Protein flexibility and drug design: how to hit a moving target

Heather A Carlson

The most advanced methods for computer-aided drug design and database mining incorporate protein flexibility. Such techniques are not only needed to obtain proper results; they are also critical for dealing with the growing body of information from structural genomics.

Addresses
Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, 428 Church Street, Ann Arbor, Michigan 48109-1065, USA; e-mail: carlsonh@umich.edu

Current Opinion in Chemical Biology 2002, 6:447–452

1367-5931/02/$ – see front matter

Published online 6 June 2002

Abbreviations
FIV feline immunodeficiency virus
MPS multiple protein structures

Introduction

In the 2001 issue of *Current Opinion in Chemical Biology* addressing 'Next-generation therapeutics', Abagyan and Totrov [1] presented an excellent review of computational methods for high-throughput docking and virtual ligand screening. For those applications, quick methods are needed to examine compound libraries that often contain hundreds of thousands of molecules. A standard timesaving practice is to treat the receptor as a rigid object, but neglecting protein flexibility is known to cause errors in computational studies [2•]. The act of binding a ligand into an enclosed active site requires that at least part of the receptor be flexible.

Protein flexibility is a result of the high dimensionality of the system [3]. To gauge the inherent flexibility in a protein, Betts and Sternberg [4] compared 12 pairs of independently solved structures of identical proteins (uncomplexed) and found that the RMSD of the backbone carbon atoms (C_α) ranged 0.1–0.5 Å. They found that side chains had larger RMSD values, particularly the surface residues. Zhao *et al.* [5] studied 123 structures of 77 apo-proteins to document the flexibility of side chains. Comparing different crystal structures of the same protein revealed that 10% of buried side chains differ in their χ-orientations by 11° or more. Again, the side chains on the surface of the proteins were much more flexible. Najmanovich *et al.* [6] specifically looked for large rearrangements upon binding; one χ angle of a side chain had to change by 60° or more to be included in the analysis. In a set of 980 bound and 353 unbound ligand–protein complexes, the probability of these extreme changes ranged from 1% for phenylalanine to 38% for lysine (eight residues had probabilities higher than 10%). Conformational changes within a binding site are the reason that some computer models fail to properly identify inhibitory compounds. If 40% of all side chains (8 of the 20 amino acids) have a 10–40% chance of undergoing a large rearrangement, it is obvious that a rigid protein structure will be inadequate for most structure-based drug-discovery applications.

Fortunately, faster computers are making it easier to include protein flexibility in computer-aided drug design. Though these methods are slower than the docking techniques reviewed by Abagyan and Totrov [1], they are more accurate. Over the past few years, several methods have appeared, and the most promising are presented below.

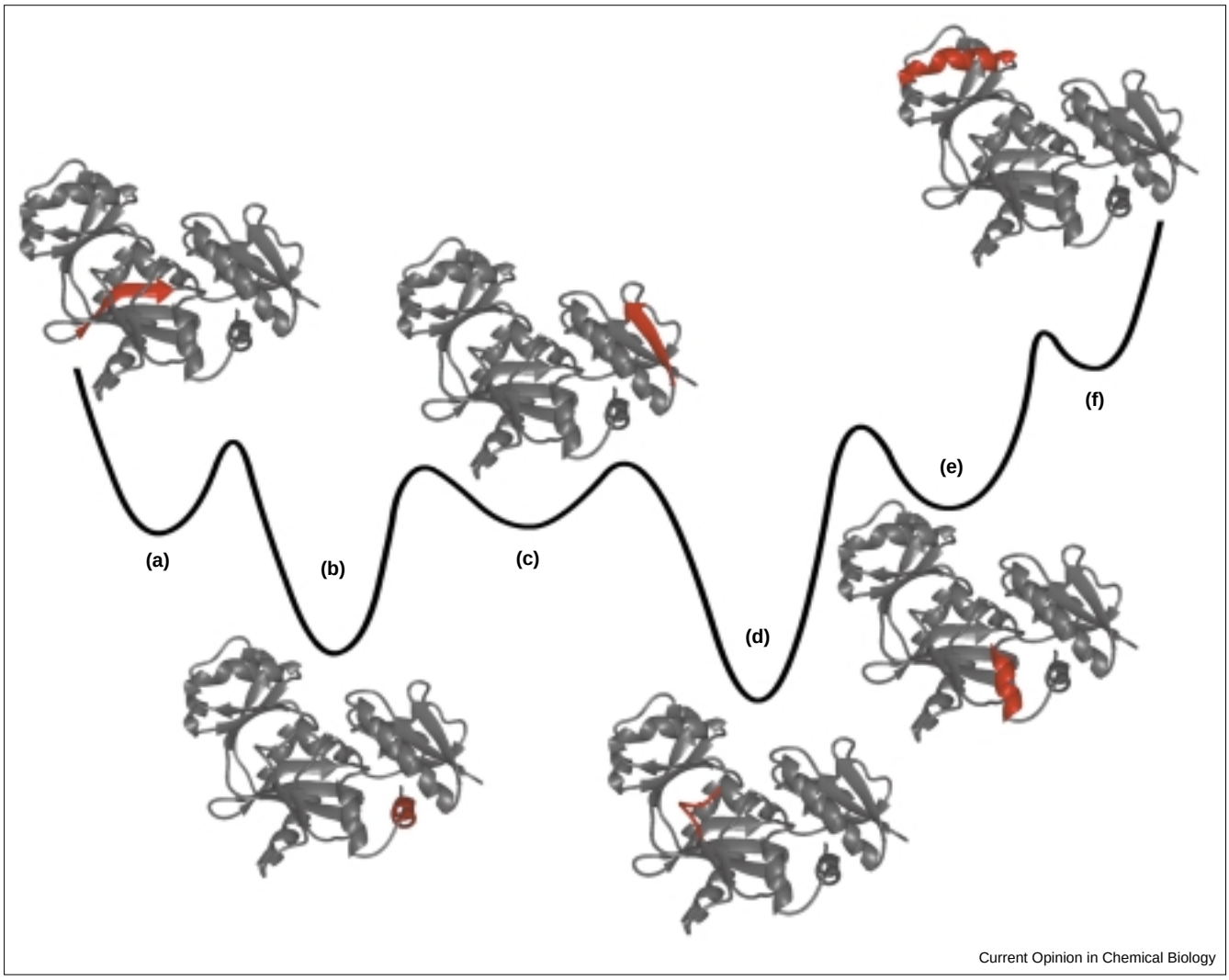
Theory

The basic premise behind the use of small-molecule drugs is that, despite the millions of years of evolution that tailor a receptor to recognize a specific ligand, we can identify molecules that bind with even higher affinity. Underlying this concept is the assumption that a binding site can accommodate many molecules and, in fact, that it *should* rearrange itself with relatively little penalty to complement the small-molecule drug. That reorganization could be as slight as small torsional motions of the side chains or as large as displacements of entire domains of the protein.

Luque and Freire [7••] have noted an interesting characteristic of many active sites. They find that binding sites appear to have a dual nature characterized by regions of high stability and regions of high flexibility (often the highest degree of flexibility of the entire protein). A certain degree of flexibility must exist in all binding sites simply to allow exchange of ligands, but the reason for areas of low flexibility is unclear. Freire and co-workers note that catalytic residues are usually in relatively stable regions of the structure. This may be necessary to ensure specificity of a system. We can predict that the relatively rigid regions of the receptors will be the most sensitive to errors when attempting structure-based drug design. Fortunately, those regions may be relatively easy to identify because they are frequently toward the center of the protein where the tighter packing density of the core leads to reduced flexibility [8].

The theories behind protein–ligand binding are moving away from the historic lock-and-key and induced-fit theories. The current paradigm describes proteins in a pre-existing ensemble of conformational states [2•,7••,9,10,11••]; Figure 1 illustrates this concept. Technically, ligands can bind to any conformation of the unbound protein. This implies that we must consider ligands of various sizes, molecules both larger and smaller than the known ligands [11••]. To have good binding affinity, a ligand could fall into two categories: a moderate binder with preference for the conformation of the receptor with the lowest free energy or a high-affinity binder that is specific for a less populated conformational state

Figure 1



A protein exists in a number of conformations in solution. (a–f) Sections of the protein highlighted in red symbolize regions where alternate conformations or displacements of sheets/helices give rise to a new conformation. In a high-resolution crystal structure, the two most favorable states (b,d) could be resolved, but in a low-resolution structure, the weighted average would fall roughly between them, possibly resembling (c) a higher free-energy state.

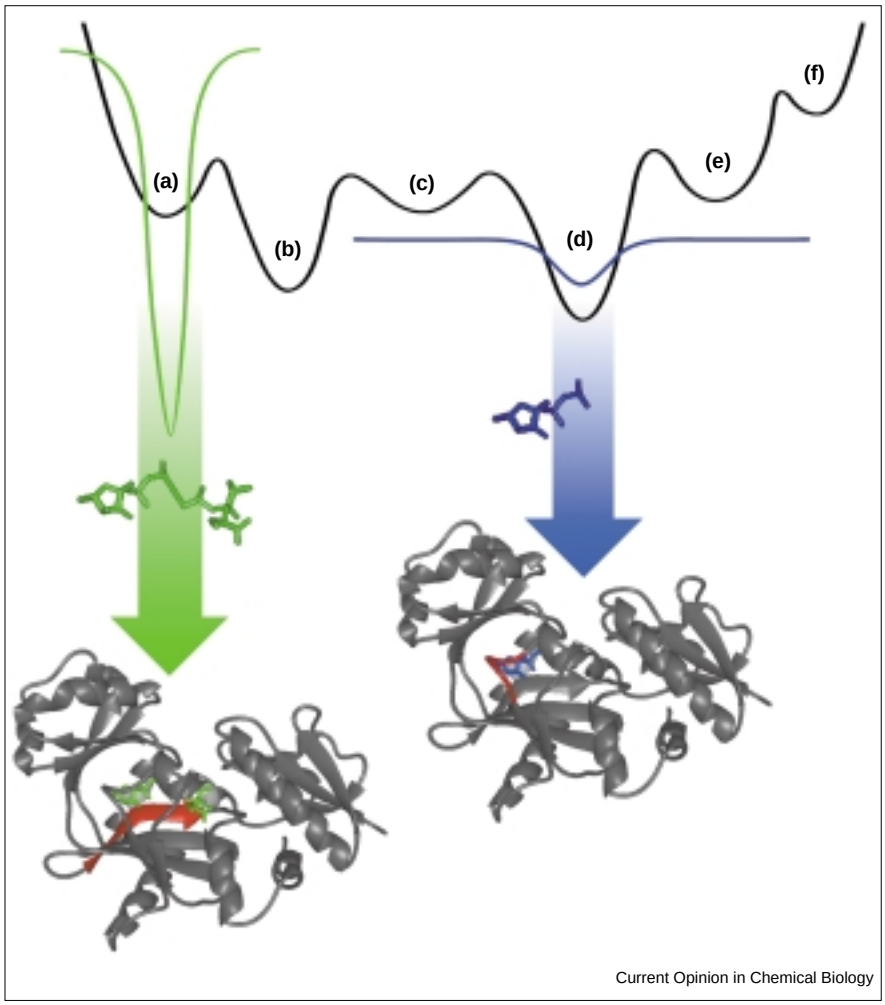
(see Figure 2). Of course, the most favorable situation is a ligand with very tight binding, specific for the most populated conformation of the unbound protein, but this is the rare instance of a lock-and-key system.

Methods for including flexibility of the target receptor

The first attempt to explicitly include protein flexibility made use of a library of discrete rotameric states for each type of side chain [12]. Since then, improved techniques have been proposed that optimize the orientation of the side chains [1,2•]. However, minimizing the energy of the side chains in the gas phase can lead to inappropriate conformations that would not be observed in the presence of solvent, and focusing on the side chains neglects any possible changes in the backbone of the protein. The use of multiple protein structures (MPS) is the best option to take advantage of the full flexibility of the receptor [2•]. These structures can come from NMR studies, multiple crystal structures, or multiple conformations generated by computational routines. The following summarizes the very few reports in the literature that use several protein conformations in structure-based drug design. In general, the techniques can be divided into methods that employ experimentally determined structures and those that use computer-generated conformations. The computational methods that have been used include molecular-dynamics routines [13], low-frequency normal modes [14,15], simulated annealing [16], and other techniques.

Figure 2

Given the conformational profile in Figure 1, two examples of favorable binding between a protein and ligand are shown. The binding free energy of the green ligand is much greater than the difference in free energy between states (a) and (d); this would correspond to an induced-fit system where the interaction with the ligand shifts the populations within the ensemble. The blue ligand is smaller, and does not fill the binding pocket as well. Its specificity and binding is weaker; however, it binds preferentially to the most populated conformation of the protein.



Five years ago, Kuntz and co-workers [17] presented the first study employing MPS. This seminal work is a must-read within the field. The authors examined both crystallographic and NMR structures as sources for protein conformations. Interaction grids were generated from each of the protein structures and then two methods for averaging the information were used to create composite grids for the DOCK program. The techniques were carefully applied to four protein systems with a dramatic improvement in accuracy and speed of the calculations.

FlexE is a recently introduced extension of the FlexX software, which incorporates protein flexibility through the use of MPS from crystallographic structures [18•]. When combining the structures, the backbone and side chains that are in good agreement are averaged, but the disordered regions (varied orientations of flexible side chains) are retained much like a rotamer library. New conformations of the receptor are generated by mixing between the MPS, choosing different rotamers within the flexible regions. Using FlexE, 83% of the ligands for the test system were placed within 2.0 Å RMSD of their crystallographic position. Bursavich and Rich [19] have proposed techniques similar to the FlexE approach. Using the aspartic protease family as an example, the authors generated alternate conformations of the binding sites by manually mixing between the side chain orientations of the homologous proteases. Though the authors were successful with the renin active site, some conformers required concerted motion of three side chains. This is difficult to do without previous knowledge, and the authors pointed to the need to automate the process.

Kastenholz *et al.* [20] have used a method based on GRID to examine the crystal structures of a family of serine proteases. Consensus principle component analysis (GRID/CPCA) was used to analyze the interaction grids, working with sets of MPS rather than pairs of structures as they had done previously. The consensus regions were used to analyze for similarities but, more importantly, the authors note that the differences were related to the structural aspects that define the specificity of the enzymes.

The most recent study using MPS was made by Österberg *et al.* [21•] in 2002. The study used 21 crystal structures

of HIV-1 protease complexed with peptidomimetic inhibitors. AutoDock was used, and the most interesting feature of the study was that the authors systematically compared four choices for combining the interaction grids. When docking the 21 inhibitors to the combined grids, a grid of mean values was shown to be too restrictive because of the dominance of the repulsive terms (steric clashes), and a minimum-energy grid overestimated the favorable regions (ignoring some regions of steric hindrance). Two methods for producing weight-averaged grids performed well for every one of the 21 inhibitors.

My own studies used the first MPS method based on computer-generated conformations of a protein [22•], and the method was later expanded to examine crystal structures [23]. Several snapshots were taken from a molecular dynamics simulation of an uncomplexed form of HIV-1 integrase. Small-molecule probes were docked to the protein structure from each snapshot. The probe molecules mapped the most favorable positions for certain functional groups to complement the surface of the receptors. The maps were overlaid to identify regions where probes were conserved over many of the conformations of the receptor. The MPS model was based on those conserved interactions (focusing on the most rigid and unforgiving features of the receptor). The composite model was used to search the *Available Chemicals Directory* for potential inhibitors. Subsequent experimental testing of the identified compounds showed that approximately one third of the compounds were inhibitory. To date, this is the only MPS method that has been experimentally verified.

Broughton [24] also introduced a technique using MD simulations to generate many conformations, but this study made use of protein–inhibitor complexes. A composite grid for docking studies was created by overlaying the structures with respect to the bound inhibitors and calculating a weighted-average of the individual grids. This method showed significant improvement over using a single crystal structure to evaluate known ligands of dihydrofolate reductase.

Kuhn and co-workers have played a central role in developing docking techniques. Their original code, SPECITOPE, was one of the first to allow for induced complementarity between a flexible ligand and flexible side chains in the receptor, and they were the first to use this technique when screening databases [25,26]. More recently, they have introduced SLIDE, an improved database-screening tool [27]. Instead of interaction grids, SLIDE uses template points in docking. The template points are similar to the complementary maps created with our technique of docking small-molecule probes, but the points are generated with geometric rules for hydrophobic regions, hydrogen-bond donors, and hydrogen-bond acceptors. Their most recent improvement to SLIDE has been the addition of backbone flexibility [28•]. Flexible regions of the backbone are identified with a graph theory technique (the FIRST algorithm), and sets of alternate positions for the backbone are generated through random sampling. The backbone motion is limited to a reasonable range by the local van der Waals interactions and the hydrogen-bonding network.

Anderson *et al.* [29] have presented an algorithm for identifying SOFTSPOTS, regions of high flexibility, from a single protein structure. This technique is combined with second routine, called PLASTIC, which provides a collection of possible conformations based on rotameric libraries. The methods were found to reduce the bias in a receptor conformation induced by a bound ligand in a co-crystal.

A fascinating method has been introduced by Ota and Agard [30••]. The technique is described as multi-conformation simulated annealing-pseudo-crystallographic refinement (MCSA-PCR). Multiple, independent cycles of 'growing' a ligand into a binding site were completed using simulated annealing for conformational sampling. This produces an ensemble of docked protein–ligand conformations, which are used to generate a corresponding electron-density map. That map is then back-analyzed with conventional crystallographic refinement tools to fit a single, optimal structure of the complex into the density. An added benefit of the method is that the system is solvated during the simulated annealing and many of the water positions seen in the crystal structures of the test system were appropriately reproduced. The only other method discussed above that has accurately incorporated water positions between the ligand and the receptor is SLIDE [27].

Future directions

In November 2000, *Nature Structural Biology* dedicated a supplemental issue to the area of structural genomics. As part of that issue, Norvell and Machalek [31] announced that the US National Institute of General Medical Sciences had recently funded seven research centers as part of its Protein Structure Initiative. Drug discovery will be revolutionized with the explosion of information to come [32], but only if we have the forethought to develop tools that combine many related protein structures in a way that is useful for structure-based drug design.

One area that will immediately benefit from the techniques described above is the use of homology models in structure-based drug design. Homology models are not accurate enough for drug design because they contain errors that often result in binding sites with the wrong size and shape. Methods that allow for protein flexibility will help correct some of those errors by allowing crosstalk between the receptor and the ligand in question. Current methods use approximate protein models with soft-docking, surface-matching, or shape-matching techniques [33]. Schafferhans and Klebe [34•] have recently suggested the use of multiple homology models based on several crystallized proteins similar to the target. As a test case, they created many homology models of thrombin using the

structures of homologous serine proteases (28–40% sequence identity). Combining the homology models led to a significant improvement in reproducing the proper binding orientations for known inhibitors of thrombin (average RMSD was 1.4 Å). It is encouraging to see that MPS methods have been successfully extended to homology models.

Murcko and co-workers [35] have discussed the potential benefits of combining the information from an entire protein family in the pursuit of small-molecule drugs. Chemical libraries, reagents, reactions and assays can be 'reused' to speed drug discovery. Other groups have also pointed to the promise of designing combinatorial libraries based on related protein structures [36] and to the application of those libraries across a family of related enzymes [37]. Essential residues and general folds are conserved across families of proteins, providing a certain measure of the 'flexibility' of a receptor based on the observed evolutionary differences. Those differences could be exploited in the pursuit of inhibitors with enhanced specificity, or it may be possible to design broad-spectrum inhibitors by using a set of homologous protein structures. This latter approach would be advantageous in the development of antibiotics where activity against many pathogens is desirable [38,39].

Recently, inhibitors have been developed based on the protease from feline immunodeficiency virus (FIV) [40•]. Six of the mutated residues that result in resistance to HIV-1 protease inhibitors are found naturally in FIV protease (Lys20Ile, Val32Ile, Ile50Val, Asn88Asp, Leu90Met and Gln92Lys). By creating inhibitors that complement FIV protease, Elder, Wong and co-workers have developed inhibitors that are also effective against HIV-1 protease and its resistance mutants. This breakthrough underscores the promise of broad therapeutics created from homologous enzymes.

A final area that would benefit greatly from the inclusion of protein flexibility is the recent use of docking techniques to screen a molecule across a database of proteins (as opposed to the usual application of screening a database of thousands of small molecules against a single protein). This new development is being used to determine the potential for toxicity and side effects of a lead compound [41,42]. These negative consequences can result from the lead compound binding to a site in an unexpected protein. Accurate results from these applications will be even more dependent upon including the plasticity of the binding site because of the need to anticipate the receptor's new and unusual modes of molecular recognition.

Conclusions

The most recent developments for including protein flexibility in computer-aided drug discovery have been presented. Only a few years ago, ligand molecules were always held rigid during docking studies, but now, ligands are routinely treated with partial or full flexibility [1]. One can imagine that in the near future, it will be routine to also treat the receptor with full flexibility. Computer-aided drug discovery will be even more reliable with this advance, and the extension of these tools to proteomics and structural genomics is on the horizon.

Update

Verkhivker *et al.* [43] have recently reviewed protein–ligand recognition, focusing on the basic theory of energy landscapes and ensembles of conformational states. Also, Lazaridis *et al.* [44] have calculated the various components of the free energy of binding between biotin and streptavidin. Though their study is not a drug-discovery application, like most of the work reviewed here, they quantify the penalty for reorganizing the protein to be 10–30 kcal mol^{-1}. Such a significant penalty is surprising for a system with femtomolar-level affinity.

Acknowledgements

I would like to thank Prof J Andrew McCammon for guidance during my postdoctoral studies [22•,23] and for his continued encouragement. This work has been supported in part by the National Institutes of Health (GM 65372) and the Beckman Young Investigator program.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

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

2. • Carlson HA, McCammon JA: Accommodating protein flexibility in computational drug design. *Mol Pharmacol* 2000 57:213-218.
A review that covers the entire span of methods for including protein flexibility: implicit methods such as 'soft docking', the use of side-chain rotamers, generation of conformational ensembles and methods for predicting movement of domains.

3. Taverna DM, Goldstein RA: Why are proteins marginally stable? *Proteins* 2002, 46:105-109.

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

5. Zhao S, Goodsell DS, Olson AJ: Analysis of a data set of paired uncomplexed protein structures: new metrics of side-chain flexibility and model evaluation. *Proteins* 2001, 43:271-279.

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

7. •• Luque I, Freire E: Structural stability of binding sites: consequences for binding affinity. *Proteins* 2000, S4:63-71.
A fascinating technique is used to estimate and rank an ensemble of conformational states. The flexibility is mapped onto the individual residues, showing networks of flexible and rigid regions. The method can also be used to map the regions of rigidification and conformational instability propagated through the system upon binding a ligand.

8. Halle A: Flexibility and packing in proteins. *Proc Natl Acad Sci USA* 2002, 99:1274-1279.

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

10. Pan H, Lee JC, Hilser VJ: Binding sites in ***Echerichia coli*** dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc Natl Acad Sci USA* 2000, 97:12020-12025.

11. •• 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.
This is one of the most recent reviews from a group at the center of the growing theory of pre-existing conformational ensembles of proteins. Here, the

authors specifically address binding events and their influence on an ensemble of protein conformations.

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

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

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

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

16. Verkhivker GM, Rejto PA, Bouzida D, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Marrone T, Rose PW: Parallel simulated tempering dynamics of ligand-protein binding with ensembles of protein conformations. *Chem Phys Lett* 2001, 337:181-189.

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

18. • Claußen H, Buning C, Rarey M, Lengauer T: FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* 2001, 308:377-395.
This method uses the docking program FLEXX. The authors present an interesting and promising method for combining protein structures.

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. Kastenholz MA, Pastor M, Cruciani G, Haaksma EEJ, Fox T: GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem* 2000, 43:3033-3044.

21. • Österberg 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.
This study carefully examines four routines for combining interaction grids from MPS. Their insights should be kept in mind when evaluating new methods for combining MPS.

22. • Carlson HA, Masukawa KM, Jorgensen WL, Lins RD, Briggs JM, McCammon JA: Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem* 2000, 43:2100-2114.
This is the first paper to use computer-generated conformations of a protein, and it is the only method that has been experimentally verified. This work focuses on an unbound protein structure, unlike the other applications that used ligand–protein complexes. Rather than employing interaction grids, this method uses receptor-based pharmacophore models for evaluating ligand interactions.

23. Carlson HA, Masukawa KM, McCammon JA: Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J Phys Chem A* 1999, 103:10213-10219.

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

25. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA: Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* 1998, 33:74-87.

26. Schnecke V, Kuhn LA. Flexibility screening for molecules interacting with proteins. In *Rigidity in Theory and Applications.* Edited by Thorpe MF, Duxbury PM. New York: Plenum Publishing; 1999:385-400.

27. Schnecke V, Kuhn LA: Virtual screening with solvation and ligand-induced complementarity. *Persp Drug Disc Design* 2000, 20:171-190.

28. • Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF: Protein flexibility predictions using graph theory. *Proteins* 2001, 44:150-165.
This method can provide similar insights as the method in [7••], but it is based on a different underlying theory. The protein is treated as a mechanical system, and the connectivity within the system is evaluated to identify structurally underdetermined (flexible) regions. The regions can then be sampled to create alternate conformations within those flexible regions.

29. 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.

30. •• Ota N, Agard DA: Binding mode prediction for a flexible ligand in a flexible pocket using multi-conformation simulated annealing pseudo crystallographic refinement. *J Mol Biol* 2001, 314:607-617.
The authors create multiple conformations of a ligand–protein complex using simulated annealing. The ensemble is then combined to generate a corresponding electron-density map that is then fit with standard crystallographic refinement tools. The final model is determined in a fashion similar to the analysis of the experimental data from crystallography and may provide the best comparison between computer-generated models and experimental structures.

31. Norvell JC, Machalek AZ: Structural genomics programs at the US National Institute of General Medical Sciences. *Nat Struct Biol* 2000, 7:S931.

32. Dean PM, Zanders ED, Bailey DS: Industrial-scale, genomics-based drug design and discovery. *Trends Biotechnol* 2001, 19:288-292.

33. Wojciechowski M, Skolnick J: Docking of small ligands to low-resolution and theoretically predicted receptor structures. *J Comput Chem* 2002, 23:189-197.

34. • Schafferhans A, Klebe G: Docking ligands onto binding site representations derived from proteins built by homology modelling. *J Mol Biol* 2001, 307:407-427.
This work presents a promising method for using homology models in drug design. An interesting point is that combining several homology models based on multiple homologous proteins yields significant improvement, much like the use of MPS to represent protein flexibility.

35. Caron PR, Mullican MD, Mashal RD, Wilson KP, Su MS, Murcko MA. Chemogenomic approaches to drug discovery. *Curr Opin Chem Biol* 2001, 5:464-470.

36. Böhm H-J, Stahl M: Structure-based library design: molecular modelling merges with combinatorial chemistry. *Curr Opin Chem Biol* 2000, 4:283-286.

37. Lamb ML, Burdick KW, Toba S, Young MM, Skillman AG, Zou X, Arnold JR, Kuntz ID: Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* 2001, 42:296-318.

38. Moir DT, Shaw KJ, Hare RS, Vovis GF: Genomics and antimicrobial drug discovery. *Antimicrob Agents Chemother* 1999, 43:439-446.

39. Russell RB, Eggleston DS: New roles for structure in biology and drug discovery. *Nat Struct Biol* 2000, 7:S928-S930.

40. • Lee T, Le V-D, Lim D, Lin Y-C, Morris GM, Wong AL, Olson AJ, Elder JH, Wong C-H: Development of a new type of protease inhibitors, efficacious against FIV and HIV variants. *J Am Chem Soc* 1999, 121:1145-1155.
This paper is a fascinating example of using an homologous enzyme to develop a broad-spectrum inhibitor for protease enzymes from immunodeficiency viruses.

41. Chen YZ, Ung CY: Prediction of potential toxicity and side effect protein targets of a small molecule by ligand–protein inverse docking approach. *J Mol Graph Model* 2001, 20:199-218.

42. Chen YZ, Zhi DG: Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* 2001, 43:217-226.

43. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW: Complexity and simplicity of ligand–macromolecule interactions: the energy landscape perspective. *Curr Opin Struct Biol* 2002, 12:197-203.

44. Lazaridis T, Masunov A, Gandolfo F: Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins* 2002, 47:194-208.