.45
LIG:H15	3.42	3.12	4.31

"Native" and "Pose 7" denote the calculated chemical shifts for the native binding pose and the docking pose of Pose 7, respectively.

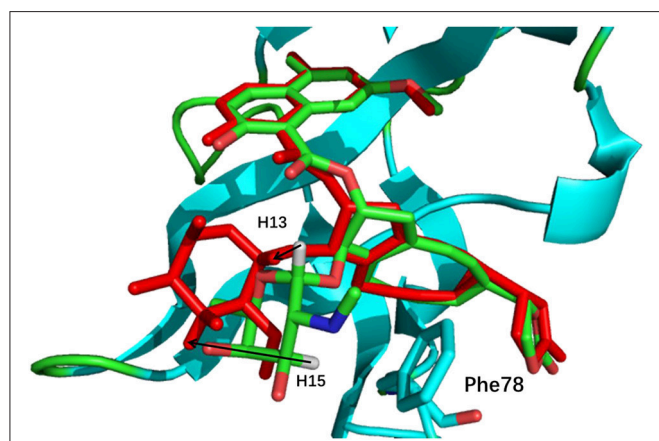


FIGURE 7 | The comparison between the binding structure of Pose 7 and the native binding pose. The green stick model denotes the native binding pose, and the red stick model denotes the binding structure of Pose 7 from Glide docking program. The aminosugar group in the chromophore was displaced in the docking pose 7 as compared to the experimental structure.

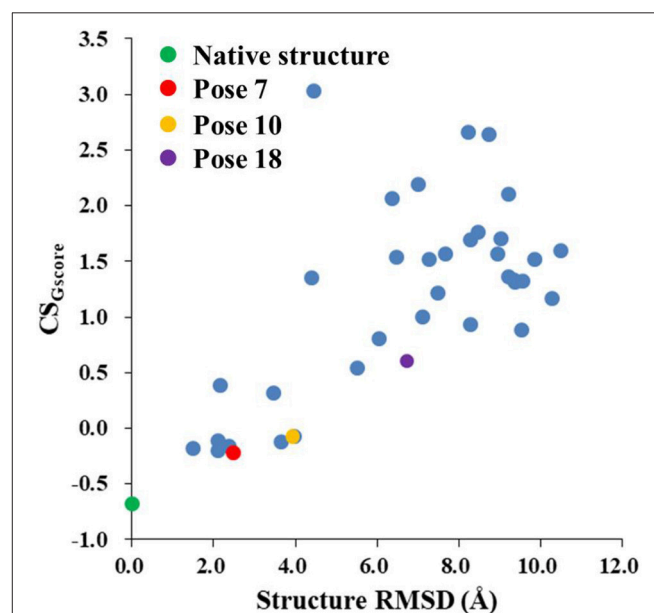


FIGURE 8 | The rankings of the native (green) and docking structures (red: pose 7; yellow: pose 10; violet: pose 18; and blue: other docked poses, predicted by Glide) calculated by CS_{Gscore} .

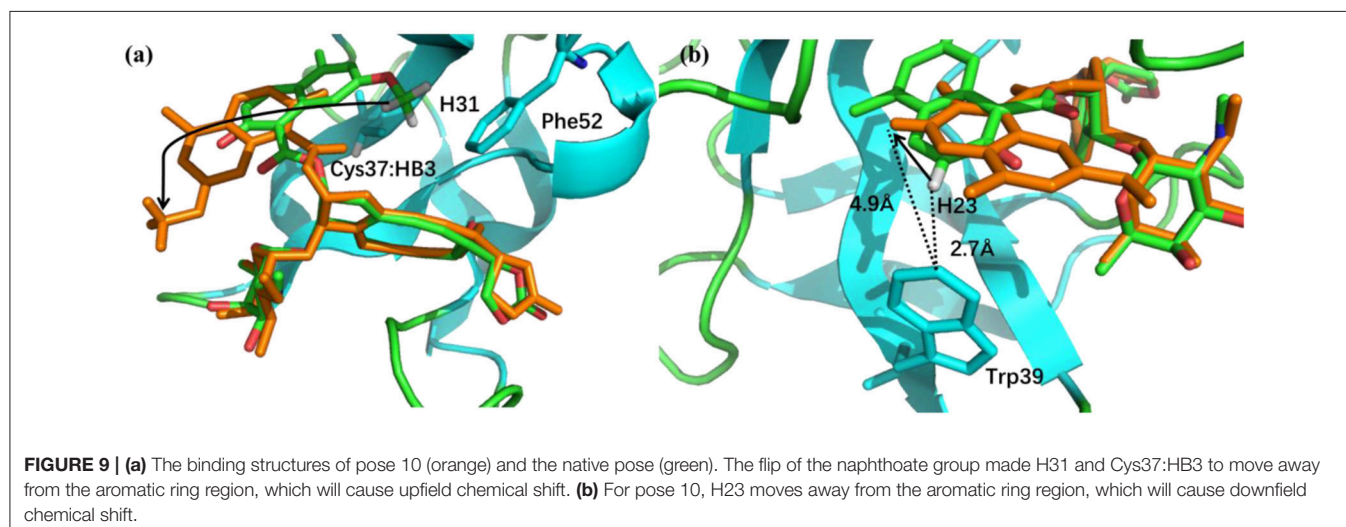
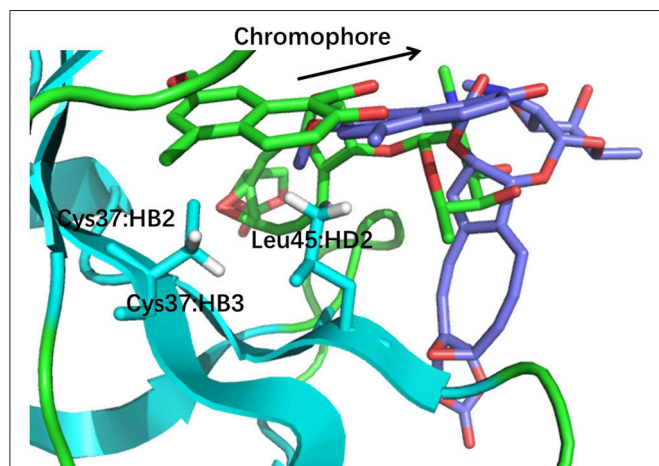


FIGURE 9 | (a) The binding structures of pose 10 (orange) and the native pose (green). The flip of the naphthoate group made H31 and Cys37:HB3 to move away from the aromatic ring region, which will cause upfield chemical shift. (b) For pose 10, H23 moves away from the aromatic ring region, which will cause downfield chemical shift.

TABLE 2 | Comparison between the experimental and calculated chemical shifts on the native binding model and pose 10 (in ppm).

	Experiment	Native	Pose 10
LIG:H23	7.40	7.91	6.14
LIG:H31	2.96	3.58	4.92
Cys37:HB3	2.17	2.08	3.19

**FIGURE 10** | The structures of pose 18 (violet) and the native state (green). The chromophore in pose 18 shifted away from the position of the native state, making that Cys37:HB2, Cys37:HB3, and Leu45:HD2 move away from the aromatic rings of the chromophore.**TABLE 3** | Comparison between the experimental and calculated chemical shifts on the native state and the docking pose 18 (in ppm).

	Experiment	Native	Pose 18
Cys37:HB2	1.98	1.52	2.60
Cys37:HB3	2.17	2.08	2.82
Leu45:HD2	-0.90	-1.25	0.69

to the apo state, which resulted in high CS_{score} score, but the interaction energy between the chromophore and NCS would be obviously weaker than that of the native state, and the Glide score for the pose 18 is substantially higher than the native binding pose (see **Figure 6A**). Therefore, in the hybrid scoring function CS_{Gscore} , the ranking of pose 18 is clearly lower than the native state, which correctly reflects that the rankings decrease as the structural RMSDs become larger.

CONCLUSION

In this work, we applied the automated fragmentation method for QM/MM calculation of NMR chemical shifts for protein-ligand binding complexes. In the AF-QM/MM approach, the atomic NMR chemical shifts were obtained by dividing the protein automatically into residue-centric fragments. In order to

reduce the computational cost for the ligand, the chromophore that contains 81 atoms was also divided into three smaller fragments to make the QM size for ligand calculation comparable to the protein fragments. The AF-QM/MM approach with the implicit solvation treatment is computationally efficient and linear-scaling with a low pre-factor. Moreover, the approach is massively parallel and can be applied to routinely calculate the *ab initio* NMR chemical shifts for protein-ligand complexes of any size.

The 1H chemical shifts calculated by the AF-QM/MM approach at the DFT level are in good agreement with large-sized system calculation, where the entire ligand and its buffer region are treated by QM, and the remaining atoms of the protein are described by background charges. The MUE between AF-QM/MM and large-sized system calculation is 0.046 ppm. Furthermore, the MUEs between calculated and experimental 1H chemical shifts in the binding pocket of apo and holo NCS are 0.45 and 0.44 ppm, respectively. Our results demonstrate that the AF-QM/MM approach is capable of reproducing the large-sized system *ab initio* calculations of NMR chemical shifts for protein-ligand complexes, and the calculated chemical shifts are in good agreement with the experimental results.

The results of CS_{score} scores show that chemical shifts could be utilized as molecular probes to detect the binding conformation of the protein-ligand complex. The experimental structure has the clear leading score as compared to the decoy binding poses. By investigating the CSP patterns of decoy structures, the position changes of the ligand could be detected by variations of chemical shifts in different local chemical environment.

In this study, we further proposed the hybrid scoring function CS_{Gscore} which combines CS_{score} and the energy-based scoring function of Glide score. The hybrid CS_{Gscore} scoring function can help to distinguish the native ligand structure from the decoy docking poses. CS_{Gscore} can also clearly separate the scores of decoy structures, which have significantly large structural RMSD values and give relatively low CS_{score} scores, from the native docking pose. The CS_{Gscore} incorporates both the experimental NMR chemical shift information and the energy-based scoring method, which could better determine the binding site structure of the protein-ligand complex. Therefore, the AF-QM/MM approach provides an accurate and efficient platform for protein-ligand binding structure prediction based on NMR derived information.

AUTHOR CONTRIBUTIONS

XH designed research; XJ, TZ, and XH performed research; XJ, TZ, JZ, and XH analyzed data; and XJ and XH wrote the paper.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (Grant No. 2016YFA0501700), National Natural Science Foundation of China (Nos. 21673074, 21761132022 and 21433004), Youth Top-Notch Talent Support Program of

Shanghai, NYU-ECNU Center for Computational Chemistry at NYU Shanghai, and Shanghai Putuo District (Grant 2014-A-02). We thank the Supercomputer Center of East China Normal University for providing us with computational time.

REFERENCES

- Adeniyi, A. A., and Soliman, M. E. S. (2017). Implementing QM in docking calculations: is it a waste of computational time? *Drug Discov. Today* 22, 1216–1223. doi: 10.1016/j.drudis.2017.06.012
- Aguirre, C., ten Brink, T., Cala, O., Guichou, J.-F., and Krimm, I. (2014). Protein–ligand structure guided by backbone and side-chain proton chemical shift perturbations. *J. Biomol. NMR* 60, 147–156. doi: 10.1007/s10858-014-9864-9
- Bratholm, L. A., and Jensen, J. H. (2017). Protein structure refinement using a quantum mechanics-based chemical shielding predictor. *Chem. Sci.* 8, 2061–2072. doi: 10.1039/C6SC04344E
- Case, D. A. (1998). The use of chemical shifts and their anisotropies in biomolecular structure determination. *Curr. Opin. Struct. Biol.* 8, 624–630. doi: 10.1016/S0959-440X(98)80155-3
- Case, D., Darden, T., Cheatham, T. E. III., Simmerling, C., Wang, J., Duke, R., et al. (2012). *AMBER 12*. San Francisco, CA: University of California.
- Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proc. Nat. Acad. Sci. U.S.A.* 104, 9615–9620. doi: 10.1073/pnas.0610313104
- Christensen, A. S., Linnet, T. E., Borg, M., Boomsma, W., Lindorff-Larsen, K., Hamelryck, T., et al. (2014). Protein structure validation and refinement using amide proton chemical shifts derived from quantum mechanics. *PLoS ONE* 8:e84123. doi: 10.1371/journal.pone.0084123
- Cioffi, M., Hunter, C. A., Packer, M. J., Pandya, M. J., and Williamson, M. P. (2008). Use of quantitative ^1H NMR chemical shift changes for ligand docking into barnase. *J. Biomol. NMR* 43, 11–19. doi: 10.1007/s10858-008-9286-7
- Cui, Q., and Karplus, M. (2000). Molecular properties from combined QM/MM methods. 2. chemical shifts in large molecules. *J. Phys. Chem. B* 104, 3721–3743. doi: 10.1021/jp994154g
- Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., and Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* 11, 425–445.
- Englebienne, P., and Moitessier, N. (2009). Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J. Chem. Inf. Model.* 49, 2564–2571. doi: 10.1021/ci900251k
- Ferreira, L., dos Santos, R., Oliva, G., and Andricopulo, A. (2015). Molecular docking and structure-based drug design strategies. *Molecules* 20, 13384–13421. doi: 10.3390/molecules200713384
- Flaig, D., Beer, M., and Ochsenfeld, C. (2012). Convergence of electronic structure with the size of the QM region: example of QM/MM NMR shieldings. *J. Chem. Theor. Comput.* 8, 2260–2271. doi: 10.1021/ct300036s
- Frank, A., Möller, H. M., and Exner, T. E. (2012). Toward the quantum chemical calculation of NMR chemical shifts of proteins. 2. Level of theory, basis set, and solvents model dependence. *J. Chem. Theor. Comput.* 8, 1480–1492. doi: 10.1021/ct200913r
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430
- Frisch, M., Trucks, G., Schlegel, H. B., Scuseria, G., Robb, M., Cheeseman, J., et al. (2009). *Gaussian 09*. Wallingford, CT: Gaussian Inc.
- Gao, Q., Yokojima, S., Fedorov, D. G., Kitaura, K., Sakurai, M., and Nakamura, S. (2010). Fragment-molecular-orbital-method-based ab initio NMR chemical-shift calculations for large molecular systems. *J. Chem. Theor. Comput.* 6, 1428–1444. doi: 10.1021/ct100006n
- Gao, Q., Yokojima, S., Kohno, T., Ishida, T., Fedorov, D. G., Kitaura, K., et al. (2007). Ab initio NMR chemical shift calculations on proteins using fragment molecular orbitals with electrostatic environment. *Chem. Phys. Lett.* 445, 331–339. doi: 10.1016/j.cplett.2007.07.103
- Garay, P. G., Martin, O. A., Scheraga, H. A., and Vila, J. A. (2014). Factors affecting the computation of the ^{13}C shielding in disaccharides. *J. Comput. Chem.* 35, 1854–1864. doi: 10.1002/jcc.23697
- Gohlke, H., Hendlich, M., and Klebe, G. (2000). Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* 295, 337–356. doi: 10.1006/jmbi.1999.3371
- Grinter, S., and Zou, X. (2014). Challenges, applications, and recent advances of protein–ligand docking in structure-based drug design. *Molecules* 19, 10150–10176. doi: 10.3390/molecules190710150
- Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *J. Med. Chem.* 47, 1750–1759. doi: 10.1021/jm030644s
- Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* 50, 43–57. doi: 10.1007/s10858-011-9478-4
- Hartman, J. D., and Beran, G. J. O. (2014). Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals. *J. Chem. Theor. Comput.* 10, 4862–4872. doi: 10.1021/ct500749h
- He, X., Wang, B., and Merz, K. M. Jr. (2009). Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B* 113, 10380–10388. doi: 10.1021/jp901992p
- He, X., Zhu, T., Wang, X., Liu, J., and Zhang, J. Z. (2014). Fragment quantum mechanical calculation of proteins and its applications. *Acc. Chem. Res.* 47, 2748–2757. doi: 10.1021/ar500077t
- Huang, S.-Y., and Zou, X. (2006). An iterative knowledge-based scoring function to predict protein–ligand interactions: I. derivation of interaction potentials. *J. Comput. Chem.* 27, 1866–1875. doi: 10.1002/jcc.20504
- Jakalian, A., Jack, D. B., and Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* 23, 1623–1641. doi: 10.1002/jcc.10128
- Jin, X., Zhu, T., Zhang, J. Z., and He, X. (2016). A systematic study on RNA NMR chemical shift calculation based on the automated fragmentation QM/MM approach. *RSC Adv.* 6, 108590–108602. doi: 10.1039/C6RA22518G
- Jorgensen, W. L., and Jenson, C. (1998). Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.* 19, 1179–1186. doi: 10.1002/(SICI)1096-987X(19980730)19:10:3.CO;2-J
- Lodewyk, M. W., Siebert, M. R., and Tantillo, D. J. (2012). Computational prediction of ^1H and ^{13}C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* 112, 1839–1862. doi: 10.1021/cr200106v
- McCoy, M. A., and Wyss, D. F. (2002). Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.* 124, 11758–11763. doi: 10.1021/ja026166c
- Medek, A., Hajduk, P. J., Mack, J., and Fesik, S. W. (2000). The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *J. Am. Chem. Soc.* 122, 1241–1242. doi: 10.1021/ja993921m
- Meiler, J. (2003). PROSHIFT: protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* 26, 25–37. doi: 10.1023/A:1023060720156
- Meiler, J., and Baker, D. (2003). Rapid protein fold determination using unassigned NMR data. *Proc. Nat. Acad. Sci. U.S.A.* 100, 15404–15409. doi: 10.1073/pnas.2434121100
- Merz, K. M. (2014). Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* 47, 2804–2811. doi: 10.1021/ar5001023

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2018.00150/full#supplementary-material>

- Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., and DesJarlais, R. L. (2005). Docking: successes and challenges. *Curr. Drug Metab.* 11, 323–333. doi: 10.2174/1381612053382106
- Mohanty, S., Sieker, L. C., and Drobny, G. P. (1994). Sequential ^1H NMR assignment of the complex of aponecarzinostatin with ethidium bromide and investigation of protein-drug interactions in the chromophore binding site. *Biochemistry* 33, 10579–10590. doi: 10.1021/bi00201a003
- Moon, S., and Case, D. A. (2007). A new model for chemical shifts of amide hydrogens in proteins. *J. Biomol. NMR* 38, 139–150. doi: 10.1007/s10858-007-9156-8
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662. doi: 10.1002/(SICI)1096-987X(19981115)19:143.0.CO;2-B
- Murray, C. W., Auton, T. R., and Eldridge, M. D. (1998). Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of bayesian regression to improve the quality of the model. *J. Comput. Aided Mol. Des.* 12, 503–519.
- Myers, A. G., Proteau, P. J., and Handel, T. M. (1988). Stereochemical assignment of neocarzinostatin chromophore. Structures of neocarzinostatin chromophore-methyl thioglycolate adducts. *J. Am. Chem. Soc.* 110, 7212–7214.
- Neal, S., Nip, A. M., Zhang, H., and Wishart, D. S. (2003). Rapid and accurate calculation of protein ^1H , ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR* 26, 215–240. doi: 10.1023/A:1023812930288
- Riedinger, C., Endicott, J. A., Kemp, S. J., Smyth, L. A., Watson, A., Valeur, E., et al. (2008). Analysis of chemical shift changes reveals the binding modes of isoindolinone inhibitors of the MDM2-p53 interaction. *J. Am. Chem. Soc.* 130, 16038–16044. doi: 10.1021/ja8062088
- Robustelli, P., Kohlhoff, K., Cavalli, A., and Vendruscolo, M. (2010). Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* 18, 923–933. doi: 10.1016/j.str.2010.04.016
- Rocchia, W., Alexov, E., and Honig, B. (2001). Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* 105, 6507–6514. doi: 10.1021/jp010454y
- Schau, S. E., Cavalieri, D., and Myers, A. G. (2001). Gene transcription analysis of *Saccharomyces cerevisiae* exposed to neocarzinostatin protein–chromophore complex reveals evidence of DNA damage, a potential mechanism of resistance, and consequences of prolonged exposure. *Proc. Nat. Acad. Sci. U.S.A.* 98, 11075–11080. doi: 10.1073/pnas.191340698
- Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* 48, 13–22. doi: 10.1007/s10858-010-9433-9
- Shen, Y., and Bax, A. (2015). Homology modeling of larger proteins guided by chemical shifts. *Nat. Methods* 12, 747–750. doi: 10.1038/nmeth.3437
- Śledź, P., and Cafilisch, A. (2018). Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* 48, 93–102. doi: 10.1016/j.sbi.2017.10.010
- Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi: 10.1124/pr.112.007336
- Swails, J., Zhu, T., He, X., and Case, D. A. (2015). AFNMR: automated fragmentation quantum mechanical calculation of NMR chemical shifts for biomolecules. *J. Biomol. NMR* 63, 125–139. doi: 10.1007/s10858-015-9970-3
- Takashima, H., Yoshida, T., Ishino, T., Hasuda, K., Ohkubo, T., and Kobayashi, Y. (2005). Solution NMR structure investigation for releasing mechanism of neocarzinostatin chromophore from the holoprotein. *J. Biol. Chem.* 280, 11340–11346. doi: 10.1074/jbc.M411579200
- Tan, H.-J., and Bettens, R. P. (2013). Ab initio NMR chemical-shift calculations based on the combined fragmentation method. *Phys. Chem. Chem. Phys.* 15, 7541–7547. doi: 10.1039/c3cp50406a
- Ten Brink, T., Aguirre, C., Exner, T. E., and Krimm, I. (2015). Performance of protein-ligand docking with simulated chemical shift perturbations. *J. Chem. Inf. Model.* 55, 275–283. doi: 10.1021/ci500446s
- Victoria, A., Möller, H. M., and Exner, T. E. (2014). Accurate ab initio prediction of NMR chemical shifts of nucleic acids and nucleic acids/protein complexes. *Nucleic Acids Res.* 42, e173–e173. doi: 10.1093/nar/gku1006
- Vila, J. A., Arnautova, Y. A., Martin, O. A., and Scheraga, H. A. (2009). Quantum-mechanics-derived $^{13}\text{C}^\alpha$ chemical shift server (CheShift) for protein structure validation. *Proc. Nat. Acad. Sci. U.S.A.* 106, 16972–16977. doi: 10.1073/pnas.0908833106
- Wang, B., and Merz, K. M. Jr. (2010). Importance of loop dynamics in the neocarzinostatin chromophore binding and release mechanisms. *Phys. Chem. Chem. Phys.* 12, 3443–3449. doi: 10.1039/b924951f
- Wang, B., Westerhoff, L. M., and Merz, K. M. (2007). A critical assessment of the performance of protein–ligand scoring functions based on NMR chemical shift perturbations. *J. Med. Chem.* 50, 5128–5134. doi: 10.1021/jm070484a
- 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, 1157–1174. doi: 10.1002/jcc.20035
- Williamson, M. P., and Craven, C. J. (2009). Automated protein structure calculation from NMR data. *J. Biomol. NMR* 43, 131–143. doi: 10.1007/s10858-008-9295-6
- Xu, X.-P., and Case, D. A. (2001). Automated prediction of ^{15}N , $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ and $^{13}\text{C}'$ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* 21, 321–333. doi: 10.1023/A:1013324104681
- Yu, Z., Li, P., and Merz, K. M. (2017). Using ligand-induced protein chemical shift perturbations to determine protein–ligand structures. *Biochemistry* 56, 2349–2362. doi: 10.1021/acs.biochem.7b00170
- Zhu, T., He, X., and Zhang, J. Z. (2012). Fragment density functional theory calculation of NMR chemical shifts for proteins with implicit solvation. *Phys. Chem. Chem. Phys.* 14, 7837–7845. doi: 10.1039/C2CP23746F
- Zhu, T., Zhang, J. Z. H., and He, X. (2013). Automated fragmentation QM/MM calculation of amide proton chemical shifts in proteins with explicit solvent model. *J. Chem. Theor. Comput.* 9, 2104–2114. doi: 10.1021/ct300999w
- Zhu, T., Zhang, J. Z. H., and He, X. (2014). Correction of erroneously packed protein's side chains in the NMR structure based on ab initio chemical shift calculations. *Phys. Chem. Chem. Phys.* 16, 18163–18169. doi: 10.1039/C4CP02553A
- Zhu, T., Zhang, J. Z. H., and He, X. (2015). “Quantum calculation of protein NMR chemical shifts based on the automated fragmentation method,” in *Advance in Structural Bioinformatics*, eds D. Wei, Q. Xu, T. Zhao and H. Dai (Dordrecht: Springer), 827, 49–70.

Conflict of Interest Statement: 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.

Copyright © 2018 Jin, Zhu, Zhang and He. 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 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.