Front. Agr. Sci. Eng. 2022, 9(1): 150–154 https://doi.org/10.15302/J-FASE-2021419

PERSPECTIVE

HIGH-PERFORMANCE COMPUTATION AND ARTIFICIAL INTELLIGENCE IN PESTICIDE DISCOVERY: STATUS AND OUTLOOK

Li ZHANG (✉)[1], Jialin CUI[1], Qi HE[1], Qing X. LI (✉)[2]

1 Innovation Center of Pesticide Research, Department of Applied Chemistry, College of Science, China Agricultural University, Beijing 100193, China.

2 Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA.

Received August 1, 2021.

Correspondences: zhang_li@cau.edu.cn, qingl@hawaii.edu

1 INTRODUCTION

Pesticides are extremely important in global crop, food and ecological safety. Use of pesticides can prevent around 20% to 30% of global crop loss every year. The continuous growth of population and the limited acreage of agricultural land require increasing crop yields to feed people. Pesticides are important for protecting crops and preventing yield losses. With the increasing concerns over environmental protection, eco-friendly pesticides are needed to replace some higher risk ones. It is slow and difficult to successfully discover new pesticides with the standard random screening. New technologies and methods, such as high-performance computation and artificial intelligence (AI), can dramatically facilitate in the pesticide discovery process.

In recent years, high-performance computation and AI technologies have been rapidly developed for image recognition, protein structure simulation, drug design, organic synthesis, and other innovations. Although the investment in new technologies for pesticides is behind that of medicine, with the accumulation of data on biological targets, biological activity, toxicity and metabolism of agricultural small molecules, the research on pesticide discovery aided with high-performance computation and AI technology has begun. This article briefly summarizes recent advances of high-performance computation and AI in four areas along with our perspectives on the topic.

2 DESIGN OF PESTICIDE CANDIDATES

Ligand-based and target-based virtual screening has become common in molecular design of pesticide candidates. Yang's group has developed several online high-performance computational platforms that can assist pesticide molecular design, including the auto *in silico* ligand directing evolution server[1], pesticide and drug fragments database[2], the pharmacophore-linked fragment virtual screening (PFVS) method, the auto core fragment *in silico* screening[3] and auto *in silico* consensus inverse docking[4]. Applications of PFVS yielded the first picomolar-range Q(o) site inhibitors of the cytochrome bc(1) complex[5]. The first selective herbicide, quinotrione, has been developed and commercialized for the weed control in sorghum crops[6]. The novel succinate ubiquinone oxidoreductase inhibitors, compounds **62** (flubeneteram) and **43**, have excellent field efficacies against mulberry powdery mildew, rust, sheath blight and other diseases. They are now in the registration and industrial development stage[7].

3 MECHANISM OF ACTION AND MECHANISM OF RESISTANCE

Specific interactions between the target and ligand define the molecular recognition and potency of the ligand. Structure-

based pesticide design seeks to identify and optimize such interactions between ligands and their host target proteins[8]. The innovation of the pesticide lead candidates can be achieved by the biorational design based on the known mechanism of action between protein and ligand. The identification of the mechanism of action is, therefore, particularly important. Molecular dynamics simulations can help understanding dynamic interactions between ligand and protein. The protein-ligand interaction server (LARMD) can visually analyze the molecular dynamics results and generate a variety of outputs, including tables, heat maps, histograms, dynamic maps and 3D structure images[9]. It is noteworthy that the target of pesticides does not refer to the specific biological macromolecule of one specie. Yang and Xi proposed a pesticide targetome approach to design a potential pesticide, which considers all targets including those in crops, target pests, and non-target organisms to achieve the selectivity, efficacy and safety of pesticides[10]. Recently, Tunyasuvunakool et al. used AlphaFold2 to predict the 3D structures of almost whole human proteome[11]. AlphaFold DB showed the potential of AI to accelerate scientific progress[11], which suggests possibilities of pesticide design based on targetome. In addition, molecular and quantum mechanics methods are used for theoretical explanations of pesticide toxicity mechanisms. Rathnayake et al. used a molecular-quantum mechanics method to study the interactions between six organophosphorus insecticides and acetylcholinesterase, which is helpful to understand the mechanism of hydrolysis of these insecticides and toxicity mechanisms between the insecticides and the enzyme active site[12].

Pesticide resistance remains a huge challenge for pesticide development, but it offers an opportunity for improvement. The mechanisms of pesticide resistance are diverse. Target mutation is a main mechanism of pesticide resistance, causing pesticides to significantly lose efficacy or even become completely ineffective. Therefore, resistance risk is considered as an important criterion for the selection of pesticide candidate in the lead compound discovery stage. Those with low resistance risk should be selected for the next development step. The selection can be aided with reliable resistance risk prediction methods. Yang's group has developed the computational mutation scanning method to study the resistance mechanisms and structure-resistance relationship of the protoporphyrinogen oxidase inhibitors[13]. A web server (auto *in silico* macromolecular mutation scanning) suite AIMMS was built to predict mutation-induced drug resistance[14].

4 TOXICITY PREDICTION

Toxicological properties are important information required in pesticide registration and safety evaluation. Many pesticide candidates cannot be registered because of their high toxicity to mammals or non-target organisms. Therefore, it is extremely important to evaluate the toxicological profiles in an early stage of pesticide development. However, due to time and cost-effectiveness and other reasons, toxicity evaluation tests still cannot meet the needs of lead candidate discovery and optimization. *In silico* prediction of toxicological profiles has become a viable alternative.

Impacts of pesticides on bees have always been a focus of non-target toxicities. Some studies have adopted k-nearest neighbors (k-NN), random forest (RF), decision tree, support vector machine (SVM), artificial neural networks (ANN) graph attention convolutional neural networks (GACNN) and other methods to develop models for predicting bee toxicity[15,16]. Among them, Wang et al.[16] used the GACNN method to construct a bee toxicity classification model on a training set of 720 pesticides, having an accuracy of nearly 84%. The website (http://beetox.cn) is public accessible for predicting bee toxicity.

Aquatic toxicity is also an important part of environmental assessment of pesticides. At present, SVM, ANN, RF and partial least squares have been used to establish toxicity models for rainbow trout[17], *Lepomis*[17], *Raphidocelis subcapitata*[18], *Daphnia magna*[18–20], fish[20], *Pimephales promelas*[18] and other aquatic organisms. The data set size of the training model ranged from 46 to 1484. Among them, He et al.[19] used the SVM method and extended fingerprint to establish a ternary classification toxicity model containing 515 pesticides based on the EC_{50} toxicity data of *Daphnia magna*, and the overall accuracy (Q_{total}) of the test set reached 0.848.

Mammalian toxicity is another focus in pesticide discovery. The current AI and computational prediction mainly focus on the acute oral toxicity of rats. According to the existing LD_{50} data, researchers have established a series of acute oral toxicity models in rats[21–24], according to the methods such as RF, k-NN, SVM, naïve Bayesian, AdaBoosted decision trees, hierarchical clustering and deep learning. The number of compounds in the training set ranged from 44 to 8613, including a regression model and a classification model. There were also studies on organ toxicity (e.g., drug-induced liver injury[25], eye injury and eye corrosion [26]) models and genomic toxicity (e.g., Ames mutagenesis[27] and carcinogenesis[28]) models. Among them, Yang et al.[4]

provided a public-accessible website "admetSAR" for evaluating the ADMET properties of compounds (http://lmmd.ecust.edu.cn/admetsar2), which can be used to evaluate mammalian toxicity.

5 SYNTHETIC ROUTE DESIGN, CRYSTAL PACKING PREDICTION AND FORMULATION CHOICE

There have been many studies on AI in the design, crystal packing prediction and formulation optimization of pharmaceuticals. It can provide a reference to use AI technology in pesticide discovery. Hasic and Ishida[29] used the template-free approach to predict the single-step retrosynthesis, which achieves over 47% top-1 accuracy for the single-step retrosynthesis task on the US Patent Office data set.

AI also assists in formulation design[2,6,30–33] and crystal packing prediction[34]. Lin et al.[6] used machine learning methods to build a solid dispersant physical stability prediction model on a data set containing 646 solid dispersant stability data points. The accuracy of the RF method reached nearly 83%. They also developed a new formulation prediction platform of solid dispersion: PharmSD[30]. This platform provides efficient functionalities for the prediction of physical stability, dissolution type and dissolution rate of solid dispersion independently. Thirty-one crystal packings of aminosalicylic acid were predicted through ab initio methods by Meenashi et al.[34].

6 OUTLOOK

Large pharmaceutical companies have successfully incorporated AI and high-performance computation into drug discovery platforms. Artificial intelligence for drug research and development has entered a period of rapid growth[35]. The integration of high-performance computing platforms and AI technology is bound to become an indispensable means for the discovery and development of eco-friendly pesticides during the implementation of the Fourteenth Five-Year Plan in China. We need to diversify pesticide development and discovery strategies by (1) combining chemical biology technology with AI to develop potential and important pesticide targets and promote the discovery of eco-friendly pesticides; (2) constructing molecules of different pesticides target prediction network with deep learning, and designing multiple targets pesticide to reduce the risk of pest resistance; and (3) using AI to build a multilevel and multiscale prediction network of pesticides and different biological targets. Computation and AI are undoubtedly essential tools for pesticide research and development, particularly in selectivity, potency, pest control spectrum and environmental safety.

Acknowledgements

This work was funded in part by the National Natural Science Foundation of China (21977114) and the USDA (Hatch project HAW5032-R).

REFERENCES

1. Wu F, Zhuo L, Wang F, Huang W, Hao G, Yang G. Auto *in silico* ligand directing evolution (AILDE) to facilitate the rapid and efficient discovery of drug lead. *iScience*, 2020, **23**(6): 101179
2. Zhao Q, Miriyala N, Su Y, Chen W, Gao X, Shao L, Yan R, Li H, Yao X, Cao D, Wang Y, Ouyang D. Computer-aided formulation design for a highly soluble lutein-cyclodextrin multiple-component delivery system. *Molecular Pharmaceutics*, 2018, **15**(4): 1664–1673
3. Hao G F, Jiang W, Ye Y N, Wu F X, Zhu X L, Guo F B, Yang G F. ACFIS: a web server for fragment-based drug discovery. *Nucleic Acids Research*, 2016, **44**(W1): W550–W556
4. Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, Li W, Liu G, Tang Y. admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics*, 2019, **35**(6): 1067–1069
5. Hao G F, Wang F, Li H, Zhu X L, Yang W C, Huang L S, Wu J W, Berry E A, Yang G F. Computational discovery of picomolar Q(o) site inhibitors of cytochrome bc1 complex. *Journal of the American Chemical Society*, 2012, **134**(27): 11168–11176
6. Lin H, Chen X, Chen J, Wang D, Wu F, Lin S, Zhan C, Wu J, Yang W, Yang G. Crystal Structure of 4-hydroxyphenylpyruvate dioxygenase in complex with substrate reveals a new starting point for herbicide discovery. *Research*, 2019: 2602414
7. Xiong L, Li H, Jiang L N, Ge J M, Yang W C, Zhu X L, Yang G

F. Structure-based discovery of potential fungicides as succinate ubiquinone oxidoreductase inhibitors. *Journal of Agricultural and Food Chemistry*, 2017, **65**(5): 1021–1029

8. Liang Z, Li Q X. π-Cation interactions in molecular recognition: perspectives on pharmaceuticals and pesticides. *Journal of Agricultural and Food Chemistry*, 2018, **66**(13): 3315–3323
9. Yang J F, Wang F, Chen Y Z, Hao G F, Yang G F. LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. *Briefings in Bioinformatics*, 2020, **21**(6): 2206–2218
10. Yang G. Chemical biology-oriented molecular design of green pesticide. *Bulletin of National Natural Science Foundation of China*, 2020, **34**(4): 495–501 (in Chinese)
11. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt G J, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl S A A, Potapenko A, Ballard A J, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior A W, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. *Nature*, 2021. doi: 10.1038/s41586-021-03828-1
12. Rathnayake L K, Northrup S H. Structure and mode of action of organophosphate pesticides: a computational study. *Computational & Theoretical Chemistry*, 2016, **1088**: 9–23
13. Hao G F, Tan Y, Xu W F, Cao R J, Xi Z, Yang G F. Understanding resistance mechanism of protoporphyrinogen oxidase-inhibiting herbicides: insights from computational mutation scanning and site-directed mutagenesis. *Journal of Agricultural and Food Chemistry*, 2014, **62**(29): 7209–7215
14. Wu F X, Wang F, Yang J F, Jiang W, Wang M Y, Jia C Y, Hao G F, Yang G F. AIMMS suite: a web server dedicated for prediction of drug resistance on protein mutation. *Briefings in Bioinformatics*, 2018, **21**(1): 318–328
15. Li X, Zhang Y, Chen H, Li H, Zhao Y. Insights into the molecular basis of the acute contact toxicity of diverse organic chemicals in the honey bee. *Journal of Chemical Information and Modeling*, 2017, **57**(12): 2948–2957
16. Wang F, Yang J F, Wang M Y, Jia C Y, Shi X X, Hao G F, Yang G F. Graph attention convolutional neural network model for chemical poisoning of honey bees prediction. *Science Bulletin*, 2020, **65**(14): 1184–1191
17. Li F, Fan D, Wang H, Yang H, Li W, Tang Y, Liu G. *In silico* prediction of pesticide aquatic toxicity with chemical category approaches. *Toxicology Research*, 2017, **6**(6): 831–842
18. Khan K, Benfenati E, Roy K. Consensus QSAR modeling of toxicity of pharmaceuticals to different aquatic organisms: ranking and prioritization of the DrugBank database compounds. *Ecotoxicology and Environmental Safety*, 2019, **168**: 287–297
19. He L, Xiao K, Zhou C, Li G, Yang H, Li Z, Cheng J. Insights into pesticide toxicity against aquatic organism: QSTR models on Daphnia Magna. *Ecotoxicology and Environmental Safety*, 2019, **173**: 285–292
20. Lunghini F, Marcou G, Azam P, Enrici M H, van Miert E, Varnek A. Consensus QSAR models estimating acute toxicity to aquatic organisms from different trophic levels: algae, *Daphnia* and fish. *SAR and QSAR in Environmental Research*, 2020, **31**(9): 655–675
21. Li X, Chen L, Cheng F, Wu Z, Bian H, Xu C, Li W, Liu G, Shen X, Tang Y. *In silico* prediction of chemical acute oral toxicity using multi-classification methods. *Journal of Chemical Information and Modeling*, 2014, **54**(4): 1061–1069
22. Sun G, Zhang Y, Pei L, Lou Y, Mu Y, Yun J, Li F, Wang Y, Hao Z, Xi S, Li C, Chen C, Zhao L, Zhang N, Zhong R, Peng Y. Chemometric QSAR modeling of acute oral toxicity of Polycyclic Aromatic Hydrocarbons (PAHs) to rat using simple 2D descriptors and interspecies toxicity modeling with mouse. *Ecotoxicology and Environmental Safety*, 2021, **222**: 112525
23. Minerali E, Foil D H, Zorn K M, Ekins S. Evaluation of assay central machine learning models for rat acute oral toxicity prediction. *ACS Sustainable Chemistry & Engineering*, 2020, **8**(42): 16020–16027
24. Zhu H, Martin T M, Ye L, Sedykh A, Young D M, Tropsha A. Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. *Chemical Research in Toxicology*, 2009, **22**(12): 1913–1921
25. Mulliner D, Schmidt F, Stolte M, Spirkl H P, Czich A, Amberg A. Computational models for human and animal hepatotoxicity with a global application scope. *Chemical Research in Toxicology*, 2016, **29**(5): 757–767
26. Wang Q, Li X, Yang H, Cai Y, Wang Y, Wang Z, Li W, Tang Y, Liu G. *In silico* prediction of serious eye irritation or corrosion potential of chemicals. *RSC Advances*, 2017, **7**(11): 6697–6703
27. Xu C, Cheng F, Chen L, Du Z, Li W, Liu G, Lee P W, Tang Y. *In silico* prediction of chemical Ames mutagenicity. *Journal of Chemical Information and Modeling*, 2012, **52**(11): 2840–2847
28. Li X, Du Z, Wang J, Wu Z, Li W, Liu G, Shen X, Tang Y. *In silico* estimation of chemical carcinogenicity with binary and ternary classification methods. *Molecular Informatics*, 2015, **34**(4): 228–235
29. Hasic H, Ishida T. Single-step retrosynthesis prediction based on the identification of potential disconnection sites using molecular substructure fingerprints. *Journal of Chemical Information and Modeling*, 2021, **61**(2): 641–652
30. Dong J, Gao H, Ouyang D. PharmSD: a novel AI-based computational platform for solid dispersion formulation design. *International Journal of Pharmaceutics*, 2021, **604**: 120705
31. Birru W A, Warren D B, Han S, Benameur H, Porter C J H, Pouton C W, Chalmers D K. Computational models of the gastrointestinal environment. 2. Phase behavior and drug solubilization capacity of a type i lipid-based drug formulation after digestion. *Molecular Pharmaceutics*, 2017, **14**(3): 580–592
32. Metwally A A, Hathout R M. Computer-assisted drug formulation design: novel approach in drug delivery. *Molecular Pharmaceutics*, 2015, **12**(8): 2800–2810

33. Sou T, Soukarieh F, Williams P, Stocks M J, Cámara M, Bergström C A S. Model-informed drug discovery and development in pulmonary delivery: biopharmaceutical pharmacometric modeling for formulation evaluation of pulmonary suspensions. *ACS Omega*, 2020, **5**(40): 25733–25746

34. Meenashi R, Selvaraju K, Stephen A D, Jelsch C. Theoretical crystal structure prediction of aminosalicylic acid: charge density topological and electrostatic analyses. *Journal of Molecular Structure*, 2020, **1213**: 128139

35. Cyranoski D. AI drug discovery booms in China. *Nature Biotechnology*, 2021, **39**(8): 900–902