Prediction of protein–protein interactions by docking methods

Graham R Smith* and Michael JE Sternberg*†

Recently, developments have been made in predicting the structure of docked complexes when the coordinates of the components are known. The process generally consists of a stage during which the components are combined rigidly and then a refinement stage. Several rapid new algorithms have been introduced in the rigid docking problem and promising refinement techniques have been developed, based on modified molecular mechanics force fields and empirical measures of desolvation, combined with minimisations that switch on the short-range interactions gradually. There has also been progress in developing a benchmark set of targets for docking and a blind trial, similar to the trials of protein structure prediction, has taken place.

Addresses
*Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
†Department of Biological Sciences, Biochemistry Building, Imperial College of Science, Technology & Medicine, London SW7 2AZ, UK; e-mail: m.sternberg@ic.ac.uk

Abbreviations

ACE	atomic contact energy
BD	Brownian dynamics
FFT	fast Fourier transform
GA	genetic algorithm
MD	molecular dynamics
NOE	nuclear Overhauser enhancement
PDB	Protein Data Bank
rmsd	root mean square deviation

Introduction

In this review, we survey recent developments in those techniques that aim to predict the structure of a protein–protein complex in atomic detail, starting from the atomic coordinates, determined by X-ray crystallography or NMR spectroscopy, of the separate components ('unbound') (the related problems of protein–DNA interactions will not be discussed). Protein–protein interactions play a central role in biochemistry, because the formation of the complex has a functional consequence (e.g. signal transduction). The task of predicting the structure of the complex is one that we are not infrequently faced with, because of difficulties in crystallising complexes; of the more than fifteen thousand files in the Protein Data Bank (PDB), many of which are able to form protein–protein complexes, there are only a few hundred nonobligate protein–protein complexes. (Obligate complexes are permanent multimers whose components are not capable of independent existence.) Accurate predictive docking methods could therefore provide substantial structural knowledge about complexes, from which functional information could be inferred or experiments designed to obtain it. Most development has

been, and continues to be, targeted towards those complexes for which the conformational change on docking is fairly small, thus enabling the use of methods that are based on shape and chemical complementarity of the unbound components. This can be considered the 'canonical case' of protein–protein docking and will be the focus of this review.

The character of the interfaces in known protein–protein complexes has been thoroughly investigated in several reviews [1–3]. Whilst obligate complexes do have an interface that is slightly more hydrophobic than the rest of the molecular surface, there is very little indication of this in nonobligate complexes. The subset of complexes for which both components have been crystallised separately unbound provides the main test-bed for developments in protein–protein docking methods. There are about 30 of these complexes and the majority are in the protease–inhibitor or antibody–antigen class. The relative sizes of backbone and sidechain conformational changes on binding have been studied [4], and the protease–inhibitor and antibody–antigen complexes compared [5]. These and many other important classes of complexes, as well as the thermodynamics and kinetics of their formation, are discussed in a recently published book on protein–protein recognition [6•].

There are two parts to the docking problem: developing a scoring function/energy function that can discriminate correctly or near-correctly docked orientations from incorrectly docked ones, and developing a search method that will be able to 'find' a near-correctly docked orientation with reasonable likelihood. The simplest, yet still powerful, scoring function is shape complementarity. To use this, it is necessary to describe the surface shape of the protein. This may be done by discretising the molecule onto a grid in space and considering which cells ('voxels') are occupied, or by using some sort of 'surfacing algorithm', which calculates the solvent-accessible or solvent-excluded surface, and a point set that triangulates it. In carrying out this calculation, many special cases of geometry need to be considered [7,8]. The triangulation reflects the geometry of the surface: surface critical points (extrema), normals and curvature. In [9•], the distinction is drawn between this type of 'shape-explicit' method and 'shape-implicit' methods, which use only the occupied voxels to define the shape of the molecule.

But shape complementarity, although sufficient to recombine the separated components of a known complex, is not usually sufficient to dock unbound components. From a thermodynamic point of view, the native complex is at the global minimum of ΔG, the free energy change (ΔG) of formation of the complex relative to its separated components (usually taken to be unique, though there may, in rare cases, be more than one mode of binding [10]). The calculation

of ΔG is complicated and still the subject of research. As well as the direct electrostatic and van der Waals interactions between the proteins, there are water-mediated interactions; these affect the electrostatics via the dielectric effect and also produce the 'hydrophobic effect', primarily entropic in origin.

The role of electrostatics in protein–protein interactions has been reviewed by Sheinerman *et al.* [11] and Tobias [12], and was explored from a more physical point of view by Elcock *et al.* [13•]. To treat the desolvation of charged groups in the interfaces accurately, it is necessary to solve the full Poisson–Boltzmann equation for each different orientation of the components that is to be examined. This is an expensive calculation and so various approximation schemes are used: these include the 'semi-Coulombic' or 'test charge' approximation, or the use of Coulomb's law with a distance-dependent dielectric, for example, proportional to r. Then, to any of these schemes may also be added an expression for the hydrophobic energy. Though theoretically complicated, it is found empirically to be approximated as a sum of the form $\gamma_i A_i$, where A_i is the solvent-accessible or solvent-excluded area of atom type i [14]. A further developed version of this, parameterised from contacts in protein interiors, is the 'atomic contact energy' or ACE [15]. There are also the dispersion and hard-core interactions, described by standard Lennard–Jones potentials.

Some 'softening' of the energy function is required (a removal of the divergences that electrostatic and Lennard–Jones energies have at r=0), otherwise, even in near-native dockings, these overwhelm the complementarity that remains. However, this softening necessarily reduces the capacity of the energy function to discriminate the correctly docked orientation from incorrect dockings, producing the many false positives that bedevil docking algorithms.

In practice, then, the above considerations frequently lead to a two- or three-stage approach to docking, as outlined in Figure 1. One begins by treating the proteins as rigid bodies, perhaps with some surface softness, searching the comparatively small (six-dimensional) space of relative protein orientations (translational and rotational) and identifying a set of candidate structures using some simple scoring function, with shape complementarity playing a major role. Then these structures are rescored using a more expensive energy function that is better at discriminating near-native orientations. In the third stage, we deal explicitly with a model in full atomic detail (if we did not before) and allow movement of the sidechains and possibly backbone, minimising a (possibly yet more complicated) energy function. The second and third stages may be combined. The energy/score landscape is rough and so it is clearly desirable to make the search as effective as possible by the use of efficient optimisation algorithms [16,17]. If extra biological information about the location of the interface is available, it can also be used as early as possible to simplify the search. Many of these considerations apply

Figure 1
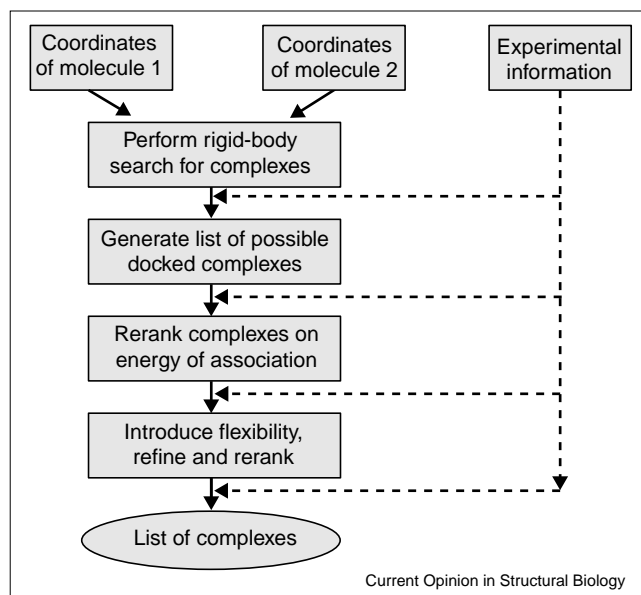
The stages of protein–protein docking.

to methods for docking small-molecule ligands to proteins and any developments will be mentioned if they may be relevant to protein–protein docking.

First stage of docking: global search

One widely used technique in the first stage of docking is the fast Fourier transform (FFT) method. In the FFT method, the molecules are discretised onto a voxel grid. It uses the fact that, if the interaction (scoring function) between the molecules can be put into the form [18]:

$$\Sigma_{ij}\, p_i\, q_j \text{ or } \Sigma_{ij}\, p_i\, G(r_{ij})\, q_j$$

then the score can be evaluated for all relative translations of the molecules in only $O(N^3 \ln N)$ operations (where N is the size of the grid). It is still necessary to search the three-dimensional space of relative orientations, which is usually done exhaustively. The functional forms given allow the approximate evaluation of electrostatic energies (the full Poisson–Boltzmann equation cannot be correctly solved) and shape complementarity using step functions that penalise overlaps between the cores of the molecules, but favour orientations in which the occupied voxels of one molecule lie in a 'surface layer' just outside the other.

This approach was first used in molecular docking by Katchalski-Katzir, Vakser and co-workers [19] and subsequently in the programs FTDock [20], 3D-Dock [21], GRAMM [22] and ZDOCK. It has proved successful in blind trials of protein docking [23]. More recently, it has been used in the program DOT [24•] and in work on identifying binding sites [25]. DOT uses 50 000 relative rotations and a 1 Å grid spacing, though 10 000 rotations

are normally used with FTDock/3D-Dock. In the unbound case, a surface 'softness' of the order of 2 Å can often be used. Very low resolutions may be used and still enable the interface region to be identified if the trial solutions are spatially clustered afterwards [26], though the correct relative orientation of the components is unlikely to be found in this case.

A significant new technique was introduced by Ritchie and Kemp [27••] in their program HEX — an expansion of the molecular surface and electric field in spherical harmonics. Fourier correlations between the expansion coefficients are used to simplify the problem of calculating the complementarity between the surfaces in different orientations to that of look up in a table of precalculated overlap integrals.

Another new method has been developed by Palma *et al.* [28•] in their program BIGGER, a surface-implicit method in which the surfaces are discretised and represented by values of 0 and 1 on two grids, the surface and core grids. These are then combined in real space, but high speed can be achieved by using fast bit-manipulation routines, as a single integer variable can hold many voxels.

Geometric hashing [29] is a shape-explicit algorithm. Surface critical points on one molecule (the ligand) are used to define local coordinate frames, in which the positions of nearby critical points are used as indices in a hash table, which stores the current coordinate frame. The process is then repeated for the other molecule (the receptor), the local coordinate frames are superimposed and the positions of the critical points near to the receptor are used to look up the corresponding ligand reference frames in the hash table of the ligand. A large number of correspondences to a particular frame indicates a strong local similarity in the shape of the surfaces, which is then checked. Again, when docking unbound proteins, a surface 'softness' must be used, corresponding to a 'tolerance' of about 1.5 Å in the coordinates. Searching is replaced by table look up, so the method is very fast. A recent development of this has been provided in [9•].

Two groups have published methods using optimisation of a real-space search with a genetic algorithm (GA) [30•,31•]. In [31•], the protein is represented by a solvent-accessible surface with normals, surface curvature and associated hydrogen-bonding character. The scoring function uses surface area, complementarity of curvature and normals, and penalises overlap of protein interiors. In [30•], a highly physically detailed scoring function, the Charmm molecular mechanics force field, is used from the beginning. In both cases, the 'chromosome' of the GA consists of the relative positions and orientations of the molecules (six 'genes').

Second stage of docking: rescoring
We now have typically a few hundred or thousand docked complexes after the first stage. We find, for example, that,

using FTDock/3D-Dock to dock unbound components, the largest surface complementarity score of an incorrect docking is typically around 1.5–2 times as great as the score of the complex that has the lowest rmsd from the crystallographic complex.

One method of rescoring is to use the statistics of residue–residue contacts across the interfaces of complexes in the PDB to define 'statistical potentials' expressing how much more probable it is that residues will interact than would be expected merely from random contacts between residues with the observed global frequencies of occurrence. The actual contacts in a complex can then be used to rescore it. In [32], such potentials were calculated using a few hundred nonhomologous nonhomodimeric protein–protein interfaces. Results were derived for residue–residue contacts and for atom–atom contacts of different chemical types, with the residue potentials proving the more powerful. Similar methods were used in [33], though with differences in the interfaces and random model, and in [34•], in which specific versions of the potentials were derived for the special case of antibody–antigen binding. Similar calculations are used to derive effective potentials between atom types for ligand–protein docking [35]

Other terms used in rescoring are electrostatics, hydrogen bonding [36], desolvation, lack of buried charges and physical unity of the interface. These were used in [28•,34•] and in the work of Vajda, Camacho and co-workers, who have concentrated not on the initial search, but on algorithms for rescoring complexes [37•] and refining them (see below). The different rescoring terms are commonly applied as a series of sequential filters or, as in [28•], combined with weights found by a neural network.

Third stage of docking: introduction of flexibility
At this stage of docking, the rigid-body approximation is abandoned and flexibility is introduced, at least in the protein sidechains and possibly in the backbone too. A molecular mechanics force field is normally used for the protein; water is still not usually explicitly included, but extra electrostatic/desolvation terms may be included instead.

As sidechain rearrangements dominate many docking problems, a common way to simplify the search is to restrict it to sampling just the known populated rotamers of the sidechains. Even so, the number of combinations of rotamers to be tested is usually too large to tackle exhaustively. One method is to use an iterative mean-field approach [38]. More recently, it has been proposed to prune the list of rotamers with dead-end elimination and then tackle the rest by either a branch-and-cut algorithm or a faster heuristic tree-based algorithm [39•]. Algorithms of this sort are also used in some ligand–protein docking algorithms, such as FlexE [40] and Dock 4.0 [41]. GAs are also well suited to the problem [42]. We remark that programs designed for protein–ligand docking can be used for the

Table 1

Programs for protein–protein docking.

Program*	Algorithm	Laboratory	URL	Details
Downloadable				
3D-Dock	Global: FFT; rescoring: residue potentials; refinement: mean-field sidechain multicopy	Imperial Cancer Research Fund/Imperial college (Sternberg)	www.bmm.icnet.uk/ docking/	Free to academic; distribution mostly source code (C); refinement sgi/linux executables
HEX	Global: Fourier correlation of spherical harmonics	Aberdeen University (Ritchie)	www.biochem.abdn. ac.uk/hex/	Free to academic; sgi/sun/ linux executables
GRAMM	Global: FFT clustering and rescoring decoys also available	SUNY/MUSC (Vakser)	reco3.ams.sunysb. edu/gramm/	Free to academic; sgi/sun/ ibm/dec/linux/win32 executables
PPD	Global: geometric hashing; rescoring: multiple filters	Columbia (Honig)	ftp://flash62.bioc. columbia.edu/pub/other	Free to academic; sgi executables
DOT	Global: FFT for shape complementarity and approximate Poisson–Boltzmann electrostatics	University of California San Diego (Ten Eyck)	www.sdsc.edu/CCMS/DOT	Free to academic; parallelised under mpi; source (C/fortran) and sun/sgi/dec/ibm executables
BIGGER (Chemera)	Global: bit mapping; rescoring: multiple filters	Universidade Nova de Lisboa	www.dq.fct.unl.pt/bioin/ chemera/ www.biotecnol.com/Paginas /Chemera.htm	Free to academic; win32 executables
MERL refinement protocol	Constrained minimisation desolvation	University of Boston (MERL) (Vajda)	engpub1.bu.edu/bioinfo/ MERL/software/decoys.html	Free; Charmm input scripts, source code (f77)
DOCK	Global: grid-based energy function; flexible docking: random search plus incremental construction	University of California San Francisco (Kuntz)	www.cmpharm.ucsf.edu/ kuntz/dock.html	Free to academic; sgi executables
AutoDock	Grid-based empirical potential flexible docking via Monte Carlo search and incremental construction	Scripps Institute (Olson)	www.scripps.edu/pub/ olson-web/download.html	Free to academic; source code and executables for sgi/ dec/sun
FlexX	Fragment assembly energy function: (Boehm potential)	GMD-SCAI (Lengauer)	cartan.gmd.de/flexx/	Licence required from Tripos.com

Program	Algorithm	Laboratory	Contact
Contact authors			
DARWIN	GA Charmm force field	University of Pennsylvania	burnett@wistar.upenn.edu
ZDOCK	FFT for complementarity, electrostatics and residue potential	University of Boston (Weng)	sullivan.bu.edu/~rong/dock/download.shtml

*Programs given in italics are designed principally for protein–ligand docking, but may be usable in the protein–protein case. Programs for MD, BD and continuum electrostatics calculations are also useful.

protein–protein problem (e.g. [43]), though there has been no systematic comparison.

Recent work by Vajda's group [37•,44••] appears to have been a notable advance. An algorithm has been developed that, by concentrating on the reduction of the roughness due to van der Waals interactions and then controlling its re-introduction as docking proceeds, seems to converge reliably from an initial rmsd of 10 Å, which may be commonly produced by docking algorithms, to 2 Å, which is enough for accurate structural studies.

Larger conformational changes than the above mentioned have generally not been tackled successfully by docking. However, the most common known class of larger conformational changes (48 out of 100 entries in the Database of Molecular Movements [45]) is the class of hinge-like interdomain movements. An algorithm capable of tackling this problem was described in [46], but seems not to have been tested beyond initial investigations on the reassembly of bound ligand–protein complexes.

Filtering

By this, we mean the use of biological information to constrain the structure of possible complexes. Though we have deferred discussion until now, it can be used early in docking. Commonly, it is known that some specific residues must lie in the interface, either because of the type of complex (e.g. one of the partners may be an antibody, in which case we know that at least some of the complementarity-determining regions must be involved) or as a result of mutagenesis information. Provision is made
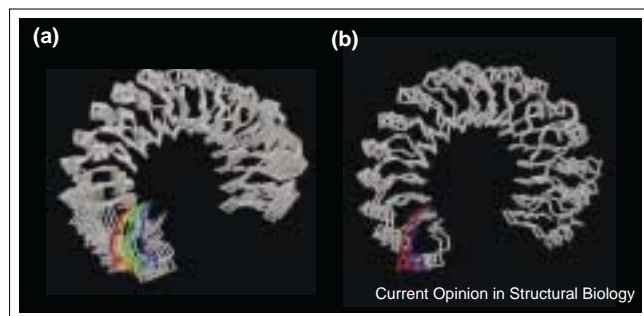
Figure comparing (a) the unbound structure of ribonuclease inhibitor and projections along the lowest eigenvector from principal components analysis of MD simulations with (b) the unbound (red) and bound (blue) X-ray structures of the inhibitor (PDB codes 2bnh and 1dfj, respectively). The coloured region forms the majority of the interface with the ribonuclease.

to use filtering information of this kind in programs such as 3D-Dock and HEX.

It is also possible to use the 'natural mutagenesis' information available from the alignment of homologous proteins to extract similar information; this has been the subject of much recent work, covered in detail in the reviews by Lichtarge and Sowa (pp 21–27) and by DeLano (pp 14–20) in this section. Generally, a cluster ('hot spot') of conserved residues, particularly polar residues, identifies a binding site with some confidence ([47•,48]; reviewed in [49]). Use of this information is aided by the concomitant development of structural alignment programs [50].

An interesting example of detailed constraints for filtering is that which comes from NMR spectroscopy of the complex. Even a small number of nuclear Overhauser enhancements (NOEs) and dipolar couplings (a few tens) suffice to determine the structure of the complex completely to high precision if the unbound components are also known. This is a sufficiently common and useful case that three similar studies have been carried out recently [51•,52•,53], using docking programs modified to take the NMR constraints as additional input. One of these methods [52•] can work with an unassigned ^{1}H spectrum, rather than NOEs, but has not been tested on unbound components.

Many of the programs described above are available to the community for their own research. Because they are computationally intensive, they must be downloaded from the World Wide Web and run on the user's own computer. Table 1 provides a list of many of the available programs.

The physics of docking: kinetics and funnel landscapes

The docking algorithms mentioned above do not generally reflect the physical process of docking. The picture

emerging (see the review by Schreiber [pp 41–47] in this section and also [54,55]) is first of a diffusional search and then the formation of an 'encounter complex', which passes though some intermediate states of increasing desolvation to the docked complex. In the diffusional search, it is well known from kinetics studies that components that have charge complementarity and a strong dipole associate much faster than would be expected from random diffusional encounters. This is most readily visualised simply as a 'funnel' in the spatial and orientational degrees of freedom, directing the incoming component towards the interface that will bind it and orientating it correctly. Desolvation may help here too [56]. A common way to simulate this process is Brownian dynamics (BD) [57] (see the review by Elcock *et al.* [13•] and the forthcoming review in this journal by Gabdoulline and Wade [58]). BD simulations have, in the past, used a simplified model of proteins, but recently full atomistic simulations have been introduced, which, though requiring an empirical definition of when the encounter complex is formed, give very good agreement with experiment ([59•] reviews recent work). Some studies have gone beyond the encounter complex, using surface area and ACE desolvation terms [60•].

As the components approach and sidechains begin to rearrange, the 'binding funnel' changes from an object in low-dimensional (Cartesian) space. It becomes more complex, with more dimensions, corresponding to the flexible sidechain and backbone movements, and more rugged, as desolvation forces and finally van der Waals forces dominate over the smoothly changing electrostatics. The structure of this part of the funnel has been investigated by Zhang *et al.* [43]. This 'binding funnel' has been the subject of several other discussions. In [61], it is predicted that proteins that have a rougher energy landscape will be more promiscuous in their binding. The binding funnel is linked conceptually to the 'folding funnel' of modern protein folding theory via a discussion of disorder/order transitions. The success of the refinement method [44••] described above suggests that this close-range binding funnel remains if the energy function is modified.

In recent work from Vakser's group [62], the spatial distribution and clustering of high-scoring orientations generated by their GRAMM docking program are used to define those proteins for which the docking landscape has a funnel. A majority (about 70%) of the complexes have a funnel-like landscape by this criterion.

Do the conformational changes on docking reflect the conformational flexibility shown by the isolated components? Molecular dynamics (MD)-based work focusing on the sidechain movements of the unbound components of four complexes has been carried out [63•]; it was found that there is some tendency to move towards the conformations found in the complex. We have been engaged in related work (GR Smith, MJE Sternberg, unpublished data) and included small backbone conformational changes. We

performed a principal components analysis of 1 ns duration MD trajectories and compared the eigenvectors of lowest frequency, describing conformational changes, with the known conformational changes on docking, using a sliding window along the protein backbone. We found appreciable overlap in many cases; an example that works particularly well is the ribonuclease inhibitor (PDB code 2bnh). In this case, the inner product and overlap of the first MD eigenvector, and the known conformational change (over an eight-residue moving window) in the interface region of the inhibitor 2bnh is very high at 0.78 (with a peak of 0.99). This indicates that the chains are moving in the same direction, as shown in Figure 2. This result suggests that the conformational changes reflect the natural conformational propensities of the unbound components, for some complexes at least.

Future directions and conclusions

The protocol of fast docking followed by rescoring and refinement is well established, and is effective in many cases of small conformational change upon docking, especially if biological information is available. Without this information, it is still discouragingly difficult to get the near-native complex from unbound components at top rank and we still tend to settle for a good chance of getting one in the top ten (say). New fast methods for the initial docking have been developed and there have been advances in methods for rescoring and refinement. Ideas from the kinetics of binding and the role of binding funnels provide insight, and are already beginning to have a practical effect on docking algorithms. In particular, by emphasising the generation of docked complexes using surface complementarity followed by rescoring, the 'classical' methods may make the refinement stage more difficult by forcing it to be carried out in the part of the search space dominated by many deep local minima due to van der Waals interactions between the components.

Turning from highly accurate and computationally expensive modelling to cases in which speed is more important, the flow of data from structural genomics projects over the next few years will also provide a challenge to docking. Probably, relatively few complexes will be solved, but data about interactions (albeit with large numbers of false positives and negatives) will come from, for example, high-throughput yeast two-hybrid screens. It will be useful to try to make sense of this data by docking the components or homology models of them (as structural genomics should make many globular proteins accessible to homology modelling). However, at the moment, only preliminary work has been carried out even on testing the performance of docking programs when docking homology models of varying degrees of accuracy.

The guidance of docking by the inclusion of the increasing amount of data available on protein–protein interactions (as reviewed in this section), for example, to give information on binding sites, interaction partners and conformational changes, will also be important.

Docking approaches to protein interactions will ultimately need to be placed in the context of the cell, where the promiscuity of interactions, the role of crowding [64] and the occurrence of weak (approximately millimolar) interactions that are nevertheless of great biological importance will need to be addressed. An examination of weak docking interactions using BD has recently been made [65]. The role of water in interfaces [2,66] also requires further investigation, if the fine details of sidechain contacts are to be reproduced. Clearly, however, the lack of explicit water does not prevent a correct prediction in many cases.

At a less speculative level, there has been some agreement on the need for the development of methods for the assessment of current docking procedures — agreed measures of the quality of a complex by rmsd or percentage of native contacts and so on — and the construction of a standard 'test set' of unbound components (see http://sullivan.bu.edu/~rong/dock/benchmark.shtml) and 'decoy sets' (false positives from docking) of various rmsd from the correct structure, but all passing the more obvious requirements of acceptable surface and electrostatic complementarity [67].

An arrangement is also being put in place for blind tests of docking methods when new protein–protein complexes are crystallised, under a similar protocol to the successful CASP (Comparative Assessment of Structure Prediction) assessment exercises (see http://predictioncenter.llnl.gov). The first such trial has recently taken place, on two antibody–viral receptor complexes and one kinase–substrate complex (see http://capri.ebi.ac.uk). Groups took part using many of the protein–protein docking programs described above, as well as protein–ligand docking programs. Trials such as this are essential for the evaluation of the current status of computer methods to predict protein–protein interaction by docking.

Acknowledgements

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.   Jones S, Thornton JM: **Principles of protein-protein interactions.** *Proc Natl Acad Sci USA* 1996, **93**:13-20.

2.   Lo Conte L, Chothia C, Janin J: **The atomic structure of protein–protein recognition sites.** *J Mol Biol* 1998, **285**:2177-2198.

3.   Norel R, Petrey D, Wolfson HJ, Nussinov R: **Examination of shape complementarity in docking of *unbound* proteins.** *Proteins* 1999, **36**:307-317.

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

5.   Jackson RM: **Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem.** *Protein Sci* 1999, **8**:603-613.

6. Kleanthous C: *Protein–Protein Recognition*. Oxford: Oxford University
  • Press; 2000.
This book contains useful information on the thermodynamics and kinetics of protein–protein interaction, analysis of the interfaces and chapters on most of the best-studied families (electron transfer proteins, integrins, antibody–antigens, MHC-TCRs, signal transduction pathways, serine proteases and protease inhibitors, and nucleases and nuclease inhibitors, in order from the most weakly interacting to the strongest).

7. Connolly ML: **Analytical molecular-surface calculation**. *J Appl Crystallogr* 1983, **16**:548-558.

8. Sanner MF, Olson AJ, Spehner JC: **Reduced surface: an efficient way to compute molecular surfaces**. *Biopolymers* 1996, **38**:305-320.

9. Goldman BB, Wipke WT: **QSD quadratic shape descriptors. 2.**
  • **Molecular docking using quadratic shape descriptors (QSDock)**. *Proteins* 2000, **38**:79-94.
In previous versions of geometric hashing, the local coordinate frames were defined using triplets of critical points, or two points and the average normal; these authors use a single point's normal and the two tangential directions associated with the principal curvatures of the molecular surface. The application is to protein–ligand docking.

10. Decanniere K, Transue TR, Desmyter A, Maes D, Muyldermans S, Wyns L: **Degenerate interfaces in antigen-antibody complexes**. *J Mol Biol* 2001, **313**:473-478.

11. Sheinerman FB, Norel R, Honig B: **Electrostatic aspects of protein–protein interactions**. *Curr Opin Struct Biol* 2000, **10**:153-159.

12. Tobias DJ: **Electrostatics calculations: recent methodological advances and applications to membranes**. *Curr Opin Struct Biol* 2001, **11**:253-261.

13. Elcock AH, Sept D, McCammon JA: **Computer simulation of**
  • **protein–protein interactions**. *J Phys Chem B* 2001, **105**:1504-1518.
This paper reviews the calculation of the kinetics and energetics of protein–protein association in detail, with special focus on BD simulations.

14. Jackson RM, Sternberg MJE: **A continuum model for protein-protein interactions: application to the docking problem**. *J Mol Biol* 1995, **250**:258-275.

15. Zhang C, Vasmatzis G, Cornette JL, DeLisi C: **Determination of atomic desolvation energies from the structures of crystallized proteins**. *J Mol Biol* 1997, **267**:707-726.

16. Diller DJ, Verlinde CLMJ: **A critical evaluation of several global optimization algorithms for the purpose of molecular docking**. *J Comput Chem* 1999, **20**:1740-1751.

17. Foreman KW, Phillips AT, Rosen JB, Dill KA: **Comparing search strategies for finding global optima on energy landscapes**. *J Comput Chem* 1999, **20**:1527-1532.

18. Harrison RW, Kourinov IV, Andrews LC: **The Fourier-Greens function and the rapid evaluation of molecular potentials**. *Protein Eng* 1994, **7**:359-369.

19. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA: **Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques**. *Proc Natl Acad Sci USA* 1992, **89**:2195-2199.

20. Gabb HA, Jackson RM, Sternberg MJE: **Modelling protein docking using shape complementarity, electrostatics and biochemical information**. *J Mol Biol* 1997, **272**:106-120.

21. Moont G, Sternberg MJE: **Modelling protein-protein and protein-DNA docking**. In *Bioinformatics - From Genomes to Drugs*, vol 1. Edited by Lengauer T. Weinheim: Wiley-VCH; 2001:361-404.

22. Vakser IA: **Protein docking for low-resolution structures**. *Protein Eng* 1995, **8**:371-377.

23. Strynadka NCJ, Eisenstein M, Katchalski-Katzir E, Shoichet B, Kuntz I, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F *et al*: **Molecular docking programs successfully determine the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase**. *Nat Struct Biol* 1996, **3**:233-238.

24. Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E,
  • Tsigelny I, Ten Eyck LF: **Protein docking using continuum electro-statics and geometric fit**. *Protein Eng* 2001, **14**:105-113.
The authors describe the program DOT, which uses FFTs to calculate surface overlap and electrostatics (semi-Coulombic) during protein docking. The program is parallel and highly optimised.

25. Bliznyuk AA, Gready JE: **Simple method for locating possible ligand binding sites on protein surfaces**. *J Comput Chem* 1999, **20**:983-988.

26. Vakser IA, Matar OG, Lam CF: **A systematic study of low-resolution recognition in protein-protein complexes**. *Proc Natl Acad Sci USA* 1999, **96**:8477-8482.

27. Ritchie DW, Kemp GJL: **Protein docking using spherical polar**
  •• **Fourier correlations**. *Proteins* 2000, **39**:178-194.
The authors present a global search method based on spherical polar Fourier correlations. The method was tested on 18 test cases, of which 9 were unbound–unbound. Between 2 and 7 cases (depending on tightness of filter) found a docking with an rmsd less than 3 Å in the first 10 by rank.

28. Palma PN, Krippahl L, Wampler JE, Moura JJG: **BiGGER: a new**
  • **(soft) docking algorithm for predicting protein interactions**. *Proteins* 2000, **39**:372-384.
Docking was done with a fast bit-mapping algorithm and rescoring using a mixture of surface complementarity, electrostatics (using a softened Coulomb's law), desolvation energy and empirically derived pair potentials (including contacts from the interior of globular proteins, however), combined with weights found by a neural network.

29. Fischer D, Lin SL, Wolfson HJ, Nussinov R: **A suite of molecular docking processes**. *J Mol Biol* 1995, **248**:459-477.

30. Taylor JS, Burnett RM: **DARWIN: A program for docking flexible**
  • **molecules**. *Proteins* 2000, **41**:173-191.
The authors demonstrated docking using a GA and the Charmm force field (with solvent effects treated via a distance-dependent dielectric). It was arranged to run efficiently in parallel. However, it has been tested only on protein–oligosaccharide ligand docking.

31. Gardiner EJ, Willett P, Artymiuk PJ: **Protein docking using a genetic**
  • **algorithm**. *Proteins* 2001, **44**:44-56.
The authors demonstrated docking using a GA. The test set is extensive, including unbound–unbound examples, and in 30 out of 34 cases a near-native solution ranked in the top 100.

32. Moont G, Gabb HA, Sternberg MJE: **Use of pair potentials across protein interfaces in screening predicted docked complexes**. *Proteins* 1999, **35**:364-373.

33. Glaser F, Steinberg DM, Vakser IA, Ben-Tal N: **Residue frequencies and pairing preferences at protein-protein interfaces**. *Proteins* 2001, **43**:89-102.

34. Norel R, Sheinerman F, Petrey D, Honig B: **Electrostatic**
  • **contributions to protein-protein interactions: fast energetic filters for docking and their physical basis**. *Protein Sci* 2001, **10**:2147-2161.
Complexes were initially produced with a geometric hashing algorithm. Several stages of scoring were then performed: a penalty was applied to highly segmented interfaces. Hydrogen-bonding potentials and statistical residue–residue potentials were used, along with a penalty for buried charges that is shown to correlate strongly with exact (not semi-Coulombic) Poisson–Boltzmann calculations. The derivation of the statistical potential reveals interesting distinctions between different classes of complex. Results are excellent for bound cases and encouraging for unbound.

35. Muegge I, Martin YC: **A general and fast scoring function for protein-ligand interactions: a simplified potential approach**. *J Med Chem* 1999, **42**:791-804.

36. Gardiner EJ, Willett P, Artymiuk PJ: **Graph-theoretic techniques for macromolecular docking**. *J Chem Inf Comput Sci* 2000, **40**:273-279.

37. Camacho CJ, Gatchell DW, Kimura SR, Vajda S: **Scoring docked**
  • **conformations generated by rigid-body protein-protein docking**. *Proteins* 2000, **40**:525-537.
The authors present a method for rescoring using successive filters; configurations must score highly both by ACE [15] and by approximate electrostatics, using Coulomb's law with epsilon = 4r. The complexes were then further refined by flexible minimisation in Charmm, using the sum of the bonded interactions, the electrostatic energy and ACE energy, and the van der Waals energy. Good results were obtained for eight test cases, a mixture of serine proteases and antibody–antigens.

38. Jackson RM, Gabb HA, Sternberg MJE: **Rapid refinement of protein interfaces incorporating solvation: application to the docking problem**. *J Mol Biol* 1998, **276**:265-285.

39. Althaus E, Kohlbacher O, Lenhof HP, Muller P: **A combinatorial**
  • **approach to protein docking with flexible side-chains**. In *RECOMB 2000 - Proceedings of the Fourth International Conference on Computational Molecular Biology: 2000 April 8–11; Tokyo*. Edited by ACM. ACM; 2000:15-24.
These workers used an AMBER force field for bonded interactions, a surface area term for hydrophobic energies and the full solution of the Poisson–Boltzmann equation with a nonuniform dielectric of 2 (protein) and 80 (water), and PARSE charges for electrostatic interactions for protein docking. They obtained good results on three unbound serine protease test cases.

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

41. Ewing TJ, Makino S, Skillman AG, Kuntz ID: **DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases**. *J Comput Aided Mol Des* 2001, **15**:411-428.

42. Jones G, Willett P, Glen RC, Leach AR, Taylor R: **Development and validation of a genetic algorithm for flexible docking**. *J Mol Biol* 1997, **267**:727-748.

43. Zhang C, Chen J, DeLisi C: **Protein-protein recognition: exploring the energy funnels near the binding sites**. *Proteins* 1999, **34**:255-267.

44. Camacho CJ, Vajda S: **Protein docking along smooth**
•• **association pathways**. *Proc Natl Acad Sci USA* 2001, **98**:10636-10641.
A very interesting refinement protocol. The components are first separated slightly along the centres of mass, then the complex is refined using the ACE while the van der Waals energy is included, weighted by a scaling factor that is gradually increased from 0 to 1 as the docking proceeds. Small random changes are made, then rigid body and sidechain minimisation performed at each step. Small random steps are taken, then the effective energy minimised by a combination of rigid-body simplex minimisation and sidechain minimisation with the Charmm force field. A small ensemble of about ten structures is refined simultaneously for each target structure, so this is a GA-like scheme. Excellent results are obtained for eight test cases, including both protease–inhibitor and antibody–antigen.

45. Krebs WG, Gerstein M: **The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework**. *Nucleic Acids Res* 2000, **28**:1665-1675.

46. Sandak B, Nussinov R, Wolfson HJ: **A method for biomolecular structural recognition and docking allowing conformational flexibility**. *J Comput Biol* 1998, **5**:631-654.

47. Aloy P, Querol E, Aviles FX, Sternberg MJE: **Automated structure-**
• **based prediction of functional sites in proteins – applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking**. *J Mol Biol* 2001, **311**:395-408.
This paper develops and benchmarks an automated approach to identify which part of a protein is involved in its activity. The approach is applied to predict the region of a protein that is likely to be involved in complex formation. Evaluation shows that this prediction provides a comparable level of reduction in the number of putative docked complexes to that likely to be obtained using biochemical information about one of the components. Thus, one can have a fully automated approach to predict with high discrimination protein–protein (and some protein–DNA) complexes based on sequence and structure without requiring biochemical constraints.

48. Hu Z, Ma B, Wolfson H, Nussinov R: **Conservation of polar residues as hot spots at protein interfaces**. *Proteins* 2000, **39**:331-342.

49. Ma B, Wolfson HJ, Nussinov R: **Protein functional epitopes: hot spots, dynamics and combinatorial libraries**. *Curr Opin Struct Biol* 2001, **11**:364-369.

50. Leibowitz N, Fligelman ZY, Nussinov R, Wolfson HJ: **Automated multiple structure alignment and detection of a common substructural motif**. *Proteins* 2001, **43**:235-245.

51. Clore GM: **Accurate and rapid docking of protein-protein**
• **complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization**. *Proc Natl Acad Sci USA* 2000, **97**:9021-9025.
The authors demonstrate docking using minimisation guided by NMR data. On the (single) unbound–unbound test case, using nine NOEs and dipolar couplings for all residues, and a weak distance constraint on active site residues, a structure only 1.3 Å from the known complex is obtained after flexible sidechain refinement. [53] is similarly successful.

52. Kohlbacher O, Burchardt A, Moll A, Hildebrandt A, Bayer P,
• Lenhof HP: **Structure prediction of protein complexes by an NMR-based protein docking algorithm**. *J Biomol NMR* 2001, **20**:15-21.
This paper describes an NMR-guided docking algorithm that uses an unassigned proton spectrum rather than NOEs; however, it has not been tested on unbound components.

53. Morelli XJ, Palma PN, Guerlesquin F, Rigby AC: **A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data**. *Protein Sci* 2001, **10**:2131-2137.

54. Janin J: **The kinetics of protein-protein recognition**. *Proteins* 1997, **28**:153-161.

55. Schreiber G, Fersht AR: **Rapid, electrostatically assisted association of proteins**. *Nat Struct Biol* 1996, **3**:427-431.

56. Camacho CJ, Weng Z, Vajda S, DeLisi C: **Free energy landscapes of encounter complexes in protein-protein association**. *Biophys J* 1999, **76**:1166-1178.

57. Ermak DL, McCammon JA: **Brownian dynamics with hydrodynamic interactions**. *J Chem Phys* 1978, **69**:1352-1360.

58. Gabdoulline RR, Wade RC: **Biomolecular diffusional association**. *Curr Opin Struct Biol* 2002, **12**:in press.

59. Gabdoulline RR, Wade RC: **Protein-protein association:**
• **investigation of factors influencing association rates by Brownian dynamics simulations**. *J Mol Biol* 2001, **306**:1139-1155.
Results are presented for BD simulations of five protein–protein complexes (two of which have been reported previously) using full atomistic simulations and (a method developed by these workers) effective charges to approximate the effect of the spatially varying dielectric. The effect of ionic strength on association rate is calculated and excellent agreement is obtained with experiment.

60. Camacho CJ, Kimura SR, DeLisi C, Vajda S: **Kinetics of desolvation**
• **mediated protein-protein binding**. *Biophys J* 2000, **78**:1094-1105.
A study of the effect of desolvation on docking in a case in which the components are not strongly charged. BD simulations were carried out with the components fairly close, a distance typical of the 'encounter complex'. Using a hydrophobic energy proportional to the buried surface area leads to non-specific aggregation, but using the ACE energy gave an on-rate within an order of magnitude of experiment.

61. Tsai CJ, Kumar S, Ma BY, Nussinov R: **Folding funnels, binding funnels, and protein function**. *Protein Science* 1999, **8**:1181-1190.

62. Tovchigrechko A, Vakser IA: **How common is the funnel-like energy landscape in protein-protein interactions?** *Protein Sci* 2001, **10**:1572-1583.

63. Kimura SR, Brower RC, Vajda S, Camacho CJ: **Dynamical view of**
• **the positions of key side chains in protein-protein recognition**. *Biophys J* 2001, **80**:635-642.
Six unbound proteins were investigated by MD simulation. It is found that, in three cases, sidechains that lie in the interface in the bound form visit the rotamers that they occupy in the complex during the simulation. However, in only one case does the sidechain spend the majority of its time in the rotamer that it occupies in the complex. The work also introduces a novel simulation technique in which some water is present explicitly and some treated via a continuum method.

64. Ellis RJ: **Macromolecular crowding: obvious but underappreciated**. *Trends Biochem Sci* 2001, **26**:597-604.

65. Elcock AH, McCammon JA: **Calculation of weak protein–protein interactions: the pH dependence of the second virial coefficient**. *Biophys J* 2001, **80**:613-625.

66. Janin J: **Wet and dry interfaces: the role of solvent in protein–protein and protein-DNA recognition**. *Structure Fold Des* 1999, **7**:R277-R279.

67. Gatchell DW, Dennis S, Vajda S: **Discrimination of near-native protein structures from misfolded models by empirical free energy functions**. *Proteins* 2000, **41**:518-534.