REVIEW

Progress in molecular docking

Jiyu Fan[1], Ailing Fu[2], Le Zhang[1,3,4,*]

[1] School of Computer Science, Sichuan University, Sichuan 610065, China
[2] College of Pharmaceutical Sciences, Southwest University, Chongqing 400715, China
[3] Medical Big Data Center of Sichuan University, Chengdu 610065, China
[4] Chongqqing Zhongdi Medical Information Technology Co., Ltd., Chongqing 401320, China

Background: In recent years, since the molecular docking technique can greatly improve the efficiency and reduce the research cost, it has become a key tool in computer-assisted drug design to predict the binding affinity and analyze the interactive mode.

Results: This study introduces the key principles, procedures and the widely-used applications for molecular docking. Also, it compares the commonly used docking applications and recommends which research areas are suitable for them. Lastly, it briefly reviews the latest progress in molecular docking such as the integrated method and deep learning.

Conclusion: Limited to the incomplete molecular structure and the shortcomings of the scoring function, current docking applications are not accurate enough to predict the binding affinity. However, we could improve the current molecular docking technique by integrating the big biological data into scoring function.

Keywords: molecular docking; numerical analysis; optimization; data mining

Author summary: Currently, molecular docking has become a key tool in computer-assisted drug design. Therefore, this review introduces the basic theories of molecular docking and compares the commonly used docking software. And then, we list the inspiring applications and latest progress in molecular docking. Finally we discuss the drawbacks of existing molecular docking techniques and the future research direction.

INTRODUCTION

Molecular docking [1] is such a structure-based drug design method that simulates the molecular interaction and predicts the binding mode and affinity between receptors and ligands. In recent years, this technology has been widely used in drug design research field. Using the compounds database to screen the potential pharmacophores is not only convenient for researchers to purchase, synthesize and complete follow-up pharmacological tests, but also greatly improves the efficiency and reduces the research cost. In addition, the emergence of the reverse molecular docking technology [2] could significantly improve the drug target predictive capacity and understand the related molecular mechanism for drug design. Finally, this review briefly introduces the latest progress

and applications of molecular docking technology.

THE INTRODUCTION OF MOLECULAR DOCKING TECHNOLOGY

The basic theory of molecular docking

Molecular docking is to simulate the optimal conformation according to the complementarity and pre-organization, which could predict and obtain the binding affinity and interactive mode between ligand and receptor [1].

Figure 1A shows the first proposed "lock-and-key model" [3], which refers to the rigid docking of receptors and ligands to find the correct orientation for the "key" to open up the "lock". This model emphasizes the importance of geometric complementarity.
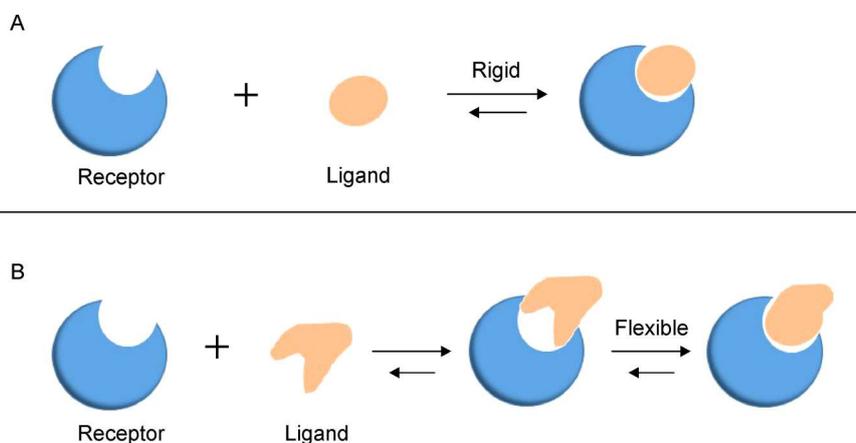
Figure 1. **Two models of molecular docking.** (A) A lock-and-key model. (B) Induced fit model.

However, the real docking process is so flexible that receptors and ligands have to change their conformation to fit each other well. Thus, we develop "induced fit model" (Figure 1B) [4]. Based on geometric complementarity, the energy complementarity and pre-organization guarantee that receptors and ligands would obtain the most stable structure in such a manner that minimizes the free energy [5].

As shown in Figure 2, the molecular docking software can help us to find the optimal conformation and orientation according to complementarity and pre-organization with specific algorithm, followed by applying a scoring function to predict the binding affinity and analyze the interactive mode. Figure 3 shows the protein-DNA docking with Autodock Vina [6] displayed in PyMOL [7].
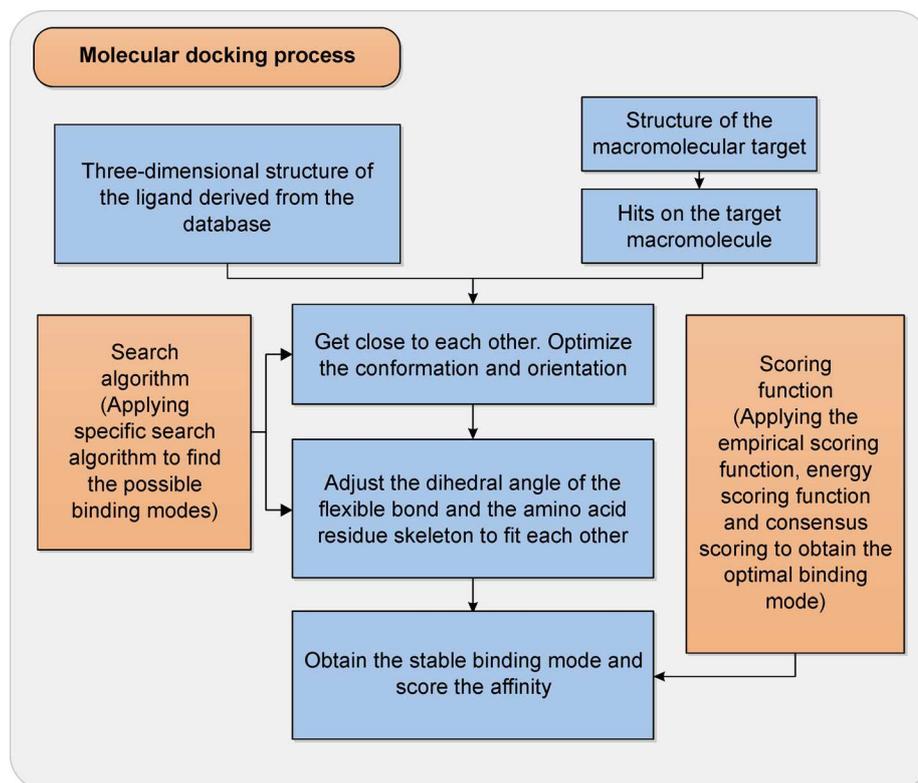

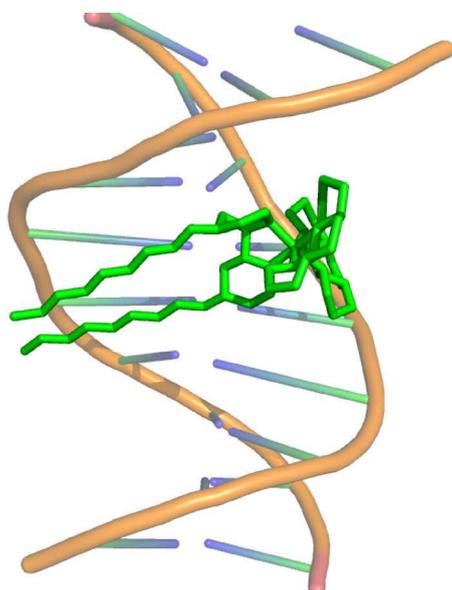**Figure 2.** **Molecular docking process.**

Figure 3. The protein-DNA docking.

Molecular docking software

Figure 4 lists the main three types of software for molecular docking. Flexible-rigid docking has been widely used. However, since flexible docking is usually more accurate, the relevant researches have become the hot studying spot in recent years. Table 1 lists the widely-used molecular docking software and its algorithms, evaluation methods, features and application areas.

Molecular docking databases

The most popular protein structure database is the public database Protein Data Bank (PDB) [8]. Also, the public databases such as PubChem Compound Database [9] and ZINC [10] are free to use. Besides, there are many important commercial databases, such as Compound Database (AcD) [11], Cambridge Structural Database (CSD) [12].

THE APPLICATIONS OF MOLECULAR DOCKING

Virtual screening to discover the lead compound and hit compound

Virtual screening [13] is to find the lead compound and hit compound from the molecular databases according to the scoring function, which has tremendously improved the screening efficiency compared with the traditional screen method (Figure 5).

The applications of virtual screening are commonly used. Notably, given the exponential growth of high-throughput [14], high-performance computing [15], machine learning [16] and deep learning [17] techniques, the integrated method flourishes quickly. For example, Pereira *et al.* [18] applied deep learning approach in virtual screening, which extracts relevant features from molecular docking data to create the distributed vector representations for protein-ligand complexes. Also, Pyzerknapp *et al.* [19] proposed the virtual high-throughput screening.

Prediction of potential targets

It is noted that the aforementioned methods are all general docking methods which use the different ligands in the database to dock with the same receptor. However, current commonly used reverse docking technique is different from them. Here, we employed Figure 6 to describe the reverse docking technique [2]. The reverse
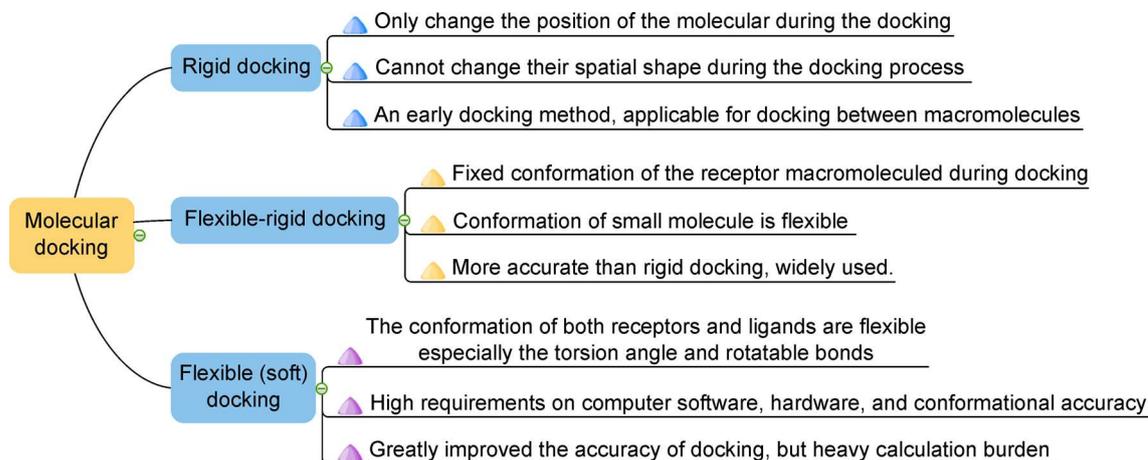

Figure 4. Molecular docking software classification.

Table 1   Representative software for molecular docking

Name	Search algorithm	Evaluation method	Speed	Features & Application areas
Flex X [33]	Fragmentation algorithm	Semi-empirical calculation on free energy	Fast	Flexible-rigid docking. It can be used for virtual screening of small molecule databases by using incremental construction strategy
Gold [34]	GA (genetic algorithm)	Semi-empirical calculation on free energy	Fast	Flexible docking. It is a GA-based docking program. The accuracy and reliability of this software have been highly evaluated
Glide [35]	Exhaustive systematic search	Semi-empirical calculation on free energy	Medium	Flexible docking. This software uses domain knowledge to narrow the searching range and has XP(extra precision), SP (standard precision) and high throughout virtual screen modes
AutoDock [36]	GA (genetic algorithm) LGA (lamarckian genetic algorithm)	Semi-empirical calculation on free energy	Medium	Flexible-rigid docking. This software is always used with Autodock-tools and it is free for academic use
ZDOCK [37]	Geometric complement-arity and molecular dynamics	Molecular force field	Medium	Rigid docking. Chen et al. [37] propose a new scoring function which combines pairwise shape complementarity(PSC) with desolvation and electrostatic and develop the ZDOCK server [38]
RDOCK [39]	GA(genetic algorithm) MC (monte carlo) MIN (Simplex minimization)	Molecular force field	Medium	Rigid docking. The CHARMm-based procedure for refinement and scoring. Besides predicting the binding mode, it is especially designed for high throughput virtual screening (HTVS) campaigns
LeDOCK [40]	Simulated annealing (SA) Genetic algorithm (GA)	Molecular force field	Fast	Flexible docking. LeDock is a new molecular docking program. From the results of the present study [41], since it is fast and exhibits a high accuracy, it is recommended for the virtual screen task
Dock [42]	Fragmentation algorithm	Molecular force field	Fast	Flexible docking. It is widely applicable and is always used in docking between flexible proteins and flexible ligands
Autodock Vina [6]	GA (genetic algorithm)	Semi-empirical calculation on free energy	Fast	Flexible-rigid docking. AutoDock Vina employs an iterated local search global optimizer and it is faster than the AutoDock 4

docking technique identifies the novel targets by assigning a single small-molecule ligand as the probe to dock with multiple receptors to discover potential binding cavities. In this way, the potential targets of drug can be predicted. For example, Grinter et al. [20] explored the potential target oxidized squalene cyclase (OSC) of PRIMA-1 by using the reverse docking software package Mdock. Also, Chen et al. [21] applied reverse docking technique to discover targeted proteins of marine compounds with anti-tumor activity. Furthermore, Chen et al. [21] also indicated that reverse docking can be complementary with in vitro assays as an effective method of target fishing. Finally, we considered that exploring relevant mechanism of action or side effect profile by structural biology analysis [22] such as the pocket analysis [23], could significantly benefit the novel drug design.

CONCLUSION

Considering the approximation capacity of the scoring function and incomplete collection of conformations, the
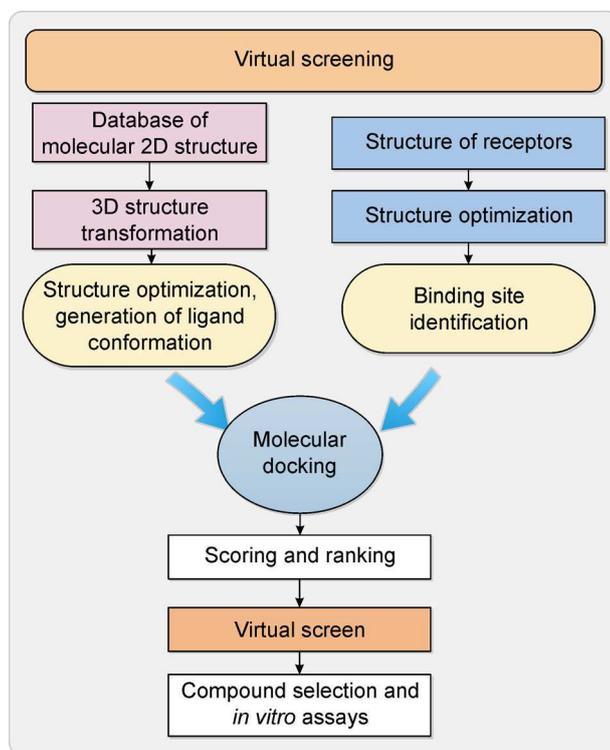
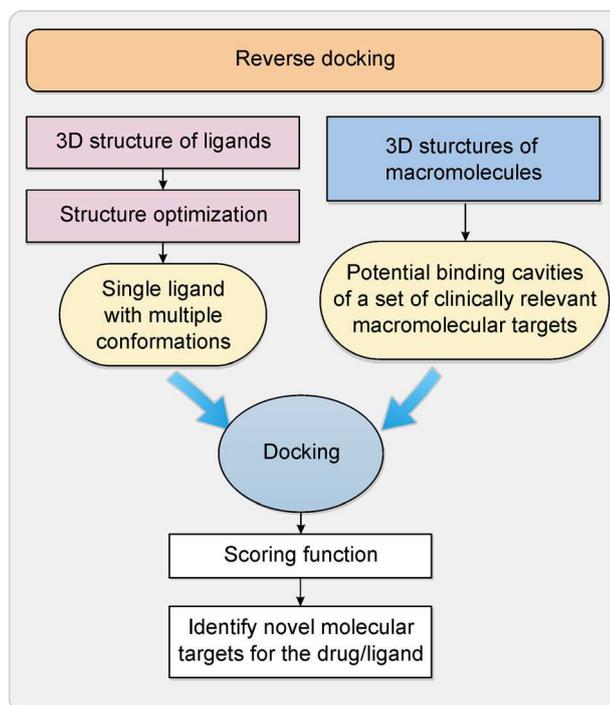**Figure 5.   The process of virtual screen.**

Figure 6.   The reverse docking technique.

molecular docking score of inactive molecules will be improperly so high that implicates false positive [24–26]. Furthermore, if the actual compound and the compound in database are significantly different in physical properties, the molecular docking score will be abnormal [27]. Therefore, it is necessary to take thermodynamic features into account [28], or use retrospective verification to evaluate the reliability of the prediction of affinity [29]. In addition, as the three-dimensional structure used for molecular docking will be away from its original environment resulting in a change in conformation, the docking result cannot truly reflect the state of the experimental docking. In the distant future, we are optimizing the conformational search algorithm by taking more flexible bonds, solvent states and integrating recent biological data mining algorithms [30–32] in consideration. In general, we believe that molecular docking technique will become such a reliable drug-design tool that integrate the big biological data by optimizing the scoring function and upgrading the relevant search algorithms.

ACKNOWLEDGEMENTS

COMPLIANCE WITH ETHICS GUIDELINES

The authors Jiyu Fan, Ailing Fu and Le Zhang declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

1. Morris, G. M. and Lim-Wilby, M. (2008) Molecular docking. Methods Mol. Biol., 443, 365–382

2. Chen, Y. Z. and Zhi, D. G. (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Proteins, 43, 217–226

3. Morrison, J. L., Breitling, R., Higham, D. J. and Gilbert, D. R. (2006) A lock-and-key model for protein-protein interactions. Bioinformatics, 22, 2012–2019

4. Koshland Jr, D. E. (2010) The key–lock theory and the induced fit theory. Angew. Chem. Int. Ed., 33, 2375–2378

5. Audie, J. and Scarlata, S. (2007) A novel empirical free energy function that explains and predicts protein-protein binding affinities. Biophys. Chem., 129, 198–211

6. Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 31, 455–461

7. DeLano, W. L. (2002) Pymol: an open-source molecular graphics tool. Ccp4 Newslett. Protein Crystallogr., 40,11

8. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. and Bourne, P. (2000) The Protein Data Bank, 1999 – . In: International Tables for Crystallography, Eds. Rossmann, M. G. and Arnold, E. 67, 675–684

9. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., et al. (2016) PubChem substance and compound databases. Nucleic Acids Res., 44, D1202–D1213

10. Irwin, J. J. and Shoichet, B. K. (2005) ZINC–a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model., 45, 177–182

11. Martin, G. E., Hadden, C. E., Russell, D. J., Kaluzny, B. D., Guido, J. E., Duholke, W. K., Stiemsma, B. A., Thamann, T. J., Crouch, R. C., Blinov, K., et al. (2010) Identification of degradants of a complex alkaloid using NMR cryoprobe technology and ACD/ structure elucidator. J. Heterocycl. Chem., 39, 1241–1250 .

12. Groom, C. R., Bruno, I. J., Lightfoot, M. P. and Ward, S. C. (2016) The Cambridge Structural Database. Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mater., 72, 171–179

13. Kitchen, D. B., Decornez, H., Furr, J. R. and Bajorath, J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov., 3, 935–949

14. Joseph-McCarthy, D., Baber, J. C., Feyfant, E., Thompson, D. C. and Humblet, C. (2007) Lead optimization via high-throughput molecular docking. Curr. Opin. Drug Discov. Devel., 10, 264–274

15. Ge, H., Wang, Y., Li, C., Chen, N., Xie, Y., Xu, M., He, Y., Gu, X., Wu, R., Gu, Q., et al. (2013) Molecular dynamics-based virtual screening: accelerating the drug discovery process by high-performance computing. J. Chem. Inf. Model., 53, 2757–2764

16. Melville, J. L., Burke, E. K. and Hirst, J. D. (2009) Machine learning in virtual screening. Comb. Chem. High Throughput Screen., 12, 332–343

17. Gawehn, E., Hiss, J. A. and Schneider, G. (2016) Deep learning in drug discovery. Mol. Inform., 35, 3–14

18. Pereira, J. C., Caffarena, E. R. and Dos Santos, C. N. (2016) Boosting docking-based virtual screening with deep learning. J. Chem. Inf. Model., 56, 2495–2506

19. Pyzerknapp, E. O., Suh, C., Gómezbombarelli, R., Aguileraiparraguirre, J. and Aspuruguzik, A. (2015) What is high-throughput virtual screening? A perspective from organic materials discovery. Annu. Rev. Mater. Res., 45, 45:195–216

20. Grinter, S. Z., Liang, Y., Huang, S. Y., Hyder, S. M. and Zou, X. (2011) An inverse docking approach for identifying new potential anti-cancer targets. J. Mol. Graph. Model., 29, 795–799

21. Chen, F., Wang, Z., Wang, C., Xu, Q., Liang, J., Xu, X., Yang, J., Wang, C., Jiang, T. and Yu, R. (2017) Application of reverse docking for target prediction of marine compounds with anti-tumor activity. J. Mol. Graph. Model., 77, 372–377

22. Xie, H., Lee, M. H., Zhu, F., Reddy, K., Huang, Z., Kim, D. J., Li, Y., Peng, C., Lim, D. Y., Kang, S., et al. (2013) Discovery of the novel mTOR inhibitor and its antitumor activities in vitro and in vivo. Mol. Cancer Ther., 12, 950–958

23. Liu, Z. P., Liu, S., Chen, R., Huang, X. and Wu, L. Y. (2017)

Structure alignment-based classification of RNA-binding pockets reveals regional RNA recognition motifs on protein surfaces. BMC Bioinformatics, 18, 27

24. Ferreira, R. S., Simeonov, A., Jadhav, A., Eidam, O., Mott, B. T., Keiser, M. J., McKerrow, J. H., Maloney, D. J., Irwin, J. J. and Shoichet, B. K. (2010) Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. J. Med. Chem., 53, 4891–4905

25. Zhang, L., Qiao, M., Gao, H., Hu, B., Tan, H., Zhou, X. and Li, C. M. (2016) Investigation of mechanism of bone regeneration in a porous biodegradable calcium phosphate (CaP) scaffold by a combination of a multi-scale agent-based model and experimental optimization/validation. Nanoscale, 8, 14877–14887

26. Zhang, L.,Zheng, C., Li, T., Xing, L., Zeng, H., Li, T., Yang, H., Cao, J., Chen, B. and Zhou, Z. (2017) Building up a robust risk mathematical platform to predict colorectal cancer. Complexity, 8917258

27. Enyedy, I. J. and Egan, W. J. (2008) Can we use docking and scoring for hit-to-lead optimization? J. Comput. Aided Mol. Des., 22, 161–168

28. Tame, J. R. (1999) Scoring functions: a view from the bench. J. Comput. Aided Mol. Des., 13, 99–108

29. Yabuuchi, H., Niijima, S., Takematsu, H., Ida, T., Hirokawa, T., Hara, T., Ogawa, T., Minowa, Y., Tsujimoto, G. and Okuno, Y. (2011) Analysis of multiple compound-protein interactions reveals novel bioactive molecules. Mol. Syst. Biol., 7, 472

30. Zhang, L., Liu, Y., Wang, M., Wu, Z., Li, N., Zhang, J. and Yang, C. (2017) EZH2-, CHD4-, and IDH-linked epigenetic perturbation and its association with survival in glioma patients. J. Mol. Cell Biol., 9, 477–488

31. Zhang, L., Xiao, M., Zhou, J. and Yu, J. (2018) Lineage-associated underrepresented permutations (LAUPs) of mammalian genomic sequences based on a Jellyfish-based LAUPs analysis application (JBLA). Bioinformatics, 34, 3624–3630

32. Zhang, L. and Zhang, S. (2017) Using game theory to investigate the epigenetic control mechanisms of embryo development: Comment on: "Epigenetic game theory: How to compute the epigenetic control of maternal-to-zygotic transition" by Qian Wang et al. Phys. Life Rev., 20, 140–142

33. Kramer, B., Rarey, M. and Lengauer, T. (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins, 37, 228–241

34. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD. Proteins, 52, 609–623

35. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T. and Banks, J. L. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J. Med. Chem., 47, 1750–1759

36. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem., 30, 2785–2791

37. Chen, R., Li, L. and Weng, Z. (2003) ZDOCK: an initial–stage protein–docking algorithm. Proteins, 52, 80–87 .

38. Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T. and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics, 30, 1771–1773

39. Li, L., Chen, R. and Weng, Z. (2010) RDOCK: refinement of rigid-body protein docking predictions. Proteins, 53, 693–707

40. Zhao, H. and Caflisch, A. (2013) Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. Bioorg. Med. Chem. Lett., 23, 5721–5726

41. Wang, Z., Sun, H., Yao, X., Li, D., Xu, L., Li, Y., Tian, S. and Hou, T. (2016) Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Phys. Chem. Chem. Phys., 18, 12964–12975

42. Therese, P.L., Brozell, S. R., Sudipto, M., Pettersen, E. F., Meng, E. C., Veena, T., Rizzo, R. C., Case, D. A., James, T. L., Kuntz, I. D. (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. Rna, 15, 1219–1230