Managing protein flexibility in docking and its applications

Chandrika B-Rao, Jyothi Subramanian and Somesh D. Sharma

Discovery Informatics, Piramal Life Sciences Limited, 1 Nirlon Complex, Off Western Express Highway, Goregaon (E), Mumbai 400063, India

Docking, virtual screening and structure-based drug design are routinely used in modern drug discovery programs. Although current docking methods deal with flexible ligands, managing receptor flexibility has proved to be challenging. In this brief review, we present the current state-of-the-art for computationally handling receptor flexibility, including a novel statistical computational approach published recently. We conclude, from a comparison of the different approaches, that a combination of methods is likely to provide the most reliable solution to the problem of finding the right protein conformation for a given ligand.

Computational modeling and simulation have become integral components of drug discovery programs in Pharma R&D [1–4]. A key structure-based drug design (SBDD) method – docking of small molecules to protein binding sites – was pioneered during the early 1980s [5]. Since then, it has been used as a tool not only for hit identification and lead optimization [6–8] but also for analysis of drug metabolism and toxicity [9,10]. Despite its potential for a wide range of applications in drug discovery, methodological problems continue to afflict docking [6,11,12]. Various issues related to scoring functions, such as handling metals and water molecules, covalent interactions and flexibility in the ligand and receptor, are still active areas of research [12].

The earliest reported docking method [5] was based on the lock-and-key theory of ligand–protein binding and treated the ligand and protein as rigid bodies. At the next stage, docking of flexible ligands to rigid proteins was addressed and these remain the most popular methods in use [12–18]. Considering the importance of protein flexibility on ligand binding, recent research has explored ways of addressing the situation where the ligand and protein are flexible [19–25]. Docking of rigid and flexible ligands to rigid protein has been reviewed thoroughly [6,11,12] and will not be discussed here. In this review, we summarize some of the methods currently being investigated to model protein flexibility and discuss their strengths and limitations.

Corresponding author: B-Rao, C. (chandrika.rao@piramal.com)

Modeling protein flexibility

While conformational variability in different apo structures of the same protein suggests intrinsic disorder, variability in different ligand-bound states seems to be related to the presence of the ligand. The debate over whether the conformation of the protein in the ligand-bound state already existed before binding by ligand and was selected from an ensemble of different conformations by the ligand [26,27], or whether it was induced by the ligand [28], remains unresolved [29], although many researchers are currently convinced of the former explanation [30,31]. In either case, it seems clear that the ligand plays a central role in determining the conformation of the protein to which it is bound [32,33]. For simplicity, we shall refer to conformational changes in the cocrystal structure of protein as ligand-induced conformational changes.

Although the existence of ligand-induced flexibility has been known for almost half a century [26,28], it was generally considered to be rare or inconsequential; however, a recent statistical analysis of the PDB revealed that 85% of the proteins contain one to three flexible residues in the active site [34]. It is also reported that single rigid receptor dockings predict incorrect binding pose for 50–70% of all ligands [35]. Flexible residues are subject to conformational changes ranging from simple side-chain movements to backbone-loop movements to major domain rearrangements [31–37]. The complexity is highlighted by a sample of some of the kinds of conformational changes observed in X-ray crystal structures (Fig. 1). Many factors have hindered the development of methods for dealing with ligand-induced protein conformational

1359-6446/06/$ - see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.drudis.2009.01.003

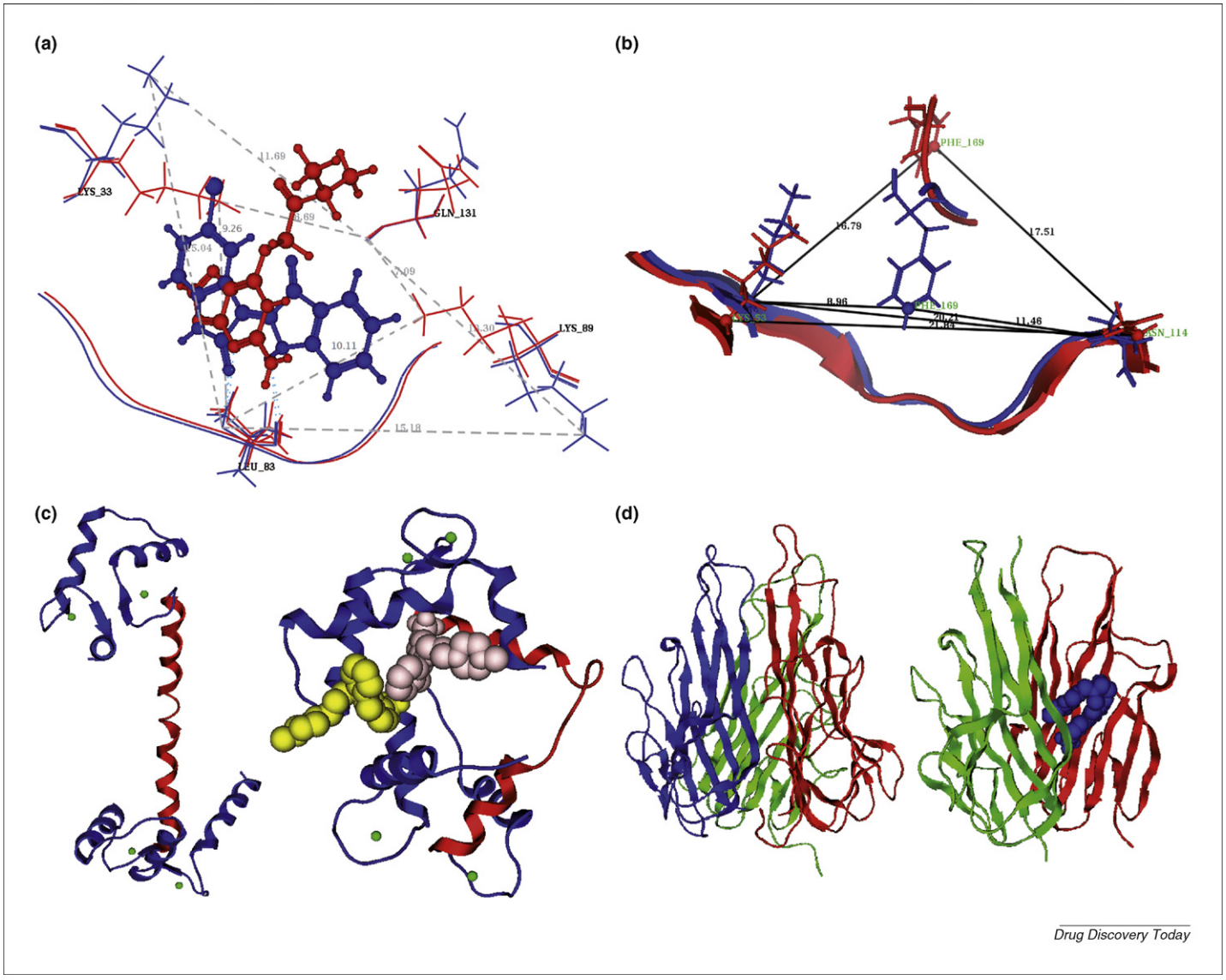

FIGURE 1

Conformational changes in proteins on ligand binding. **(a)** Different conformations of Lys33 and Lys89 side-chains in CDK2 on binding to two different ligands [24]. The ligands are shown as ball-and-stick models. For the sake of clarity, only the active site of CDK2 is represented by line models. The protein conformations (in red and blue) correspond to the conformations on binding to the ligands depicted in red and blue, respectively. For the quantification of flexibility, distances of Lys33 and Lys89 from invariant residues Leu83 and Gln131 of CDK2 are shown. The areas of the active site defined by these four residues for the conformation in blue is double that for the conformation in red [24]. **(b)** Different conformations of the DFG-loop and the hinge region in p38MAPK on binding to different ligands. The ligands are omitted in this figure for the sake of clarity. Backbone DFG-loop movements in p38MAPK and their distances from invariant residues Lys53 and Asn114 are shown by lines to quantify the changes. Red is a DFG-in and blue is a DFG-out conformation [25]. **(c)** Domain readjustments in the calcium-binding protein, calmodulin, on binding to two molecules of the ligand trifluoperazine. The dumbbell-shaped, unliganded conformation is shown on the left with the central linker helix in red and the two lobes in blue (PDB ID 3CLN). The conformation of calmodulin on binding to two molecules of trifluoperazine is shown on the right (PDB ID 1A29). The two ligand molecules are shown as yellow and pink spheres. The central linker helix has collapsed in this case. In these figures, the calcium ions are shown as green balls. **(d)** Absence of a monomer in the TNF-α trimer in the ligand-bound state. The figure on the left is the apo TNF-α trimer (PDB ID 1TNF). Upon binding to a small-molecule inhibitor there is accelerated dissociation of a monomer and the TNF-α dimer in complex with the small-molecule inhibitor is shown on the right (PDB ID 2AZ5). The inhibitor is represented in a blue sphere model. Ligands are shown as ball-and-stick models (a) or spheres (c and d) or not shown (b).

changes. Among these is a lack of a clear understanding of what kinds of changes may be induced by a ligand in a given protein. Experimental and computational efforts at resolving this are currently being undertaken [33,36]. The numerous degrees of freedom involved in incorporating protein flexibility during docking have posed computational challenges [31,32,35,36]. Difficulties in separating the amount of protein effects on ligand conformation and simultaneous ligand effects on protein conformation, and how to develop algorithms to handle such complex interactions, are other issues that needed to be addressed. The lack of adequate understanding of receptor flexibility owing to the paucity of experimental data has also obstructed the development of theoretical and computational models.

As a first approximation, any holo form of the protein was thought to take care of the conformational changes that may be induced by any other ligand adequately [38]. Ligands with

diverse chemical scaffolds and of different sizes are, however, known to induce different types, or different extents, of changes in protein conformation [24,25,32], and these cannot be adequately taken care of by this approach, because even a slight change in the protein conformation can have a significant effect on the results of docking [39]. The results of many cross-docking experiments have also demonstrated that docking accuracy invariably falls when a ligand is docked, not to its own true protein conformation but to an alternate conformation, an average conformation or an apo conformation of the protein; and this drop in docking accuracy mirrors the degree to which the positions of the protein active site residues are affected by ligand binding [40].

Soft docking

The earliest attempts to accommodate small changes in receptor conformation used a rigid protein for the docking but soft scoring functions to evaluate the binding [20,21]. These soft functions tolerate some overlap between the ligand and the protein, which accounts for a small amount of plasticity of the receptor. Soft docking has the advantage of being computationally efficient (only the scoring function parameters need to be changed, everything else remains the same as in the case of rigid receptor docking) and is still being actively explored [12,31,33,36,37]. Soft potentials can take care of only minor side-chain movements, but not backbone movements or other major structural rearrangements (Table 1).

Survey methods – rotamer exploration and multiple protein structures (MPS)

Survey methods offer an alternative way to model receptor flexibility and are similar in spirit to the approach of exploring multiple ligand conformations for simulating ligand flexibility [11]. They come in a variety of implementations. One of these is the exploration of rotamer libraries [12,31,33,37,41,42] to simulate side-chain movements (Table 1). Other types of conformational changes cannot be handled by this approach and, hence, this method fails to simulate ligand-induced flexibility fully.

Another simple implementation of the survey method is the docking of the ligand to multiple receptor structures [31,43–45]. While virtually simulating protein conformation selection by ligand, this becomes computationally very expensive as the number of conformations increases. It allows for all kinds of conformational flexibility, but is restricted by what is represented in the MPS set being used. As the number of structures used increases, the computational effort increases, and the chances of ambiguity caused by multiple equally good docked poses and of false positive results also increase. A variety of ways of generating or selecting the multiple receptor structures have been proposed [22,31,33,36], some of which are presented here, albeit briefly (Box 1).

Efforts to overcome the computational problems of multiple dockings include cross-docking experiments aimed at reducing the set of crystal structures to be used for docking (e.g. see Ref. [44]). The number and choice of protein structures is, however, an issue that remains to be resolved [43] as different training sets produce different shortlisted sets of conformations. Also, being based on the set of cocrystallized ligands in the training set, there is no guarantee that it includes the conformation appropriate for a query ligand having a completely different chemical structure. Cross-docking methods do not take properties of specific ligands into consideration for protein selection.

Ensemble docking methods that use an average receptor grid [35,46] generated from MPSs take into account the range of flexibility seen in the multiple structures while significantly reducing the computational cost of multiple dockings. This approach, however, also has its pitfalls, notably the difficulty in including significant structural diversity in the average grid and the danger of artificial ligand poses that match only the artificial average grid structure and not any real conformation of the protein [36]. ''*In situ* cross-docking'' approaches [47,48] can overcome these issues to some extent while retaining the computational advantage, but have limitations in the number of structures that can be examined at a time [36].

Modeling protein flexibility using ligand properties

None of the methods discussed above, except perhaps rotamer exploration, considers ligand-induced flexibility in a quantitative manner. The first report of quantitative characterization of ligand-induced receptor movements [49], used classical molecular interaction potentials (cMIP), root mean squared deviations (RMSDs) and volumes to quantify the binding sites of proteins. A principal component analysis of the cross-correlation matrix of the cMIP was used to cluster the binding sites into similar and dissimilar ones and to identify representative structures for docking and, finally, ensemble docking was recommended [49]. No effort was made to associate the properties of individual ligands with differences in the binding site conformation, and ligand properties were not used in the selection of the protein structures for docking.

We have recently proposed [24,25] a knowledge-based approach for handling receptor flexibility, which interfaces chemistry (small molecule properties) with biology (protein structure) using data analysis and modeling techniques. We refer to it as quantitative structure-induced conformation relationship (QSiCR) analysis because it is similar in spirit to the quantitative structure–activity relationship (QSAR) approach. The difference between the two approaches is that QSiCR predicts the effect of ligand on the *structural conformation* of key residues in the active site of *the protein* instead of the *biological activity* of *the ligand* or the concentration of ligand causing a 50% change in activity of protein, predicted by QSAR models.

In the QSiCR approach, the conformational differences that can be observed in the active site of mainly ligand-bound forms of crystal structures of proteins are first identified, typically at the level of residues. Suitable position-invariant residues in the active site are then identified from the same set of structures, to define a frame of reference. Conformational changes are defined in terms of distances (or other geometric quantities) of variable residues from invariant residues. These geometric quantities are modeled as functions of ligand properties using statistical–computational methods. Relevant ligand properties (structural, as well as physico-chemical) that play a role in protein conformational changes are inferred [24,25]. Relevant properties of a query ligand are calculated, the QSiCR model is applied and the expected protein conformation binding to the ligand is predicted. The predicted structure is then used for docking.

For two proteins, cyclin-dependent kinase 2 (CDK2) and p38 MAP kinase (p38 MAPK), chosen as case studies, it has been shown

TABLE 1

Approaches to incorporate protein flexibility in docking.

Approach	Merits	Limitations	Refs
Docking with soft potentials	Low computational cost Easy to interpret Can detect subtle changes that other methods overlook	Only subtle side-chain changes detected Completely novel conformation cannot be found Low level of ligand effects No quantification of conformational changes Biased by input conformation	[20,21]
Rotamer exploration	Moderate to high computational cost depending on size of rotamer library Can find novel conformation if included in rotamer library Moderate level of ligand effects Ideal when active site shows only side-chain rotations	Only side-chain conformational changes Multiple solutions may lead to ambiguity in interpretation May be biased by rotamer library May fail if desired rotamer not included in library	[41,42]
Multiple protein structures (MPS) multiple dockings	Side-chain and backbone changes but restricted by ensemble set used Ligand effects included to moderate level Multiple ways of generating multiple structures	High computational cost if many protein structures Novel conformation cannot be found Ambiguity in interpretation of multiple equally good solutions No quantification of conformational changes for specific ligand Biased by input set of conformations May fail if specific ligand induces novel conformation not included in MPS set	[44,53]
Multiple protein structures (MPS) ensemble docking	Side-chain and backbone changes but restricted by ensemble set used Easy to interpret Computationally better than multiple structures for VS Ligand effects included to moderate level	Moderate computational cost A small chance of finding novel conformation No quantification of conformational changes May fail if ensemble is biased Average protein structure may not give a real conformation	[46,54,56]
Current hybrid methods	Side-chain and backbone movements Can find novel conformation to some extent Moderate to high level of ligand effects Low bias due to input protein conformation	Moderate to high computational cost Interpretation may be difficult if multiple equally good solutions found May miss large conformational changes	[19,50]
Molecular dynamics simulations (MDS) after docking	Side-chain and backbone movements but restricted by simulation time Can find novel conformation to limited extent Moderate to high level of ligand effects Quantification of conformational changes Can handle solvent and ionic effects and membrane-bound proteins	High computational cost Analysis of huge output effort-intensive Biased by starting structure and initial velocity May fail owing to short times of simulations and may miss best solution	[22,68]

that side-chain, as well as large backbone, conformational changes can be handled well by a QSiCR approach [24,25] (Fig. 1a,b). As in the case of a QSAR analysis, different multivariate statistical techniques including regression and classification methods can be explored to model the receptor conformational changes as functions of ligand properties, to obtain a well-validated model. Hence, QSiCR is an approach rather than a single method.

For drawing a relationship between ligand properties and protein conformations, 2D and/or 3D descriptors are used. Three-dimensional descriptors of protein-bound ligands in their crystal structure conformations can be easily used in developing the model, but they pose problems for predicting the changes 'induced' by a query ligand whose 3D protein-bound conformation is not known. Hence, models based on 2D descriptors are preferred. The cases of CDK2 and p38 MAPK suggest that 2D descriptors of cocrystallized ligands are sufficient to capture most of the variation. In the case of CDK2, important descriptors in the model indicated that bulky and negatively charged groups shifted the Lys33 side-chain to the *trans* position, whereas bulky H-bond donor groups kept Lys33 in gauche. Hydrophobicity of the ligand played a major role in the Lys89 side-chain position [24]. In the case of p38 MAPK, molecular access system (MACCS) fingerprints, which relate to topological properties of the ligand, were found to play a major role [25].

Model validation is an important step before its application to protein selection; the QSiCR approach is rejected if no validated model is found. For a query ligand the QSiCR models can interpolate between the conformations found in crystal structures of the training set. For p38MAPK, it was observed that intermediate DFG-loop conformations were predicted for some novel ligands [25]. In general, QSiCR models cannot be used to extrapolate to new conformations that fall outside the scope of the training set.

Because protein structure prediction is done before docking, QSiCR circumvents the added computational complexity of incorporating receptor flexibility during docking (e.g. GOLD [15] and FlexE [21]). This model is expected to provide an estimate of the end-state conformation of the protein after binding to ligand without actually going through the intermediate steps until that state is achieved. This constitutes a major computational advantage of QSiCR over multiple dockings to different conformations and MDS methods.

BOX 1

Generation and selection of multiple protein structures

An MPS set can be obtained by experimental methods like NMR and X-ray diffraction models [55] and techniques like Mossbauer spectroscopy, neutron scattering, time-resolved Laue X-ray diffraction and hydrogen exchange along with MDS [33], or computationally by the application of a rotamer library, (Bio) molecular dynamics simulation (MDS), normal mode analysis (NMA) and Monte Carlo sampling [33,36,58]. Here, we will describe only the MDS and NMA approaches.

MDS is considered to be the method that most rigorously attempts to study receptor flexibility for docking [22,30]. In MDS, the time-dependent behavior of a molecular system is calculated. Integration of the equations of Newton's second law of motion yields a trajectory that describes the positions of the atoms as they vary with time. Although early attempts at using MDS in drug design were restricted to obtaining biophysical insights into protein folding and flexibility [59,60], in recent times it has found many more applications – for exploring receptor flexibility before docking, to include solvent effects, to simulate induced-fit; to refine docked complexes and to calculate binding free energies [22,31]. Pico- and nano-second MDS generate multiple low-energy states of protein, which include minor readjustments in the active-site atoms, and provide a set of MPS for docking a ligand [30,36,53,61]. Predictions from a model are only as good as the inputs used to develop the model. Incomplete knowledge and understanding of the forces determining protein flexibility are liable to hamper the quality of predictions from MDS [30]. The computational power needed to run an MDS is also a major hindrance preventing its routine use. The computationally feasible times of simulation are often unrealistically short (typically pico- to nano-seconds although coarse-grained MDS can reach micro- to milli-second levels) as compared to the known real timescales of protein folding/unfolding events (seconds to minutes). Hence, larger structural changes may be missed.

As an alternative to MDS, researchers have examined the simpler technique of normal mode analysis (NMA) to identify flexible regions in the receptor at a much lower computational cost [61,62]. The eigenvectors of the Hessian matrix of the potential energy are the normal modes. The protein movement is represented as a superposition of normal modes, fluctuating around a minimum energy conformation. Although NMA was pioneered in the early 1980s [61–63], it gained widespread use only after the introduction of coarse-grained representations of the protein [64–66]. A recent systematic comparative study [67] of NMA with MDS has shown that the important space defined by the first most relevant NMA eigenvectors provides a correct picture of the flexibility of proteins in aqueous solution and, hence, NMA can replace MDS for the generation of MPS at a lower computational cost [23]. See Ref. [30] for a comprehensive review of the application of different flavors of NMA and MDS to deal with receptor flexibility before docking.

Other computational methods include the approaches in IFREDA [50] and Glide IfD [19], which involve pilot docking of ligands allowing for small clashes with protein residues and then *in silico* generation of the protein conformations that have adjusted themselves to the ligands.

Hybrid methods

A popular commercial software program, Schrodinger/Glide IfD [19], uses a combination of soft potentials, rotamer exploration and/or active-site mutation to simulate induced-fit effects. The ligand is first docked to the active site of the protein by using Glide with soft potentials, and multiple poses of the ligand are selected. For each of these poses, a shell of 5 Å radius is generated and a super-shell, which is the union of all shells, is created. For each pose, side-chains of the residues to which the ligand has very close contacts are removed and rebuilt by rotamer exploration using optimized procedures in Schrodinger/Prime, keeping the ligand rigid in its position in the active site; this simulates induced-fit effects of ligand on protein. Then, the complex is energy minimized to allow for backbone rearrangements. At the next step, the ligand is re-docked to the optimized protein conformation. This procedure is repeated for all the selected poses, the results are compared and the best-docked pose is selected. A similar combination of multiple ligand pose generation with soft potentials, complex-relaxation to allow receptor to adjust to ligand, multiple dockings of each ligand to selected receptor conformations, selection of best-docked structures and further refinement of the complexes, was earlier used in IFREDA [50]. Hybrid methods have the advantage of taking care of all types of conformational changes. Although they need more computational time and effort than rigid body docking, the cost:benefit ratio is sometimes better than that for multiple dockings.

Molecular dynamics simulations

Some researchers have used MDS after docking to examine the stability of the docked conformation and the strength of binding [22,31], which also allows for receptor and ligand rearrangements to obtain lower energy conformations of the docked complex. Apart from the high computational cost of the simulation, the conformations explored by MDS are dependent on the starting solution provided to the simulator as well as the initial velocity (Table 1). This dependence may be overcome by multiple MDS runs [51], but this puts further strain on computational resources. The solutions arrived at with all this effort may or may not include the most likely conformation. Another approach could be to obtain a good starting solution by QSiCR or a hybrid approach and then carry out MDS.

Discussions

Conformational changes in protein structures are often defined only at the macromolecular level, in terms of changes in backbone conformation, hinge bending movements and other domain rearrangements. B-factors, backbone torsion angles, minimum RMSD between the backbone or C_α atoms, deformation index, among others, are used as measures of these conformational changes [52]. Even small or subtle changes, like side-chain movements and shifts in single residue positions, alter the volume and recognition properties of the binding site and, hence, affect the results of docking and SBDD [33,36,49]. Therefore, computational drug discovery scientists define conformational changes in terms of rotational movements of side-chains, minor shifts at residue levels and also backbone changes. The importance of receptor conformational changes to docking and SBDD is evident from the number of papers that have been published on the subject over the past decade [19–25,29–33,35–37,41–56] and their treatment in general reviews on docking [11,12,31,43]. Some of these methods have also been incorporated into docking software [15,19,21,43,53,54,56].

Most of the MPS methods can be used with computationally generated protein conformations even when no ligand-bound crystal or NMR structure is available, but the cMIP and QSiCR methods require an adequate number of representative cocrystal

structures to be used for training the models. Errors in the crystal structure models of the protein would affect the reliability of these methods. The quality of prediction obtained by QSiCR is highly dependent on the quality of the training set used.

One of the reasons for the failure of structure-based virtual screening (VS) is that it is computationally expensive to allow for protein flexibility during high-throughput docking [57]. The MPS ensemble docking method has a lower computational cost than docking to MPSs, but inaccuracies creep in as a result of the same structure being used for all ligands, irrespective of the individual preferences of each ligand. The QSiCR model for a protein can be applied to a database of ligands to select the best protein conformation for each ligand, partition the database according to the conformation and use these to improve the reliability of VS. This procedure of using different single protein structures for different classes of ligands based on ligand properties is expected to show better accuracy without increasing the computational effort for docking.

Concluding remarks

A major part of the research towards tackling ligand-induced receptor flexibility has focused on docking to MPSs, and many methods have been devised for rational selection of the structures to be used. The main advantage of the approach is that it virtually simulates the process of protein conformation selection by ligand, which is currently believed to be the true natural process and, in some implementations, it can take care of the side-chain and backbone flexibility. The main drawback is the computational effort required for doing multiple dockings and, in some implementations, also for selection of the protein structures to dock to.

The QSiCR approach offers a computationally efficient way of dealing with receptor flexibility for docking, and also provides an estimate of reliability of the prediction. A well-validated model provides a single 'best' conformation of protein to be used for docking a given query ligand. The selection of the single protein structure is determined by the specific ligand's properties, before docking and, hence, saves on the computational effort of multiple dockings and avoids the potential artificial nature of ensemble docking. Although it also provides an insight into the ligand-based determinants of the induced-fit, its applicability is limited to crystallizable proteins with good quality crystal structures that are representative of the range of possible ligand-induced conformational changes.

Using soft potentials in the scoring functions, rather than generating multiple protein conformations for docking, is an option that is being actively pursued because of its computational simplicity. Much research is, however, needed in this direction to resolve issues like accounting for large conformational changes.

All methods have their own merits and shortcomings. A combination of methods may help in improving the reliability of docking to flexible proteins, and hence for SBDD. Some such attempts have been made [19,50] but there is still scope for investigating many other combinations. It would be interesting to find out whether a single ideal combination exists that would work well for all proteins. For VS of large compound libraries, however, computationally simple methods like soft docking, ensemble docking and QSiCR seem to be more suitable. The last word on handling ligand-induced protein conformational changes has yet to be written.

Acknowledgement

We would like to thank E Bharathy for some useful discussions.

References

1 Alvarez, J. and Shoichet, B., eds) (2005) *Virtual Screening In Drug Discovery*, CRC Press
2 Kumar, N. *et al.* (2006) Applying computational modeling to drug discovery and development. *Drug Discov. Today* 11, 806–811
3 Stahl, M. *et al.* (2006) Integrating molecular design resources within modern drug discovery research: the Roche experience. *Drug Discov. Today* 11, 326–333
4 Muster, W. *et al.* (2008) Computational toxicology in drug development. *Drug Discov. Today* 13, 303–310
5 Kuntz, I.D. *et al.* (1982) A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 161, 269–288
6 Krovat, E.M. *et al.* (2005) Recent advances in docking and scoring. *Curr. Comput. Aided Drug Des.* 1, 93–102
7 Joseph-McCarthy, D. *et al.* (2007) Lead optimization via high-throughput molecular docking. *Curr. Opin. Drug Discov. Dev.* 10, 264–274
8 Orry, A.J. *et al.* (2006) Structure-based development of target-specific compound libraries. *Drug Discov. Today* 11, 261–266
9 Recanatini, M. *et al.* (2004) In silico antitarget screening. *Drug Discov. Today* 1, 209–215
10 Du, L. *et al.* (2007) A novel structure-based virtual screening model for the hERG channel blockers. *Biochem. Biophys. Res. Commun.* 355, 889–894
11 Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931
12 Moitessier, N. *et al.* (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* 153, S7–S26
13 Rarey, M. *et al.* (1996) A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489
14 Morris, G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J. Comp. Chem.* 19, 1639–1662
15 Jones, G. *et al.* (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748
16 Friesner, R.A. *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
17 McGann, M.R. *et al.* (2003) Gaussian docking functions. *Biopolymers* 68, 76–90
18 Perola, E. *et al.* (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* 56, 235–249
19 Sherman, W. *et al.* (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* 49, 534–553
20 Jiang, F. and Kim, S.H. (1991) Soft docking: matching of molecular surface cubes. *J. Mol. Biol.* 219, 79–102
21 Claussen, H. *et al.* (2001) FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308, 377–395
22 Alonso, H. *et al.* (2006) Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* 26, 531–568
23 Sander, T. *et al.* (2008) Prediction of the receptor conformation for iGluR2 agonist binding: QM/MM docking to an extensive conformational ensemble generated using normal mode analysis. *J. Mol. Graph. Model.* 26, 1259–1268
24 Subramanian, J. *et al.* (2006) A novel computational analysis of ligand-induced conformational changes in the ATP binding sites of cyclin dependent kinases. *J. Med. Chem.* 49, 5434–5441
25 Subramanian, J. *et al.* (2008) Modeling and selection of flexible proteins for structure-based drug design: backbone and side chain movements in p38 MAPK. *ChemMedChem* 3, 336–344
26 Monod, J. *et al.* (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12, 88–118
27 Tsai, C.J. *et al.* (1999) Folding funnels, binding funnels and protein function. *Protein Sci.* 8, 1181–1190

28 Koshland, D.E. (1958) Applications of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 44, 98–104

29 Xu, Y. *et al.* (2008) Induced-fit or preexisting dynamics? Lessons from protein crystallography and MD simulations on acetylcholinesterase and implications for structure-based drug design. *Protein Sci.* 17, 601–605

30 Marco, E. and Gago, F. (2007) Overcoming the inadequacies or limitations of experimental structures as drug targets using computational modeling tools and molecular dynamics simulations. *ChemMedChem* 2, 1388–1401

31 Carlson, H.A. (2002) Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* 6, 447–452

32 Teague, S.J. (2003) Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* 2, 527–540

33 Barril, X. and Fradera, X. (2006) Incorporating protein flexibility into docking and structure-based drug design. *Expert Opin. Drug Discov.* 1 (4),

34 Najmanovich, R. *et al.* (2000) Side-chain flexibility in proteins upon ligand binding. *Proteins* 39, 261–268

35 Totrov, M. and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* 18, 178–184

36 Cozzini, P. *et al.* (2008) Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* 51, 6237–6255

37 Cavasotto, C.N. *et al.* (2005) The challenge of considering receptor flexibility in ligand docking and virtual screening. *Curr. Comput. Aided Mol. Des.* 1, 423–440

38 McGovern, S.L. and Shoichet, B.K. (2003) Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* 46, 2895–2907

39 Kitchen, D.B. *et al.* (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 3, 935–949

40 Erickson, J.A. *et al.* (2004) Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* 47, 45–55

41 Leach, A.R. (1994) Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235, 345–356

42 Alberts, I.L. *et al.* (2005) Receptor flexibility in de novo ligand design and docking. *J. Med. Chem.* 48, 6585–6596

43 Carlson, H.A. and McCammon, J.A. (2000) Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57, 213–218

44 Sutherland, J.J. *et al.* (2007) Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. *J. Chem. Inform. Model.* 47, 2293–2302

45 Barril, X. and Morley, S.D. (2005) Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* 48, 4432–4433

46 Knegtel, R.M. *et al.* (1997) Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266, 424–440

47 Macchiarulo, A. *et al.* (2004) Ligand selectivity and competition between enzymes in silico. *Nat. Biotechnol.* 22, 1039–1045

48 Zentgraf, M. *et al.* (2006) Addressing protein flexibility and ligand selectivity by ''in situ cross-docking''. *ChemMedChem* 1, 1355–1359

49 Fradera, X. *et al.* (2002) Ligand-induced changes in the binding sites of proteins. *Bioinformatics* 18, 939–948

50 Cavasotto, C.N. and Abagyan, R. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* 337, 209–225

51 Caves, L.S. *et al.* (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* 7, 649–666

52 Nigham, A. and Hsu, D. (2008) Protein conformational flexibility analysis with noisy data. *J. Comp. Biol.* 15, 1–16

53 Amaro, R.E. *et al.* (2008) An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* 22, 693–705

54 Osterberg, F. *et al.* (2002) Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* 46, 34–40

55 Damm, K.L. and Carlson, H.A. (2007) Exploring experimental sources of multiple protein conformations in structure-based drug design. *J. Am. Chem. Soc.* 129, 8225–8235

56 Huang, S.-Y. and Zou, X. (2007) Ensemble docking of multiple protein structures: considering protein structure variations in molecular docking. *Proteins: Struct. Funct. Bioinform.* 66, 399–421

57 Waszkowycz, B. (2008) Towards improving compound selection in structure-based virtual screening.. *Drug Discov. Today* 13, 219–226

58 Jacobs, D.J. *et al.* (2001) Protein flexibility predictions using graph theory. *Proteins: Struct. Funct. Genet.* 44, 150–155

59 McCammon, J.A. *et al.* (1977) Dynamics of folded proteins. *Nature* 267, 585–590

60 Lautz, J. *et al.* (1987) Molecular dynamics simulations of cyclosporin A: the crystal structure and dynamic modeling of a structure in apolar solution based on NMR data. *J. Comput. Aided Mol. Des.* 1, 219–241

61 Brooks, B. and Karplus, M. (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6571–6575

62 Go, N. *et al.* (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. U. S. A.* 80, 3696–3700

63 Levitt, M. *et al.* (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181, 423–447

64 Keskin, O. (2002) Comparison of full-atomic and coarse-grained models to examine the molecular fluctuations of c-AMP dependent protein kinase. *J. Biomol. Struct. Dyn.* 20, 333–345

65 Schuyler, A.D. and Chirikjian, G.S. (2003) Normal mode analysis of proteins: a comparison of rigid cluster modes with C_α coarse graining. *J. Biomol. Struct. Dyn.* 21, 395–405

66 Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *J. Mol. Graph. Model.* 22, 183–193

67 Rueda, M. *et al.* (2007) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* 15, 565–575

68 Zhao, W.-N. *et al.* (2005) Molecular dynamics simulation on complex of HPPK and substrate HP. *J. Mol. Struct. Theochem.* 731, 23–28