Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Pharmacology

Computer-aided drug-discovery techniques that account for receptor flexibility

Jacob D Durrant[1] and J Andrew McCammon[2,3,4]

Protein flexibility plays a critical role in ligand binding to both orthosteric and allosteric sites. We here review some of the computer-aided drug-design techniques currently used to account for protein flexibility, ranging from methods that probe local receptor flexibility in the region of the protein immediately adjacent to the binding site, to those that account for general flexibility in all protein regions.

Addresses
[1] Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, CA 92093, United States
[2] Department of Chemistry & Biochemistry, NSF Center for Theoretical Biological Physics, National Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093, United States
[3] Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, United States
[4] Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093, United States

Corresponding author: Durrant, Jacob D (jdurrant@ucsd.edu)

Current Opinion in Pharmacology 2010, **10**:770–774

This review comes from a themed issue on
New technologies – the importance of protein dynamics
Edited by Dennis R Livesay

Available online 29th September 2010

1471-4892/$ – see front matter

DOI 10.1016/j.coph.2010.09.001

Introduction

Protein receptor flexibility plays an important role in ligand binding. The lock-and-key model of binding, first proposed by Emil Fischer in 1894 [1], provided much insight; however, the assertion that a protein receptor exists in a single conformational state that is perfectly amenable to ligand binding without the need for conformational rearrangement is demonstrably false. With the advent of x-ray crystallography, comparisons between bound and unbound protein–ligand complexes have consistently demonstrated that proteins undergo a wide range of motions upon ligand binding, from small changes in binding-site residues to large-scale motions of entire protein domains [2–4]. Indeed, in some cases conformational rearrangement is so great that binding is best seen as linked to protein refolding. NMR studies capable of directly measuring protein motions have further confirmed that these macromolecules are highly dynamic.

The lock-and-key model of ligand binding, while didactically useful, was eventually supplanted by the induced-fit model, first proposed by Koshland in 1958 [5]. This model, which suggests that ligand binding itself induces conformational changes in the protein receptor, is supported by much crystallographic data. Crystal structures of bound protein–ligand complexes routinely demonstrate that 70–100% of the ligand is buried in the protein binding site, suggesting that binding-site residues 'wrap around' the ligand after the initial binding event [6].

This crystallographic evidence aside, the induced-fit model cannot explain all binding phenomena. In the late 1990s, researchers began to envision a population-based mechanism of ligand binding [7–10]. The protein receptor is thought to fluctuate between multiple conformational states, even in the unbound state. The frequency with which these various states are occupied is governed by their relative free energies according to the Boltzmann factor. Only a certain subset of these conformations is amenable to ligand binding. When a ligand binds to an amenable conformation, the binding energy stabilizes that conformation, making it more energetically favorable, and the population of all conformations consequently shifts. The population-shift model easily explains why some proteins can be activated or deactivated, depending on the bound ligand. If the bound ligand is an agonist, it probably stabilizes an active conformation; if it is an antagonist, it probably stabilizes a less active conformation.

The induced-fit and population-shift theories of ligand binding are not mutually exclusive. To varying degrees, it is likely that both effects contribute to ligand binding [11]. A ligand in solution encounters a highly dynamic protein that fluctuates between multiple, low-energy states. After initial binding, the ligand stabilizes a certain subpopulation of those states. Following binding, smaller induced-fit conformational changes may occur that further optimize protein–ligand interactions.

With the development of these theories, the critical role that protein flexibility plays in ligand binding has become apparent. Medicinal chemists engaged in computer-aided drug design (CADD) must account for this flexibility if they wish to successfully identify small-molecule ligands *in silico*. Traditionally, a single static protein structure has been used in CADD projects. While this single structure may perchance be amenable to the binding of some ligands, the assumption that a single structure can accommodate all true ligands is equivalent to the acceptance of the now antiquated lock-and-key model of binding.

Recognizing the weaknesses of methods that fail to account for protein flexibility, computational chemists have envisioned several ways of incorporating receptor flexibility into their methodologies [6,12–15]. For example, in 2004 Schames *et al.* used a molecular dynamics simulation to identify a novel, cryptic binding trench in HIV integrase that was not evident in any of the crystal structures [16]. This flexible trench was subsequently exploited pharmacologically, leading to the development of raltegravir (Isentress), approved by the FDA in 2007. A very recent computational study of the raltegravir–integrase complex, with a detailed treatment of the key divalent metal ions, confirmed that the breathing motions of the trench allow for orientationally distinct binding poses [17]. A similar study recently described the identification of a novel cryptic binding pocket adjacent to the enzymatic site of cruzain, the main cysteine protease of *Trypanosoma cruzi*. Future studies may identify novel antichagastic therapeutics that exploit this cryptic pocket as well [18].

We here review some of the CADD techniques currently used to account for protein receptor flexibility, ranging from methods that probe local receptor flexibility in the region of the protein immediately adjacent to the binding site, to those that account for general flexibility in all protein regions.

Methods that probe local receptor flexibility

A number of methods have been developed that account for the flexibility of those residues immediately adjacent to the ligand-binding site. These methods range from those that are merely forgiving of steric clashes, essentially 'soft lock and key' methods, to those that allow for local side-chain and backbone movements [19,20].

'Soft docking' was one of the earliest methods developed to account for protein flexibility [21]. Most force-field scoring functions use the Lennard-Jones potential to approximate the van der Waals force. This potential increases rapidly to infinity as the interatomic distance approaches zero; consequently, even minor steric clashes carry enormous energy penalties. If a true ligand does not fit perfectly into a static model of the protein binding site, it may be mistakenly rejected as a candidate binder.

In soft docking, the Lennard-Jones potential is replaced by a more forgiving function that does not approach infinity as the interatomic distance approaches zero. Consequently, candidate inhibitors need not fit perfectly into the target binding site; rather, some minor steric clashes are tolerated, as if protein atoms were allowed to flexibly distance themselves from the ligand upon binding. This faux flexibility is clearly limited, however; large protein rearrangements are not permitted.

To account for greater protein flexibility, methods based on rotamer libraries have been used [22–24]. The rotatable bonds of binding-site residues are first identified, and libraries of discrete rotameric states are generated for each by systematically rotating around these bonds. The rotameric states most amenable to ligand binding are then identified. Rotamer libraries can range from simple, containing only the rotamers of hydrogen-bond donors, to more complex, allowing for full side-chain rotation.

Energy refinement techniques further account for flexibility by allowing a full spectrum of backbone and/or side-chain motions, as opposed to discrete side-chain rotamers only [25–27]. Following docking, geometry relaxation procedures optimize the position of the ligand and/or protein atoms. This local energy minimization is thought to simulate the protein motions of an induced-fit effect. However, while successful in some cases, the energy minimum closest to the initial docked pose is not always the global optimum, weaknesses in force fields aside.

Methods that probe global flexibility

A number of global methodologies have been developed to overcome the limitations of methods that probe only local receptor flexibility [28••,29]. Rather than considering a single protein structure or conformation, methods that account for global receptor flexibility typically rely on multiple, conformationally diverse structures. These multiple structures can be derived experimentally from x-ray crystallography or NMR [30•,31–34], or computationally from Monte Carlo or molecular dynamics simulations [31–36].

Computational methods like molecular dynamics simulations are particularly appealing because they generate a full continuum of structures. However, as they are typically limited to at most the low-microsecond timescale, there is some concern that they may not sample all possible conformations. Recent efforts have focused on modifying the underlying free-energy profiles of molecular dynamics simulations in order to facilitate transitions between protein conformations that might otherwise be energetically unfavorable. These so-called accelerated molecular dynamics simulations address concerns about inadequate conformational sampling and may prove useful in future drug-design efforts [37,38].

Hybrid experimental/computational approaches for generating multiple structures have also been envisioned. For example, one algorithm, FlexE, compares multiple experimental structures. Those regions that do not vary among the multiple structures are considered rigid and are simply averaged. Those regions that do differ are considered flexible. Composite protein conformations are generated by mixing and matching the flexible regions from various experimental structures while maintaining the average structure of rigid regions, thus producing novel structures that may be pharmacologically relevant [39].

Several protocols have been developed that incorporate binding information gleaned from multiple structures, whatever their source, into existing docking and pharmacophore methodologies. In the relaxed complex scheme, candidate ligands are docked into multiple structures so that each compound is associated with a whole spectrum of binding scores, rather than a single score from docking into a single structure. The ligands are then ranked by different properties of this docking-score spectrum (e.g., the ensemble-average or ensemble-best score) [28••,31,40•,41,42•]. Aside from accounting for protein flexibility, this method has the advantage of relying not on one docking score, with its associated inaccuracies, but rather on multiple docking scores.

A second method generates a single composite energy grid by averaging grids calculated for multiple structures. Candidate ligands are then docked into this ensemble-average energy grid [43,44]. However, we note that true ligands may not bind to an average structure; they may instead bind to and stabilize rare protein conformations that differ substantially from the average. Additionally, the average-grid methodology relies on a single docking score, rather than on a consensus of multiple scores. As docking scores are notoriously inaccurate, reliance on a single score may be ill advised.

A third method used to predict ligand binding from multiple structures is called the dynamic pharmacophore. Multiple protein conformations, typically extracted from a molecular dynamics simulation, are characterized according to their active-site binding regions (i.e. regions that could potentially contribute to ligand binding through hydrogen bonds, aliphatic contacts, etc.). A composite pharmacophore model based on these many characterizations is then generated, and ligand databases are searched for compounds with complementary chemical features [12,15,34].

Finally, normal mode analysis has also been used to incorporate protein flexibility into computer-aided ligand-identification protocols. Low-frequency normal modes representing large-scale protein dynamics are first identified, typically from a molecular dynamics simulation. Simple parameters, included as variables in the docking algorithm, specify the extent to which the protein is deformed along these normal modes during the docking process [45,46,47•,48–50].

Emerging methods

Several new methods for predicting and scoring ligand binding have been recently developed. Among these, metadynamics, a method for exploring entire free-energy landscapes, is particularly notable. A molecular dynamics simulation of the ligand and the protein is performed. After having sampled a given region of the free-energy profile sufficiently, a Gaussian repulsive potential is placed in that region, thereby biasing the simulation towards new free-energy regions in a history-dependent manner. When the simulation is completed, the free-energy profile of the system can be reconstructed from the sum of the added Gaussians, thus allowing the identification of both the best docked pose at the free-energy minimum as well as the free energy of binding itself. Additionally, the general topography of the free-energy surface can provide insights into the binding mechanism [51]. If metavariables like those used in traditional docking algorithms (translation, rotation, and ligand conformation) are employed, a thorough exploration of the variable space is computationally intractable. However, if the molecular dynamics simulation is allowed to 'equilibrate the fast degrees of freedom,' fewer, more general metavariables can be chosen [51].

A second method called four-dimensional docking is also notable [52•,53]. Like the relaxed complex scheme, four-dimensional docking employs multiple protein conformations. Rather than performing a complete docking into each structure, however, the structure used is determined during the docking process itself, as one of the variables in the docking algorithm. This method, while less computationally demanding than the relaxed complex scheme, still maintains the advantages of multiple-receptor docking.

Conclusion

Here in, we enumerate the important roles that protein flexibility plays in ligand binding and describe several computational methods designed to account for receptor flexibility. Only a few years ago, computational chemists performing virtual screens routinely ignored ligand flexibility, though accounting for such flexibility is almost a universal feature of all modern docking algorithms. In the near future, accounting for full protein receptor flexibility, though far more computationally demanding, may likewise become routine, leading to significantly improved computer-aided identification of small-molecule ligands.

Acknowledgements

This work was supported partly by funding from NIH GM31749, NSF MCB-0506593, and MCA93S013 to JAM. Additional support from the Howard Hughes Medical Institute, the National Center for Supercomputing Applications, the San Diego Supercomputer Center, the W.M. Keck Foundation, the National Biomedical Computational Resource, the Center for Theoretical Biological Physics, and the NSF supercomputer centers is gratefully acknowledged.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Fischer E: **Einfluss der Configuration auf die Wirkung der Enzyme**. *Ber Dtsch Chem Ges* 1894, **27**:2985-2993.

2. Zhao S, Goodsell DS, Olson AJ: **Analysis of a data set of paired uncomplexed protein structures: new metrics for**

side-chain flexibility and model evaluation. *Proteins* 2001, **43**:271-279.

3. Betts MJ, Sternberg MJ: **An analysis of conformational changes on protein-protein association: implications for predictive docking**. *Protein Eng* 1999, **12**:271-283.

4. Najmanovich R, Kuttner J, Sobolev V, Edelman M: **Side-chain flexibility in proteins upon ligand binding**. *Proteins* 2000, **39**:261-268.

5. Koshland DE: **Application of a theory of enzyme specificity to protein synthesis**. *Proc Natl Acad Sci U S A* 1958, **44**:98-104.

6. Teague SJ: **Implications of protein flexibility for drug discovery**. *Nat Rev Drug Discov* 2003, **2**:527-541.

7. Ma B, Kumar S, Tsai CJ, Nussinov R: **Folding funnels and binding mechanisms**. *Protein Eng* 1999, **12**:713-720.

8. Kumar S, Ma B, Tsai CJ, Wolfson H, Nussinov R: **Folding funnels and conformational transitions via hinge-bending motions**. *Cell Biochem Biophys* 1999, **31**:141-164.

9. Tsai CJ, Kumar S, Ma B, Nussinov R: **Folding funnels, binding funnels, and protein function**. *Protein Sci* 1999, **8**:1181-1190.

10. Ma B, Shatsky M, Wolfson HJ, Nussinov R: **Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations**. *Protein Sci* 2002, **11**:184-197.

11. Bui JM, McCammon JA: **Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism**. *Proc Natl Acad Sci U S A* 2006, **103**:15451-15456.

12. Carlson HA, McCammon JA: **Accommodating protein flexibility in computational drug design**. *Mol Pharmacol* 2000, **57**:213-218.

13. May A, Zacharias M: **Accounting for global protein deformability during protein-protein and protein–ligand docking**. *Biochim Biophys Acta* 2005, **1754**:225-231.

14. Abagyan R, Totrov M: **High-throughput docking for lead generation**. *Curr Opin Chem Biol* 2001, **5**:375-382.

15. Carlson HA: **Protein flexibility and drug design: how to hit a moving target**. *Curr Opin Chem Biol* 2002, **6**:447-452.

16. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA: **Discovery of a novel binding trench in HIV integrase**. *J Med Chem* 2004, **47**:1879-1881.

17. Perryman AL, Forli S, Morris GM, Burt C, Cheng Y, Palmer MJ, Whitby K, McCammon JA, Phillips C, Olson AJ: **A dynamic model of HIV integrase inhibition and drug resistance**. *J Mol Biol* 2010, **397**:600-615.

18. Durrant JD, Keranen H, Wilson BA, McCammon JA: **Computational identification of uncharacterized cruzain binding sites**. *PLoS Negl Trop Dis* 2010, **4**:e676.

19. Bursavich MG, Rich DH: **Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles**. *J Med Chem* 2002, **45**:541-558.

20. Sowdhamini R, Srinivasan N, Guruprasad K, Rufino S, Dhanaraj V, Wood SP, Emsley J, White HE, Blundell T: **Protein three-dimensional structure and molecular recognition: a story of soft locks and keys**. *Pharm Acta Helvetiae* 2000, **69**:185-192.

21. Jiang F, Kim SH: **''Soft docking'': matching of molecular surface cubes**. *J Mol Biol* 1991, **219**:79-102.

22. Leach AR: **Ligand docking to proteins with discrete side-chain flexibility**. *J Mol Biol* 1994, **235**:345-356.

23. Anderson AC, O'Neil RH, Surti TS, Stroud RM: **Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking**. *Chem Biol* 2001, **8**:445-457.

24. Yang AY, Kallblad P, Mancera RL: **Molecular modelling prediction of ligand binding site flexibility**. *J Comput Aided Mol Des* 2004, **18**:235-250.

25. Cavasotto CN, Abagyan RA: **Protein flexibility in ligand docking and virtual screening to protein kinases**. *J Mol Biol* 2004, **337**:209-225.

26. Totrov M, Abagyan R: **Flexible protein–ligand docking by global energy optimization in internal coordinates**. *Proteins* **(Suppl. 1)**:1997:215-220.

27. Meiler J, Baker D: **ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility**. *Proteins* 2006, **65**: 538-548.

28. •• Amaro RE, Baron R, McCammon JA: **An improved relaxed complex scheme for receptor flexibility in computer-aided drug design**. *J Comput Aided Mol Des* 2008, **22**:693-705.
In the relaxed complex scheme, candidate ligands are docked into multiple structures so that each compound is associated with a whole spectrum of binding scores, rather than a single score from docking into a single structure. The ligands are then ranked by different properties of this docking-score spectrum (e.g. the ensemble-average or ensemble-best score).

29. Totrov M, Abagyan R: **Flexible ligand docking to multiple receptor conformations: a practical alternative**. *Curr Opin Struct Biol* 2008, **18**:178-184.

30. • Damm KL, Carlson HA: **Exploring experimental sources of multiple protein conformations in structure-based drug design**. *J Am Chem Soc* 2007, **129**:8225-8235.
NMR and x-ray crystallography are important sources of experimentally derived protein structures. This review compares these two sources and evaluates their respective utility in a multi-structure, dynamic-pharmacophore protocol that accounts for global protein flexibility.

31. Lin JH, Perryman AL, Schames JR, McCammon JA: **Computational drug design accommodating receptor flexibility: the relaxed complex scheme**. *J Am Chem Soc* 2002, **124**:5632-5633.

32. Wong CF, Kua J, Zhang Y, Straatsma TP, McCammon JA: **Molecular docking of balanol to dynamics snapshots of protein kinase A**. *Proteins* 2005, **61**:850-858.

33. Mangoni M, Roccatano D, Di Nola A: **Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation**. *Proteins* 1999, **35**:153-162.

34. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA: **Developing a dynamic pharmacophore model for HIV-1 integrase**. *J Med Chem* 2000, **43**:2100-2114.

35. Broughton HB: **A method for including protein flexibility in protein–ligand docking: improving tools for database mining and virtual screening**. *J Mol Graph Model* 2000, **18**: 247–257, 302–244.

36. Perryman AL, Lin JH, McCammon JA: **Optimization and computational evaluation of a series of potential active site inhibitors of the V82F/I84V drug-resistant mutant of HIV-1 protease: an application of the relaxed complex method of structure-based drug design**. *Chem Biol Drug Des* 2006, **67**: 336-345.

37. Hamelberg D, de Oliveira CA, McCammon JA: **Sampling of slow diffusive conformational transitions with accelerated molecular dynamics**. *J Chem Phys* 2007, **127**:155102.

38. Hamelberg D, Mongan J, McCammon JA: **Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules**. *J Chem Phys* 2004, **120**:11919.

39. Claussen H, Buning C, Rarey M, Lengauer T: **FlexE: efficient molecular docking considering protein structure variations**. *J Mol Biol* 2001, **308**:377-395.

40. • Amaro RE, Schnaufer A, Interthal H, Hol W, Stuart KD, McCammon JA: **Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei***. *Proc Natl Acad Sci U S A* 2008, **105**:17278-17283.
A real-world application of the relaxed complex scheme. Multiple protein structures extracted from a molecular dynamics simulation were used to identify novel inhibitors of a key enzyme in *Trypanosoma brucei*, the etiological agent of African sleeping sickness.

41. Cheng LS, Amaro RE, Xu D, Li WW, Arzberger PW, McCammon JA: **Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase**. *J Med Chem* 2008, **51**:3878-3894.

42. Durrant JD, Urbaniak MD, Ferguson MA, McCammon JA:
• **Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4′-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness**. *J Med Chem* 2010, **53**:5025-5032.
Another application of the relaxed complex scheme. In this example, multiple molecular dynamics structures were used to identify novel inhibitors of another key enzyme in *Trypanosoma brucei*.

43. Knegtel RM, Kuntz ID, Oshiro CM: **Molecular docking to ensembles of protein structures**. *J Mol Biol* 1997, **266**:424-440.

44. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS: **Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock**. *Proteins* 2002, **46**:34-40.

45. Zacharias M, Sklenar H: **Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: application to DNA minor groove ligand complex**. *J Comput Chem* 1999, **20**: 287-300.

46. Keseru GM, Kolossvary I: **Fully flexible low-mode docking: application to induced fit in HIV integrase**. *J Am Chem Soc* 2001, **123**:12708-12709.

47. May A, Zacharias M: **Protein–ligand docking accounting for**
• **receptor side chain and global flexibility in normal modes: evaluation on kinase inhibitor cross docking**. *J Med Chem* 2008, **51**:3499-3506.
Normal mode analysis has been used to incorporate protein flexibility into computer-aided ligand-identification protocols. Low-frequency normal modes representing large-scale protein dynamics are first identified, typically from a molecular dynamics simulation. Simple parameters, included as variables in the docking algorithm, specify the extent to which the protein is deformed along these normal modes during the docking process.

48. Cavasotto CN, Kovacs JA, Abagyan RA: **Representing receptor flexibility in ligand docking through relevant normal modes**. *J Am Chem Soc* 2005, **127**:9632-9640.

49. Tatsumi R, Fukunishi Y, Nakamura H: **A hybrid method of molecular dynamics and harmonic dynamics for docking of flexible ligand to flexible receptor**. *J Comput Chem* 2004, **25**:1995-2005.

50. Zacharias M: **Rapid protein–ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP**. *Proteins* 2004, **54**: 759-767.

51. Gervasio FL, Laio A, Parrinello M: **Flexible docking in solution using metadynamics**. *J Am Chem Soc* 2005, **127**:2600-2607.

52. Bottegoni G, Kufareva I, Totrov M, Abagyan R: **Four-dimensional**
• **docking: a fast and accurate account of discrete receptor flexibility in ligand docking**. *J Med Chem* 2009, **52**:397-406.
Like the relaxed complex scheme, four-dimensional docking employs multiple protein conformations. Rather than performing a complete docking into each structure, however, the structure used is determined during the docking process itself, as one of the variables in the docking algorithm. This method, while less computationally demanding than the relaxed complex scheme, still maintains the advantages of multiple-receptor docking.

53. Huang SY, Zou X: **Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking**. *Proteins* 2007, **66**:399-421.