



A New Metric Quantifying Chemical and Biological Property of Small Molecule Metabolites and Drugs

Chuanbo Huang, Yuan Zhou, Jichun Yang, Qinghua Cui* and Yanhui Li*

MOE Key Laboratory of Cardiovascular Sciences, Department of Biomedical Informatics, Department of Physiology and Pathophysiology, Center for Noncoding RNA Medicine, School of Basic Medical Sciences, Peking University, Beijing, China

One prominent class of drugs is chemical small molecules (CSMs), but the majority of CSMs are of very low druggable potential. Therefore, it is quite important to predict drug-related properties (druggable properties) for candidate CSMs. Currently, a number of druggable properties (e.g., logP and pKa) can be calculated by *in silico* methods; still the identification of druggable CSMs is a high-risk task, and new quantitative metrics for the druggable potential of CSMs are increasingly needed. Here, we present normalized bond energy (NBE), a new metric for the above purpose. By applying NBE to the DrugBank CSMs whose properties are largely known, we revealed that NBE is able to describe a number of critical druggable properties including logP, pKa, membrane permeability, blood–brain barrier penetration, and human intestinal absorption. Moreover, given that the human endogenous metabolites can serve as important resources for drug discovery, we applied NBE to the metabolites in the Human Metabolome Database. As a result, NBE showed a significant difference in metabolites from various body fluids and was correlated with some important properties, including melting point and water solubility.

OPEN ACCESS

Edited by:

Lei Deng, Central South University, China

Reviewed by:

Haixiu Yang, Harbin Medical University, China Qinghua Jiang, Harbin Institute of Technology, China

*Correspondence:

Qinghua Cui cuiqinghua@hsc.pku.edu.cn Yanhui Li liyanhui@bjmu.edu.cn

Specialty section:

This article was submitted to Metabolomics, a section of the journal Frontiers in Molecular Biosciences

Received: 28 September 2020 Accepted: 02 November 2020 Published: 15 December 2020

Citation:

Huang C, Zhou Y, Yang J, Cui Q and Li Y (2020) A New Metric Quantifying Chemical and Biological Property of Small Molecule Metabolites and Drugs. Front. Mol. Biosci. 7:594800. doi: 10.3389/fmolb.2020.594800 Keywords: drug, chemical small molecule, metabolites, druggable property, normalized bond energy

INTRODUCTION

Research and development of pharmaceuticals is a resource-consuming and long process with a variety of challenging risks (Szewczak et al., 2020). Chemical small molecules (CSMs) represent a big class of drugs which mainly function by binding with disease-related target molecules (Wishart et al., 2018a). Given the huge space of target molecules and CSMs, evaluating the druggable potential of both targets (Jung and Kwon, 2015; Liu et al., 2016; Floris et al., 2018; Sztuba-Solinska et al., 2019) and CSMs (Sun et al., 2016; Ashenden et al., 2017; Chitre et al., 2019; Heitmeier et al., 2019; Bhattacharjee et al., 2020) is thus one of the key points of drug discovery. For CSMs, it is known that a number of drug-related properties (druggable properties) affect their druggable potential, for example, human intestinal absorption (HIA), blood–brain barrier (BBB) penetration (Blake, 2000), and some pharmacokinetic properties (Ferreira and Andricopulo, 2019). Therefore, it is crucial to accurately predict druggable properties for an early-phase candidate CSM and large-scale druggable CSM screening.

For the above purpose, a number of *in silico* methods or metrics have already been proposed. For example, properties of logP, logD, logS, logW, and pKa can be calculated using the free

1

online tool ALOGPS (Tetko and Tanchuk, 2002). ChemAxon, a tool that provides solutions and services for chemistry and biology (ChemAxon, 2020), can be used to predict druggable properties such as water solubility, polar surface area (PSA), H bond acceptor count, H bond donor count, and pKa. In addition, given that CSM transport is a key attribute toward better drug potential, several metrics have been presented to evaluate CSM transport properties, including octanol/water partition coefficient, molecular size and shape, hydrogenbonding capabilities, and topological PSA (van De Waterbeemd et al., 1996; van de Waterbeemd, 1998; Winiwarter et al., 1998; Ertl et al., 2000). These in silico methods or metrics provide support for quickly quantifying CSM properties in order to evaluate their druggable potential. However, due to the huge complexity of both biology and chemistry, these methods or metrics are still far from solving all problems in drug research and development. For example, when chemical structures are diverse and complex, the molecular-transport-related physicochemical metric descriptors introduced above may not be reliable enough to predict molecular transport properties (Artursson et al., 1996). Thus, it is necessary to present new in silico methods or metrics to quantify druggable properties of CSMs.

We previously revealed that the free energy of the RNA secondary structure has a significant contribution to the importance score of both protein-coding RNA molecules (mRNAs) and noncoding RNAs (lncRNAs and miRNAs) (Zeng et al., 2018; Song et al., 2019). Based on the above observations, we hypothesized that the energy status of CSMs could also represent some properties of these molecules. To confirm this hypothesis, here, we present normalized bond energy (NBE), a new metric. Moreover, we reveal here that the NBE score can significantly represent some critical druggable properties-such as logP, pKa, permeability, BBB penetration, and HIA. Additionally, given that the human endogenous metabolites could be explored as a resource for drug discovery (Bofill et al., 2019), we calculated the NBE scores for CSMs in the human metabolome and performed a comprehensive bioinformatic analysis for the relations between NBE score and other properties of these endogenous metabolic small molecules.

MATERIALS AND METHODS

Datasets of CSMs

We obtained the structural data in SDF format for CSMs from the DrugBank database (Wishart et al., 2018a) (Version 5.0), which include approved small molecule drugs and experimental drugs. Biological macromolecular drugs were excluded from the dataset. We obtained the structural data in SDF format for small molecule metabolites from the Human Metabolome Database (HMDB) (Wishart et al., 2018b) as well. Experiment-derived property (e.g., melting point, logP, and pKa) data of CSMs were also curated from the DrugBank and HMDB. For the property of water solubility, CSMs with terms like "insoluble," "almost insoluble," "low soluble," "mostly insoluble," "non-soluble," "not soluble," and "poorly soluble" were assigned as the insoluble group, whereas CSMs with terms like "soluble," "easily soluble," "completely soluble," "freely soluble," "highly soluble," and "very soluble" were assigned as the soluble group. In addition, we obtained the Caco-2 monolayer permeability data of 690 CSMs from the study done by van De Waterbeemd et al. (van De Waterbeemd et al., 1996; Palm et al., 1998; Pham The et al., 2011), the BBB penetration data of 1,638 CSMs from the study reported by Kelder et al. (Kelder et al., 1999; Shen et al., 2010), and the HIA data of 598 CSMs from the study done by Shen et al. (Palm et al., 1997; Shen et al., 2010). The structure files of these CSMs in SMILES format were obtained as well.

Calculation of NBE

For a representative CSM, we first extracted its bonds from its molecular structure using the RDKit library (Floris et al., 2018) (Version 2017.09.1, 2017) and then determined the energy (kJ/mol) of each bond by matching it with the bond energy table (Supplementary Table 1) through the bond type and the atom type. The bond type here includes single bond, double bond, and triple bond, which denotes the number of shared electron pairs between the two corresponding atoms. Bond energy is defined as a parameter that physically quantifies the strength of a chemical bond and can be measured by the amount of energy required to break a bond. In general, bond energy is the average value of bond dissociation energy of a mole of molecule in the gas phase, typically at a temperature of 298 K. Given that the bigger molecules usually have more bonds and thus would have larger bond energy, we next defined NBE. Here, the original bond energy is normalized using molecular weight (MW), which was calculated using RDKit. The algorithm for the procedure is shown in Figure 1.

Then, NBE can be calculated using the following equation.

$$NBE = \frac{\sum_{i=1}^{n} Bond \ Energy(i)}{MW}$$
(1)

where *Bond Energy* (i) is the bond energy of bond *i*, *n* is the number of bonds, and *MW* is the molecular weight of the CSM.

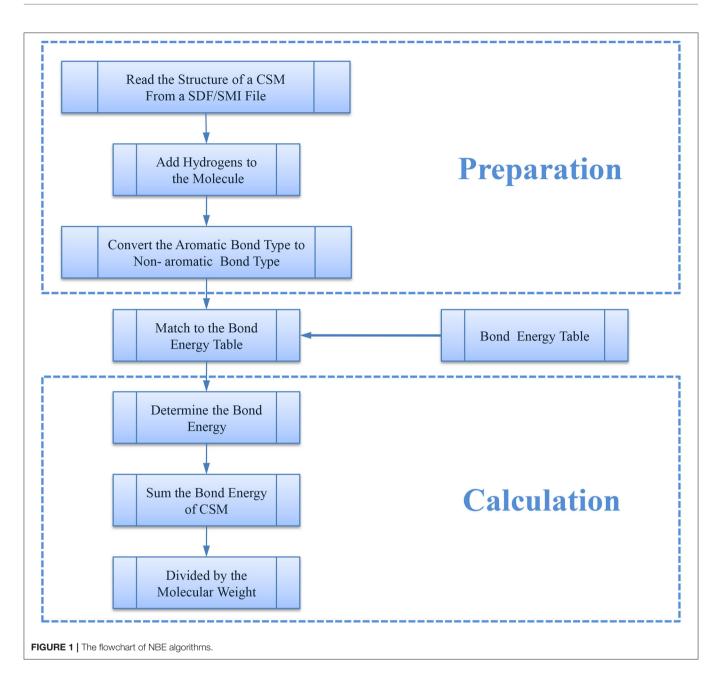
Statistical Computation

We implemented the algorithm of NBE (http://www.cuilab.cn/ nbe/nbe.zip) using Python. Spearman's correlation analysis, *t*test, and Wilcoxon test were performed using R studio.

RESULTS

Global Distribution of NBE Scores

The whole framework of this study is shown in **Figure 2**. We calculated NBE scores for 10,426 CSMs (2,444 are approved drugs and 7,982 are experimental drugs) from DrugBank and 113,878 human metabolic CSMs from HMDB. The distributions of the DrugBank NBE scores are shown in **Figure 3A**. The approved drugs have greater NBE scores than the unapproved drugs (*p*-value = 1.17e-13, Wilcoxon test; **Figure 3B**).



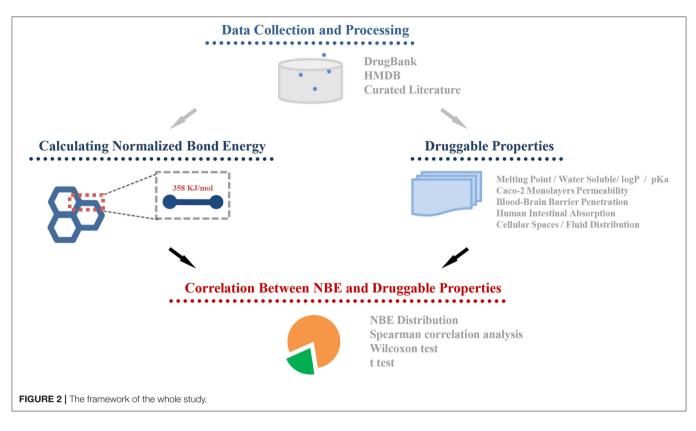
Correlations of NBE Scores With Experimentally Identified Properties of CSMs in DrugBank

As a result, we found that NBE score is significantly associated with a number of experimentally identified properties of CSMs in DrugBank. First, we evaluated if there is a difference in NBE scores of the soluble CSMs and the insoluble ones. The results showed that the soluble CSMs have smaller NBE scores compared to the insoluble ones (mean: 53.3 vs. 56.7, *p*-value = 0.009, *t*-test; **Figure 4A**). Further, we investigated the relations of NBE score with the melting point, logP, and pKa. We found that NBE score shows a significantly negative correlation with melting point (Rho = -0.19, *p*-value = 1.89e-13, **Figure 4B**)

but shows a positive correlation with logP (Rho = 0.36, *p*-value = 1.25e-47, **Figure 4C**) and pKa (Rho = 0.38, *p*-value = 9.08e-19, **Figure 4D**).

NBE Is Correlated With Permeability, BBB Penetration, and HIA

Permeability, BBB penetration, and HIA are the three critical properties which greatly affect the transport properties of a CSM. We observed significant correlations between NBE and Caco-2 monolayer permeability (Rho = 0.80, *p*-value = 0.01, **Figure 5A**; Rho = 0.71, *p*-value = 0.003, **Figure 5B**; Rho = 0.22, *p*-value = 9.186e-09, **Figure 5C**), between NBE and HIA (Rho = 0.44, *p*-value = 0.05, **Figure 5D**, Rho = 0.12, *p*-value = 3.567e-03,



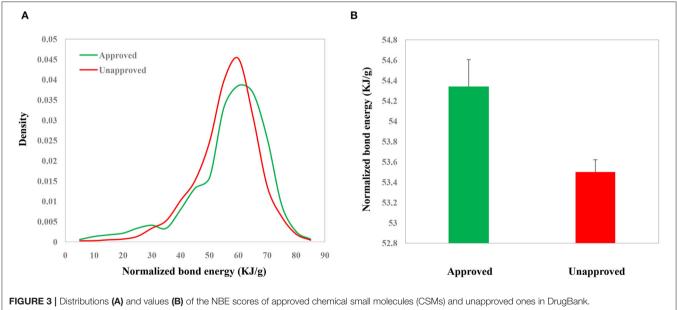


Figure 5E), and between NBE and BBB penetration (Rho = 0.55, p-value = 9.39e-05, **Figure 5F**). Moreover, using the FA% value (the oral drug absorption in humans) threshold of 30%, we divide CSMs of the HIA dataset into absorbable (HIA+) terms and nonabsorbable (HIA-) terms. We observed that HIA+ CSMs have bigger NBE scores than the HIA- CSMs (mean: 57.0 vs.

52.4, *p*-value = 0.0064, *t*-test; **Figure 5G**). Whether CSMs can penetrate the BBB has been recognized as one of the most critical issues for designing drugs targeting the central nervous system. We also observed that BBB penetrable (BBB+) CSMs have bigger NBE scores than the impenetrable (BBB-) CSMs (mean: 58.7 vs. 54.2, *p*-value = 1.08e-15, *t*-test; **Figure 5H**).

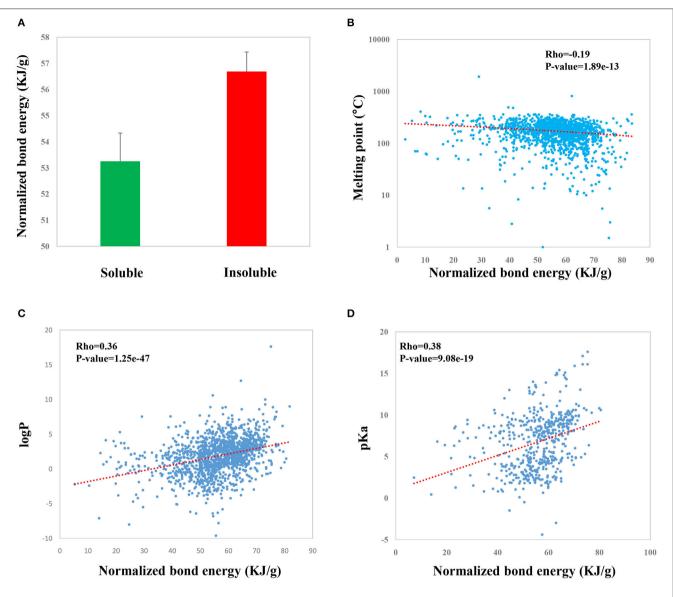
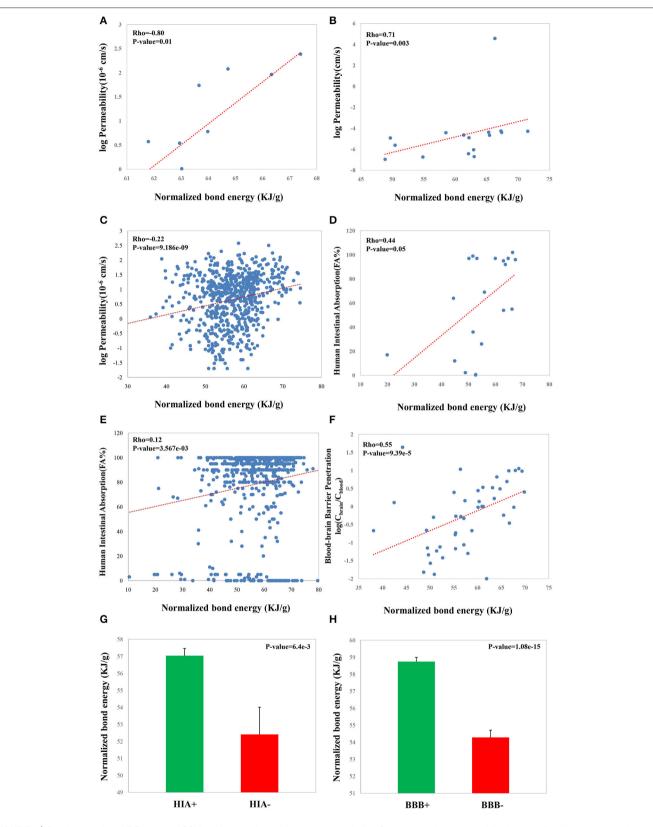


FIGURE 4 | Correlations of the NBE scores of CSMs in DrugBank with druggable properties including water solubility (A), melting point (B), logP (C), and pKa (D).

NBE Is Correlated With Properties of Human Metabolites

Natural products represent one important class of CSMs with high druggable potential. The human endogenous metabolites a type of natural product—could be explored as a resource for drug discovery. It is thus important to investigate whether NBE can describe some properties of human metabolites. We previously revealed biased subcellular distributions for miRNA target genes and sex-biased genes. We found that miRNAs prefer to target genes located within the inner cellular space compared to genes located in the outer cellular space (Cui et al., 2006). Female-biased genes are enriched in the outer cellular space, whereas male-biased genes are enriched in the inner cellular space (Guo et al., 2018). It was considered of much interest to investigate whether there exists a difference in NBE for metabolites in different cellular spaces. The results showed that metabolites in the outer cellular space (metabolites in the extracellular space and/or membrane) have greater NBE scores than those in the inner cellular space (metabolites in the cytoplasm and/or nucleus) (*p*-value = 0, Wilcoxon test; **Figure 6A**). Moreover, metabolites derived from different body fluids showed a significant difference in NBE scores (*p*-value = 0.0, ANOVA, **Figure 6B**). Metabolites from feces and saliva showed the highest NBE scores (**Figure 6B**). In addition, NBE scores of metabolites were correlated with melting point (Rho = -0.29, *p*-value = 7.11e-138; **Figure 6C**) and water solubility





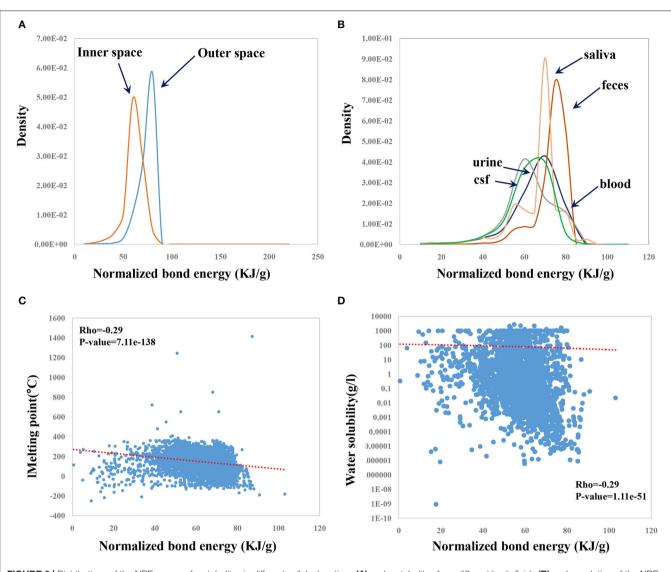


FIGURE 6 | Distributions of the NBE scores of metabolites in different cellular locations (A) and metabolites from different body fluids (B) and correlation of the NBE scores with melting point (C) and with water solubility (D) in HMDB.

(Rho = -0.29, *p*-value = 1.11e-51; Figure 6D), which is consistent with the results on CSMs in DrugBank.

DISCUSSION

In this study, we presented a new *in silico* metric, NBE, for quantifying some properties of CSMs. The results showed that NBE is able to describe some critical druggable properties of CSMs. We found that the NBE is correlated with some drug features including logP and pKa, membrane permeability, BBB penetration, and HIA. These metrics are usually used for quantifying the druggable properties of small molecules. For example, logP (octanol–water partition coefficient) is used in drug design as a measure of molecular hydrophobicity, and pKa is related to lipophilicity and the rate/extent of membrane penetration. And other properties including membrane permeability, BBB penetration, and HIA are also utilized for presenting permeability of small molecules.

The BBB separates the brain from the systemic blood circulation and maintains the homeostasis of the central nervous system. Thus, the blood-brain distribution of a CSM is a key characteristic for determining whether it is potentially druggable for the central nervous system or not. HIA is related to the rate of a particular compound crossing the intestinal wall to reach the portal blood circulation. The significant correlated relationships between NBE and BBB penetration and between NBE and HIA provide a simple but efficient metric to quickly judge the potential of a CSM to move into the brain from the circulation and the potential of a CSM to move into the circulation from the intestine.

In addition, we have similar and consistent observations for human metabolites. Interestingly, NBE distributions have a bias for different cellular locations and for different body fluids, which could provide some valuable clues toward metabolite-based drug discovery. However, we found that NBE is not a metric that can be used to judge the druggable potential of candidate small molecule only by a threshold. To use this metric to quantify the druggable potential, researchers of pharmaceuticals need to compare more small molecules with their NBE scores to draw conclusions. In addition, due to the limitation of data sources, the number of CSMs in some datasets used in this study is small and may introduce bias in the accuracy of the results. In summary, this study presented a simple but efficient metric to describe druggable properties of CSMs. The utility of NBE may be improved by combining it with other *in silico* methods or metrics in the future.

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

JY and QC designed this project. CH implemented the algorithm and wrote the paper. YL and YZ provided data or suggestions in

REFERENCES

- Artursson, P., Palm, K., and Luthman, K. (1996). Caco-2 monolayers in experimental and theoretical predictions of drug transport. Adv. Drug Deliv. Rev. 22, 67–84. doi: 10.1016/S0169-409X(96)00415-2
- Ashenden, S. K., Kogej, T., Engkvist, O., and Bender, A. (2017). Innovation in small-molecule-druggable chemical space: where are the initial modulators of new targets published? *J. Chem. Inf. Model.* 57, 2741–2753. doi: 10.1021/acs.jcim.7b00295
- Bhattacharjee, A., Hossain, M. U., Chowdhury, Z. M., Rahman, S. M. A., Bhuyan, Z. A., Salimullah, M., et al. (2020). Insight of druggable cannabinoids against estrogen receptor beta in breast cancer. *J. Biomol. Struct. Dyn.* 1–10. doi: 10.1080/07391102.2020.1737233. [Epub ahead of print].
- Blake, J. F. (2000). Chemoinformatics predicting the physicochemical properties of 'drug-like' molecules. *Curr. Opin. Biotechnol.* 11, 104–107. doi: 10.1016/S0958-1669(99)00062-2
- Bofill, A., Jalencas, X., Oprea, T. I., and Mestres, J. (2019). The human endogenous metabolome as a pharmacology baseline for drug discovery. *Drug Discov. Today* 24, 1806–1820. doi: 10.1016/j.drudis.2019.06.007
- ChemAxon (2020). InstantJChem 20.8.0.
- Chitre, N. M., Moniri, N. H., and Murnane, K. S. (2019). Omega-3 fatty acids as druggable therapeutics for neurodegenerative disorders. CNS Neurol. Disord. Drug Targets 18, 735–749. doi: 10.2174/18715273186661911140 93749
- Cui, Q., Yu, Z., Purisima, E. O., and Wang, E. (2006). Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.* 2:46. doi: 10.1038/msb4100089
- Ertl, P., Rohde, B., and Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J. Med. Chem. 43, 3714–3717. doi: 10.1021/jm000942e
- Ferreira, L. L. G., and Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discov. Today* 24, 1157–1165. doi: 10.1016/j.drudis.2019.03.015

the revision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the PKU-Baidu Fund (Grant Number 2019BD014 to QC), National Natural Science Foundation of China (Grant Numbers 62025102, 81670462, 81970440, and 81921001 to QC); Peking University Basic Research Program (Grant Number BMU2020JC001 to QC).

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at bioRxiv (Huang et al., 2020). We thank Mogoedit for language editing of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb. 2020.594800/full#supplementary-material

- Floris, M., Olla, S., Schlessinger, D., and Cucca, F. (2018). Genetic-driven druggable target identification and validation. *Trends Genet.* 34, 558–570. doi: 10.1016/j.tig.2018.04.004
- Guo, S., Zhou, Y., Zeng, P., Xu, G., Wang, G., and Cui, Q. (2018). Identification and analysis of the human sex-biased genes. *Brief Bioinform* 19, 188–198. doi: 10.1093/bib/bbw125
- Heitmeier, M. R., Hresko, R. C., Edwards, R. L., Prinsen, M. J., Ilagan, M. X. G., Odom John, A. R., et al. (2019). Identification of druggable small molecule antagonists of the *Plasmodium falciparum* hexose transporter PfHT and assessment of ligand access to the glucose permeation pathway via FLAG-mediated protein engineering. *PLoS ONE* 14:e0216457. doi: 10.1371/journal.pone.0216457
- Huang, C., Yang, J., and Cui, Q. (2020). A simple and efficient metric quantifying druggable property of chemical small molecules. *bioRxiv* [*Preprint*]. doi: 10.1101/2020.07.13.199752
- Jung, H. J., and Kwon, H. J. (2015). Target deconvolution of bioactive small molecules: the heart of chemical biology and drug discovery. *Arch. Pharm. Res.* 38, 1627–1641. doi: 10.1007/s12272-015-0618-3
- Kelder, J., Grootenhuis, P. D., Bayada, D. M., Delbressine, L. P., and Ploemen, J. P. (1999). Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* 16, 1514–1519. doi: 10.1023/A:1015040217741
- Liu, X., Baarsma, H. A., Thiam, C. H., Montrone, C., Brauner, B., Fobo, G., et al. (2016). Systematic identification of pharmacological targets from small-molecule phenotypic screens. *Cell. Chem. Biol.* 23, 1302–1313. doi: 10.1016/j.chembiol.2016.08.011
- Palm, K., Luthman, K., Ungell, A. L., Strandlund, G., Beigi, F., Lundahl, P., et al. (1998). Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors. J. Med. Chem. 41, 5382–5392. doi: 10.1021/jm98 0313t
- Palm, K., Stenberg, P., Luthman, K., and Artursson, P. (1997). Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* 14, 568–571. doi: 10.1023/A:1012188625088

- Pham The, H., Gonzalez-Alvarez, I., Bermejo, M., Mangas Sanjuan, V., Centelles, I., Garrigues, T. M., et al. (2011). *In silico* prediction of Caco-2 cell permeability by a classification QSAR approach. *Mol. Inform.* 30, 376–385. doi: 10.1002/minf.201000118
- Shen, J., Cheng, F., Xu, Y., Li, W., and Tang, Y. (2010). Estimation of ADME properties with substructure pattern recognition. J. Chem. Inf. Model. 50, 1034–1041. doi: 10.1021/ci100104j
- Song, F., Cui, C., Gao, L., and Cui, Q. (2019). miES: predicting the essentiality of miRNAs with machine learning and sequence features. *Bioinformatics* 35, 1053–1054. doi: 10.1093/bioinformatics/bty738
- Sun, L. L., Wu, H., Zhang, Y. Z., Wang, R., Wang, W. Y., Wang, W., et al. (2016). Design, synthesis and preliminary evaluation of the anti-inflammatory of the specific selective targeting druggable enzymome cyclooxygenase-2 (COX-2) small molecule. *Pharm. Biol.* 54, 2505–2514. doi: 10.3109/13880209.2016.1160939
- Szewczak, L., Hazuda, D., and Birnbaum, M. (2020). Looking to the future for pharma and the drug development ecosystem. *Cell* 181, 15–18. doi: 10.1016/j.cell.2020.03.016
- Sztuba-Solinska, J., Chavez-Calvillo, G., and Cline, S. E. (2019). Unveiling the druggable RNA targets and small molecule therapeutics. *Bioorg. Med. Chem.* 27, 2149–2165. doi: 10.1016/j.bmc.2019.03.057
- Tetko, I. V., and Tanchuk, V. Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J. Chem. Inf. Comput. Sci.* 42, 1136–1145. doi: 10.1021/ci025515j
- van de Waterbeemd, H. (1998). Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. J. Drug Target 15:490.
- van De Waterbeemd, H., Camenisch, G., Folkers, G., and Raevsky, O. A. (1996). Estimation of Caco-2 cell permeability using calculated molecular

descriptors. Quant. Struct. Activity Relat. 15, 480-490. doi: 10.1002/qsar. 19960150604

- Winiwarter, S., Bonham, N. M., Ax, F., Hallberg, A., Lennernas, H., and Karlen, A. (1998). Correlation of human jejunal permeability (*in vivo*) of drugs with experimentally and theoretically derived parameters. A multivariate data analysis approach. J. Med. Chem. 41, 4939–4949. doi: 10.1021/jm9810102
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018a). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–d82. doi: 10.1093/nar/gk x1037
- Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vazquez-Fresno, R., et al. (2018b). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 46, D608–D617. doi: 10.1093/nar/gkx1089
- Zeng, P., Chen, J., Meng, Y., Zhou, Y., Yang, J., and Cui, Q. (2018). Defining essentiality score of protein-coding genes and long noncoding RNAs. *Front. Genet.* 9:380. doi: 10.3389/fgene.2018. 00380

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.

Copyright © 2020 Huang, Zhou, Yang, Cui and Li. 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.