Protein–Protein Docking Dealing With the Unknown

IRINA S. MOREIRA, PEDRO A. FERNANDES, MARIA J. RAMOS
REQUIMTE, Departamento de Química, Faculdade de Ciências da Universidade do Porto,
Rua do Campo Alegre 687, Porto 4169-007, Portugal

Received 27 June 2008; Revised 20 February 2009; Accepted 24 February 2009
DOI 10.1002/jcc.21276
Published online 21 May 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Protein–protein binding is one of the critical events in biology, and knowledge of proteic complexes three-dimensional structures is of fundamental importance for the biochemical study of pharmacologic compounds. In the past two decades there was an emergence of a large variety of algorithms designed to predict the structures of protein–protein complexes—a procedure named docking. Computational methods, if accurate and reliable, could play an important role, both to infer functional properties and to guide new experiments. Despite the outstanding progress of the methodologies developed in this area, a few problems still prevent protein–protein docking to be a widespread practice in the structural study of proteins. In this review we focus our attention on the principles that govern docking, namely the algorithms used for searching and scoring, which are usually referred as the docking problem. We also focus our attention on the use of a flexible description of the proteins under study and the use of biological information as the localization of the hot spots, the important residues for protein–protein binding. The most common docking softwares are described too.

© 2009 Wiley Periodicals, Inc. J Comput Chem 31: 317–342, 2010

Key words: protein–protein docking; searching; scoring; flexibility; CAPRI

Introduction

Protein–protein binding is one of the critical events in biology. It is crucial for increasing the knowledge of many biological phenomena, and thus is of supreme significance in pharmaceutical or/and medicinal sciences.[1] In the last years, experimental and theoretical work has been devoted to unravel the principles of protein-protein interactions.[2] It is extremely valuable to obtain structural information for a complete understanding of both the biochemical nature of the process for which the components come together, and to facilitate the design of compounds that might influence it. However, due to the greater difficulty in crystallizing protein–protein complexes, there is relatively little structural information available about them compared to the proteins that exist as single chains or form permanent oligomers.[3] Hence, experimental studies are faced with outstanding technical difficulties and the number of solved complexes deposited in the Protein Data Bank (PDB; www.rcsb.org/pdb) is still orders of magnitude smaller than those of experimental information on protein interactions and of structures of individual proteins.[2] This disparity reflects the fact that it is very difficult to prepare complexes suitable for structural studies and to determine their structures.[2]

Despite the practical difficulties for a better understanding of the biological function of a protein, knowledge of its three-dimensional structure is fundamental.[4] Protein structures have been mainly achieved by two methods so far: X-ray crystallography and nuclear magnetic resonance (NMR). X-ray and NMR encounter difficulties in dealing with structures of complexes. In fact, by X-ray, the dynamics of the complex formation makes the crystallization difficult, while high molecular weight complexes are difficult to deal with NMR.[4] Thus, in the past two decades there was an emergence of a large variety of theoretical algorithms designed to predict the structures of protein–protein and protein–ligand complexes—a procedure named docking.[5] Interest in protein docking is growing within the scientific community, and many interdisciplinary approaches are being applied to model, predict, and understand protein–protein interactions, a major challenge in structural biology.[6]

Protein docking studies, that is, the task of assembling two separate protein components as the ones seen in Figure 1a and b into their biologically relevant complex structure (Figure 1c) are therefore important as an aid to our understanding of the ways in which proteins bind.[7,8] Computational methods, if accurate and reliable, can therefore play an important role, both to infer functional properties and to guide new experiments. So, generating models of molecular complexes is of indisputable significance and may provide additional insight into the nature of macromolecular recognition. It is a demanding problem, which has

Correspondence to: M. J. Ramos; e-mail: mjramos@fc.up.pt

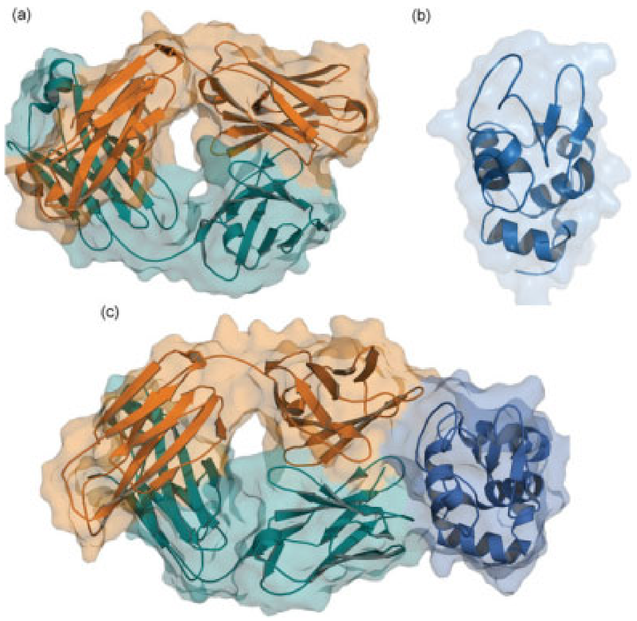

Figure 1. X-ray structure of (a) FAB Hyhel63 antibody (PDBID: 1DQQ), (b) HEW lysozyme (PDBID: 3LZT) and (c) of the complex formed between the two (PDBID: 1DQJ).

attracted a vast deal of attention due to its potential applications in rational drug design and protein engineering.[9,10]

The first protein–protein docking algorithm was developed by Janin and Wodak in 1978.[11] Although protein–protein docking procedures should be a helpful guide for genetic and biochemical experiments, they must first be tested and their validity evaluated.[12,13] Thus, it is required to obtain objective estimates of the model quality and of the performance of docking methods.[14] Albeit important successes, docking procedures remain hampered by the prediction of false positives and negatives.[15] Because of the complexity of the problem, protein–protein docking is still largely at the theoretical stage and there is still considerable scope for the development of methodology.[16]

The objective of protein–protein docking is to predict the three-dimensional arrangement of a protein–protein complex from the coordinates of its component molecules, being an accurate prediction the one that will point out most of the residue-residue contacts involved in the target interaction.[17–23] Usually, this involves an exhaustive search of the rotational and translational space of one protein with respect to the other, resulting in a six-dimensional search. One of the most important difficulties of protein docking is that the interface residues of both the receptor and the ligand may undergo a conformational change on complex formation. Although often the conformational change is limited to side-chains, large backbone movements are sometimes also observed.[24] To develop protein–protein docking algorithms, a perfect test case should be formed by the unbound three-dimensional structures of both the receptor and the ligand, as well as by the complex structure that is used only for assessing the algorithm performance. The PDB contains only a limited number of such test cases. In Figure 2 we can observe the number of X-ray structures deposited in the PDB since 1976. As we can see, 6458 new structures became available last year. Although every year more and more structures are becoming available, the Protein–Protein Docking Benchmark,[25] consists of only 124 cases for which high-resolution crystal structures are available in both the unbound and bound states. The number of the years between the deposition of the first of the three struc-

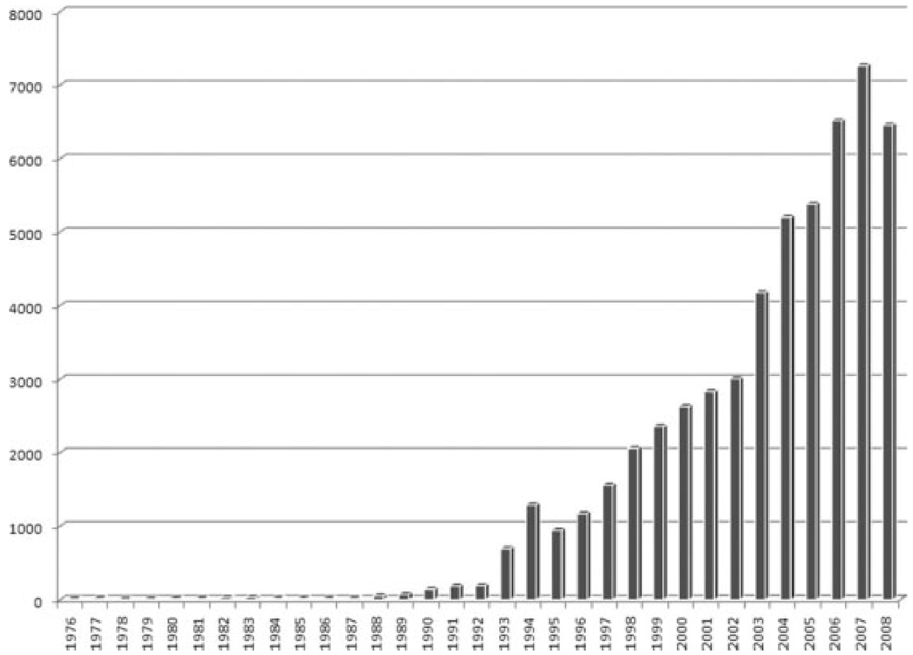

Figure 2. Number of X-ray structures deposited in PDB per year.

tures of the unbound-unbound test case and the last can vary between 0 and 4 years but unfortunately, values such as 19 are still found. We can just hope that this situation will be altered in the next few years. Despite the fact that unbound-unbound constitute a perfect test case, it is also adequate if only one unbound structure is available, to use an unbound-bound test case, in which the bound structure of the other molecule is used as it appears in the complex and the other was crystallized as a free protein.

Hence, there are three key ingredients in the docking: representation of the system, conformational space search, and ranking of potential solutions.[21] Although these can vary the protein–protein docking contains certain problems common to all procedures: "searching and scoring."[16] The first refers to how accurately the energy function of a given protein–protein complex is described, and the latter is concerned with obtaining the global minimum energy structure of the complex using that same energy function.[21]

Searching

Nowadays, a number of programs perform "*ab initio*" protein–protein docking using the same approach: one protein is fixed in space and the second one is rotated and translated around the first one. The disadvantage of these methods is that the search through the entire conformational space of the complex geometry makes the calculation expensive, and therefore it is important to consider the best orientation of their side chains leading to the minimum energy and the best side chain contacts.[4] Thus, a simple systematic search is usually impracticable even if the molecules are treated as rigid bodies, in which the degrees of freedom are limited to translational and rotational ones because the searching algorithm entails evaluating in the order of billions (10^9) of distinct possibilities. More elaborate search techniques are required and should be both accurate and efficient.[16]

The docking method is generally based on the idea of complementarity between the interacting molecules, which may be geometric, electrostatic or hydrophobic, or all three. Geometric complementarity of the protein surfaces is the filtering criterion most commonly used to eliminate a large number of solutions with poor surface matching.[16,20] Although usually both proteins are treated as rigid bodies, since to allow flexibility may be computationally tricky, many shape-based docking algorithms have been proposed. These can include other kind of information such as hydrophobicity, electrostatics, which are used in combination with shape matching or as a following filter.[26] Thus, the complexity of computational docking increases in the following order: rigid body docking (extremely basic model that considers the two proteins as two rigid solid bodies), semiflexible docking (one of the molecules, typically the smaller ligand, is the only one considered flexible), flexible docking (both molecules are considered flexible, while obviously the degree of flexibility of either or of both is unavoidably limited or simplified).[21,22] In protein–protein docking different methods have been used to search the conformational space: matching the position of surface spheres and surface normals,[27] application of real space,[28] or Fourier correlation techniques such as the fast fourier transform (FFT) algorithm, introduced by Katchalski-Katzir,[29] which evaluates the surface-surface contacts between the candidate proteins while penalizing for protrusions into the protein core.[29–33] A correlation approach using spherical polar basis functions is also possible.[33]

Other algorithms use matching of surface cubes[34] or geometric hashing, i.e. the identification and matching of convex and concave protein surface regions.[35–40] Genetic algorithms,[41] Brownian dynamics simulations,[42] and combinations of Brownian dynamics and energy minimization[43] have also been applied to deal with the protein docking problem.[44]

Representation of the System

The basic description of the protein surface is the atomic representation of exposed residues, which can be usually achieved by mathematical models, such as geometrical shape descriptors or a grid.[21] Geometric shape descriptors as sphere representations of amino acids, negative sphere images of the binding site, molecular surface cubes, surface normals at sparse critical points, and cross-sectional slices represented as polygons are widely used.[1] One of the most commons ways is to represent the surface by its geometric features such as the Connolly surface, which consists of the part of the van der Waals surface of the atoms that is accessible to the probe sphere (contact surface) connected by a network of convex, concave, and saddle shape surfaces that smoothes over the crevices and pits between the atoms.[21]

The geometric descriptors could partly be associated with further properties of physicochemical meaning or as an alternative grid representations may be used for the macromolecular structure. This commonly refers to affinity grids, which are calculated on the basis of force field potentials for van der Waals and electrostatic interactions.[1] The protein interior, the surface and the outer space can be differentiated by the use of grid-based molecular representations in combination with Fourier correlation algorithms. Albeit the representation of the macromolecular target frequently involves similar descriptive method for the ligand, the ligand may be also treated in a fully atomic detail. For this it is essential to use the energy grid techniques. A few alternatives of the grid approach also allow an atomic representation of the binding site and represent only the bulk of the receptor protein as a grid.[1]

Biological Information

Biological information available from experiments or from computational methods on the likely regions or residues involved in the interaction can confine the search of allowed complex configurations or filter out wrong solutions.[21,45] Protein–protein binding site identification can be achieved by potential hydrogen bonding groups, enzyme clefts and charged sites on a protein surface as well as structural comparisons with molecules with known binding sites. Since binding sites are at least partially flexible, searches for part-flexible part-rigid sites have also produced hopeful results. Algorithms that predict the location of hinges and modes of motions, or those that carry out structural comparisons of the protein family are also very helpful.[21,45]

Most docking studies focus on enzyme-inhibitor complexes and antibody-antigen complexes since they present significant differences in the interface residue composition, hydrophobicity and electrostatics.[46–50] For example, the catalytic triad of the serine proteases (His, Asp, Ser) and the complementarity defining regions of immunoglobulins are both well characterized.[51] In general, a protease-inhibitor interface is more static and consequently more easily predicted than an antibody-antigen interface.[26,51]

Hot spots may also be incorporated in the scoring process. Hot spots have been defined as those sites where alanine mutations cause a significant increase in the binding free energy of at least 2.0 kcal/mol.[48,52,53] To have a strong impact in protein binding the binding free energy should be higher than 4 kcal/mol (three orders of magnitude in the binding affinity constant). However, residues whose mutation results in such a large difference are quite unusual, and the threshold for the hot spots had to be lowered to 2 kcal/mol in order to get enough data for statistical analysis.[54,55] Therefore, in a protein–protein interface, a small subset of the buried amino acids typically contributes to the majority of binding affinity as determined by the change in the free energy of binding ($\Delta\Delta G_{\text{binding}}$) upon mutation of the residue to an alanine.[56,57] It has been observed that hot spots are preferentially located either on protrusions ("knobs") or in depressions ("holes") of the protein surfaces and they are coupled across the interface in tight fitting regions that exclude solvent molecules.[43] Interestingly, hot spot residues appear to undergo little conformational changes upon binding, a property that might facilitate their identification in the unbound state.[45] Nevertheless there are no general rules to predict a binding interface. Therefore, machine learning techniques are being used to predict automatically interfaces using a combination of various factors: e.g. buried surface areas, desolvation and electrostatic interaction energies, hydrophobicity scores, and residue conservation scores.[58–66] A variety of different techniques are now being explored such as the evolutionary trace approach that exploits the fact that functionally important residues are often conserved across species,[67] sequence-base approaches that locate correlated mutations in multiple sequence alignments for pairs of interacting proteins across different organisms,[68] NMR data such as chemical shift perturbations and residual dipolar couplings are expressed in terms of ambiguous interactions restrains by the software HADDOCK.[4]

As domain interactions frequently determine protein function, an understanding of how domains combine and assemble is clearly necessary. So, a related problem of considerable importance is domain docking meaning the prediction of the structure of a multidomain protein from the structures of its component domains.[45]

Flexibility

Structures of protein complexes reveal intricate shape complementarity, which seemingly confirms the initial lock-and-key description of protein interaction, first introduced by Fischer in 1894.[69] However, this model did not take into account the conformational changes that can occur upon binding. Daniel Koshland's (1958)[70] induced fit model admits a certain degree of plasticity and postulates a mutual adaptation of the two proteins. A range of studies including molecular, NMR, and single-molecule FRET experiments, have questioned the extent to which conformational change can be considered induced by the binding partner. Therefore, a third model, inspired by the MWC mechanism of allosteric regulation (Monod et al., 1965)[71] and later adapted to protein interactions by Kumar et al., a statistical mechanics emerged recently as a result of view of protein binding.[72–74] This model of conformational selection, postulates that proteins exist as ensembles of conformations, and that the binding of the partner to a specific conformation shifts the equilibrium toward the specific binding conformation.[75] A fourth model is a hybrid conformer selection/induced fit (CS/IF) model proposed by Grunberg et al.[76] where binding is a two-stage process that begins with conformational selection to form an encounter complex, followed by an IF or 'refolding' step that leads to the final bound conformation. The different conformations can present various degrees of flexibility, which may consist of subtle side-chain adjustments or may involve domain, tertiary, and/or secondary structure motions.[77]

In protein–protein docking, because of the large number of atoms and degrees of configurational and conformational freedom involved, it would be impracticable to treat molecular flexibility in an explicit way with the current computers available.[44] This situation contrasts with small ligand-protein docking in which usually the binding site is already known, and due to the restricted nature of the problem and the small size of the ligand, the flexibility can be taken into account later.[78] Although easily treated, flexibility still constitutes a major challenge in protein–ligand docking due to the computational time needed[79] Usually, protein docking methods are based on rigid- or semirigid-body treatment of the molecules that reduces the proteic complex to a nonflexible structure resulting in a radical simplification of the search process. However, this procedure can lead to wrong solutions especially if the interacting molecular structures acquire different conformations as they are in the unbound form or complexed to the other protein.[80] For example, in the case of antibodies, a diversity of phenomena have been observed, ranging from adjustments of single amino acid side chains, over loop rearrangements, to entire domain movements.[1] The main defy in computational protein–protein docking is to merge high-accuracy energy calculations, speed and sampling power, and the ability to handle induced conformational changes at the interface.[81]

Since it is infeasible to explore all possible conformations, protein flexibility is introduced into docking protocols in a variety of ways being the most common procedure the use of "soft" scoring functions that accommodate flexibility while others explicitly include domain hinging movements or side-chain flexibility in the docking.[3] Both backbone and side-chain flexibility are also being introduced using molecular dynamics (MD) in combination with some form of rigid-body docking, either before or after the MD simulations.[3,15] Thus, as it is not feasible to execute extensive conformational searches during docking, unless the binding site is known, it has been generally adopted the two-stage approach. Initially the receptor and ligand are treated as rigid bodies and a fully exploration of the six-dimensional rotational and translational degrees of freedom is made. At a second stage, a much smaller number of structures acquired in the initial stage are refined and reranked by more scrupulous energy functions that include small backbone and side-chain movements as well as rigid-body adjustments to take into account conformational changes.[82] As it is difficult to perform the three at the same time they can be done sequentially.[83] A number of algorithms have been developed for this purpose. We will focus in some of the most commonly used.

Soft Docking

In the majority of the known protein complexes, the complex structure of the proteins is only very vaguely changed, compared to the free forms, with most of the conformational changes confined to the side chain atoms of surface amino acids. Therefore, with slight modifications of the basic rigid body procedure, some techniques may tolerate a limited degree of molecular flexibility by using a "soft" representation of the molecular surface.[84–89] The soft docking concept, originally proposed by Jiang and Kim, describes the molecular surface and volume as a cube representation, which implies implicit conformational changes by way of size/shape complementarity, close packing and liberal steric overlap.[84] Ritchie and Kemp introduced a 'soft' model of electrostatic complementarity in the algorithm.[89] Palma et al. proposed a surface implicit method in which the surface is represented by values 0 and 1 on two grids, the surface and core grids.[31,88] This digitization introduces the first level of 'softness' in the algorithm.[80,90]

Therefore, soft docking methods generally use rigid-body docking and smooth the protein surfaces or allow some degree of interpenetration.[79] They only deal with side-chain flexibility and can be divided into brute force techniques,[80,84] randomized methods[7,91] and shape complementarity methods.[92]

Side-Chain Flexibility

Side-chain flexibility can permit in favorable cases an efficient docking if some interfacial side-chains are in incorrect conformations.[8] Totrov et al., 1994 published one of the first successful *ab initio* predictions of a complex that combined pseudo Brownian Monte Carlo minimization with a biased-probability global side-chain placement procedure. They showed that side-chain optimization was fundamental for discrimination of near-native conformations from false positives.[93,94] The majority of the methods adjust side chain conformations explicitly during a refinement stage following the rigid-body search, which is characteristically performed only for a selected set of protein side chains close to the putative binding site and side chain conformations are represented as a discrete set of rotamers from libraries. These libraries are derived from statistical analysis of side-chain conformations in known high-resolution protein structures.[83] Originally it was proposed a method that uses pre-generated side-chain rotamer libraries, which are subjected to optimization during a ligand docking procedure via the dead-end elimination algorithm, and subsequently scored in order to rank the lowest energy combination of side-chain and ligand conformers.[95] Side-chain flexibility was introduced in the Mining Minima optimizer, which allows conformations of user-selected side-chains in the active site to be optimized along with the conformation and position of the ligand.[96,97] Some algorithms are also utilized such as SOFTSPOTS that makes use of a knowledge-based function, which identifies active site residues most likely to undergo conformational change upon ligand binding. The associated PLASTIC algorithm generates the side-chain rotamers, or a minimal conformational manifold, prior to docking calculations.[8,77]

All the amino acids do not show the same degree of freedom. Amongst the protein complexes arginine, lysine, glutamate, and metionine preset the highest frequency and amplitude of movements between the structures of free and co-crystallized proteins.[80] In contrast, many of the smaller polar or charged residues, such as asparagine, aspartate and histidine, and the large aromatics, phenylalanine, tyrosine and tryptophan, are markedly inflexible. At the binding interface, the degree of flexibility follows the order lysine > arginine > methionine > glutamine > glutamate > isoleucine > leucine > aspargine > threonine > tyrosine > serine > histidine > aspartate > cysteine > tryptophane > phenylalanine. Thus, the lysine side chains flex 25 times more often than do phenylalanine side chains.[98–100]

Backbone Flexibility

As already mention, a comparison of bound and unbound structures can reveal significant changes in backbone conformation upon binding, which represents the greatest challenge to predictive protein docking. The main challenge is the incorporation of full backbone flexibility into protein-protein docking simulations, due to the enormous complexity of size and degrees of freedom of the conformational space of the backbone. Backbone flexibility can be treated at different stages of the docking procedure and in different combinations.[83] Sometimes, ensembles of conformations or backbone minimization are obtained prior to the docking procedure because backbone deformation is more important to the global structure than side-chain deformation and because side-chain conformations may depend of the backbone torsional angles. As reviewed by Andrusier et al., 2008[83] the flexible docking can be divided into three parts: (i) analysis of flexibility of proteins by ensemble analysis, rigidity theory, Gaussian network model, normal mode analysis, MD or essential dynamics; (ii) flexible docking in which subdomains can be docked separately and ensembles can be generated and docked using cross-docking or the Mean Field Approach; (iii) refinement of backbone (by normal modes minimization), side-chains (mean field approach, iterative elimination, graph theory algorithms)and rigid-body orientation (by a variety of minimization methods).

The various biophysical models suggest distinct conformational sampling strategies in flexible protein docking. For example, the key-lock model successful at predicting complexes for proteins that undergo minimal backbone conformational change upon binding, is the underlying idea for the original grid-based and FFT techniques, and is included, among others, in ZDOCK/RDOCK, ClusPro and RosettaDock.[101] In Koshland's induced fit model the backbone conformational space must be sampled explicitly during docking in response to local energetics of the interface, which can be achieved by MD, energy minimization, or gradient-based methods in MC minimization. In the conformational selection model for docking, backbone flexibility is modeled implicitly as a pregenerated ensemble of rigid structures generated from the unbound structure. The ensembles can be achieved by using different solved three-dimensional structures from X-ray or NMR studies of diverse conformations of the same protein. From a computational point of view, depending on the time scales and the energy barrier heights, MD and Monte Carlo simulations that generate full receptor ensembles have also been used to incorporate protein flexibility in docking.[19,76] However, MD can be used only to model small-scale movements, in a nanosecond time scale, and sometimes the simulations can result in unrealistically docked substrates.[67,83] This can be overcome by restricting the degrees of freedom to the torsional space and by using simulated annealing methods and scaling/rescaling ligand-receptor interaction potential methods. Multiple-copy simultaneous search methods can help to speed up energy-based searches because they use numerous ligand copies, which are transparent to each other but subjected to the full influence of the protein. The docking algorithm based on the hybrid CS/IF model would have combined both ensemble docking and explicit backbone flexibility during docking. For example, HADDOCK combines both implicit and explicit backbone flexibilities while incorporating biochemical information, and it is capable of using ensembles from a wide variety of sources including MD, homology modeling, or NMR structures.[4]

Normal-mode analysis can also be used to calculate the normal modes (a set of basis vectors) that are related with the flexibility of the protein and therefore may be used to modeled large global motions.[101] Two of the most common models used for normal modes calculations are the Anisotropic Network Model and the Gaussian Network Model. Dobbins et al. used normal modes, which are related to the thermal motion of individual proteins, to gain insights into the nature of conformational changes in protein–protein interactions by identifying the mobile regions and reproducing the directions of conformational changes. Only backbone motion is studied because the model used to calculate the normal modes considers Cα atoms only.[75] Principal component analysis of MD trajectories is also being used to generate conformations for docking by capturing the main flexible degrees of freedom of a protein.[102]

In life we can find multiple domain proteins, which are predominant in eukaryotic proteomes and have been shown to play a critical role in their structure and function.[103–105] Therefore, it is crucial to have the right computational tools to construct these complex macromolecular structures even if biophysical data is not available. Multimeric docking algorithms are being usually developed by several groups (Table 5) in which some degree of symmetry is applied to candidate dimmers, rejecting those that produce intolerable steric clashes.[23,106–110] During complex formation, flexible segments that separate domains of multidomain proteins or subdomains of proteins can move—hinge-bending motions. There are some methods capable of dealing with this type of flexibility. For example, the FlexDock algorithm from the group that developed PatchDock, is initiated by the automatic detection of the hinges by the HingeProt algorithm, an NMA-based approach for the identification of hinge regions and rigid parts given a single protein structure.[111] Eisenstein et al. also incorporate domain movement in their software Molfit in which individual domains of the molecules are treated as soft rigid objects in multibody, multistage docking protocols.[112]

Unfortunately, even with some recent improvements, the treatment of flexibility is still a major problem for the docking community as it can be seen by the CAPRI results, where in cases with significant conformational changes the predictions were disappointing.

Modeling Interfacial Water

Even though an O-ring hypothesis[113] states that the majority of the interfaces are generally occluded from the solvent, still many interfaces present bound water molecules. Therefore, although solvation and desolvation effects are crucially important in the thermodynamics of complex formation, most docking algorithms neglect to take into account the specific interactions with water molecules that occur in some interfaces. Recently, the HADDOCK software has been introducing explicit solvent in the calculation in which the starting structure is solvated with a 5.5 Å solvent shell in a short MD run. This protocol is a very promising methodology that result in considerably better scores and RMS deviations than unsolvated docking for the majority of the 10 complexes studied, which included examples of both wet and dry interfaces.[4]

Scoring

Generally speaking, the protein–protein docking problem can be classified as one of the global optimization problems, since its key principle is to evaluate the energies of protein–protein docking poses so as to identify the pose with the lowest energy as the predicted binding mode.[20] So, the fundamental point of any docking method is to be computationally efficient, having a scoring scheme able of evaluating a huge number of solutions and discriminating the correct binding modes from the decoy complex structures.[21] With the development of the Fourier correlation approach,[29] it became computationally practicable to generate and evaluate billions of possible docked conformations by simple scoring functions.

The geometry of the complex corresponding to the lowest free energy of binding must be found, which it is not easy to do considering the macromolecular nature of the protein, with its high dimensionality of the coordinate space and considerable complexity of the energetics governing the interaction.[1] In the most favorable case, the best prediction is reasonably close to the crystal structure, but none of the docking procedures achieves this on all test complexes.[13] As docking methods, force fields, and degrees of refinement vary widely, the goal of having a single scoring function for every model regardless of its source might not be easily accomplished.[114–117] This way, the docking process should be able to discriminate between native-like and wrong docked structures within a reasonable computation time.[21,115]

Most of the docking algorithms developed so far use the extent of geometric complementarity of the protein surfaces as an initial filter to eliminate a large number of solutions with poor surface matching. It is, however, usually recognized that a criterion based exclusively on geometric complementarity is far-off from being enough to distinguish among native and non-native docked geometries, except for a very a small number of cases.[28] Numerous criteria have been implemented with different levels of success: steric complementarity of the shapes of the interaction sites, electrostatic interactions, and hydrogen bonding. Furthermore, exclusion of the solvent from the interface and the associated solvent entropy change play an important role in stabilization of protein interactions, and can be estimated from empirical potentials or data base derived functions.[2,13] These scoring parameters can be divided into two groups: collective parameters (refer to a property that characterizes the entire molecule) and individual parameters (refers to a specific atom or residue).[18] However, despite the method used for predicting the most favorable interaction mode, none of the individual energetic contributions that have been evaluated proves to be sufficient, *per se*, to distinguish between native and misdocked structures for all tested complexes.[18] The most natural scoring scheme, the free energy of binding $\Delta G_{\mathrm{binding}}$, is not easily accessible but other scoring functions that model $\Delta G_{\mathrm{binding}}$ as accurately as possible, i.e. provide good correlations with experimental binding affinities, must be used.[16] Some scoring functions involve solvation potentials, empirical atom-atom or residue-residue contact energies, and continuum electrostatics. However, the empirical free energy and the molecular mechanics potential alone cannot provide a valid discrimination of the true solution because molecular mechanics potential is just part of the binding free energy, and the entropy is not taken into account. MMPBSA (Molecular Mechanics Poisson Boltzmann Surface Area), or the free energy perturbation method may thus be used to discriminate the correct docking structure.[20] Hence, scoring functions can be roughly classified into three distinct categories: knowledge-based, empirical and forcefield-based.[77] Docking algorithms can be classified by the phase of scoring in the algorithm flow into two groups: integrated and edge functions. In integrated algorithms, scoring is integrated into the search stage and filter emerging solutions. In edge algorithms, scoring is applied at the end of the search stage. The major difference is that the scoring function forms part of the design of the solutions in integrated algorithms but not in edge algorithms.[21]

The final score can be determined with regard to other solutions, a known structure, or the solution itself. A solution with a low rmsd from the complex is considered to be the correct one. Usually, other solutions that differ only slightly from it should be found, and therefore, a comparison of each solution with other solutions, by a direct comparison or by clustering (dense populations of predicted conformations), is beneficial. In principle, the cluster size may also be used as a parameter in a scoring function.[21] Therefore, scoring functions can include: heuristic scores based on residue contacts, shape complementarity of molecular surfaces ("stereochemistry"), free energies, phylogenetic desirability of the interacting regions, and clustering coefficients. It is usual to combine one or more categories above in a weighted sum whose weights are optimized on cases from the benchmark. The benchmark cases used to optimize the weights must not overlap with the cases used to make the final test of the score to avoid bias.

Recent developments in scoring functions have focused on the estimates of geometric and energetic complementarity during rigid-body search using terms derived from statistical analysis of known interfaces or Machine Learning techniques,[117] or more sophisticated force fields used in subsequent refinement steps.[118] Bernauer et al., 2007 developed a scoring function based on a Voronoi tessellation of the protein three-dimensional structures, which provides a convenient low-resolution description of protein structure and protein–protein interfaces.[119] Using the ROGER statistical learning procedure, they ranked the models obtained by two different docking algorithms, improving in almost all cases the rank of native-like solutions.[119] Nevertheless, scoring continues to be one of the biggest challenges in the docking procedure.

Critical Assessment of Prediction of Interactions

The international Critical Assessment of Prediction of Interactions (CAPRI) experiment was designed to evaluate current computational approaches that address protein–protein docking.[120] The CAPRI is a community-wide experiment designed according to the model of the Critical Assessment of Techniques for Protein Structure Prediction (CASP).[14] It was designed in June 2001 at the Conference on Modeling Protein Interactions in Genomes organized in Charleston, SC, by Ilya Vakser (Medical University of South Carolina) and Sandor Vajda (Boston University). Unlike CASP, which has a fixed time schedule and targets are single proteins, CAPRI targets are protein–protein complexes and it is data-driven, meaning that starts whenever an experimentalist offers an adequate target and ends 6–8 weeks later with the submission of predicted structures.[14,121,122] Computational researchers are given the three-dimensional coordinates of the unbound structures before the co-crystallized complexes are published. The researchers are then given a few weeks to dock the two structures together, and can use biological information and literature searches. In just a couple of years, CAPRI challenge has provided the docking community with a unique blind setting of simultaneously assessing all docking algorithms, and has led to significant advances in the field.[123,124]

Therefore, the CAPRI experiment concentrates on a relatively small number of prediction targets, with unknown three-dimensional structures and all the groups participating in the experiment study the same targets. This way, comparison of the performance of different algorithms becomes easier because the selection of targets, which may be harder or easier for prediction, is eliminated and the bias, which is naturally introduced when the predictor knows the expected results is also eradicated.[9] One of the most imperative ambitions of CAPRI is to comprehend how well present methods can unravel unbound-unbound docking problems, which proved to be far from trivial.[125] The participants can use any biochemical or structural information available in the literature, which can facilitate the search toward correct solutions even for targets that would be very difficult to predict without such additional information.[125]

The CAPRI protein–protein targets present in Table 1 provided so far a good combination of problems with diverse levels of complexity. T16 and T18 had to be cancelled, and predictions

Table 1. CAPRI Protein–Protein Targets.

Round	Target	Complex	Type of Complex	Reference
1	T01	HPr kinase/HPr	Unbound-unbound	126
	T02	Rotavirus VP6/Fab	Unbound-bound	127
	T03	Flu hemagglutinin/Fab	Unbound-bound	128
2	T04	Amylase/camel VHH	Unbound-bound	129
	T05	Amylase/camel VHH	Unbound-bound	129
	T06	Amylase/camel VHH	Unbound-bound	129
	T07	Superantigen/TCRb	Unbound-unbound	130
3	T08	Nidogen/laminin	Unbound-bound	131
	T09	LicT dimer	Unbound-unbound	132
4	T10	TBE virus E trimer	Unbound-unbound	133
	T11	Cohesin/dockerin (unbound)	Unbound-homology model	134
	T12	Cohesin/dockerin (bound)	Unbound-bound	134
	T13	SAG1/Fab	Unbound-bound	135
5	T14	Phosphatase 1d/MYPTI	Homology model-bound	136
	T18	Xylanase/TAXI	Unbound-bound	137
	T19	Ovine prion/Fab	Homology model-bound	138
6	T20	PrmC/RF1	Unbound-homology model	139
7	T21	Orc1/Sir1	Unbound-unbound	140
8	T22	U5-15K/U5-52K	Unbound-unbound	141
	T23	GBP1 GTPase domain	Homodimer	142
9	T24	Arf1-GTP/ArfBD (unbound)	Unbound-homology model	143
	T25	Arf1-GTP/ArfBD (bound)	Unbound-bound	143
10	T26	TolB/Pal	Unbound-unbound	144
11	T27	E2-25KDA/UBC9	Unbound-unbound	145
12	T28	NEDD4L dimer	Homology model	146
13	T29	TRM8-TRM82	Homology model	147
14	T30	Rnd1-GTP/RBD dimer	Unbound/unbound	148
15	T31	Rac1-GTP/RBD monomer	Unbound/unbound	149
	T32	Subtilisin Savinase/Alpha-amylase	Unbound/unbound	150
	T33	Rlma2 methyltransferase/RNA complex	Homology model/unbound	
	T34	Rlma2 methyltransferase /RNA complex	Homology model/bound	
	T35	Pair of covalently linked modules: CBM22 and GH10	Homology model/homology model	
	T36	Pair of covalently linked modules: CBM22 and GH10	Homology model/bound	

of T15 were interrupted prematurely, as the corresponding experimental structure was published by the authors before the deadline for submitting the predictions. T22 and T23 were also canceled except for the docking servers. In order of preference, the ideal would be to start with the coordinates of two unbound molecules or with one bound and one unbound set of coordinates, and finally if only the bound coordinates are available, to start with the coordinates of the backbone.[125] In CAPRI the unbound-unbound targets appear to be intrinsically more difficult than the bound-unbound targets because, as already mentioned, molecules suffer conformational changes from the unbound to the bound structures. There are three types of conformational changes: involving small-scale, fast motions, involving large-scale, slow domain motions, and the third outcoming of protein "disorder." In such cases, the native state has a small hydrophobic core, or the molecule (or its disordered domain) contains uncompensated buried charges.[125]

To assess the quality of the models, after a least-square superimposition of the receptor in the model and target, three aspects should be analyzed: the RMS distance L_{rms} between Cα atoms of the ligand (L) in the model and target, as well as the rotation angle θ_{L} and translation d_{L} needed to further superimpose L; the interface RMS distance I_{rms}, calculated with the Cα's of the epitopes only; and the fraction of native contacts $f_{\mathrm{nc}} = n_{\mathrm{c}}/N_{\mathrm{c}}$, where N_{c} is the number of residue pairs in contact in the target, and n_{c} the number of those native contacts that are present in the model. These parameters I_{rms}, L_{rms}, and f_{nc} are then combined to rank models. In models of the "high-quality" and "medium" categories, f_{nc} is higher than 0.3, I_{rms} is lower than 2 Å, L_{rms} is lower than 5.0 Å and the misorientation is lower than 20° depending on the shape and size of the molecules and the interface. Models with 10–30% of the native contact pairs and I_{rms} between 2 Å and 4 Å, are placed in the "acceptable" category. Although their geometry is poor, they should still be useful for site-directed mutagenesis and other experiments, because a large part of the epitopes must be correctly identified to yield $f_{\mathrm{nc}} \geq 0.1$.[12] Table 2 summarizes the criteria available for ranking the CAPRI predictions.[13]

Vajda et al.[125] have demonstrated that the best predictors of success in docking are the conformational change upon binding,

Table 2. Criteria for Ranking CAPRI Predictions.

Rank	f_{nc}	L_{rms}	Or I_{rms}
High	≥ 0.5	≤ 1.0	Or ≤ 1.0
Medium	≥ 0.3	1.0–5.0	Or 1.0–2.0
Acceptable	≥ 0.1	5.0–10.0	Or 2.0–4.0
Incorrect	<0.1		

the change in the solvent accessible surface area, and the hydrophobicity of the interface. They have calculated the change in the solvent accessible surface area (ΔASA) by $\Delta\text{ASA} = \text{ASA}_{\text{complex}} - \text{ASA}_{\text{receptor}} - \text{ASA}_{\text{ligand}}$, where ASA denotes the solvent accessible surface area of the proteins in the CAPRI contest. They have measured also the hydrophobicity of the interface as the free energy of desolvation upon association ΔG_{des}, calculated using the atomic contact potential (ACP),[150] an atom-level extension of the Miyazawa and Jernigan potential[151] as well as the $C\alpha$ RMSD, the α-carbon root mean square deviation between free and bound proteins. With this study they have divided the protein–protein complexes in five different types present in Table 3.[125]

Contrasting with small molecule docking, which has become a custom computational instrument in rational drug design, protein–protein docking has remained mainly an academic exercise.[152] A communitywide blind prediction helps to prove the value of the prediction methods and assess their reliability, before transferring the technology to a wider circle of users.[13] The CAPRI challenge expose the work of the structural biologists who submit targets and offers the opportunity to evaluate and compare different methods and protocols for protein–protein docking. The results of the CAPRI rounds have shown that there is a single method capable of docking each and every complex, although acceptable predictions are made for most complexes.[9,153]

The Software

The comparison of different docking programs and to rank their relative performance is very important but extremely difficult because most investigators have access to merely a partial number of methods for evaluation, the test cases tend to be limited in the number and type of targets evaluated, and each algorithm merges a particular search strategy and a particular scoring function.[154,155] The algorithms mostly differ in the method for searching the six-dimensional-transformation space that they apply and in their evaluation of the resolved complexes, and are computationally too expensive for large-scale experiments.[23] As different algorithms may perform better for different types of complexes, a methodical exploration of all algorithms may reveal directions of enhancement.[51] It is essential to stress that recent progress in docking algorithms and computer hardware makes it possible to implement such procedures as automated Web servers, which greatly improves the utility of the docking approaches in the biological community.[156] In this part of our review we will focus in some softwares most frequently used to perform protein–protein docking.

Attract

Attract is based on energy minimization of the binding proteins and a reduced protein model (consisting of up to three pseudoatoms per amino acid residue) to allow systematic docking with several thousands of initial configurations.[8] The interaction between amino acids in the reduced model considers dissimilarities in physicochemical character of the side-chains such that complex formation is driven not only by surface complementarity but also by the physicochemical character of the interacting protein surfaces. The use of a reduced model allows a much faster calculation of protein–protein interactions, the number of distinct docking minima is much smaller compared with an atomic detail representation of the protein surfaces. During docking, both partners are considered as rigid, and a systematic docking process consisting of a series of minimizations is performed. The first three minimizations include a harmonic distance restraint between the two partners in order to generate an initial tight complex, followed by free minimization towards the closest minimum configuration.[44,157]

Flexibility of the partner structures is taken into account by representing flexible surface side-chains as multiple conformational copies.[8,44,157,158] Experimental data and knowledge of hot spots can be taken into account at various stages of the docking procedure.[159] Usually the presence of flexible surface loops, which must adapt to the steric and electrostatic properties of a partner, presents a major obstacle. Attract allows large loop movements during a systematic exploration of the possible arrangements of the two partners in terms of position and rotation by taking into account an ensemble of possible loop conformations by a multicopy representation within a reduced protein model.[122] Thus, Attract is based on a reduced protein model and energy minimization of docked complexes. The program

Table 3. Classification of Proteic Complexes on the Basis of Docking Difficulty.

Type	ΔASA (Å^2)	ΔG_{des} (kcal/mol)	$C\alpha$ RMSD	Expected difficulty of docking
I	1400–2000	<-4.0		Easy, unless key side-chains are in the wrong conformations
II	2000–3000		<2.0	Moderated difficulty
III	1400–2000	>-4.0		Very difficult almost unpredictable
IV	<1400			Very difficult
V	>2500		>2.0	Rigid-body methods always seem to fail

includes side chain flexibility (by a multicopy approach) and global flexibility using normal mode variables during systematic docking runs (in contrast to most programs that include flexibility only in a postprocessing or refinement step of rigidly docked complexes).

BIGGER

BIGGER is a software that incorporates a soft-docking algorithm to predict three-dimensional-structures of proteic complexes. First a three-dimensional matrix (volume matrix) composed of small cubic cells of 1 Å size, which represents the complex shape of each molecule is generated. Then, a grid-like search algorithm is used to systematically search the binding space of both molecules, and a structure is selected on the basis of the geometric complementarity and amino acid pairwise affinities between the two molecular surfaces. No information about the binding site is used, and so all the proteic surfaces are subjected to a detailed search. The scoring function comprise terms related with the geometric complementarity of the surfaces, explicit electrostatic interactions, desolvation energy, and pairwise propensities of the amino acid side chains to contact across the molecular interface.[28,56]

ClusPro

The web server, ClusPro,[160] includes a rigid-body search, a rapid estimation of the knowledge-based potentials, such as the Atomic Contact Potential and electrostatic energies for filtering, a ranking based on the clustering properties of low free energy complexes, and a brief side-chain minimization using CHARMm to remove clashes from the docked interface. The current version includes two FFT-based docking programs, DOT[161] and ZDOCK[82] as its front-end performing a rigid-body search.[116] DOT runs retain 20,000 docked conformations, while ZDOCK runs retained 2000 structures. Both DOT and ZDOCK use FFT but DOT uses a shape complementarity score, whereas ZDOCK scoring function includes a combination of shape complementarity, Coulombic electrostatics, and desolvation free energy based on the Zhang et al.[150,162] atomic contact potential.[58,123] Through the introduction of the Fourier correlation method, it is now possible to evaluate billions of putative complex structures.[164] The algorithm filters the docked conformations by selecting the ones with favorable desolvation and electrostatics properties, clusters the retained structures using a hierarchical pairwise RMSD algorithm, and selects the centers of the most populated clusters as predictions of the unknown complex. The free energy filters select complexes with lowest desolvation and electrostatic energies.[116,124] It is possible to prohibit binding to certain regions as well as select residues that should be close in the proteic complex.[163]

The server has also been further developed to allow modeling of multimeric assemblies. Given the number of monomers forming a multimeric complex and the structure of one monomer, the method predicts the symmetry and structure of the complex. The method was designed to scan all possible interactions, and select the models with the broadest free energy funnels that also satisfy the symmetry constraints without steric overlaps.[123]

FastContact' is a server (http://structure.pitt.edu/servers/fastcontact/) that estimates the direct electrostatic and desolvation component of the free energy, based on a classic distance dependent dielectric and an empirical contact potential for the desolvation contribution. This server was used in conjugation with Cluspro in the latest targets of the CAPRI contest, and discriminated near-native predictions from docked conformations.[164]

Recently, some of the original authors of this software developed a new docking program PIPER based on the FFT correlation approach. The PIPER program was used with a new class of structure-based potentials called DARS (Decoys As the Reference State),[78] based on the inverse Boltzmann approach. The structures of the top-ranked clusters are further refined using SDU, a stochastic global optimization method, which exploits and utilizes the funnel-like behavior of the free energy function ΔG in the regions of the conformational space defined by separate clusters.[165,166]

3D-Dock

The 3D-Dock algorithm[167] is initiated by a global scan of translational and rotational space of possible positions of the two binding partners, limited by surface complementarity and an electrostatic filter using a standard FFT to search for putative conformations.[168–170] The scoring criterion is based on shape complementarity but also incorporates an approximate electrostatics scoring. The protein is described at the atomic level by electrostatic and van der Waals interactions. Multiple copy representation according to a rotamer library on a fixed peptide backbone is used to model side-chain conformations. The biological filter screens the conformational space of the complexes by restricting the distance between different groups of residue of the two interfaces.[3,170,171] An energy minimization is made to remove steric clashes on the side-chains of the interface, and thus it models the effects of side-chain conformational change and the rigid-body movement of the interacting proteins during refinement. Thus, the program uses a rotamer library to reposition the side-chains; it refines the position of the backbone in a deterministic manner toward local energy minima, and provides an interaction energy value that we use as a score to rank the putative conformations.[3,172] At the end, rigid-body energy minimization is performed to relax the interface.[3] Use of biological information as distance constraints to filter complexes appears to be essential generating three-dimensional-Dock predictions with an RMSD lower than 1 Å.[168]

Recently, Sternberg et al. have created a server: three-dimensional-Garden, Global And Restrained Docking Exploration Nexus (http://www.sbg.bio.ic.ac.uk/∼3dgarden), which uses an ensemble generation based on the marching cubes algorithm[173] with a conformational refinement engine to form a comprehensive framework for flexible docking. Three-dimensional-Garden's scoring procedure is more basic than most other docking protocols since it includes no clustering step and no explicit solvation term.[174]

Bates et al. uses a modified version of the FT_Dock (a component of the three-dimensional-Dock). The main difference is the sampling process that also includes MD, which allows the consideration of backbone flexibility. It also provides a side-chain multicopy flexible refinement force field, and it uses an orientation filter to treat oligomerization symmetry.[3]

DOT

DOT performs an FFT systematic search of basically rigid molecules, over usually a 1 Å translation step and a 4° rotation step. It can permit a limited amount of presumed flexibility by allowing a certain number of atoms to collide in a particular run. Symmetry can be treated by reducing the conformational search, which can be achieved by limiting the set of rotations to be consistent with such symmetry. The DOT energy model consists of an electrostatic component and a nonbonded contact term, but not explicit desolvation. Thus, DOT evaluates the energy of interaction for many orientations of the moving molecule by summing the Boltzmann factor and the van der Waals contribution over all rotations at each grid point.[161,175]

The DOT authors participated in the CAPRI contest with a methodology that involves starting from 100,000 DOT results, a geometric complementarity analysis performed for the top 1500 DOT results by using the FADE program, proximity analysis by a distance cutoff filter and FADE to determine the amount of contact and shape complementarity between regions of interest, and cluster analysis performed on the top 1000 results returned by DOT. Finally manual inspection using computer graphics or three-dimensional printing excluded those that were clearly not possible due to unfavorable charge-charge interactions. Those with more favorable interactions, including charge-charge, hydrogen bonding, nonpolar-nonpolar are then selected.[161,176,177]

Gramm-X

The Gramm-X implements a rigid-body fine-grid FFT search with a smoothed Lennard-Jones potential to accommodate conformational changes, followed by local minimization and rescoring. The resulting local minima are re-ranked according to the weighted sum of Lennard-Jones potential, pairwise residue-residue statistical preferences, cluster occupancy, and the degree of the evolutionary conservation of the predicted interface.[156] Flexibility was not taken into account in the CAPRI contest. To take into account the degree of change at the interface area, bound conformations of the most flexible interface side chains were substituted into the unbound structures from the benchmark complexes. It was demonstrated that the knowledge of the conformational change upon binding of only three critical interface side chains per complex would provide a 40% improvement of the benchmarking results, and beyond that other factors such as backbone changes and force field accuracy would dominate.[179] In predictive docking experiments the interface side-chains and the extent of their conformational change upon binding are not know but could be estimated by rotamer library analysis or force field simulation (Monte Carlo or MD). Biological information such as interacting residue data from mutational studies and general knowledge about the interaction being studied is not used during the docking algorithm. Thus, the proteic complex should be manually inspected.[178] Dockground[179] project (http://dockground.bioinformatics.ku.edu) focuses on generation of comprehensive and sophisticated datasets for developing and validating new docking methodologies. One of the key aspects are unbound (experimental and simulated) protein structures, corresponding to complexes of known structure, which are much more limited than those for the bound sets, because only a limited number of proteins are crystallized in both bound and unbound form.[181]

HADDOCK

HADDOCK starts with a randomization of orientations and rigid body energy minimization, followed by semirigid simulated annealing in torsion angle space, and final refinement in Cartesian space with explicit solvent. During the last two phases, the amino acids at the interface (side chains and backbone) are allowed to move to optimize the interface packing.[4] Biochemical and/or biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments or mutagenesis data information about protein-protein interface residues is often used to reduce the conformational search space or filter the solutions. One of the most fundamental differences in comparison with other algorithms is that HADDOCK translates information about the interface into highly ambiguous intermolecular distance restraints used to directly drive the docking process.[4,181]

Flexibility is introduced at several levels in the algorithm: by docking from ensembles of structures and taking all possible pairwise combinations; by introduction of flexibility in the side chain at the interface; and by allowing both side-chain and backbone flexibility in the final refinement stage. Flexibility also considerably improved the ranking of structures. In addition to the explicit inclusion of flexibility in the refinement stage, it is also implicitly included in the rigid-body docking stage by starting from ensembles of structures obtained from short MD simulations in explicit solvent.[182] The final structures are clustered using the pairwise backbone RMSD at the interface and analyzed according to their average interaction energies (sum of electrostatic, van der Waals and Ambiguous Interaction Restraints that are derived from any kind of experimental information available concerning the residues involved in the intermolecular interaction) and their average buried surface area.[4] Recently, explicit inclusion of interfacial water was incorporated in the docking protocol and incorporated in the latest CAPRI predictions.[183] A consistent improvement is observed for Fnat after water refinement, casing an aceeptable prediciton to become a medium prediciton.[184] Recently this group lauched a new web server that can found at http://haddock.chem.uu.nl/Haddock/haddockserver-file.php.

ICM-DISCO

The ICM-DISCO protein–protein docking method is a direct stochastic global energy optimization from multiple starting positions of the ligand, which shows the ability to distinguish near-native rigid-body geometries in a relatively low number of alternative docking poses.[6,185] The rigid-body uses ICM pseudo-Brownian[185] optimization of binding potentials precalculated on a three-dimensional grid: van der Waals potential, an electrostatic potential corrected for the solvation effect, the hydrogen-bonding potential, and a hydrophobicity potential. The solvation energy based on atomic solvent-accessible surfaces was added to the total energy to re-evaluate the docking solutions obtained from unbound subunits. It is followed by a refinement of the solutions and final scoring, which includes specific filtering criterion on a case-by-case basis. A new desolvation descriptor, based on atomic solvation parameters (ASPs) derived from octanol-water transfer experiments, was optimized for rigid-body docking. Symmetry is imposed as an intrinsic feature of the model, so that only symmetric configurations can be realized throughout the course of the docking procedure.[6] The algorithm handles the induced changes of surface side-chains but is less successful if the backbone undergoes large scale rearrangements.[185,186] Recently some of the developers of ICM-DISCO created a new protocol called pyDock, based on FFT generation of rigid-body docking solutions, with a scoring function consisting on electrostatics and desolvation energy terms.[187]

Molfit

Molfit[188] starts by a weighted-geometric search, in which contacts involving specified parts of the surfaces of either one or both molecules are up-weighted or down-weighted, and in which the whole rotation-translation space is scanned (global scan) or a part of it (partial scan). The molecules are represented by three-dimensional grids that carry information on the shape and the chemical character of the molecular surfaces. The grids are correlated using FFT.[189] The weights are based on available structural and biochemical data or on sequence analyses. In addition, a geometric scan should be performed to get an estimate of the outcome of weighing. The solutions in each scan are sorted by their complementarity scores.[72] The top ranking solutions from each scan are filtered, clustered, and manually analyzed. The manual viewing serves to eliminate severe clashes and to estimate qualitatively the possibility of ion pair and hydrogen bond formation across the interface. At the end of the procedure, the best solutions are refined by small rigid body rotations of 2° around the position obtained in the scan.[72,188–190]

PatchDock

PatchDock[23,191] is a geometry-based molecular docking algorithm, which divides the Connolly dot surface representation of the molecules into concave, convex and flat patches and matches complementary patches in order to generate candidate transformations. Small-scale flexibility is taken into account implicitly by allowing some extent of steric clashes.[191,192] Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy. So, candidates are ranked according to a geometric shape complementarity score, where surface contact is scored positively and "acceptable" steric clashes are penalized.[191] Finally, an RMSD clustering is applied to the candidate solutions to remove superfluous solutions. PatchDock does a fast transformational search, which is driven by local feature matching and utilizes advanced data structures and spatial pattern detection techniques, such as geometric hashing and pose clustering.[23] PatchDock enables integration of external information concerning potential binding sites such as restricting the matching stage to patches that include residues important for binding.[192]

Although PatchDock does not perform side-chain refinement, recently a new server was launched by the same authors: FireDock (http://bioinfo3d.cs.tau.ac.il/FireDock), which includes optimization of side-chain conformations and rigid-body orientation, and allows performing a high-throughput refinement.[193]

RosettaDock

RosettaDock uses real-space Monte Carlo minimization (MCM) on both rigid-body and side-chain degrees of freedom to identify the lowest free energy of the docked proteic complex.[91] After a low-resolution search, explicit side chains are added to the protein backbones using a backbone-dependent rotamer packing algorithm. The sampling problem is attacked with supercomputing clusters to create very large numbers of decoys, which are discriminated using a scoring function including van der Waals and solvation interactions, hydrogen bonding, residue-residue pair statistics, and rotamer probabilities. Decoys are then ranked, clustered, manually inspected, and selected. Algorithm convergence, as measured by solution degeneracy after decoy clustering, is used as a final criterion in decoy selection.[74,91] Recent modifications of the protocol have improved side-chain modeling by enhancing side-chain conformational sampling through gradient based off-rotamer optimization first introduced by Abagyan et al.,[194] and also by including information from the unbound structures.[147] Prediction are usually performed without including any *a priori* biological information, being the energy of a model the primary criterion for the selection of the submissions. However, in some cases, biological information constraints can be used.[91] To select the final models, the largest clusters and best scoring decoys can be examined manually to look for special features such as specific contacts (i.e., close contacts, hydrogen bonds, or hydrophobic packing), chemical environment (exposed hydrophobic groups or buried polar groups), overall fit (size and shape of interface or the presence of voids at the interface), and general arrangement (the number of complementarity determining region loops interacting with the antigen).[91]

Baker uses a modified version of the RosettaDock that includes an additional local refinement of models, the energy filters were transformed into target-specific at each step, resulting in maximal enrichment of low-energy models in the global run.[195] The increased sampling of side-chain conformations is achieved through an additional step that includes off-rotamer, gradient-based minimization (RTMIN: Rotamer Trial with Minimization in torsion space). Additionally, side-chain conformations of the free monomers are added to the rotamers from the backbone dependent library. For the prediction of the structure of homomultimers, there is a search for the optimal conformation within the space of symmetric conformations. The homomultimer is created from symmetry operations based on the ordinates, ensuring full sampling of possible symmetric conformations.[196]

Recently, a RosettaDock server (http://rosettadock.graylab.jhu.edu) was developed, which allows the identification of low-energy conformations of a protein–protein interaction, near a given starting configuration, by optimizing rigid-body orientation and side-chain conformations.[197]

SKE-DOCK

Initially this server used the benzene cluster (BC) fitting as a searching method but, as the results of earlier CAPRI rounds showed that it is impossible to obtain a fully correct docking structure, if the docking structures generated by BC fitting proved to be wrong, that searching method would be changed by a geometric docking method.[198] Two of the main advantages of the geometric docking, which superimposes a pair of quadrangular pyramids representing the local shape feature of the receptor or ligand, are its speed or alternatively the number of samples processed. To remove side-chain clashes it was used the automated homology modeling program FAMS, which is based on database searches for homologous structures and simulated annealing energy minimization, and includes main chain adjustments. They use a knowledge-based scoring function to calculate the model quality from the side-chain environment of each amino acid residue from three parameters: the fraction of the molecular surface area of the side-chain covered by the polar atom, the fraction of the side-chain area buried by some other atoms, and the secondary structure.[199,200]

SmoothDock

SmoothDock is an algorithm that includes four steps: rigid body docking using the FFT-based program DOT,[162] reranking of the structures according to a free energy estimate that includes both desolvation and electrostatics, filtering of the complexes using a pairwise RMSD criterion, and subjection of the 25 largest clusters to a smooth docking discrimination algorithm described in Camacho and Vajda[201] where van der Waals forces are taken into account.[120] Usually no constrains on the binding area are imposed.[120] As an alternative method for refinement, Camacho and Vajda were able to exploit the distance dependencies of various energy functions to optimize the geometric position of the ligand with respect to the receptor. They have shown that the electrostatic component dominates the energy equation while the proteins are far apart, guiding the correct interfaces towards each other. As the proteins approach each other, the desolvation component plays a larger role in the energetics of interaction. Finally, once the interfaces become fully desolvated, the van der Waals energy plays a much more dominant role. They consider these distant-dependent properties of protein–protein binding and linearly increase the vdW contribution to binding with respect to the electrostatic and desolvation components of the free energy as the software progresses. This refinement has the capability of refining clusters of protein complexes from 10 Angstroms (Å) from the native complex to fewer than 4 Å from the native. Lastly, each cluster is refined by linearly varying the weights of the van der Waals forces, atomic contact potential (ACP), and electrostatic components to the free energy of binding.[40]

ZDOCK

ZDOCK is a rigid body FFT based algorithm that combines shape complementarity, desolvation, and electrostatics. It searches the rotational space explicitly being the translational space searched by using an FFT algorithm.[82,124] ZDOCK should be use in combination with RDOCK that is an energy minimization algorithm for refining and reranking ZDOCK results.[90,202–204] Biological information should be used on some level to discourage contacts between certain residues in the ZDOCK predictions. Blocking of the chosen atoms of the residues may be done.[202–204] ZDOCK is usually used in conjugation with RDOCK. RDOCK is implemented as a protocol in CHARMm involving the following three steps: remove clashes that occur from the soft-shape complementarity parameter in ZDOCK that allows for small conformational change; optimize polar interactions; and optimize charge interactions. The key factor of RDOCK is a three stage energy minimization scheme, followed by the evaluation of electrostatic and desolvation energies. RDOCK represents a simple approach toward refining unbound docking predictions.[202–204] After the use of RDOCK the predictions are rescored using both electrostatics and desolvation terms, and the new scores are use to re-rank the top ZDOCK predictions.[82,90,202] Thus, the method used by the authors include: ZDOCK, RDOCK, clustering of the top predictions after RDOCK to reduce structural redundancy, contact filtering and manual inspection.[82,90,202,204]

As the RDOCK minimization step can be lengthy and its success is limited by the number of near-native structures produced by ZDOCK, the authors developed a new program: ZRANK (Zlab Rerank) that quickly reranks the rigid body docking results from ZDOCK.[205,206]

Table 4 resumes all the crucial characteristics of the softwares described earlier. We focus essentially in the algorithm used for searching and scoring, the use of biological information, flexibility and symmetry. Nowadays the majority of the methods include a step of side-chain modeling. RosettaDock[91] as well as ICM[185] and the Bates three-dimensional-Dock[3] group uses more than one strategy to handle side-chain flexibility such as rotamer library, energy minimization, pseudo-Brownian Monte Carlo minimization or multiple copy refinement techniques. HADDOCK[4] uses simulated annealing starting from several side-chain conformations for each residues and Attract[44] incorporates side-chain flexibility at the docking step. Backbone flexibility is very difficult to treat. Usually a global refinement step is introduced that enables only small backbone adjustments. HADDOCK[4] and the version of three-dimensional-Dock[3] of the Bates group allow backbone flexibility being able of producing larger structural deformations than the first ones. Molfit[188] and PatchDock[23,185] can handle conformational changes of any sizes such as the ones involving movements of whole domains. A large number of the softwares introduced earlier have produced a way for docking identical subunits into symmetrical assemblies. We can also observe that there are a large number of methods for the scoring of the results, which can use different combination of terms such as shape complementarity, van der Waals, Coulomb and desolvation terms, rotamer probabilities, contact pair potentials or knowledge-based potentials. Knowledge of the binding site is very important to guide the protein–protein docking procedure, and biological and structural information is becoming widely used as experimental restrains to guide the search or in order

Table 4. Protein–Protein Docking Softwares Characteristics.

Software	Searching	Filtering at searching stage	Flexibility at searching stage	Flexibility at refinement	Symmetry	Scoring at sampling stage	Scoring at refinement stage	References
Attract	Series of minimizations with the ligand center being placed at regular positions around the receptor surface at a distance slightly larger than its biggest radius	–	Side chain flexibility (by a multicopy approach) and global flexibility using normal mode variables	–	–	van der Waals, hydrophobic potential	van der Waals and electrostatics	44
BIGGER	Binary grids	Geometric complementarity, pair-wise amino-acid contacts	Soft docking: arginine, lysine, aspartate, glutamate, and metionine flexibility	–	–	Surface matching, side chain contacts, electrostatics, and solvation energy	Global scoring function: surface matching, side chain contacts, electrostatics, and solvation energy	28
ClusPro	Rigid-body FFT docking	Filtering by empirical potentials: desolvation and electrostatics properties. Use of anchor residues to identify hot spots	–	–	Homo-*N*-mer generation by symmetry	Shape complementarity, electrostatics and desolvation energy	Shape complementarity, electrostatics and desolvation energy	160
3D-DocK	Rigid-body FFT docking	Biological structural information filtering: 3D conservation analysis to identify interfacial residues	Flexible refinement: mean-field and side-chain multicopy; representative MD structures	Side-chain optimizations and backbone refinement	–	Shape complementarity and electrostatics	Electrostatics, contact pair potential, van der Waals potential, and Langevin dipole solvation	167
DOT	Rigid body FFT search	–	–	–	Orientation filtering	Electrostatics and shape complementarity	Scoring using the sum of the Boltzmann factor and the van der Waals contribution.	161
Gramm-X	Rigid-body FFT search	–	–	–	–	A fine-grid projection of a softened Lennard-Jones potential function	Soft Lennard-Jones potential, evolutionary conservation of predicted interface, statistical residue–residue preference, volume of the minimum, empirical binding free energy and atomic contact energy	178

Table 4. (*Continued*).

Software	Searching	Filtering at searching stage	Flexibility at searching stage	Flexibility at refinement	Symmetry	Scoring at sampling stage	Scoring at refinement stage	References
HADDOCK	Randomization of orientations and rigid body energy minimization follow by semi-rigid simulated annealing in torsion angle space, and finally refinement in Cartesian space with explicit solvent.	CSP and RDC with additional NMR information such as diffusion anisotropy relaxation data or mutagenesis, H/D exchange, and 13C-labeling data transformed into AIRs	Docking from ensembles of structures, side chain at the interface flexibility	Allows both side-chain and backbone flexibility in the final refinement stage by simulated annealing MD and steepest descent minimization	Explicit symmetry restraints with Nbody docking	Clustering	Sum of intermolecular van der Waals, electrostatic, ambiguous interaction restraint energy, the buried surface area and the desolvation energy	4
ICM-DISCO	Rigid body pseudo-Brownian MC with grid-based energy function	Specific filtering criterion on a case-by-case basis	–	Flexible surface side-chain	Variable linkage of equivalent monomer	Truncated vdW, electrostatic, solvation, the hydrogen-bonding potential, and a hydrophobicity potential	Truncated vdW, electrostatic, solvation, the hydrogen-bonding potential, rotamer probablity, and a hydrophobicity potential	185
Molfit	Weighted-geometric search with FFT	–	–	–	Orientation filter and monomer assembly by symmetry; intradomain assembly	Scoring by weighted geometric complementarity, general electrostatic and hydrophobic potential	–	188
PachDock	Geometric docking: matching of local shape features and geometric hashing	Biological structural information filtering: allows specification of the binding residues in order of importance	Allows some extent of steric clashes, backbone flexibility through the incorporation of hinge-bending regions (FlexDock) and model of cyclic symmetry (SymmDock)	–	Homo-N-mer generation by symmetry	Shape complementarity	Geometric fit and atomic desolvation energy	23 and 191

Table 4. (*Continued*).

Software	Searching	Filtering at searching stage	Flexibility at searching stage	Flexibility at refinement	Symmetry	Scoring at sampling stage	Scoring at refinement stage	References
RosettaDock	Rigid-body Monte Carlo search and minimization	–	Explicit side chain minimization; rotamer library	–	–	Residue pair potential	Van der Waals potential, desolvation energy, hydrogen bond potential, electrostatics scoring	91
SKE-DOCK	Geometric docking	Structural information can be used manually	–	Automated homology modeling program FAMS, which is based on database searches for homologous structures and simulated annealing energy minimization, and includes main chain adjustments	–	Shape complementarity	Circle, which calculates the model quality from the side-chain environment of each amino acid residue	198
SmoothDock	Rigid body docking using FFT-based program DOT	Desolvation and electrostatics	Predocking MD and Adopted Basis Newton-Raphson minimzation	–	Homo-N-mer generation by symmetry	Van der Waals contribution, shape complementarity, electrostatic and desolvation energy	Shape complementarity, electrostatic and desolvation energy	120
ZDOCK	Rigid body FFT search	Biological structural information filtering: allows the definition of blocking residues, which in contrast with interfacial residues would be given zero desolvation energy	–	Optimizing full atoms internal energy and vdW	Home-N-mer generation by symmetry	Shape complementarity	van der Waals, electrostatics, desolvation energy	82

Table 5. Advantages and Disadvantages of Some Docking Algorithms.

Softwares	Advantages	Disadvantages	Web address	Institutions	Reference
Attract	The reduced protein model allows systematic docking minimization of many thousand start structures in reasonable computer time. Includes side chain and global flexibility during the systematic docking search	Scoring scheme may not be accurate enough to account for the balance between electrostatic solvation and Coulomb interaction. The reduced protein model may in principle be not as specific as an atomic resolution description	http://www.iu-bremen.de/schools/ses/mzacharias/08475/	International University Bremen, Bremen, Germany	44
Bigger	FFT method only dependent on O(*N*2.8) relative to the size of the Matrix. The computation time is also lower than that reported FFT implementations.	Does not lead to an association model at the atomic level. Candidate models must be subjected to further structure relaxation. Weak transient complexes such as those occurring between electron transfer proteins cannot be treated	http://www.cqfb.fct.unl.pt/bioin/chemera/	Universidade Nova de Lisboa, Caparica, Portugal	28
ClusPro	It is a fully automatic algorithm that rapidly docks, filters, and ranks putative protein complexes within a short amount of time using only the given structures of the component proteins and thermodynamic considerations.	As it is an automatic system it fails when it is necessary to introduce manually extra information for a correct docking	http://nrc.bu.edu/cluster/	Boston University, Boston, Ma	160
3D-Dock	A substantial contribution toward the improved predictions is the clustering of conformations, in combination with side-chain trimming, and the clustering of conserved residues on the protein surface.	Although only recently the docking flexibility has been taken into account, it is still not yet completely introduced	http://www.sbg.bio.ic.ac.uk/docking/	Imperial College, London, United Kingdom	167
DOT	Good hits for most systems within its top 1500 results	Does not account to any conformational changes induced on binding, which can make it difficult to predict the true binding site	http://www.sdsc.edu/CCMS/DOT	UCSD, La Jolla, CA	161

Table 5. (*Continued*).

Softwares	Advantages	Disadvantages	Web address	Institutions	Reference
Gramm-X	The docking server back-end analyzes the input structures and selects the best course of action automatically.	The antibody–antigen complexes are more difficult to predict than enzyme–inhibitor complexes, the limited conformational change upon binding can be tolerated with a reasonable chance of success (interface should have <3 s rmsd of conformational change), and complexes with a significant backbone movement at the interface area are typically out of scope	http://vakser.bioinformatics.ku.edu/resources/ gramm/grammx/	University of Kansas, Lawrence, KS	179S
HADDOCK	The fact that the both the side chains and backbone are allowed to move at the refinement stage increases the accuracy of the scoring compared with classical rigid body docking. The interaction space search is limited to relevant regions by imposed constrains	Without additional experimental information the scoring might not be as effective in the case of complexes lacking some kind of asymmetry in their interface. As is data-driven it depends of the level of confidence of the biological and structural information	http://www.nmr.chem.uu.nl/haddock/	Utrecht University, Utrecht, The Netherlands	4
ICM-DISCO	The procedure is global and fully automated. The algorithm handles the induced changes of surface side-chains.	Is less successful if the backbone undergoes large scale rearrangements	http://www.molsoft.com/docking.html	Molsoft, LLC, La Jolla, Ca	185
Molfit	Incorporates external data from different sources, such as biochemical and biophysical experiments or theoretical analyzes of sequence data. Handle conformational changes of any size, but preferably those involving relative movements of whole domains	The scoring function needs improvement in order to rank such solutions higher, possibly via post scan re-evaluation.	http://www.weizmann.ac.il/Chemical_Research_Support//molfit/home.html	Weizmann Institute of Science, Rehovot, Israel	188
PatchDock	Handle conformational changes of any size, but preferably those involving relative movements of whole domains. Treats flexible (hinge-bent) docking.	Does not deal with side-chain flexibility	http://bioinfo3d.cs.tau.ac.il/PatchDock/	Tel Aviv University, Tel Aviv, Israel	23 and 191

Table 5. (*Continued*).

Softwares	Advantages	Disadvantages	Web address	Institutions	Reference
Rosetta-Dock	Close correspondence of the lowest free energy structures with the X-ray structure	Problems with complexes with significant backbone conformational	http://graylab.jhu.edu/docking/rosetta	Johns Hopkins University, Baltimore, MD	91
Ske-Dock	Performs well for easy targets, with large interface areas and no conformational change.	The scoring function does not select good models consistently, even if native-like complex structures are obtained in candidate models	http://www.pharm. kitasato-u.ac.jp/bmd/files/SKE_DOCK.html	School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan	198
SmoothDock	Consistently ranked the correct model first (*i.e.*, with highest confidence)	Low-affinity complexes are hard to discriminate. Problems with the large cavities observed at the interfaces, which are most likely filled with structural water molecules that, for the most part, are neglected by the empirical free energies	http://structure.pitt.edu/servers/smoothdock/	Boston University, Boston, Massachusetts	120
ZDOCK	Particularly effective for the antibody-antigen category of test cases	Problems in the complexes with large conformational change, and in the unbound–bound targets that require modeling	http://zlab.bu.edu/zdock/index.shtml	Boston University, Boston, Ma	82

Table 6. Summary of Docking Predictions in the CAPRI Contest.

Softwares	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12	T13	T14	T18	T19	T20	T21	T22	T23
Attract	–	.	–	–	–	–	–	**	0	–	–	–	–	***	0	**	0	0	–	–
Bigger	–	0	–	0	0	**	*	0	0	0	–	0	0	0	0	0	–	–	–	–
ClusPro	–	–	–	–	–	–	–	**	0	0	0	***	*	0	0	*	0	O	*	0
3D-DOCK (Sternberg)	0	*	0	0	0	***	*	**	0	0	*	*	0	**	0	*	0	0	–	–
3D-DOCK (Bates)	–	–	–	0	0	0	***	*	0	*	**	*	0	**	**	*	0	0	–	–
DOT	*	*	0	0	0	**	0	0	0	0	0	***	***	**	0	0	0	**	–	–
Gramm-X (manual)	0	*	0	–	–	–	–	–	–	0	–	–	–	**	**	0	0	0	–	–
Gramm-X (server)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0
HADDOCK	–	–	–	–	–	–	–	–	–	**	**	0	***	***	0	0	0	**	–	–
ICM	0	0	**	0	0	***	**	**	0	*	**	***	*	***	**	**	0	0	–	–
Molfit	*	*	0	0	0	0	***	***	0	0	*	***	0	**	0	0	0	0	–	–
PatchDock (manual)	*	0	0	0	0	0	***	**	*	*	*	*	0	**	**	*	0	0	–	–
PatchDock (server)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	*	0
Rosetta (Baker)	0	0	0	0	0	**	***	–	0	0	**	***	**	***	0	***	*	0	–	–
ROSETTA (Gray)	0	0	0	0	0	**	***	***	–	–	**	***	0	0	0	**	0	**	–	0
SKE-DOCK	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0
Smooth-Dock (manual)	*	0	0	0	0	***	***	**	0	0	0	***	***	**	**	*	0	0	–	–
Smooth-Dock (server)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	*	0	0
ZDOCK & RDOCK	0	**	0	0	0	0	**	**	0	0	*	***	***	***	**	**	0	*	–	–

														Total		
Softwares	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35	T36	Summary	ALL	UU	UB
Attract	0	0	*	*	0	0	*	NA	***	0	0	0	0	*** 2 ** 2 * 3 0 12	7 (23*)	4 (4*)	1 (2*)
Bigger	–	–	–	–	–	–	–	NA	–	–	–	–	–	*** 0 ** 1 * 1 0 12	2 (3*)	1 (1*)	1 (2*)
ClusPro	0	*	0	*	0	0	0	NA	0	0	*	0	0	*** 1 ** 1 * 6 0 17	8 (11*)	1 (1*)	4 (7*)
3D-Dock (Sternberg)	0	0	0	0	0	0	0	NA	–	–	–	–	–	*** 1 ** 2 * 5 0 16	8 (12*)	1 (1*)	4 (7*)
3D-Dock (Bates)	0	*	0	0	0	0	*	NA	0	0	0	0	0	*** 1 ** 3 * 6 0 17	10 (15*)	3 (5*)	4 (5*)
DOT	0	0	0	*	0	**	0	NA	0	0	0	0	0	*** 2 ** 5 * 3 0 21	9 (17*)	3 (4*)	3 (8*)
Gramm-X (manual)	0	0	**	*	–	0	0	NA	0	0	0	0	0	*** 0 ** 3 * 2 0 17	5 (8*)	2 (3*)	2 (3*)
Gramm-X (server)	0	**	0	0	0	0	0	NA	0	0	0	0	0	*** 0 ** 1 * 0 0 14	1 (2*)	0 (0*)	1 (2*)

Table 6. (*Continued*).

														Total		
Softwares	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35	T36	Summary	ALL	UU	UB
HADDOCK	0	*	**	*	0	**	0	NA	0	0	**	0	0	*** 2 ** 6 * 2 0 11	10 (20*)	4 (7*)	2 (4*)
ICM	*	*	*	0	0	–	–	NA	–	–	–	–	–	*** 3 ** 6 * 5 0 9	14 (26*)	3 (4*)	7 (14*)
Molfit	0	***	*	*	0	*	0	NA	**	0	*	0	0	*** 4 ** 2 * 8 0 17	13 (23*)	5 (8*)	4 (10*)
PatchDock (manual)	0	0	0	*	0	0	0	NA	**	0	*	0	0	*** 1 ** 4 * 8 0 17	13 (19*)	6 (8*)	3 (5*)
PatchDock (server)	0	**	*	0	–	0	0	NA	0	0	0	0	0	*** 0 ** 1 * 2 0 11	3 (4*)	1 (1*)	1 (2*)
Rosetta (Baker)	0	0	**	*	0	0	0	NA	***	0	**	0	0	*** 5 ** 5 * 2 0 17	12 (27*)	4 (9*)	3 (7*)
Rosetta (Gray)	0	0	**	0	0	0	0	NA	***	0	*	0	0	*** 4 ** 5 * 1 0 19	10 (23*)	4 (10*)	3 (8*)
SKE-DOCK	0	**	0	0	0	0	0	NA	0	0	0	0	0	*** ** * 0	2 (3*)	1 (1*)	1 (2*)
Smooth-Dock (manual)	*	*	0	*	0	0	0	NA	**	0	*	0	0	*** 4 ** 4 * 6 0 16	14 (26*)	4 (7*)	6 (14*)
SMOOTH-DOCK (server)	0	**	0	0	0	0	0	NA	0	0	0	0	0	*** 0 ** 1 * 1 0 14	2 (3*)	1 (1*)	1 (2*)
ZDOCK & RDOCK	*	**	*	**	0	*	0	NA	0	0	0	0	*	*** 3 ** 7 ** 6 0 14	16 (29*)	4 (6*)	6 (14*)

"0" indicates that none of the submitted predictions was of acceptable quality. "—" indicates that no predictions were submitted. "NA" indicates that the results are not available, "*" indicates that at least one of the submitted predictions was in the acceptable range, "**" indicates that at least one of the submitted predictions was of medium accuracy, and "***" indicates that at least one prediction was of high accuracy. In the total columns it is possible to encounter the number of hits and the total number of stars (in brackets). UU-Unbound-unbound test cases, UB-Unbound-bound test cases. "Man" means that the softwares were used under human supervision.

to filter wrong solutions. In Table 5 it is possible to encounter the main advantages and disadvantages of the softwares. They are almost related with the time necessary to perform the docking, capacity of handling side-chain and backbone flexibility and problems with the scoring function used.[207–210]

Table 7. Summary of Docking Predictions in the CAPRI Contest.

Software	ALL	UU	UB
Attract	0,68	1,00	0,67
Bigger	0,23	0,33	0,25
3D-Dock (Bates)	0,56	0,63	0,63
3D-Dock (Stenrsberg)	0,48	0,13	0,70
DOT	0,57	0,44	0,90
Gramm-X (man)	0,40	0,43	0,75
ICM	1,13	0,57	1,40
HADDOCK	0,95	1,17	1,00
Molfit	0,77	0,89	1,00
PatchDock (man)	0,63	1,00	0,50
Rosetta (Baker)	0,93	1,00	0,78
Rosetta(Gray)	0,79	1,43	0,80
Smooth (man)	0,87	0,78	1,40
ZDOCK & RDOCK	0,97	0,67	1,40

UU, Unbound-unbound test cases; UB, Unbound-bound test cases; "Man" means that the softwares were used under human supervision.

In Table 6 we have summarized the docking predictions in the CAPRI contest taken from Wodak et al. 2003,[13] 2005[207] and 2007[208] as well as from the CAPRI web address (http://capri.ebi.ac.uk/). Although important to achieve some conclusions from the analysis of the results we cannot forget that they are not statistically very meaningful because the number of targets is still very small. We can observe that there are a lot of acceptable results from almost all the groups and that almost every group (except the servers) have at least one high prediction. If we analyze the results giving quantitative measures to the nonacceptable, acceptable, medium and high results (a value of 0, 1, 2, or 3), which are present in Table 7, we notice that globally ICM,[185] ZDOCK,[82] HADDOCK[4] and Baker (modified version of RosettaDock)[195] must be the best predictors, followed closely by Gray's version of RosettaDock,[91] Camacho group's SmoothDock[120] with manual modifications, and Molfit.[188] From the unbound-unbound test cases we have to emphasize the very good behavior of RosettaDock and HADDOCK,[4] followed by Wolfson's group PatchDock[23,191] with manual modifications, Attract, Molfit and Zdock. Regarding the unbound-bound test cases ICM,[185] ZDOCK,[82] ClusPro,[160] and Camacho group (SmoothDock[120] with manual modifications) have very good results followed by Molfit,[188] HADDOCK[4] and Dot.[161] Again we should highlight that the lower performance scores might not necessarily reflect the quality of the approach used by the different groups.

In Figure 3 we have plotted the number of citations per year of the docking programs described earlier (data took from ISI Web of Science considering the references from Table 4). From Figure 3.1 it is possible to observe that only after 2003 there was an increase of the number of citations of the protein–protein docking softwares. Since their publication the most cited softwares are HADDOCK,[4] RosettaDock,[91] three-dimensional-Dock,[167] BIGGER,[28] and Dot.[161] It is possible to observe an increase of the number of citations per year of the PatchDock,[23,191] ClusPro,[160] HADDOCK,[4] RosettaDock[91] and ZDOCK.[82] We have to stress that even though HADDOCK[4] is a recent software, it seems to be very popular presenting a clearly higher number of citations per year. If we consider just papers that apply the different softwares to specific biological problems (represented in Figure 3.2) Haddock[4] is the most popular, followed by ClusPro,[160] PatchDock[23,191] and RosettaDock.[91]

The servers are also becoming very used because they greatly improve the utility of the docking approaches in the biological community. PatchDock[23,191] since January of 2004 has more than 30,000 submissions, 12,000 of them in 2008, from around 4000 different users (Dr. Schneidman-Duhovny, private communication); ClusPro[160] since January of 2003 had around 18,600 jobs submitted by 2700 different users worldwide (Dr. Comeau, private communication); the new version of ClusPro (PIPER)[165] is still in development but had already 300 jobs submitted from about 80 different users (Dr. Brenk, private communication); Gramm-X[178] since 2006 has processed more than 12,000 jobs submitted by more than 2300 users (Dr. Tovchigrechko, private communication); RosettaDock[91] server opened in April 2007, and over 150 individuals have used the web server for more than 800 docking jobs; SKE-DOCK[198] has monthly a 1–2 submissions by 1–2 users (Dr. Terashi, private communication); and HADDOCK server since June 1st 2008 has 1519 submissions and 197 registered users.[211]

Conclusion

A comprehensive understanding of the interactions between proteins is indispensable for interpreting many biochemical phenomena and is of supreme practical relevance in pharmaceutical and medicinal sciences. Computational docking tries to predict the correct binding mode of the interacting partners, which has been demonstrated to be a difficult assignment considering the macromolecular nature of the protein. Thus, protein–protein docking is a difficult challenge especially because of the differences between the conformations of the bound and unbound molecules, which increase the dimensionality of the problem.

Usually, the protein–protein docking procedure begins by treating the proteins as rigid bodies, perhaps with some surface softness, searching the 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 responsibility. Rescoring with a better function of these structures is followed in order to discriminate near-native orientations. Then full atomic detail is added (if not before) as well as allowing the movement of the sidechains and possibly backbone, minimizing an energy function. 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. From the results of the CAPRI experiment and the software popularity we can observe that ICM,[185] ZDOCK,[82] HADDOCK[4] and ROSETTADOCK[91] seem to be some of the best predictors that are most commonly used. Softwares, such as HADDOCK4, which are capable of dealing with side-chain and backbone flexibility as well as using biological information regarding the complex in the searching stage, seem to perform better in the protein–protein docking world.

However, because of the complexity of the problem, protein–protein docking is still largely at the theoretical stage, and con-

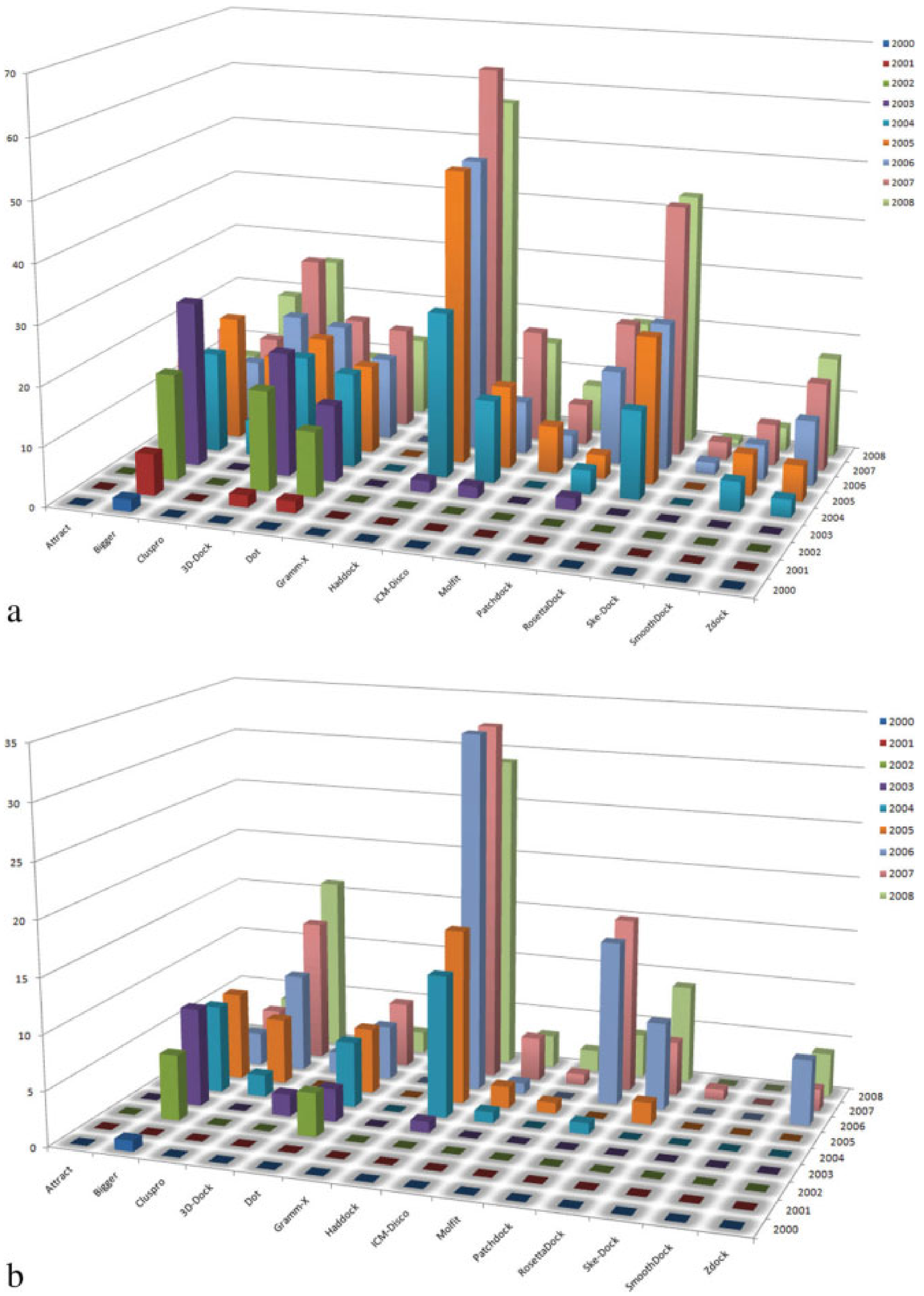

Figure 3. Number of citations per year of the docking programs described earlier. Data taken from ISI Web of Science (February of 2007) considering the references from Table 4. (a) All articles are considered.; (b) Only the articles with experimental predictions were considered.

tinues to be a significant scientific challenge to structural biologists and the biomolecular modeling community.

References

1. Sotriffer, C. A.; Flander, W.; Winger, R. H.; Rode, B. M.; Liedl, K. R.; Varga, J. M. Methods 2000, 20, 280.
2. Fahmy, A.; Wagner, G. J Am Chem Soc 2002, 124, 1241.
3. Smith, G. R.; Fitzjohn, P. W.; Page, C. S.; Bates, P. A. Proteins 2005, 60, 263.
4. Dominguez, C.; Boelens, R.; Bonvin, A. M. J Am Chem Soc 2003, 125, 1731.
5. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Proteins 2006, 65, 15.
6. Fernandez-Recio, J.; Abagyan, R.; Totrov, M. Proteins 2005, 60, 308.
7. Gardiner, E. J.; Willet, P.; Artymiuk, P. J. Proteins 2001, 44, 44.
8. Zacharias, M. Proteins 2005, 60, 252.
9. Janin, J. Protein Sci 2005, 14, 278.
10. Gray, J. J.; Moughon, S. E.; Kortemme, T.; Schueler-Furman, O.; Misura, K. M. S.; Morozov, A. V.; Baker, D. Proteins 2003, 52, 118.
11. Janin, J.; Wodak, S. J Mol Biol 1978, 125, 357.
12. Ma, X. H.; Li, C. H.; Shen, L. Z.; Gong, X. Q.; Chen, W. Z.; Wang, C. X. Proteins 2005, 60, 319.
13. Janin, J.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J. E; Vajda, S.; Vakser, I.; Wodak, S. J. Proteins 2003, 52, 2.
14. Janin, J. Proteins 2002, 47, 257.
15. Brenk, R.; Vetter, S. W.; Boyce, S. E.; Goodin, D. B.; Schoichet, B. K. J Mol Biol 2006, 357, 1449.
16. Li, C. H.; Ma, X. H.; Chen, W. Z.; Wang, C. X. Protein Eng 2003, 16, 265.
17. Camacho, C. J.; Vajda, S. Curr Opin Struct Biol 2002, 12, 36.
18. Halperin, I.; Wolfson, H.; Nussinov, R. Structure 2004, 12, 1027.
19. Smith, G. R.; Sternberg, M. J.; Bates, P.A. J Mol Biol 2005, 347, 1077.
20. Yu, Y. H.; Lu, B. Z.; Han, J. G.; Zhang, P. F. J Computer-Aided Mol Design 2004, 18, 251.
21. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Proteins, 2002, 47, 409.
22. Lee, K.; Sim, J.; Lee, J. Proteins, 2005, 60, 257.
23. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Nucleic Acids Reser 2005, 33, 363.
24. Chen, R.; Mintseris, J.; Janin, J.; Weng, Z. Proteins 2003, 52, 88.
25. Hwang, H.; Pierce, B.; Mintseris, J.; Janin, J.; Weng, Z. Proteins 2008, 73, 705.
26. Gary, J. J. Curr Opin Struct Biol 2006, 16, 183.
27. Shoichet, B. K; Kuntz, I. D. J Mol Biol 1991, 221, 327.
28. Palma, P. N.; Krippahl, L.; Wampler, J. E.; Moura, J. J. G. Proteins 2000, 39, 372.
29. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Proc Natl Acad Sci USA 1992, 89, 2195.
30. Meyer, M.; Wilson, P.; Schomburg, D. J Mol Biol 1996, 264, 199.
31. Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. J Mol Biol 1997, 272, 106.
32. Vakser, I. A. Proteins 1997, 1, 226.
33. Ritchie, D. W.; Kemp, G. J. L. Proteins 2000, 39, 178.
34. Hart, T. N.; Read, R. J. Proteins 1992, 13, 206.
35. Helmer-Citterich, M.; Tramonato, A. J Mol Biol 1994, 235, 1021.
36. Norel, R.; Lin, S. L.; Wolfson, H.; Nussinov, R. Biopolymers 1994, 34, 933.
37. Norel, R.; Wolfson, H.; Nussinov, R. Comb Chem High Throughput Screen 1999, 2, 177.
38. Fischer, D.; Lin, S. L.; Wolfson, H. L.; Nussinov, R. J Mol Biol 1995, 248, 459.
39. Kohlbacker, O.; Burchardt, A.; Moll, A.; Hildebrandt, A.; Bayer, P.; Lenhof, H. P. J Biomol NMR 2001, 20, 15.
40. Ausiello, G.; Cesareni, G.; Helmer-Citterich, M. Proteins 1997, 28, 556.
41. Gardiner, E. J.; Willet, P.; Artymiuk, P. J. Protein, 2003, 52, 10.
42. Gabdoulline, R. R.; Wade R. C. J Mol Biol 2008, 306, 1139.
43. Fernandez-Recio, J.; Totrov, M.; Abagayan, R. Protein Sci 2002, 11, 280.
44. Zacharias, M. Protein Sci 2003, 12, 1271.
45. Lise, S.; Walker-Taylor, A.; Jones, D. T., BMC Bioinformatics 2006, 7, 3.
46. Jones, S.; Thornton, J. M. Proc Natl Acad Sci USA 1996, 93, 13.
47. Betts, M. J.; Sternberg, M. J. E. Prot Eng 1999, 12, 271.
48. Lo Conte, L.; Chothia, C.; Janin, J. J Mol Biol 1999, 285, 2177.
49. Norel, R.; Petrey, D.; Wolfson, H.; Nussinov, R. Proteins 1999, 35, 403.
50. Decanniere, K.; Transue, T. R.; Desmyter, A.; Maes, D.; Muyldermans, S.; Wyns, L. J Mol Biol 2001, 313, 473.
51. Jackson, R. M. Prot Sci 1999, 8, 603.
52. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Proteins 2006, 63, 811.
53. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. J Phys Chem B 2006, 110, 10962.
54. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Theor Chem Acc 2007, 117, 99.
55. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. J Comput Chem 2007, 28, 644.
56. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Int J Quantum Chem 2007, 107, 299.
57. Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. J Phys Chem B 2007, 111, 2697.
58. Koike, A.; Takagi, T. Protein Eng Design Sel 2004, 17, 165.
59. Neuvirth, H.; Raz, R.; Schreiber G. J Mol Biol 2004, 338, 181.
60. Bordner, A. J.; Abagyan, R. Proteins 2005, 60, 353.
61. Bradford, J. R.; Westhead, D. R. Bioinformatics 2005, 21, 1487.
62. Chen, H.; Zhou, H. X. Proteins 2005, 61, 21.
63. del Sol, A.; O'Meara, P. Proteins 2005, 58, 672.
64. Fernandez-Recio, F.; Totrov, M.; Skorodumov, C.; Abagyan, R. Proteins 2005, 58, 134.
65. Chung, J. L.; Wang, W.; Bourne, P. E. Proteins 2006, 62, 630.
66. Porollo, A.; Meller, J. Proteins 2007, 66, 630.
67. Ritchie, D. W. Curr Protein Pept Sci 2008, 9, 1.
68. Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. J Mol Biol 1997, 271, 511.
69. Fischer, E. Ber Dtsch Chem Ges 1894, 27, 2985.
70. Koshland, D. E., Jr. Proc Natl Acad Sci USA 1958, 44, 98.
71. Monod, J.; Wyman, J.; Changeux, J. P. J Mol Biol 1965, 12, 88.
72. Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Protein Sci 2000, 9, 10.
73. Kumar, S.; Ma, B.; Tsai, C. J.; Wolfson, H.; Nussinov, R. Cell Biochem Biophys 1999, 31, 141.
74. Tsai, C. J.; Kumar, S.; Ma, B.; Nussinov, R. Protein Sci 1999, 8, 1181.
75. Dobbins, S. E.; Lesk, V. I.; Sternberg, M. J. Proc Natl Acad Sci USA 2008, 105, 10390.
76. Grünberg, R.; Leckner, J.; Nilges, M. Structure 2004, 12, 2125.
77. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P. DesJarlais, R. L. Curr Pharm Des 2005, 11, 323.

78. Chuang, G. Y.; Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. Biophys J 2008, 95, 4217.
79. Bonvin, A. M. Curr Opin Struct Biol 2006, 16, 194.
80. Krippahl, L.; Moura, J. J.; Palma, P. N. Proteins 2003, 52, 19.
81. Fernandez-Recio, J.; Totrov, M.; Abayan, R. Proteins 2003, 52, 113.
82. Chen, R.; Li, L.; Weng, Z. Proteins 2003, 52, 80.
83. Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. Proteins 2008, 73, 271.
84. Jiang, J.; Kim, S. H. J Mol Biol 1991, 219, 79.
85. Walls, P. H.; Sternberg, M. J. J Mol Biol 1992, 228, 277.
86. Sandak, B.; Nussinov, R.; Wolfson, H. J Comp Appl Bio Sci 1995, 11, 87.
87. Vakser, I. A. Prot Eng 1995, 8, 371.
88. Morelli, X.; Czjzek, M.; Hatchikian, C. E.; Bornet, O.; Fontecilla-Camps, J. C.; Palma, N. P.; Moura, J. J.; Guerlesquin, F. J Biol Chem 2000, 275, 23204.
89. Ritchie, D. W. Proteins 2003, 52, 98.
90. Li, L.; Chen, R.; Weng, Z Proteins 2003, 53, 693.
91. Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. J Mol Biol 2003, 331, 281.
92. Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. Curr Med Chem 2004, 11, 91.
93. Totrov, M. M.; Abagyan, R. A. Nature Struct Biol 1994, 1, 259.
94. Abagyan, R.; Totrov, M.; Kuznetsov, D. J Comput Chem 1994, 15, 488.
95. Leach, A. R. J Mol Biol 1994, 235, 345.
96. David, L.; Luo, R.; Gilson, M. K. J Comput Aided Mol Des 2001, 15, 157.
97. Head, M. S.; Given, J. A.; Gilson, M. K. J Phys Chem 1997, 101, 1609.
98. Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Proteins 2000, 39, 261.
99. Heifetz, A.; Eisenstein M. Protein Eng 2003, 16, 179.
100. Segal, D.; Eisenstein, M. Proteins 2005, 59, 580.
101. Chaudhury, S.; Gray, J. J. J Mol Biol 2008, 381, 1068.
102. Zacharias, M. Proteins 2004, 54, 759.
103. Ponting, C. P.; Russell, R. R. Annu Rev Biophys Biomol Struct 2002, 31, 45.
104. Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A. Curr Opin Struct Biol 2004, 14, 208.
105. Wriggers, W.; Chakravarty, S.; Jennings, P. A. Biopolymers 2005, 80, 736.
106. Eisenstein, M.; Shirav, I.; Koren, G.; Friesem, A. A.; Katchalski-Katzir, E. J Mol Biol 1997, 266, 135.
107. Berchanski, A.; Eisenstein, M. Proteins 2003, 53, 817.
108. Comeau, S. R.; Camacho, C. J. J Struct Biol 2004, 150, 233.
109. Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. J Mol Biol 2005, 349, 435.
110. Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. Phys Biol 2005, 2, S156.
111. Emekli, U.; Schneidman-Duhovny, D.; Wolfson, H. J.; Nussinov, R.; Haliloglu, T. Proteins 2008, 70, 1219.
112. Ben-Zeev, E.; Kowalsman, N.; Ben-Shimon, A.; Segal, D.; Atarot, T.; Noivirt, O.; Shay, T.; Eisenstein, M. Proteins 2005, 60, 195.
113. Bogan, A. A.; Thorn, K. S. J Mol Biol 1998, 280, 1
114. Camacho, C. J.; Ma, H.; Champ, P. C. Proteins 2006, 63, 868.
115. Camacho, C. J.; Gatchell, D. W.; Kimura, S. R.; Vajda, S. Proteins 2000, 40, 525.
116. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. Nucleic Acids Res 2004, 32, 96.
117. Kozakov, D.; Schueler-Furman, O.; Vajda, S. Proteins 2008, 72, 993.
118. Wang, C.; Schueler-Furman, O.; Andre, I.; London, N.; Fleishman, S. J.; Bradley, P.; Qian, B.; Baker, D. Proteins 2007, 69, 758.
119. Bernauer, J.; Azé, J.; Janin, J.; Poupon, A. Bioinformatics 2007, 23, 555.
120. Camacho, C. J.; Gatchell, D. W. Proteins 2003, 52, 92.
121. Wodak, S. J.; Janin, J. Adv Protein Chem 2002, 61, 9.
122. Wodak, S. J.; Mendez, R. Curr Opin Struct Biol 2004, 14, 242.
123. Comeau, S. R.; Vajda, S.; Camacho, C. J. Proteins 2005, 60, 239.
124. Chen, R.; Tong, W.; Mintseris, J.; Li, L.; Weng, Z. Proteins 2003, 52, 68.
125. Vajda, S. Proteins 2005, 60, 176.
126. Fieulaine, S.; Morera, S.; Poncet, S.; Mijakovic, I.; Galinier, A.; Janin, J.; Deutscher, J.; Nessler, S. Proc Natl Acad Sci USA 2002, 99, 13437.
127. Thouvenin, E.; Schoehn, G.; Rey, F.; Petitpas, I.; Mathieu, M.; Vaney, M. C.; Cohen, J.; Kohli, E.; Pothier, P.; Hewat, E. J Mol Biol 2001, 307, 161.
128. Barbey-Martin, C.; Gigant, B.; Bizebard, T.; Calder, L. J.; Wharton, S. A.; Skehel, J. J.; Knossow, M. Virology 2002, 294, 70.
129. Desmyter, A.; Spinelli, S.; Payan, F.; Lauwereys, M.; Wyns, L.; Muylder-mans, S.; Cambillau, C. J Biol Chem 2002, 77, 23645.
130. Sundberg, E. J.; Li, H.; Llera, A. S.; McCormick, J. K.; Tormo, J.; Schlievert, P. M.; Karjalainen, K.; Mariuzza, R. A. Structure 2002, 10, 687.
131. Takagi, J.; Yang, Y.; Liu, J. H.; Wang, J. H.; Springer, T. A. Nature 2003, 424, 969.
132. Graille, M.; Zhou, C. Z.; Receveur-Brechot, V.; Collinet, B.; Declerck, N.; van Tilbeurgh, H. J Biol Chem 2005, 280, 14780.
133. Bressanelli, S.; Stiasny, K.; Allison, S. L.; Stura, E. A.; Duquerroy, S.; Lescar, J.; Heinz, F. X.; Rey, F. A. EMBO J 2004, 23, 728.
134. Carvalho, A. L.; Dias, F. M. V.; Prates, J. A. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romao, M. J.; Fontes, C. M. G. A. Proc Natl Acad Sci USA 2003, 100, 13809.
135. Graille, M.; Stura, E. A.; Bossus, M.; Muller, B. H.; Letourneur, O.; Battail-Poirot, N.; Sibai, G.; Gauthier, M.; Rolland, D.; Le Du, M. H.; Ducancel, F. J Mol Biol 2005, 354, 447.
136. Terrak, M.; Kerff, F.; Langsetmo, K.; Tao, T. Nature 2004, 429, 780.
137. Sansen, S.; De Ranter, C. J.; Gebruers, K.; Kristof Brijs, K.; Courtin, C. M.; Delcour, J. A; Rabijns, A. J Biol Chem 2004, 279, 36022.
138. Eghiaian, F.; Grosclaude, J.; Lesceu, S.; Debey, P.; Doublet, B.; Treguer, E.; Rezaei, H.; Knossow, M. Proc Natl Acad Sci USA 2004, 100, 10254.
139. Graille, M.; Heurgue-Hamard, V.; Champ, S.; Mora, L.; Scrima, N.; Ulryck, N.; van Tilbeurgh, H.; Buckingham, R. H. Mol Cell 2005, 20, 917.
140. Hou, Z.; Bernstein, D. A.; Fox, C. A.; Keck, J. L. Proc Natl Acad Sci USA 2005, 102, 8489.
141. Laggerbauer, B.; Liu, S.; Makarov, E.; Vornlocher, H. P.; Makarova, O.; Ingelfinger, D.; Achsel, T.; Luhrmann, R. RNA 2005, 11, 598.
142. Ghosh, A.; Praefcke, G. J.; Renault, L.; Wittinghofer, A.; Herrmann, C. Nature 2006, 440, 101.
143. Ménétrey, J.; Perderiset, M.; Cicolari, J.; Dubois, T.; Elkhatib, N.; El Khadali, F.; Franco, M.; Chavrier, P.; Houdusse, A. EMBO J 2007, 26, 1953.
144. Grishkovskaia, I.; Bonsor, D. A.; Kleanthous, C.; Dodson, E. J. J Am Chem Soc 2007, 129, 4800.
145. Walker, J. R.; Avvakumov, G. V.; Xue, S.; Newman, E. M.; Mackenzie, F.; Weigelt, J.; Sundstrom, M.; Arrowsmith, C. H.; Edwards, A. M.; Bochkarev, A.; Dhe-Paganon, S. A. To be published.
146. Walker, J. R.; Avvakumov, G. V.; Xue, S.; Butler-Cole, C.; Weigelt, J.; Sundstrom, M.; Arrowsmith, C. H.; Edwards, A. M.; Bochkarev, A.; Dhe-Paganon, S. To be published.

147. Tong, Y.; Chugha, P.; Hota, P.K.; Alviani, R. S.; Li, M.; Tempel, W.; Shen, L.; Park, H. W.; Buck M. J Biol Chem 2007, 282, 37215.
148. Tong, Y.; Hota, P. K.; Hamaneh, M. B.; Buck, M. Structure 2008, 16, 246.
149. Micheelsen, P. O.; Vevodova, J.; De Maria, L.; Ostergaard, P. R.; Friis, E. P.; Wilson, K.; Skjot, M. J Mol Biol 2008, 380, 681.
150. Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. J Mol Biol 1997, 267, 707.
151. Miyazawa, S.; Jernigan, R. L. J Chem Phys 2005, 122, 24901.
152. Janin, J. Proteins 2005, 60, 170.
153. Van Dijk, M.; van Dijk, A. D. J.; Hsu, V.; Boelens, R.; Bonvin, A. M. J. J. Nucleic Acids Res 2006, 34, 3317.
154. Ma, X. H.; Wang, C. X.; Li, C. H.; Chen, W. Z. Protein Eng 2002, 15, 677.
155. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. J Comput Chem 2003, 24, 1637.
156. Tovchigrechko, A.; Vakser, I. A. Proteins 2005, 60, 296.
157. May, A.; Zacharias, M. Biochim Biophys Acta 2005, 1754, 225.
158. Koehl, P.; Delarue, M. Curr Opin Struct Biol 1996, 6, 222.
159. Bastard, K.; Prevost, C.; Zacharias, M. Proteins 2006, 62, 956.
160. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. Bioinformatics 2004, 20, 45.
161. Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyi, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; Ten Eyck, L. F. Protein Eng 2001, 14, 105.
162. Zhang, C.; Cornette, J. L.; Delisi, C. Protein Sci 1997, 6, 1057.
163. Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. Bioinformatics 2004, 20, 45.
164. Champ, P. C.; Camacho, C. J. Nucleic Acids Res 2007, 35, 556.
165. Shen, Y.; Brenke, R.; Kozakov, D.; Comeau, S. R.; Beglov, D.; Vajda, S. Proteins 2007, 69, 734.
166. Comeau, S. R.; Kozakov, D.; Brenke, R.; Shen, Y.; Beglov, D.; Vajda, S. Proteins 2007, 69, 781.
167. Aloy, P.; Querol, E.; Aviles, F. X.; Sternberg, M. J. E. J Mol Biol 2001, 311, 395.
168. Smith, G. R.; Sternberg, M. J. Proteins 2003, 52, 74.
169. Smith, G. R.; Sternberg, M. J. Curr Opin Struct Biol 2002, 12, 28.
170. Carter, P.; Lesk, V. I.; Islam, S. A.; Sternberg, M. J. Proteins 2005, 60, 281.
171. Jackson, R. M.; Gabb, H. A.; Sternberg, M. J. J Mol Biol 1998, 276, 265.
172. Sternberg, M. J.; Gabb, H. A.; Jackson, R. M.; Moont, G. Methods Mol Biol 200, 143, 399.
173. Lorensen, W. E.; Cline, H. E. Comput Graphics 1987, 21, 163.
174. Lesk, V. I.; Sternberg, M. J. Bioinformatics 2008, 24, 1137.
175. Mitchell, J. C.; Kerr, R.; Ten Eyck, L. F. J Mol Graph Model 2001, 19, 325.
176. Law, D.; Hotchko, M.; Ten Eyck, L. Proteins 2005, 60, 302.
177. Law, D. S.; Ten Eyck, L. F.; Katzenelson, O.; Tsigelny, I.; Roberts, V. A.; Pique, M. E.; Mitchell, J. C. Proteins 2003, 52, 33.
178. Tovchigrechko, A.; Vakser, I. A. Nucleic Acids Res 2006, 34, 310.
179. Douguet, D.; Chen, H. C.; Tovchigrechko, A.; Vakser, I. A. Bioinformatics 2006, 22, 2612.
180. Liu, S.; Gao, Y.; Vakser, I. A. Bioinformatics 2008, 24, 2634.
181. de Vries, S. J.; van Dijk, A. D.; Bonvin, A. M. Proteins 2006, 63, 479.
182. van Dijk, A. D.; Boelens, R.; Bonvin, A. M. FEBS J 2005, 272, 293.
183. van Dijk, A. D.; Bonvin, A. M. Bioinformatics 2006, 22, 2340.
184. De Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thruseau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. Proteins 2007, 69, 726.
185. Fernandez-Recio, J.; Totrov, M.; Abayan, R. Proteins 2003, 52, 113.
186. Fernandez-Recio, J.; Totrov, M.; Abagyan, R. Pac Symp Biocomput 2002, 552.
187. Grosdidier, S.; Pons, C.; Solernou, A.; Fernández-Recio, J. Proteins 2007, 69, 852.
188. Berchanski, A.; Shapira, B.; Eisenstein, M. Proteins 2004, 56, 130.
189. Kowalsman, N.; Eisenstein, M. Bioinformatics 2007, 23, 421.
190. Heifetz, A.; Katchalski-Katzir, E.; Eisenstein, M. Protein Sci 2002, 11, 571.
191. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Proteins 2005, 60, 224.
192. Schneidman-Duhovny, D.; Inbar, Y.; Polak, V.; Shatsky, M.; Halperin, I.; Benyamini, H.; Barzilai, A.; Dror, O.; Haspel, N.; Nussinov, R.; Wolfson, H. J. Proteins 2003, 52, 107.
193. Mashiach, E.; Schneidman-Duhovny, D.; Andrusier, N.; Nussinov, R.; Wolfson, H. J. Nucleic Acids Res 2008, 36, 229.
194. Abagyan, R. A.; Totrov, M. M.; Kuznetsov, D. A. J Comp Chem 1994, 15, 488.
195. Wang, C.; Schueler-Furman, O.; Baker, D. Protein Sci 2005, 14, 1328.
196. Kim, D. E.; Chivian, D.; Baker, D. Nucleic Acids Res 2004, 32, 526.
197. Lyskov, S.; Gray, J. J. Nucleic Acids Res 2008, 36, 233.
198. Terashi, G.; Takeda-Shitaka, M.; Takaya, D.; Komatsu, K.; Umeyama, H. Proteins 2005, 60, 289.
199. Takeda-Shitaka, M.; Terashi, G.; Chiba, C.; Takaya, D.; Umeyama, H. Med Chem 206, 2, 191.
200. Terashi, G.; Takeda-Shitaka, M.; Kanou, K.; Iwadate, M.; Takaya, D.; Umeyama, H. Proteins 2007, 69, 866.
201. Camacho, C. J.; Vajda, S. Proc Natl Acad Sci USA 2001, 98, 10636.
202. Wiehe, K.; Pierce, B.; Mintseris, J.; Tong, W. W.; Anderson, R.; Chen, R.; Weng, Z. Proteins 2005, 60, 207.
203. Chen, R.; Weng, Z. Proteins 2003, 51, 397.
204. Wiehe, K.; Pierce, B.; Tong, W. W.; Hwang, H.; Mintseris, J.; Weng, Z. Proteins 2007, 69, 719.
205. Pierce, B.; Weng, Z. Proteins 2007, 67, 1078.
206. Pierce, B.; Weng, Z. Proteins 2008, 72, 270.
207. Mendez, R.; Leplae, R.; Lensink, M. F.; Wodak, S. J. Proteins 2005, 60, 150.
208. Lensink, M. F.; Mendez, R.; Wodak, S. J. Proteins 2007, 69, 704.
209. Janin, J. Proteins 2007, 69, 699.
210. Janin, J.; Wodak, S. Structure 2007, 15, 755.
211. http://haddock.chem.uu.nl/Haddock/haddock.php.