Pharmaceutical Research, Vol. 15, No. 3, 1998

Commentary

Recent Advances in Structure-Based Ligand Design Using Molecular Dynamics and Monte Carlo Methods

Peter Kollman[1,2]

Received October 15, 1997; accepted November 25, 1997
KEY WORDS: docking; scoring; ligand design; molecular dynamics; free energy calculations.

Computer based methods have become increasingly useful and powerful in aiding in ligand design, whether the detailed three dimensional structure of the target protein is unknown or known. If such a structure is unknown, methods to do 2D and 3D similarity searching on large databases, given lead compounds, allow a rapid analysis of which compounds might be of potential use to target certain receptors (1). 3D-QSAR methods such as COMFA (2) also enable one to explore in quantitative terms pharmacophoric hypotheses in such examples.

However, thanks mainly to advances in molecular biology and methods to crystallize proteins, even in some cases membrane proteins, there are more and more cases where there is a three dimensional structure of a target for ligand design that is known. The successes of pharmaceutical companies like Agouron and Vertex has been founded on an effective use of this information in an area of pharmaceutical relevance, but even with the receptor structure in hand, ligand design is still a challenging and interesting exercise.

How can one use computational methods to start ligand design de novo when a receptor structure is known? If the structure contains a bound ligand or substrate, this information is used to delineate where the binding site is. However, the challenge is often to design a completely novel ligand. With the advent of combinatorial chemistry, there are ever more possibilities to consider in this regard (3).

The challenge here is to consider as many possible ligands as possible, find their optimum geometry(ies) in the binding site and "score" their likelihood of binding tightly to the site. The first methodology to do this in a practical way has been the DOCK approach of Kuntz and co-workers (4–6). This methodology involves creating an image of the binding site in terms of sphere distances and then screening large databases (>100,000) ligands to find which of these ligands can potentially interact well with the site. By using a database such as the Available Chemicals Directory (ACD), one can immediately buy and test the most promising DOCK hits and thus, not embark on long and involved synthetic chemistry based on uncertain models.

This DOCK approach has been successful in finding novel micromolar inhibitors for a variety of targets and some have been taken into the nanomolar range. Nonetheless, to find an algorithm that is efficient in dealing with so many ligands, one must limit the conformational searching of the ligands and use a simple method to "score" their likely binding strength to the site. There are a variety of methods to improve the conformational searching of ligands, including those that consider different ligand conformations "on the fly" and those that use multiple conformations in the databases, which will increase their sizes by 10–100 fold. One is also seeking methods to rank ligand binding more accurately. More complex scoring functions and methodologies by Bohm (7), Jain (8), and Marshall *et al.* (9) seem to have promise in this regard. There have been many other exciting approaches to DOCKing also developed, including those that use genetic algorithm-based approaches. These may not be as fast as DOCK, but are likely more accurate (7, 10–16). But the challenge can be summarized thus: Docking and scoring, how can one make them more efficient and accurate in selecting tight binding ligands?

Our own view can be summarized in Fig. 1, which lists some of the computational methodologies for ligand design. One starts with DOCK, which, as noted above, is capable of considering ~ 100000 ligands in a practical amount of computer time. At the bottom of the list is the most accurate and rigorous approach—free energy perturbation/thermodynamic integration calculations (17)—which generally can only be done for 2 ligands at a time.

It is likely that the best strategy for computer assisted structure based ligand design is a divide and conquer one, where one uses a series of filters to consider possible ligands. Using organic chemical intuition is key also, because, even though one can buy compounds from the Available Chemical Directory that DOCK has suggested will bind, even if they bind, it may be too difficult to make analogs and make them into practical pharmaceuticals. Issues such as bioavailability and metabolism should be considered as early as possible.

Thus, we expect that docking approaches (3–5,7,8,10–16) will continue to become more accurate and to consider more molecules at a time, but the main purpose of this review is to acquaint the reader with some interesting and promising

[1] Professor of Chemistry and Pharmaceutical Chemistry, University of California, San Francisco, California 94143.
[2] To whom correspondence should be addressed. (e-mail: pak@cgl.ucsf.edu)

0724-8741/98/0300-0368$15.00/0

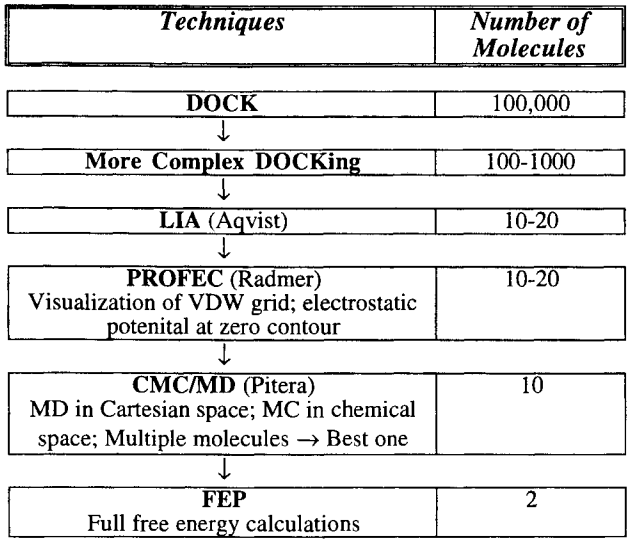

Fig. 1. Flowchart for divide and conquer strategy in structure-based drug design showing techniques with numbers of molecules that can be efficiently considered. One envisages starting from the top and using the various methods as filters, concentrating the more computer intensive techniques on the most promising molecules.

methodologies that are between docking and free energy calculations in this flowchart, viz., LIA, PROFEC and CMC/MD (Chemical Monte Carlo/Molecular Dynamics).

The LIA method of Aqvist (18) uses molecular dynamics simulations on potential ligands free in solution and when bound to the active site (19–22). The average non-bonded interactions of these ligands with their environment (either water in the free case or protein and water in the bound case) are determined. A separate analysis of non-bonded van der Waals and non-bonded electrostatic interactions is done. These interaction energies converge significantly faster than the total energy even though they probably do contain considerable statistical noise. One then estimates the binding free energy for the ligand using equation (1).

$$\Delta G = \alpha \langle \Delta E_{elec} \rangle + \beta \langle \Delta E_{vdw} \rangle \quad (1)$$

where $\langle \Delta E_{elec} \rangle$ is the difference in average ligand electrostatic interaction energy with surrounding molecules when bound and free and $\langle \Delta E_{vdw} \rangle$ is the corresponding difference in the van der Waals energies.

There are two empirical parameters in this equation, the electrostatic prefactor α and the van der Waals factor β. A value for α of 0.5 can be assumed from linear response theory, although Aqvist has derived that somewhat smaller values are appropriate if the ligand contains OH groups (23). For β, the van der Waals coefficient, Aqvist finds a value of 0.16 and Ornstein 1.0 (24). Wang has also found a value of 1.0 for biotin-streptavidin. (J. Wang, unpublished)) Jorgensen has used a related approach with Monte Carlo simulations on thrombin inhibitors (25). Thus, the appropriate coefficient to use to best fit the binding data may depend on the protein. What is amazing is that equation (1) works as well as it does, since it contains no estimate of translational/rotational entropy loss on ligand binding or any estimate of change in intraligand energies upon binding. Furthermore, much longer trajectories than currently used would be necessary if the ligands had multiple binding modes on the protein. Nonetheless, one can imagine LIA being a very useful filter for novel ligands, particularly in ranking their interaction strengths or even their mode of binding in the protein site.

Complementary to LIA is a method called PROFEC, developed by Randy Radmer in my group (26). PROFEC can use the same ligand free and ligand bound trajectories as LIA, but analyzes the differences in van der Waals interaction of the protein vs water on a free energy grid, separately considering van der Waals and electrostatic interactions. PROFEC is complementary to LIA because LIA gives a semi-quantitative estimate of the absolute binding free energy of the ligand and PROFEC gives a qualitative estimate of where on the ligand and what type of functional groups might improve binding.

Another promising method for considering multiple ligands in an active site is CMC/MD (27) initially described by Tidor and further developed by Pitera at UCSF. This method carries out MD trajectories on 1 real ligand and a number of "ghost" ligands all present at once, either in water or in a binding site. By using a Monte Carlo (MC) process to exchange a dummy ligand for a real one, one can quantitate the probability of ligand binding for a series of ligands and thus, rank their binding affinity. Solvation offsets or simply offsets in energy to improve sampling can be used to improve the statistics of the process. This method has been successfully applied to both a simple tennis ball/host system (J. Pitera, unpublished) and on HIV Reverse Transcriptase, (M. Eriksson and J. Pitera, unpublished) with encouraging results.

The above just gives a flavor of some of the promising technologies in structure based ligand design. Obviously, combinatorial methods can be added to the mix, both experimentally and computationally. The future is bright that such methods will continue to have ever-increasing impact on improving the design process.

REFERENCES

1. D. J. Weld and P. Willet. *J. Chem. Inf. Comp. Sci.* **36**:159–167 (1996).
2. C. L. Waller, T. I. Opera, A. Giolitti, and G. A. Marshall. *J. Med. Chem.* **36**: 4152–4160 (1993).
3. *Acc. Chem. Res.* (March 1996).
4. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. *J. Mol. Biol* **161**:269–288 (1982).
5. I. D. Kuntz, E. C. Meng, and B. K. Shoichet. *Acc. Chem. Res.* **27**:117–123 (1994).
6. E. Kick, D. Roe, G. Skillman, G. Liu, T. Ewing, Y. Sun, I. D. Kuntz, and J. A. Ellman. *Chemistry and Biology* **4**:623–632 (1994).
7. H. J. Bohm. *J. Comp.-Aided Mol. Des.* **8**:623–632 (1994).
8. W. Welch, J. Ruppert, and A. N. Jain. *Chemistry and Biology* **3**:449–462 (1996).
9. H. J. Head, M. L. Smythe, T. I. Opera, C. L. Waller, S. M. Green, and G. A. Marshall. *J. Amer. Chem. Soc.* **118**:3959–3969 (1996).
10. G. M. Morris, D. S. Goodsell, R. Huey, and A. J. Olson. *J. Comp. Aided Mol. Des.* **10**:293–304 (1996).
11. R. P. Meadows and P. J. Hajduk. *J. Biomol. NMR* **6**:41–47 (1995).
12. C. M. Oshiro, I. D. Kuntz, and J. S. Dixon, *J. Comp.-Aided Mol. Des.* **9**:113–130 (1995).
13. R. S. Judson, Y. T. Tan, E. Mori, C. Melius, E. P. Jaeger, A. M. Treasurywala, and A. Mathiowitz. *J. Comp. Chem.* **16**:1405–1419 (1995).

14. G. M. Verkhivker, P. A. Rejto, D. K. Gehlhaar, and S. T. Freer. *Proteins-Structure, Function and Genetics* **25**:342–353 (1996).
15. G. Jones, R. C. Glen, P. Willett, A. R. Leach, and R. Taylor. *J. Mol. Biol.* **267**:727–748 (1997).
16. M. Rarey, B. Kramer, and T. Lengauer. *J. Comp. Aided Mol. Des.* **11**:369–384 (1997).
17. P. A. Kollman. *Chem. Rev.* **93**:2395–2417 (1993).
18. J. Aqvist, C. Medina, and J. Samuelsson. *Prot. Engin.* **7**:385–391 (1994).
19. J. Aqvist and S. L. Mowbray. *J. Biol. Chem.* **270**:9978–9981 (1995).
20. T. Hansson and J. Aqvist. *Prot. Engin.* **8**:1137–1144 (1995).
21. J. Aqvist. *J. Comp. Chem.* **17**:1587–1597 (1996).
22. J. Hulten, N. M. Bonham, U. Nillroth, T. Hansson, G. Zuccarello, A. Bouzide, J. Aqvist, B. Classon, U. H. Danielson, A. Karlen, I. Kvarnstrom, B. Samuelsson, and A. Halberg. *J. Med. Chem.* **40**:885–8897 (1997).
23. J. Aqvist and T. Hansson. *J. Phys. Chem.* **100**:9512–9521 (1996).
24. M. D. Paulsen and R. L. Ornstein. *Prot. Engin.* **7**:567–571 (1996).
25. D. K. Jones-Hertzog and W. L. Jorgensen. *J. Med. Chem.* **40**:1539–1549 (1997).
26. R. Radmer and P. Kollman. "Approximate Free Energy Calculations Methods and Structure Based Ligand Design", *J. Comp. Aided Mol. Des.* (in press).
27. B. Tidor. *J. Phys. Chem.* **97**:1069–1073 (1993).