Chapter 5
Docking Predictions of Protein-Protein Interactions and Their Assessment: The CAPRI Experiment

Joël Janin

Abstract Protein-protein docking was born in the 1970s as a tool to analyze macromolecular recognition. It developed afterwards into a method of prediction of the mode of association between proteins of known structure. Since 2001, the performance of docking procedures has been assessed in blind predictions by the CAPRI (Critical Assessment of PRedicted Interactions) experiment. The results show that docking routinely yields good models of the protein-protein complexes that undergo only minor changes in conformation and associate as rigid bodies. In contrast, flexible recognition accompanying large conformation changes in the components remains difficult to simulate, and structural predictions generally yield lower quality models. In recent years, a new challenge has been to predict affinity and to estimate the stability of the complex along with its structure. Over the years, CAPRI has proved to be a strong incentive to develop new flexible docking procedures and more discriminative scoring functions, and it has provided a common ground for discussing methods and questions related to protein-protein recognition.

Keywords Community-wide experiment • Protein-protein recognition • Binding affinity • Protein flexibility • Protein engineering • Docking simulation • Computational biology • DOCK • CAPRI • Rigid body docking • Cube representation • Geometric hashing algorithm • Monte-Carlo • RosettaDock • Template-based docking • CASP • Blind prediction • Flexible docking • Molecular dynamics • Flexibility • Affinity benchmark

J. Janin (✉)
IBBMC, Université Paris-Sud, 91405, Orsay, France
e-mail: joel.janin@u-psud.fr

I. Roterman-Konieczna (ed.), *Identification of Ligand Binding Site and Protein-Protein Interaction Area*, Focus on Structural Biology 8, DOI 10.1007/978-94-007-5285-6_5,
© Springer Science+Business Media Dordrecht 2013

5.1 Introduction

The specific recognition between two proteins is the physical process that governs the construction of the macromolecular machines and assemblies which carry out most biological functions in cells and living organisms. Ubiquitous and essential to life, protein-protein recognition has in recent years become a major subject of study in post-genomic molecular biology, biochemistry, structural biology, and biophysics. When structural data are available, it can also be approached computationally, by docking simulations in which a protein-protein complex is assembled from the component structures. We relate here how protein-protein docking was attempted in the early 1970s, preceding small molecule docking at a time where very few proteins had a known three-dimensional structure, and how it developed into a family of novel algorithms after 1990. Since then, docking algorithms have turned into structural prediction procedures, and their reliability has been tested in the CAPRI blind prediction experiment. An outcome of the test was that the initial model of recognition in which the proteins bind as rigid bodies, progressively evolved into one of flexible recognition. The new paradigm takes into account the structure changes that may accompany the association reaction, and offers estimates of their effect on the stability of the assembly that the reaction produces, and on the specificity of the recognition process.

5.2 An Early History of Protein Docking

The first attempt to model the self-assembly of two proteins concerned trypsin and the bovine trypsin pancreatic inhibitor (BPTI). David Blow of Cambridge, UK, and Robert Huber of Martinsried, Germany, respectively authors of the α-chymotrypsin and BPTI X-ray structures, teamed to build an atomic model of the trypsin/BPTI complex. Their paper (Blow et al. 1972) does not say how they did it, only that "when a model of the relevant part of the inhibitor was compared with the active site of α-chymotrypsin, it was evident that only one mode of binding was possible". At the time, "model" meant a physical wire model, not one a computer could handle, and no atomic coordinates of the complex remain to assess its accuracy. Beddell et al. (1976) still used a wire model to do molecular modeling at the Wellcome Research Laboratories in Kent, UK. They engineered biphenyl compounds to bind at the DPG (2,3-diphospho-glycerate) site of hemoglobin. Some of the compounds did, and they had the predicted effects on oxygen binding, possibly the first success of structure-based drug design. While the Wellcome scientists had access to hemoglobin atomic coordinates from Pr. Max Perutz, their paper says that "a more accurate representation was needed", and they chose to build a wire model. Their designs were based on interactions predicted from that model, not computation.

Nevertheless, Perutz' hemoglobin coordinates had already been used to do molecular modeling in the computer, and more specifically, to dock proteins together. Pr. Cyrus Levinthal of Columbia University, New York, had devised an

algorithm that could build a model of the sickle cell fiber from individual hemoglobin molecules (Levinthal et al. 1975). Shoshana Wodak, one of Levinthal's co-authors, joined me in Pr. Georges Cohen's laboratory at the Pasteur Institute in Paris, and we decided together to investigate what computer simulations could tell us about protein-protein recognition. For that purpose, we designed a procedure that generated all the orientations of one protein relative to another, and brought the two surfaces into contact by translation. To gain computer time and memory space, we borrowed from Michael Levitt a simplified protein model that represented each amino acid residue by a sphere of appropriate radius (Levitt 1976). We allowed a degree of penetration between the spheres, and estimated the quality of the fit by the number of intersubunit residue-residue contacts. Our test system was the same trypsin/BPTI complex as in Blow et al. (1972), but by then, Huber's lab had determined a X-ray structure of the complex (Huber et al. 1974), and issued coordinates that could serve to assess the accuracy of the docking models. In the summer of 1976, we were given access to a state-of-the-art computer in Orsay, France – one that was only about 10,000 times slower than a laptop today – during a workshop of the Centre Européen de Calcul Atomique et Moléculaire (CECAM). In about an hour of cpu time, our software (named DOCK like several others after it) generated models of the inhibitor filling the active site of the protease in 2,300 different orientations. To our satisfaction, an orientation close to Huber's X-ray structure showed a good fit, but there were several other that achieved a similar score. In other terms, the procedure had produced a native-like model of the assembly, plus some false positives. We attributed the false positives to the coarse nature of our score, which took into account the geometric complementarity of the two molecular surfaces, but ignored their chemical nature and the physics of their interaction (Wodak and Janin 1978).

Computational biology had no established status in the mid-1970s, and we had a difficult time convincing journal editors that protein-protein docking was more than a futile game. Yet, Levinthal had addressed a related question, protein folding, several years before, and ambitious attempts were already being made to solve it in the computer (Levitt and Lifson 1969; Levitt 1976; Némethy and Scheraga 1977). With rigid molecules, docking is a much simpler problem than protein folding. Whereas folding has thousands of degrees of freedom, docking has only six, and by restricting the search to the active site of trypsin, we had reduced that number to four, which had made the calculation feasible.

The next application of our software was to simulate the allosteric transition of hemoglobin. Hemoglobin is an order of magnitude larger than BPTI, but its twofold symmetry also reduces the search to four degrees of freedom, and the computation was within the reach of extant computers. It was done at the Free University, Brussels, in the summer of 1981, also during a CECAM workshop (Janin and Wodak 1981). We used a much improved version of DOCK to build hemoglobin tetramers from alpha-beta dimers in a range of orientations that covered the T and R quaternary structures described by Perutz. The results showed that the allosteric transition from R to T could not proceed along a linear pathway, due to steric hindrance at the dimer-dimer interface, and it drew an alternative pathway in excellent agreement with the classical description of Baldwin and Chothia (1979).

5.3 Protein-Protein Docking Algorithms

5.3.1 Bound vs. Unbound Docking

The hemoglobin simulation faced even more editorial skepticism than the trypsin/BPTI study. By the time it got published (Janin and Wodak 1985), small molecule docking had come of age in the hands of Kuntz et al. (1982), Goodford (1985), and a few others. Soon, it became an established procedure in drug design, while protein-protein docking remained confidential for over a decade. Meanwhile, computers became orders of magnitude faster, and crystallographers determined many new structures. The latter included a score of protease/inhibitor complexes, and the first antigen/antibody complexes (Janin and Chothia 1990). Cherfils et al. (1991) tested on those complexes the Wodak-Janin algorithm, implemented as a simulating annealing procedure to make the search more efficient. This allowed all six degrees of freedom to be explored, and most importantly, "unbound" docking to be tested for the first time. Unbound docking uses the atomic coordinates of the free proteins, bound docking, coordinates taken from the complex. Bound docking ignores the conformation changes that may accompany association, and it has no predictive value, since the solution must be known in advance. The new study yielded native-like models of all the target complexes, and a majority of those models scored near the top. However, there were many false positives, especially with the unbound proteins, and it was evident that other features than shape complementarity had to be taken into account to identify the correct docking models among all the false positives.

5.3.2 Rigid-Body Docking

The early 1990s were a period of renewed interest in protein-protein docking. Several new algorithms, all based on geometry and shape complementarity, were published almost simultaneously. Connolly (1986) had devised a procedure in which molecular surfaces were described by sets of discrete points; matching critical points (holes and pits) of two surfaces assessed their complementarity, and this could be used for docking. A related method of surface triangulation, independently developed for "computer-vision" by Pr. Haim Wolfson of Tel Aviv University in Israel, was implemented into a docking procedure through a very efficient geometric hashing algorithm (Nussinov and Wolfson 1991; Norel et al. 1994). In Berkeley, California, Jiang and Kim (1991) designed a "cube representation" of proteins specifically for docking. In that model, the surface of the proteins and their interior volume are sampled on a cubic grid, and a docking pose is generated by matching surface cubes while rejecting overlaps between volume cubes. Jiang and Kim made a very important point: docking must be "soft" to allow for minor conformation changes. The cube model, like the residue sphere model of the Wodak-Janin procedure, made for that softness by blurring the atomic details of the protein structures.

The cubic grid representation is an essential element of the FFT correlation docking algorithm published soon afterwards by Katchalski-Katzir et al. (1992) of the Weizmann Institute in Israel. To start with, one picks an orientation of a protein relative to the other, and assigns appropriate weights to grid points of the surface and the interior volume of the two molecules. The correlation between the two sets of weights is used as a score. It may be written as a convolution product, and efficiently computed for all translations at one time thanks to the Fast Fourier Transform (FFT) algorithm. Then, the orientation is changed and the calculation repeated. The method has been very successful, and it has benefited from many developments (Vakser and Aflalo 1994; Gabb et al. 1997; Ritchie and Kemp 2000; Mandell et al. 2001; Heifetz et al. 2002; Chen et al. 2003a). Whereas the original formulation of the algorithm assessed only the geometric complementarity, other molecular features can be encoded as weights on a cubic grid; for instance, an electrostatic interaction energy may be calculated by correlating the electric charges on one protein with the electric field created by the other protein. Electrostatics, hydrophobicity, and a number of other terms may be combined into a scoring function. Each of the Web sites listed in Table 5.1 has its own scoring function, and its own way to calculate its terms as FFT correlations.

5.3.3 Monte-Carlo and Related Docking Algorithms

Albeit "soft", the FFT correlation and the geometric hashing algorithms explore only the six degrees of freedom of rigid-body docking. Other algorithms developed afterwards handle other variable parameters, dihedral angles for instance, in order to simulate side chain rotations and main chain conformation changes. They take a heuristic approach to the problem, instead of performing an exhaustive search. Monte-Carlo simulated annealing, the choice method in the 1990s, allowed Totrov and Abagyan (1994) to adjust side chain conformations at the same time as the docking search. These authors employed a detailed atomic model and a standard molecular mechanics force field, which was computationally very expensive. Instead, all the later docking procedures based on simulated annealing or related algorithms, proceed in two or more steps. The first step explores the rigid-body parameter space with a simplified protein model and a coarse force field, the second carries out a detailed refinement of the local minima (Fernández-Recio et al. 2002; Zacharias 2003). The RosettaDock procedure (Gray et al. 2003) is a good example: a first Monte-Carlo search is carried out on a low-resolution protein model with residue-level potentials; it identifies many (a thousand or more) candidate solutions, which are refined afterwards using a full-atom model and the Rosetta force field. That force field, optimized on protein data, includes terms for desolvation or rotamer preferences not present in standard force fields. It performs very well in protein folding, its original application, and also in docking, at least when the conformation changes are of limited amplitude (Schueler-Furman et al. 2005).

Table 5.1 Web servers for protein-protein docking

Protein structure and benchmark sets	
Protein Data Bank (PDB)	http://www.rcsb.org/pdb/
CAPRI experiment	http://capri.ebi.ac.uk/
Docking benchmark	http://zlab.bu.edu/zdock/benchmark.shtml
Structure/affinity benchmark	http://bmm.cancerresearchuk.org/~bmmadmin/Affinity
FFT correlation and related docking algorithms	
ClusPro	http://cluspro.bu.edu/login.php
DOT	http://www.sdsc.edu/CCMS/Papers/DOT_sc95.html
FTDOCK	http://www.sbg.bio.ic.ac.uk/docking/ftdock.html
GRAMM-X	http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/
HEX	http://www.loria.fr/~ritchied/hex/
MolFit	http://www.weizmann.ac.il/Chemical_Research_Support//molfit/
ZDOCK	http://zlab.bu.edu/zdock/
Molecular dynamics, Monte-Carlo and related flexible docking algorithms	
ATTRACT	http://www.ibpc.fr/chantal/www/ptools/
HADDOCK	http://www.nmr.chem.uu.nl/haddock/
ICM-DISCO	http://www.molsoft.com/icm_pro.html
RosettaDock	http://graylab.jhu.edu/docking/rosetta/
Geometric hashing and related flexible docking algorithms	
PatchDock	http://bioinfo3d.cs.tau.ac.il/PatchDock
FireDock	http://bioinfo3d.cs.tau.ac.il/FireDock/
SymmDock	http://bioinfo3d.cs.tau.ac.il/SymmDock
FiberDock	http://bioinfo3d.cs.tau.ac.il/FiberDock/
MultiFit	http://salilab.org/multifit/ and http://bioinfo3d.cs.tau.ac.il/
3D-Garden	http://www.sbg.bio.ic.ac.uk/3dgarden
SKE-Dock	http://www.pharm.kitasato-u.ac.jp/bmd/files/SKE_DOCK.html

5.3.4 Template-Based Docking

An alternative to docking is to use a template, and build a model of a protein-protein complex by analogy to one of known structure. When both components of two complexes are close homologs with a high level of sequence identity (40% or more), it is straightforward to model build both the components and their assembly, but the method has a very limited field of application. It can be extended by accepting templates with a low level of sequence identity, or templates that have similar three-dimensional structures irrespective of their sequences, under the assumption that the mode of interaction is conserved (Lu et al. 2002; Sinha et al. 2010; Kundrotas et al. 2012). Although the limits of validity of this assumption are uncertain, genome-wide

libraries of model assemblies have been built in this way (Lu et al. 2003; Stein et al. 2011). Templates may also be selected on the basis of the local similarity of the protein surfaces: two surfaces that have a similar geometry and similar physical-chemical features may be expected to make similar interactions (Günther et al. 2007; Keskin et al. 2008), in which case the PDB may already be adequate to represent the diverse architectures observed in nature (Tuncbag et al. 2008; Kundrotas et al. 2012). Here again, the quality of the models remains to be assessed.

5.4 Assessing Docking Predictions: The CAPRI Experiment

5.4.1 CAPRI

By the turn of the century, several docking algorithms had developed into full-fledged prediction procedures (see reviews by Smith and Sternberg 2002; Camacho and Vajda 2002; Halperin et al. 2002). At that time, an entirely new field of application opened, due to the structural genomics (or proteomics) initiatives that accompanied the completion of the human genome sequence. High-throughput X-ray and NMR studies were going to determine the structure of thousands of new proteins that would include the components of many binary or larger assemblies. Docking procedures could in principle build models of these assemblies from the component structures, but should we trust the results at all? The procedures had been thoroughly tested, but most of the unbound docking tests had been done on protease/inhibitor or antigen/antibody complexes, the only ones for which the component structures were available. How would docking perform on new, possibly very different, systems, and how accurate would the models be?

These questions were discussed in Charleston, South Carolina, in June 2001, at a meeting on Modeling Protein Interactions in Genomes organized by Pr. Sandor Vajda and Ilya Vakser, and the conclusion was that a blind prediction experiment should be organized (Vajda et al. 2002). Named CAPRI (Critical Assessment of PRedicted Interactions), the experiment was modeled after CASP (Critical Assessment of Structural Predictions), an older experiment that tests methods to predict a protein fold based on its amino acid sequence (Moult et al. 1995). The targets of CAPRI would be protein-protein complexes, and the prediction start from component structures taken from the Protein Data Bank. The predictors would dock the components, and submit models to the CAPRI Website, to be assessed by comparison with a newly determined, but unpublished, experimental structure of the complex (Janin et al. 2003). A blind prediction of that sort had been done once before, on a β-lactamase in complex with a protein inhibitor. Six participant groups had submitted models of the complex that were close to the X-ray structure (Strynadka et al. 1996). Could that performance be reedited?

An answer came soon after the Charleston meeting. The first round of CAPRI, held in the summer of 2001, had three targets, three complexes whose X-ray structures had just been determined by collaborators of mine, willing to help starting the experiment. Two were viral antigen proteins in complex with monoclonal antibodies,

the third, a bacterial protein kinase co-crystallized with its substrate, the small protein HPr. Fifteen predictor groups submitted a total of 193 models, and the CAPRI assessors led by S. Wodak, compared them to the X-ray structures. The assessors found that the submissions contained good models of the two antigen/antibody complexes, but not of the HPr/kinase complex (Méndez et al. 2003). They nevertheless decided that a few of the HPr/kinase models were "acceptable": their geometry was poor, but most of the residues in the contact regions were correctly predicted, which could in principle help designing experiments. Predicting the residues in contact was not a big feat in that case, since the location of the kinase active site and the serine residue phosphorylated on HPr were known from the literature. Moreover, the poor geometry of the models had an obvious origin: in the X-ray structure of the complex, the rotation of a α-helix in the kinase modified the shape of the substrate binding site and the way it bound HPr (Fieulaine et al. 2002). Thus, rigid-body docking was able to locate the correct epitopes on the two viral antigens, and place them correctly at the antibody combining sites, but it failed on HPr/kinase due to a conformation change, albeit one of limited amplitude.

5.4.2 Success and Failure in Blind Predictions

This pattern was repeatedly observed in later prediction rounds (Méndez et al. 2005; Lensink et al. 2007; Lensink and Wodak 2010; Janin 2005, 2010). In the 10 years that followed the Charleston meeting, CAPRI has had 22 rounds, with a total of 43 targets and an average of 45 predictor groups, each submitting ten models of each target. In addition to protein-protein complexes, the targets have been a protein-RNA complex and four oligomeric proteins. For each target, the predictors were given the coordinates of the unbound components, or of an homolog protein that could be used for model building, and they had 3–6 weeks to make their prediction and submit their models. A majority (70%) of the targets obtained good quality models. Almost all those that displayed only small backbone movements did, and in most cases, the good models came from several groups using different docking procedures. Figure 5.1 shows an example. Target T37, drawn here after the X-ray structure of Isabet et al. (2009), is a complex between the G-protein Arf6, a member of the Ras family of small GTPases, and the LZ2 segment of JIP4 (JNK-interacting protein 4), an effector of Arf6. LZ2 was known to form a leucine zipper, and it had to be model built from its amino acid sequence before docking on Arf6. A standard leucine zipper yields a rather accurate model of its structure in the complex, while Arf6 undergoes little change in the interaction. Correspondingly, the submissions contained a number of good quality models of LZ2/Arf6, submitted by nine different groups (Lensink and Wodak 2010).

On the other hand, CAPRI predictions have yielded at best "acceptable" models of the targets in which the backbone changes were large, or the homology models of poor quality. Prediction yielded no valid model at all in six cases. In two, the failure could be traced to misleading biochemical information rather than the structure itself, in the other four, to large conformation changes. Moreover, some of the targets

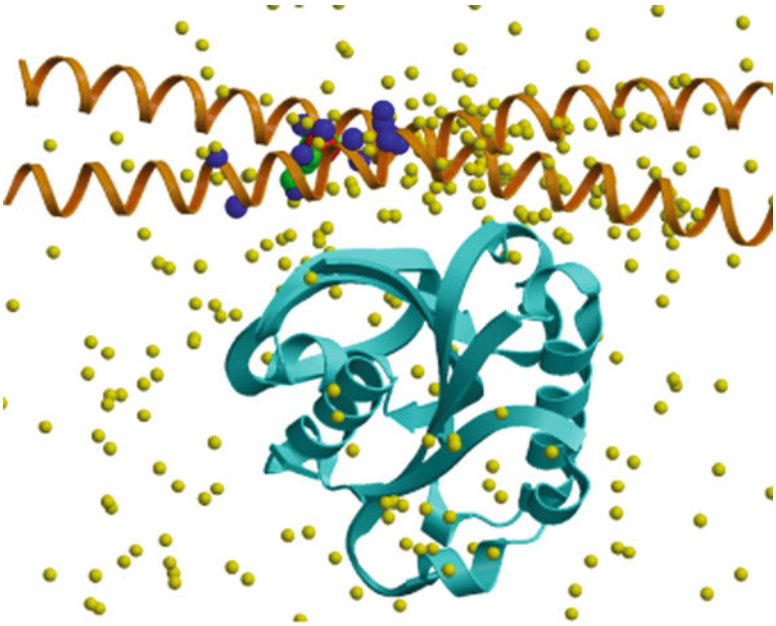

Fig. 5.1 A successful docking prediction. Target T37 was submitted for blind prediction during Round 16 of CAPRI, held in November 2008. The target, a complex of the small GTPase Arf6 with the LZ2 leucine zipper of the JIP4 effector protein, was a gift of Dr. Julie Ménétrey (Institut Curie, Paris). Predictors were given an unbound Arf6 structure, and the amino acid sequence of LZ2, which they had to model build before docking on Arf6. The figure represents the X-ray structure (PDB code 2 W83, Isabet et al. 2009) with Arf6 in cyan, LZ2 in *pink*. The dots are the centers of mass of LZ2 in the models submitted by the 39 predictor groups and the 11 scorer groups who participated in Round 16. The dots are *green* for good quality models, *blue* for "acceptable" models, and *yellow* for incorrect models. All the models can be accessed at http://www.ebi.ac.uk/msd-srv/capri/round16/ (Courtesy of Dr. Marc Lensink (Lille))

were subjected twice for prediction, first with both components unbound, then with the more flexible component in its bound conformation. The second step always yielded much better models; for instance, prediction of the protein/RNA complex failed with the unbound RNA, but there were many good models with the bound RNA, which has a very different conformation (Lensink and Wodak 2010).

5.5 Flexible Docking and the Scoring Experiment

5.5.1 Simulating Conformation Changes and Mechanisms of Recognition

Two very important objectives of CAPRI were to stimulate the development of new methods, and create a forum where they could be discussed and information would spread within the community. The experiment succeeded on both grounds.

It generated lively discussions on line through the CAPRI Website hosted by the European Bioinformatics Institute (Hinxton, UK), and face-to-face during assessment meetings that took place at regular intervals. Moreover, CAPRI is at the origin of most of the progress seen in the last 10 years. The new scoring functions, the methods to model conformation changes and the flexible docking procedures developed since 2001, owe much to the experiment. Docking searches based on simulated annealing and molecular dynamics have been adapted to reproduce loop movements or the rotation of a structural domain about a pre-defined hinge, with good results on some CAPRI targets (Janin 2005; Lensink et al. 2007; Lensink and Wodak 2010). However, a more general solution to the problem of flexible docking is to generate conformers of the two components prior to the search, and assemble them pairwise (Grünberg et al. 2004; Bonvin 2006; Lesk and Sternberg 2008; Dobbins et al. 2008; Ritchie 2008; Zacharias 2010). A recent, and valuable, application of the method is implemented in the MultiFit server (Table 5.1); it allows multiple conformers generated from a X-ray structure to be fitted in electron microscopy images that have a much lower resolution, but may display significant conformation changes relative to the atomic model (Tjioe et al. 2011).

The different approaches to the problem of conformation changes in docking correspond to different possible recognition mechanisms. Rigid-body docking mimics the specific recognition between two proteins that bear complementary surfaces, ready to interact when they come into contact. Monte-Carlo searches with variable dihedral angles simulate an induced fit mechanism in which the components first make a low-stability, low-specificity contact, and then adjust their conformation to optimize their interaction. Docking conformers pairwise closely reproduces conformer selection, an alternative to induced fit. In this mechanism, a minority of the molecules have the conformation that allows rigid-body recognition to start with, and it is the formation of a complex that causes the equilibrium to shift (Grünberg et al. 2004, 2006).

5.5.2 Scoring in CAPRI

Docking conformers pairwise multiplies the number of searches, and this is practical only with a fast docking algorithm (Schneidman-Duhovny et al. 2005; Mashiach et al. 2010). In addition to being computationally demanding, the method generates a great number of false positives, and puts a heavy load on the scoring functions. In recent years, CAPRI has been adapted to assess scoring separately from docking. The scoring part of the experiments operates in this way: after a prediction round is completed, the predictor groups are asked to upload a hundred or so of their models, which are merged into a file that may contain a thousand models, issued from ten or more different procedures. The scorer groups download the file and rank the whole set, to make their own ten-model submission. In several cases, target T37 of Fig. 5.1 for instance, the scorers' submissions contained more accurate models than the

predictors' ones. These models came from the same docking searches, but the predictors' procedures scored them low, whereas some of the schemes developed by scorer groups adequately identified them as correct (Lensink and Wodak 2010).

A scoring function can be physical-chemical (force fields, solvation energies), or empirical, combining terms from different origins with weights optimized on sets of positive and negative examples. It may include non-structural information derived from the comparison of homologous sequences, from point mutants or other genetic or biochemical experiments. However, such information is often ambiguous, and sometimes misleading. If external information is used to screen models during or after the search, it should be treated as a flexible restraint rather than a rigid constraint. The HADDOCK procedure efficiently incorporates such information into a search algorithm that can also handle data from other sources, NMR experiments for instance (Dominguez et al. 2003; de Vries et al. 2007, 2010; Stratmann et al. 2011).

5.5.3 Flexibility and the Docking Benchmark

Developing scoring functions is an active field of research in many fields of science, but in docking, the main difficulty remains flexibility. The structures deposited in the Protein Data Bank illustrate many kinds of conformation changes, the docking benchmark of Weng and colleagues, also. The benchmark is a set of PDB entries assembled to test docking procedures. It contained only 59 complexes in its first version (Chen et al. 2003b), but now has entries for 176 protein-protein complexes and their unbound components; one-third display significant backbone movements with root-mean-square amplitudes that range from 1.5 to 10 Å (Hwang et al. 2010, and Table 5.1).

The complexes of the benchmark are implicated in all sorts of biological processes. Antigen/antibody and enzyme/inhibitor complexes are no longer a majority. Signal transduction and cellular trafficking (exemplified by Arf6 in Fig. 5.1) are well represented, and the protein-protein complexes involved these processes offer many examples of flexible recognition. Conformation changes mediate signal transduction in many ways: they may change the affinity of a protein for a small ligand, another protein or DNA, enhance or inhibit a catalytic activity, the GTPase activity of a G-protein for instance, mask or reveal a group that governs the cellular localization of the protein or its attachment to a membrane. Their variety is immense, comparable in principle to the variety of macromolecular interactions seen in nature, which neither the docking benchmark nor the PDB itself, are close to cover. Moreover, entire classes of interactions are missing: those that involve membrane proteins and intrinsically disordered proteins (IDP), for instance. IDP are implicated in many macromolecular interactions (Dunker et al. 2005, 2008; Tompa et al. 2009), and they undergo disorder-to-order transitions when they interact with other components. Simulating such transitions in the context of docking will remain a challenge for many years.

5.6 Designing Interactions and Predicting Affinity

5.6.1 Engineering Novel Protein-Protein Interactions

Docking can serve other purposes than predicting structures. In Seattle, David Baker, who developed Rosetta, uses docking to engineer novel interactions. The procedure starts by selecting a pair of protein scaffold structures; a coarse-grain docking search identifies candidate complexes; they are computationally mutated at a few interface sites, the modeled mutant complexes are energy-refined, and the top-scoring solutions selected for cloning and expression in yeast. A first experiment aimed to generate a stable interaction between an ankyrin repeat protein and a set of 37 small, structurally diverse, proteins (Karanicolas et al. 2011). A second experiment targeted the stem region of the flu virus hemagglutinin, aiming to mimic the way a neutralizing antibody binds to that epitope (Fleishman et al. 2011a). Both yielded protein constructs that showed reproducible binding, and a round of in vitro evolution was sufficient to improve their affinity to K_d values below nanomolar. Moreover, two co-crystal structures showed that the binding modes had been correctly modeled, although in one, the ligand was oriented 180° away from the model (Karanicolas et al. 2011).

This remarkable piece of protein engineering demonstrates that rational design is now capable to create functional interactions de novo. However, the success rate was low. In the flu hemagglutinin experiment, computational design had culled some 260,000 docking models down to 88 candidate binders derived from 79 different protein scaffolds, but when the constructs were expressed and tested in yeast, only two actually bound (Fleishman et al. 2011a). Nevertheless, the Rosetta force field had predicted about the same binding energies for the designs that failed and for the natural complexes of the Weng docking benchmark. To improve the success rate, a more accurate force field, or a more discriminative scoring function, was clearly required.

5.6.2 The CAPRI Affinity Prediction Experiment

The Seattle group decided to put the question to the CAPRI community: given the structure of a designed complex, can one predict whether it will be stable or not? And they submitted as targets of the scoring experiment a total of 108 designs, including two that bound, during two successive CAPRI rounds held in 2010. The scorers were asked to estimate the affinity of the designed complexes, and rank them along with the complexes of the docking benchmark. When the submissions were analyzed, none of the scorers had ranked the natural complexes significantly above the designs (Fleishman et al. 2011b). Moreover, of the two designs that bound, one had been predicted to be stable by two groups, the other, by no one, a result not far from random. The obvious conclusion of this experiment was that the scoring

functions used in docking did not yield reliable binding energies. They had been developed to identify the correct mode of assembly of two proteins known to interact, not to determine whether or not they form a stable complex, and this was beyond their capacity. A parallel study showed a very poor correlation between experimental binding energies and values calculated with several scoring procedures (Kastritis and Bonvin 2010), with the same conclusion that the latter could not predict affinity.

5.6.3 A Structure Affinity Benchmark

The binding energy of a complex, or more correctly its Gibbs free energy of dissociation ΔG_d derived from the equilibrium constant K_d, is a convenient measure of affinity. K_d is known from biophysical measurements in solution for many protein-protein complexes that have been studied by crystallography, and a number of authors have attempted to derive ΔG_d from these structures. The first were Horton and Lewis (1992). They collected data on 16 protein-protein complexes of known structure (mostly protease/inhibitor complexes at that time), and found that a model based on just the size and chemical composition of the interface yielded ΔG_{calc} values that were within 1 or 2 kcal.mol^{-1} of the measured ΔG_{exp}. However, there was an exception: their model predicted a very similar affinity for BPTI binding to trypsin and trypsinogen, whereas the experimental values differed by 10 kcal.mol^{-1}. Horton and Lewis knew the reason why, and their paper discusses it. Trypsinogen, an inactive precursor of trypsin, has flexible surface loops that become ordered when BPTI binds (Bode et al. 1978). As a result, its affinity for the inhibitor is orders of magnitude less than trypsin, where no such change occurs, even though the two complexes with BPTI are nearly identical in structure.

Like trypsin, most of the proteases and inhibitors of the Horton-Lewis set bind as rigid bodies, with no major conformation change to affect their thermodynamic stability. Later studies of the affinity/structure relationship in protein-protein complexes employed larger data sets and more elaborate models of ΔG_{calc}. But as none took into account the structure of the free proteins, they all ignored the role of conformation changes, and also the large effect that experimental conditions, especially pH, can have on K_d. Not surprisingly, the correlation between ΔG_{calc} and ΔG_{exp} was poor in these studies. In addition, errors accumulated in the structure/affinity sets that served to optimize or test the models, as each study re-used data collated by previous ones. Many of the experimental values in the sets were incorrect, some grossly so; for instance, trypsinogen/BPTI and trypsin/BPTI were given the same ΔG_{exp}, a 10 kcal.mol^{-1} error. There was an obvious need for a validated test set, and in 2010, I teamed with three other groups to assemble a benchmark set of binary complexes that would have (a) experimental structures for both the complex and its components; (b) a reliable K_d measured under well-defined conditions. The 176 complexes of the Weng docking benchmark satisfied condition (a). They were an obvious starting point, and we undertook to scan the biochemical literature in search of a K_d for them.

To our great satisfaction, we could locate thermodynamic data for most of the docking benchmark, although some complexes had to be replaced by homologs that also satisfied condition (a). The K_d values, which cover a wide range from 10^{-5} to 10^{-14} M, are derived from either a titration, mostly ITC (isothermal titration calorimetry), or from the binding kinetics (surface plasmon resonance); a few are from enzymic inhibition. The present version of the structure/affinity benchmark comprises 144 complexes, and includes nine pairs that have very similar structures and very different affinities, due to differences in conformation or in sequence. For each entry, the benchmark cites PDB codes for the complex and its components, the K_d and ΔG_d values with the method and experimental conditions of their measurement, and the relevant literature references (Kastritis et al. 2011, and Table 5.1).

5.7 Conclusion

The major achievement of protein-protein docking has been its contribution to our understanding of macromolecular interaction. Docking simulations demonstrate that the shape and chemical complementarity of the molecular surfaces is the major determinant in rigid-body recognition, which is a valid approximation in a number of biological systems. Then, docking has a high predictive value, confirmed by CAPRI and by experiments in which novel interactions are rationally designed *de novo*. However, many processes of great biological importance rely on flexible recognition, in which case the molecular surfaces become complementary only as a result of conformation changes. The CAPRI targets that display flexible recognition have stimulated new developments in the field of docking. Albeit still be far from routine, methods to predict and simulate conformation changes have reached the stage where they can produce useful models, and this has relevance to other fields. In structural biology, much effort is made to fit the atomic resolution structure of assembly components into lower resolution images from cryo-electron microscopy, or an envelope derived from small-angle X-ray scattering, while allowing the structure to change. This is a typical flexible docking problem, to which some docking algorithms have already been applied. In drug design, the target proteins often make other interactions than the one of interest. This may induce conformation changes and allosteric effects that should be taken into account in the design procedure. Similarly, computational biologists may want to study how protein folding is affected by external interactions, in a homodimer for instance. Beyond the structure, we want to understand what governs the specificity of macromolecular recognition and the stability of protein assemblies. This implies that we should be able to model the thermodynamics and the mechanism of the association reaction. The recent attempt to predict affinity within the CAPRI experiment suggests that present force fields are inadequate, and new methods must be developed. The structure/affinity benchmark assembled on this occasion should help biophysicists to correlate function to structure, and remind them that the structure may change as new interactions are formed.

Acknowledgements JJ acknowledges a long term collaboration with Pr. Shoshana Wodak (Free University, Brussels, and University of Toronto), fostered in its early days by Dr. Georges Cohen (Institut Pasteur, Paris), the European Molecular Biology Organisation, and Centre Européen de Calcul Atomique et Moléculaire. The CAPRI experiment owes much to R. Mendez and M. Lensink who perform the assessment together with S. Wodak, to K. Henrick and S. Velenkar who manage its Web site at the European Bioinformatics Institute (Hinxton, UK), to J. Moult, S. Vajda, I. Vakser, M. Sternberg, and L. Ten Eyck, who assist in its management, and to all the structural biologists who have contributed targets.

References

Baldwin J, Chothia C (1979) Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. J Mol Biol 129:175–220

Beddell CR, Goodford PJ, Norrington FE, Wilkinson S, Wootton R (1976) Compounds designed to fit a site of known structure in human haemoglobin. Br J Pharmacol 57:201–209

Blow DM, Wright CS, Kukla D, Rühlmann A, Steigemann W, Huber R (1972) A model for the association of bovine pancreatic trypsin inhibitor with chymotrypsin and trypsin. J Mol Biol 69:137–144

Bode W, Schwager P, Huber R (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 A resolution. J Mol Biol 118:99–112

Bonvin AM (2006) Flexible protein-protein docking. Curr Opin Struct Biol 16:194–200

Camacho CJ, Vajda S (2002) Protein-protein association kinetics and protein docking. Curr Opin Struct Biol 12:36–40

Chen R, Li L, Weng Z (2003a) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52:80–87

Chen R, Mintseris J, Janin J, Weng Z (2003b) A protein-protein docking benchmark. Proteins 52:88–91

Cherfils J, Duquerroy S, Janin J (1991) Protein-protein recognition analyzed by docking simulation. Proteins 11:271–280

Connolly ML (1986) Shape complementarity at the hemoglobin $\alpha 1$ $\beta 1$ subunit interface. Biopolymers 25:1229–1247

de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 69:726–733

de Vries SJ, van Dijk M, Bonvin AM (2010) The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 5:883–897

Dobbins SE, Lesk VI, Sternberg MJ (2008) Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. Proc Natl Acad Sci USA. 105:10390–5

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125:1731–1737

Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272:5129–5148

Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. Curr Opin Struct Biol 18:756–764

Fernández-Recio J, Totrov M, Abagyan R (2002) Soft protein-protein docking in internal coordinates. Protein Sci 11:280–291

Fieulaine S, Morera S, Poncet S, Mijakovic I, Galinier A, Janin J, Deutscher J, Nessler S (2002) X-ray structure of a bifunctional protein kinase in complex with its protein substrate HPr. Proc Natl Acad Sci USA 99:13437–13441

Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D (2011a) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332:816–821

Fleishman SJ, Whitehead TA, Strauss EM et al., Wodak SJ, Janin J, Baker D (2011b) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol Accepted manuscript doi:10.1016/j.jmb.2011.09.031

Gabb HA, Jackson RM, Sternberg MJ (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 272:106–120

Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 28:849–857

Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 331:281–299

Grünberg R, Leckner J, Nilges M (2004) Complementarity of structure ensembles in protein-protein binding. Structure 12:2125–2136

Grünberg R, Nilges M, Leckner J (2006) Flexibility and conformational entropy in protein-protein binding. Structure 14:683–693

Günther S, May P, Hoppe A, Frömmel C, Preissner R (2007) Docking without docking: ISEARCH–prediction of interactions using known interfaces. Proteins 69:839–844

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47:409–443

Heifetz A, Katchalski-Katzir E, Eisenstein M (2002) Electrostatics in protein-protein docking. Protein Sci 11:571–587

Horton N, Lewis M (1992) Calculation of the free energy of association for protein complexes. Protein Sci 1:169–181

Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W (1974) Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 A resolution. J Mol Biol 89:73–101

Hwang H, Vreven T, Janin J, Weng Z (2010) Protein-protein docking benchmark version 4.0. Proteins 78:3111–3114

Isabet T, Montagnac G, Regazzoni K, Raynal B, Elkhadali F, Franco M, England P, Chavrier P, Houdusse A, Menetrey J (2009) The structural basis of Arf effector specificity: the crystal structure of ARF6 in a complex with JIP4. EMBO J 28:2835–2845

Janin J (2005) Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci 14:278–283

Janin J (2010) Protein-protein docking tested in blind predictions: the CAPRI experiment. Mol Biosyst 6:2351–2362

Janin J, Chothia C (1990) The structure of protein-protein recognition sites. J Biol Chem 265:16027–16030

Janin J, Wodak SJ (1981) Report on the workshop on non-bonded interactions and the specificity of protein association and folding. CECAM, Orsay, France pp 5–64

Janin J, Wodak SJ (1985) Reaction pathway for the quaternary structure change in hemoglobin. Biopolymers 24:509–526

Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ (2003) CAPRI: a critical assessment of PRedicted interactions. Proteins 52:2–9

Jiang F, Kim SH (1991) "Soft docking": matching of molecular surface cubes. J Mol Biol 219:79–102

Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D (2011) A de novo protein binding pair by computational design and directed evolution. Mol Cell 42:250–260

Kastritis PL, Bonvin AM (2010) Are scoring functions in protein-protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res 9:2216–2225, Erratum in: J Proteome Res 10:921–2, 2011

Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J (2011) A structure-based benchmark for protein-protein binding affinity. Protein Sci 20:482–491

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

Keskin O, Nussinov R, Gursoy A (2008) PRISM: protein-protein interaction prediction by structural matching. Methods Mol Biol 484:505–521

Kundrotas PJ, Zhu Z, Janin J, Vakser IA (2012) Templates are available to model nearly all complexes of structurally characterized proteins. Proc Natl Acad Sci USA.109:9438–41

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288

Lensink MF, Wodak SJ (2010) Docking and scoring protein interactions: CAPRI 2009. Proteins 78:3073–3084

Lensink MF, Méndez R, Wodak SJ (2007) Docking and scoring protein complexes: CAPRI 3rd edition. Proteins 69:704–718

Lesk VI, Sternberg MJ (2008) 3D-Garden: a system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. Bioinformatics 24:1137–1144

Levinthal C, Wodak SJ, Kahn P, Dadivanian AK (1975) Hemoglobin interaction in sickle cell fibers. I: theoretical approaches to the molecular contacts. Proc Natl Acad Sci USA 72:1330–1334

Levitt M (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104:59–107

Levitt M, Lifson S (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol 46(2):269–279

Lu L, Arakaki AK, Lu H, Skolnick J (2003). Multimeric threading-based prediction of protein-protein interactions on a genomic scale: application to the Saccharomyces cerevisiae proteome. Genome Res. 13:1146–54

Lu L, Lu H, Skolnick J (2002) MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. Proteins 49:350–364

Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF (2001) Protein docking using continuum electrostatics and geometric fit. Protein Eng 14:105–113

Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ (2010) An integrated suite of fast docking algorithms. Proteins 78:3197–3204

Méndez R, Leplae R, De Maria L, Wodak SJ (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. Proteins 52:51–67

Méndez R, Leplae R, Lensink MF, Wodak SJ (2005) Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 60:150–169

Moult J, Pedersen JT, Judson R, Fidelis K (1995) A large-scale experiment to assess protein structure prediction methods. Proteins 23(3):ii–v

Némethy G, Scheraga HA (1977) Protein folding. Q Rev Biophys 10:239–252

Norel R, Lin SL, Wolfson HJ, Nussinov R (1994) Molecular surface complementarity at protein-protein interfaces: the critical role played by surface normals at well placed, sparse, points in docking. J Mol Biol 252:263–273

Nussinov R, Wolfson HJ (1991) Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. Proc Natl Acad Sci USA 88:10495–10499

Ritchie DW (2008) Recent progress and future directions in protein-protein docking. Curr Protein Pept Sci 9:1–15

Ritchie DW, Kemp GJ (2000) Protein docking using spherical polar Fourier correlations. Proteins 39:178–194

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) Geometry-based flexible and symmetric protein docking. Proteins 60:224–231

Schueler-Furman O, Wang C, Bradley P, Misura K, Baker D (2005) Progress in modeling of protein structures and interactions. Science 310:638–642
Sinha R, Kundrotas PJ, Vakser IA (2010) Docking by structural similarity at protein-protein interfaces. Proteins 78:3235–3241
Smith GR, Sternberg MJ (2002) Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol 12:28–35
Stein A, Mosca R, Aloy P (2011) Three-dimensional modeling of protein interactions and complexes is going 'omics. Curr Opin Struct Biol 21:200–208
Stratmann D, Boelens R, Bonvin AM (2011) Quantitative use of chemical shifts for the modeling of protein complexes. Proteins 79:2662–2670
Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MN (1996) Molecular docking programs successfully predict the binding of a β-lactamase inhibitory protein to TEM-1 β-lactamase. Nat Struct Biol 3:233–239
Tjioe E, Lasker K, Webb B, Wolfson HJ, Sali A (2011) MultiFit: a web server for fitting multiple protein structures into their electron microscopy density map. Nucleic Acids Res 39:W167–W170
Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. Bioessays 31:328–335
Totrov M, Abagyan R (1994) Detailed ab initio prediction of lysozyme-antibody complex with 1.6 A accuracy. Nat Struct Biol 1:259–263
Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O (2008) Architectures and functional coverage of protein-protein interfaces. J Mol Biol 381:785–802
Vajda S, Vakser IA, Sternberg MJ, Janin J (2002) Modeling of protein interactions in genomes. Proteins 47:444–446
Vakser IA, Aflalo C (1994) Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 20:320–329
Wodak SJ, Janin J (1978) Computer analysis of protein-protein interaction. J Mol Biol 124:323–342
Zacharias M (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci 12:1271–1282
Zacharias M (2010) Accounting for conformational changes during protein-protein docking. Curr Opin Struct Biol 20:180–186