Journal of
Biomolecular Research & Therapeutics

Mishra et al., J Biomol Res Ther 2018, 7:2
DOI: 10.4172/2167-7956.1000166

Opinion Open Access

Computational Issues of Protein-Ligand Docking

Avinash Mishra[1,2]* MA Hakim Newton[1] and Abdul Sattar[1]

[1]*Institute for Integrated and Intelligent Systems, Griffiths University, Australia*
[2]*Novo Informatics Pvt. Ltd, New Delhi, India*

Protein-ligand docking [1,2] is a crucial step in *in-silico* Structure Based Drug design (SBDD). It is mainly based on the binding affinity of a small molecule with a specific protein. Molecular level docking has obtained significant progress in recent times through the development of sophisticated algorithms and advancement of computing resources. As a result, pharmaceutical industries have also started implementing computational molecular docking at a large scale with the hope of curtailing down the time and the capital investment in the drug discovery research [3,4]. Docking algorithms mainly comprise two major components:

Generating protein-ligand samples and scoring the generated samples. Both components require heavy real number mathematical calculations; which essentially raises the overall complexity level of the computational approaches. While developing a docking algorithm, sample generation can be implemented separately or in conjugation with sample scoring [5,6]. Separate sampling of only ligand is a promising way to generate mul-tiple ligand conformations through the use of translational and rotational operators [6]. Sampling proteins along with comes under a flexible docking scheme, where sets of pocket lining residues are allowed to undergo conformational changes. Although protein sampling improves accuracy, unfortunately it multiplies the computational complexity by many folds. Overall, most sampling methods such as shape matching [7,8] systematic sampling [9] and stochastic sampling [10] are combinatorial in nature and requires significant execution time. Sample generation thus remains a great challenge as to which samples are to be generated from the enormous sample space. Sample scoring functions have their own high time complexities because of the mathematical rigorousness. Moreover, many known scoring functions that are based on force field [11,12] empiricalness [13,14] entropy [15,16] or knowledge-learnt [17,18] involve all or multi atoms. The purpose of a scoring function is to allow comparison of various conformations and to select the more promising ones. A good scoring function can thus effect how further samples are generated. A pertinent issue is the granularity level of scoring, given a higher granularity level needs resources extensively. In this article, we argue for a focussed and comprehensive approach to look at the computational aspects of the docking holistically. We suggest to develop a docking method that uses knowledge-based sampling for focussed search, simplified scoring functions for accelerated calculations and constraint based approaches for sample generation and scoring interaction. Further details of these aspects are provided below. Knowledge-based approaches can help learn from protein-ligand experimental complexes and exploit the knowledge in sample generation. For example, rotational bond sampling could be streamed down to restricted bonds that involve major atom types such as hydrogen bond donors and acceptors. This can lead to simplified model of sampling as has been explored in protein folding by Ken Dill [19]. Moreover, highly probable angles learnt from restricted bonds and major atom types can be used in finding potential orientations of the ligand molecules. This would result in both guided search and in limiting the sample space. Simplified scoring functions help accelerate computations. Scoring functions are primarily distance based measures with additional information either from chemistry principle (e.g. force, field, entropy) or even from knowledge based repositories. These additional information induces a higher mathematical complexity to the algorithm. For example, force field methods can have multiple terms including van der Waal forces, electrostatics, solvation, hydrogen bonds and many more. One way to reduce the complexity of these functions is to eliminate some of their detailing parameters and make them take some elementary forms. This can also help unify techniques that are needed for both sampling and scoring. Constraint based search approaches benefit from not sampling blindly without considering scoring functions that essentially capture problem specific knowledge through constraints and objectives. While traditional search approaches often take a random or exhaustive sampling approach, their search guidance mainly comes from scoring of the samples. In contrast, constraint-based approaches analyse current samples and identify parts that cause poorness of the score. The sample generation approaches then generate further samples focussing on the identified parts. Thus the sample generation and scoring processes work in an interleaving fashion and affect each others performance. In this kind of process, new samples are often close to old ones in terms of the changes in variable values. A sophisticated change propagation engine such as Kangaroo [20] in those cases can make the sample scoring efficient as it would recompute scores incrementally taking only the changes into account. Also, when multiple scoring functions are used, in an interleaving or in an on-demand fashion, computations of unused scoring functions could be avoided taking a lazy approach.

References

1. Anderson AC (2003) The process of structure-based drug design. Chem Biol 10: 787-797.
2. Kuntz ID (1992) Structure-based strategies for drug design and discovery. Science 257: 1078-1082.
3. Shoichet BK, McGovern SL, Wei B, Irwin JJ (2002) Lead discovery using molecular docking. Curr Opin Chem Biol 6: 439-446.
4. Price M (2012) Computational Biologists: The Next Pharma Scientists. Science Careers.
5. Neves MC, Totrov M, Abagyan R (2012) Docking scoring with ICM: the benchmarking results and strategies for improvement. J Comput Aided Mol Des 26: 675-686.

***Corresponding author:** Avinash Mishra, Institute and Intelligent Systems, Griffiths University, Australia, E-mail: avish2k@gmail.com

Received October 01, 2018; **Accepted** October 15, 2018; **Published** October 20, 2018

Citation: Mishra A, MA Hakim Newton, Sattar A (2018) Computational Issues of Protein-Ligand Docking. J Biomol Res Ther 7: 166. doi:10.4172/2167-7956.1000166

6. McGann M (2011) Fred pose prediction and virtual screening accuracy. Journal of chemical information and modeling 51: 578-596.

7. Mcgann MR, Almond HR, Nicholls A, Grant JA, Brown FK, et al. (2003) Gaussian docking functions. Biopolymers: Original Research on Biomolecules 68: 76-90.

8. Jain AN (2003) Surflex:fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 46: 499-511.

9. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, 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.

10. Thomsen R, Christensen MH (2006) Moldock: a new technique for high-accuracy molecular docking. J Med Chem 49: 3315-3321.

11. Yin S, Biedermannova L, Vondrasek J, Dokholyan NV (2008) Medusascore: an accurate force field-based scoring function for virtual drug screening. J Chemical Inf Model 48: 1656-1662.

12. Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP (2006) Physics-based scoring of protein-ligand complexes: Enrichment of known inhibitors in large-scale virtual screening. J Chem Inf Model 46: 243-253.

13. Bohm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8: 243-256.

14. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (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.

15. Ruvinsky AM (2007) Calculations of protein-ligand binding entropy of relative and overall molecular motions. J Comput Aided Mol Des 21: 361-370.

16. Chang MW, Belew RK, Carroll KS, Olson AJ, Goodsell DS (2008) Empirical entropic contributions in computational docking: evaluation in APS reductase complexes. J Comput Chem 29: 1753-1761.

17. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 295: 337-356.

18. Muegge I, Martin YC (1999) A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. J Med Chem 42: 791-804.

19. Dill K A (1985) Theory for the folding and stability of globular proteins. Biochemistry 24: 1501-1509.

20. Newton MH, Pharma DN, Sattar A, Mahre M (2011) Kangaroo: An Efficient Constraint-Based Local Search System Using Lazy Propagation 645-659.