Applied Cell Biology

ISSN : 2320-1991 Online
ISSN : 2320-1983 Print
Volume 11 Issue 2
DOI: https://dx.doi.org/10.53043/2320-1991.acb11005

Commentary
Appl Cell Biol, 11(2), 2023 [70-73]

Molecular Vibration-Based Holistic Approach to Predict Drug-Target Interactions

Ziwei Zhao, Xuweisheng JI, Yun Wang*

Key Laboratory of TCM-Information Engineer of State Administration of TCM, Beijing University of Chinese Medicine, Beijing, 100102, China

***Corresponding author:** Yun Wang, Key Laboratory of TCM-Information Engineer of State Administration of TCM, Beijing University of Chinese Medicine, Beijing, 100102, China

INTRODUCTION

In drug discovery process, predicting drug–target interactions (DTI) is an preliminary step to narrow down the scope of screening candidate medications[1]. The most common-used computational methods are quantitative structure–activity relationship (QSAR) and molecular docking (MD)[2,3]. These methods are generally simple and efficient, but the performance is far from satisfaction. Since the used algorithm has constantly been developed, and deep learning has shown great powerful capacity in prediction tasks, we believe that the most impacted bottleneck currently is how to select the most appropriate features to represent drugs and targets.

Status quo of drug molecular representation

Identifying essential features for molecular representation is paramount to the accurate modeling and predicting of molecule properties and bioactivities, including DTIs. Currently, the molecular descriptors have diverse types and a massive amount of data. It can be broadly divided into four categories: string, chemical table, graph, and feature-based representations (Table 1).

It can be seen that nearly all molecular descriptors across the four types focus on meticulously picturing the structure of a molecule. They can most accurately convey the underlying structure of the molecule, but overlook whether they are essential for potential as a drug, which may cause the waste of computing power and the poor performance of the model. It is critical to ensure that model performance derives from learning medical relevant patterns instead of by exploiting confounding variables or other noise. Thus, it will be necessary to select the molecular feature prudently for the high-efficiency and continued advancement of machine learning in drug discovery.

For a deeper consideration, when drug molecule reacts with its target, they together make up an interaction system, as opposed to two separate individuals[4]. They are both the essential part of a corporate system, achieved by finding each other, then matching and binding. So it is necessary to consider the mutual impact as a united system of drugs and targets when represents them for drug discovery. In most existing models, however, extracting feature vectors from the drug and the target is processed separately. The consequence is that the features strongly contributing to the interacting system are hardly

Table 1: Four basic type of molecular representations

Molecular descriptor type		Further development
String	Generally consist of characters from the American Standard Code for Information Interchange (ASCII) character encoding standard. Such as SMILES, WLN, etc..	Combined with auto encoder (AAE, VAE), Word Embedding (word2vec), RNN, CNN, etc..
Chemical table	A chemical table (CT) listing the *x*-, *y*-, and *z*-, coordinates of each atom, and how they are bonded to each other in a molecule. Such as MDL molfile, CDXML, etc..	Combined with Network Pharmacology, Graph mining, etc..
Graph	Molecules represented as undirected graphs, with nodes as atoms and edges as bonds, etc..	Combined with GNN, CNN, GAN, etc..
Feature-based	A combination of physical or chemical molecular attributes, such as molecular weight, density, melting point, etc..	Combined with feature selection algorithm, etc..

Commentary

captured, causing unsatisfactory results of the DTI prediction. Some studies have paid attention to this problem and worked on improvement[5,6]. The attention mechanism is one of the commonly used methods to characterize the impact of targets on drugs and vice versa. But still, it could not combine the interacting features as a whole[7-10].

A holistic-based perspective to represent drugs

Given the limitation in DTI studies above, a holistic-based approach for drug-target system representation is needed. The selected characteristic attributes for drug/target representation should be able to show or at least highly related to features of the whole interacting system. Therefore, different types of drug-target systems shall be comparable and be unified into a model.

Specifically, the features highly correlated with the interaction between drugs and targets should be selected, and the feature selection should be processed based on the joint data. As a practical example, we recently published a prediction model for drug–target interactions (DTIs) affinity, with dissociation constant (Kd) and 50% of maximal effect (EC_{50}) as quantitative indicators[11]. **This model was based on molecular vibrations to represent drug overall characteristics by taking molecule-target as a whole system.**

According to the Superstring Theory, the essence of everything is vibration and its interconnectedness is drawing all things together through vibrations on specific frequencies[12]. The mutual impact between drugs and targets are also a type of resonance of vibration. And molecular vibrations are caused by vibrations of chemical bonds within molecules and are affected by various factors such as conjugation effect, induction effect, spatial effect, hydrogen bonding, vibrational coupling effect, etc.[13]. Thus, we believe that molecular vibrations can mostly reflect the overall properties of drug molecules in terms of their interaction with targets.

In view of this, we collected seven types of physicochemical properties particularly relevant to molecular vibrations, including electronegativity, π-atomic charge, total charge, and bond polarity. Then, in the process of feature screening, we further filtered the descriptors according to the importance scores to reflect their contribution to EC50 and KD. The importance score of single feature is equal to the result of oob_accuracy subtract oob_accuracy_after_perputation, where oob_acc_after_perputation refers to the accuracy of samples on the single tree count after shuffling the dimensional feature using the out-of-bag method. The top-ranking molecular descriptors were shown in Table 2.

Table 2: The top-ranking molecular descriptors in EC_{50} and KD datasets.

	Molecular descriptor	Descriptor specific meaning
EC_{50}	JGI5	Mean topological charge index of order 5
	minaaSe	Minimum atom-type E-State: aSea
	minHsSH	Minimum atom-type H E-State: -SH
	maxaaS	Maximum atom-type E-State: aSa
	maxssssSn	Maximum atom-type E-State: >Sn<
	nHdsCH	Count of atom-type H E-State: =CH-
	maxsNH2	Maximum atom-type E-State: -NH2
	maxssPH	Maximum atom-type E-State: -PH-
	ETA_Beta_s	A measure of electronegative atom count of the molecule
	maxddssSe	Maximum atom-type E-State: -=Se=-
KD	ETA_Beta_s	A measure of electronegative atom count of the molecule
	minaaSe	Minimum atom-type E-State: aSea
	ETA_BetaP_s	A measure of electronegative atom count of the molecule relative to molecular size
	mindssS	Minimum atom-type E-State: >S=
	maxHBint3	Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 3
	nddsN	Count of atom-type E-State: -N<<
	minsssNHp	Minimum atom-type E-State: >NH-+
	minssssNp	Minimum atom-type E-State: >N<+
	maxssssSn	Maximum atom-type E-State: >Sn<
	nHdsCH	Count of atom-type H E-State: =CH-

As shown in the table, we retained descriptors with importance scores greater than 0.85 in the feature screening process with maximum value of 1. A higher importance score represents the corresponding descriptor is more important for quantification of DTIs. It can be concluded that molecular descriptors with high importance were highly concentrated on E-state descriptors. E-state descriptors characterize both topological information of each atom and electronic relationships between atoms in the molecule[14]. The three molecular forces, dispersion, dipole moment and hydrogen bonding, which influence the strength of DTIs affinity, are closely related to the electronic relationships characterized by E-state descriptors[15-17]. This suggests that among hundreds of molecules vibrant characteristic, E-state descriptors are superior for analyzing and predicting DTIs affinity.

The prominent features of the model can be summarized as follow:

1. Conciseness: A defined molecular character can reduce computational cost and ensure that models learn important patterns over less noise. Vibration is able to mostly represent the essential drug properties of molecules and also distinguish the subtle differences of them.
2. Interpretability: For machine learning in drug discovery, molecular representation is better to be interpreted by human experts for further assessing the patterns learned by their models and providing a sanity check against domain knowledge. Our model is explicitly orientated and allows fine-tuning of the feature selecting based on the specific research needs.
3. Generalization: vibration-based molecular representation and our workflow can be widely used in different kinds of drugs, such as anti-cancer drugs, specific targeted drugs and so on.
4. Practicability: Compared with the complex three-dimensional structure, the vibration characteristics of the molecules are far more convenient and easier to detect, vibrational spectroscopy like Raman spectroscopy and Near-infrared spectroscopy has superior performance on portability, detection velocity and stability. Coupled with these instruments, in the future, the affinity of drug–target interactions can be evaluated real time in the scenario of drug discovery and development, drug manufacturing, etc.

New perspective for the perception of drug

Using the characteristic vibrations to represent the drugs' overall features is not only limited in the molecular level. We also practiced this hypothesis on the study of herbal medicine and made a striking progress. For side-by-side comparison, western medicine uses EC_{50}, KD, etc., to evaluate the effect of drugs on the human body, whereas traditional Chinese medicine (TCM) uses Four Qi (hot, warm, cold and cool) to describe the efficacy of herbs. If the chemical index provides precision, measurability, and assess ability, drug efficacy from TCM is more holistic and systematic, but are difficult to quantify. We adopted Raman spectra to detect the holistic characteristic vibrations of traditional Chinese medicines, and successfully built the identification model and measurement model of TCM medicinal properties[18-20].

These two works have shown that representing drugs from the vibration characteristics and using it to predict their efficacy is a breakthrough and promising idea. To go one step further, specific frequency of vibration has a long history of being used to treat disease. Chinese has used five different modes of music to cure diseases in different organs for over two millenniums[21]. While Gregorian Chants have been frequently studied since the mid-90s and been proved to benefit both the singer and listener to reduce anxiety, lower blood pressure, and to even help the fluency of speech for stroke victims[22]. Just as Nikola Tesla says, "If you want to find the secrets of the universe, think in terms of energy, frequency and vibration[23]." Many scholars have regarded everything in the universe as the result of vibration[24, 25]. The healing power of vibration can exist in different genres or through different carriers, including music, molecules, herbs, etc. In summary, the cognition of drugs and their effects on human body based on the perspective of vibration has a very expansive view; there lies more profound physical and philosophical significance than its practical value which is worth to be explored.

Prospect

In this article, we propose a molecular vibration-based holistic approach to predict DTIs, which also provides clues and guidance to elucidate the mechanism of drugs. Based on this thought, our tentative study on predicting and quantifying affinity of DTIs outperforms competitive baselines. Combined with the generative AI, it can not only generate the specific functional drug molecules, but can also create different drug-target interacting systems, which can improve the predictive property of DTI models as well as the drug discovery efficiency.

REFERENCES

1. Agamah, F.E., et al., Computational/in silico methods in drug target and lead prediction. Brief Bioinform, 2020. 21(5): p. 1663-1675.

Commentary

2. Cherkasov, A., et al., QSAR modeling: where have you been? Where are you going to? 2014. 57(12): p. 4977-5010.
3. Pinzi, L. and G.J.I.j.o.m.s. Rastelli, Molecular docking: shifting paradigms in drug discovery. 2019. 20(18): p. 4331.
4. Koshland, D.E.J.A.C., The Key–Lock Theory and the Induced Fit Theory. 1995. 33: p. 2375-2378.
5. Li, F., et al., Effective drug-target interaction prediction with mutual interaction neural network. Bioinformatics, 2022. 38(14): p. 3582-3589.
6. Zhao, T., et al., Identifying drug-target interactions based on graph convolutional network and deep neural network. Brief Bioinform, 2021. 22(2): p. 2141-2150.
7. Tsubaki, M., K. Tomii, and J. Sese, Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics, 2019. 35(2): p. 309-318.
8. Karimi, M., et al., DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics, 2019. 35(18): p. 3329-3338.
9. Chen, W., et al., Predicting Drug-Target Interactions with Deep-Embedding Learning of Graphs and Sequences. J Phys Chem A, 2021. 125(25): p. 5633-5642.
10. Chen, L., et al., TransformerCPI: improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. Bioinformatics, 2020. 36(16): p. 4406-4414.
11. Wang, X.R., et al., Quantitative prediction model for affinity of drug-target interactions based on molecular vibrations and overall system of ligand-receptor. BMC Bioinformatics, 2021. 22(1): p. 497.
12. Vyas, R. and M. Joshi, Superstring Theory: A Candidate of Theory of Everything, Quantum Gravity and A Door to Pre-Big Bang Era. 2017, Christ College Proceedings, ISBN No: 9788193347553, Page.
13. Okabayashi, N., et al., Vibrations of a molecule in an external force field. Proc Natl Acad Sci U S A, 2018. 115(18): p. 4571-4576.
14. Souza, E.S., et al., Estimating the octanol/water partition coefficient for aliphatic organic compounds using semi-empirical electrotopological index. Int J Mol Sci, 2011. 12(10): p. 7250-64.
15. Wang, C., et al., Predicting drug-target interactions with electrotopological state fingerprints and amphiphilic pseudo amino acid composition. 2020. 21(16): p. 5694.
16. Roy, K. and I. Mitra, Electrotopological state atom (E-state) index in drug design, QSAR, property prediction and toxicity assessment. Curr Comput Aided Drug Des, 2012. 8(2): p. 135-58.
17. Butina, D., Performance of Kier-Hall E-state descriptors in quantitative structure activity relationship (QSAR) studies of multifunctional molecules. Molecules, 2004. 9(12): p. 1004-9.
18. JI Xu-wei-sheng, et al., Identification of Neutral Chinese Medicine Based on Raman Spectroscopy. Mod Chin Med, 2022.12. 24(12): p. 2364-2370.
19. Sha HAN, et al., Identification of sweetness in traditonal Chinese medicine based on Raman spectroscopy Central South Pharmacy, 2022. 20(8): p. 1825-1828.
20. LIANG Hao, et al., Discrimination of Cold and Heat Medicinal Properties Based on Raman Spectroscopy of TCM Drugs. World Chinese Medicine, 2023. 18(1): p. 65-69.
21. Zhang, Y. and S. Gao, Chinese traditional five-tone music therapy. 2022. 52(6): p. 323-327.
22. Seidel, A.M., Sacred Sound: Tuning the Cosmic Strings of the Subtle Dancing Body. Congress on Research in Dance Conference Proceedings, 2012. 2012: p. 140-145.
23. Planck, M., Energy is the inherent effort of every multiplicity to become unity.
24. Hawking, S., The universe in a nutshell. 2001: Bantam.
25. Greene, B., The elegant universe: Superstrings, hidden dimensions, and the quest for the ultimate theory. 2000, American Association of Physics Teachers.

Citation: Zhao Z, Xuweisheng JI, Wang Y (2023) Molecular Vibration-Based Holistic Approach to Predict Drug-Target Interactions. Appl Cell Biol, 11(2), 2023 [70-73]