

## Chapter 2

# Multiscale Protein and Peptide Docking

Mateusz Kurcinski, Michał Jamroz, and Andrzej Kolinski

**Abstract** The number of functional protein complexes in a cell is larger by an order of magnitude than the number of proteins. The experimentally determined three-dimensional structures exist for only a very small fraction of these complexes. Thus, the methods for theoretical prediction of structures of protein assemblies are extremely important for molecular biology. Association of two (or more proteins) always induces conformational changes of the individual components. In many cases, these induced changes are relatively small and involve mostly the side chains at the association interface. In such cases, the approaches of rigid-body docking of two (or more) structures are quite successful. Quite frequently, however, the docking-induced conformational changes are significant. In such cases, prediction of the resulting structures is extremely challenging. The cases, where experimental structures of some components do not exist, are yet even more difficult. In this chapter, we briefly overview the existing *in silico* docking methods and describe a multiscale strategy of unrestricted flexible docking of proteins and peptides.

### 2.1 Introduction

In eukaryotic cells, an average protein can participate in several protein–protein (or protein–nucleic acid) complexes. The number of such complexes is larger by an order of magnitude than the number of proteins. Since the number of experimentally solved protein structures (about 60,000) is a small fraction of all proteins, the fraction of structurally annotated protein complexes is very small. Thus, the theoretical, *in silico*, prediction of molecular structures of multimeric protein assemblies is one

---

A. Kolinski (✉)

Faculty of Chemistry, University of Warsaw, Warsaw, Poland  
e-mail: kolinski@chem.uw.edu.pl

of the most important task of bioinformatics and computational biology (Wodak and Janin 1978; Valencia and Pazos 2002; Salwinski and Eisenberg 2003; Aloy and Russell 2004). There are relatively dependable computational methods for so-called rigid docking. These methods are applicable, provided that the structures of individual components are known and the conformational changes of these components induced upon docking are small. For proteins of known structures, the last requirement is approximately fulfilled quite frequently (Ritchie 2008). Then the problem reduces to generation of a large number of possible poses, according to the shape complementarity of the components and scoring of binding poses by interaction patterns of the interfaces. The latter task is by no means trivial since at least some of the side chains at the interface certainly change their conformations with respect to the conformations seen in the monomeric state or in different complexes. At the moment, the knowledge-based statistical potentials, either atom-wise or united-atom-wise ones, seem to be most productive in scoring of protein–protein interactions.

The ultimate goal of the protein docking could be described as follows: having just a set of sequences, find the structure (structures) of the possible assemblies. In general, this may appear to be not feasible, at least at present, but maybe this is not so hopeless. Firstly, with the constant progress in protein structure prediction, mainly via rectification of comparative modeling, now it is possible to predict monomeric structures at least for a half of protein sequences and at least with a moderate resolution. Secondly, provided that bioinformatics methods are developed for identification of structure fragments that may change upon the docking, it should be possible to design methods for semi-flexible docking that accounts for the allowed conformational changes of fragments of the components' structures. A step toward such a docking methodology is described in this chapter. The method employs a multiscale modeling of proteins and peptides. It is based on CABS modeling software. CABS is a high-resolution, coarse-grained, protein modeling tool (the acronym stands for the united atoms representing a residue in a polypeptide chain: CA alpha carbon of the main chain, CB -beta carbon, and S – the center of side group). The CABS protein structure representation is based on united-atom description of protein structure, where a single residue is represented by several (three or four, depending on the size of side chains) united atoms. The conformational space of CABS polypeptide chains is sampled by means of very efficient Monte Carlo schemes. Details of the CABS design are described in the first chapter of this book and in previous publications (Kolinski 2004). The spatial resolution of CABS allows for quite precise reconstruction of atomic details. This reconstruction process (Gront et al. 2007) for main-chain atoms is very fast and accurate within range of few tens of Angstrom. The reconstruction of side groups is less accurate and depends on the achieved accuracy of the C $\alpha$ -trace fold.

Below, we overview briefly the techniques for rigid docking problem, docking with a highly limited flexibility of some structural elements, and then we outline the more flexible (and fully flexible) docking based on a multiscale approach in which the CABS-based structure assembly is the key step of a molecular complex building procedures.

## 2.2 Rigid Docking Procedures

Suppose we know the three-dimensional structures of two proteins that form a dimer, although we do not know how these two proteins are posed in the complex, and which residues form the protein–protein interface. Finding the structure of the resulting complex is not a trivial task. Classical rigid docking consists of two or three fundamental steps (Vajda and Kozakov 2009). The first one is the generation of a large number of binary structures. The second one is scoring the structures according to the shape complementarity and interactions at the interface. Finally, one may perform rectification of the best structures by adjustments of conformations of the side chain at the interface. Finding plausible poses in rigid docking is not trivial – this requires a very effective search algorithm. Fast Fourier Transform makes it possible to reduce the six-dimensional problem to a one-dimensional problem. A number of algorithms have been developed for this purpose (Katchalski-Katzir et al. 1992; Vakser and Aflalo 1994; Vakser 1995; Mandell et al. 2001; Del Carpio-Muñoz et al. 2002; Chen et al. 2003; Carter et al. 2005; Kozakov et al. 2006; Sternberg et al. 1998). Alternatively, various geometric hashing procedures could be used (Fischer et al. 1995). The resulting poses, usually several thousands of them, need to be scored in order to produce a small number of plausible structures. Scoring functions span a wide range, from a simple shape complementarity (Chen et al. 2003), through the knowledge-based statistical potentials (Kozakov et al. 2006; Tobi and Bahar 2006; Zhang et al. 2005; Cerutti et al. 2005), physics-based force fields (Koehl 2006; Sheinerman et al. 2000; Jiang et al. 2002) to data-driven docking, supported by available biochemical information (Res and Lichtarge 2005; Res et al. 2005; Lichtarge et al. 1996; Dominguez et al. 2003; Nilges 1995; Anand et al. 2003; van Dijk et al. 2005). It has been also noted that rigid docking could be achieved in a different way. Instead of *ab initio* computing the assembly structure, sometimes it is more effective to predict binding interfaces of the interacting proteins and then perform docking in a limited conformational space. In some sense, this is yet another variant of data-driven docking (Jones and Thornton 1997; Burgoyne and Jackson 2006).

Efficiency of various approaches to protein docking is systematically evaluated within the framework of community-wide experiments of Critical Assessments of PRediction of Interactions (CAPRI) (Carter et al. 2005; Janin et al. 2003).

## 2.3 Flexible Docking

Usually, although not always, protein association induces conformational changes of the components (Bonvin 2006; Camacho and Vajda 2001; May and Zacharias 2005). In many cases, these conformational changes are essentially limited to the interface side chains (Andrusier et al. 2008). RosettaDock algorithm is well suited to deal with such cases (Gray et al. 2003; Daily et al. 2005; Wang et al. 2005). The procedure starts from rigid docking and then the side chains are optimized using

either a rotamer library or free-space side-chain optimization. Such an approach proven to be very successful in blind predictions within CAPRI (Wang et al. 2007; Schueler-Furman et al. 2005). Recently, Rosetta modeling technology has been applied to fully flexible docking, or rather “folding and docking” (Das et al. 2009), of small homo-oligomeric protein assemblies. The method utilizes available experimental data: nuclear magnetic resonance (NMR) chemical shifts and residual dipolar coupling (RDC). Somewhat similar strategy for semi-flexible docking is adapted in ATTRACT (the name of the algorithm comes from attractive interactions of the interface residues) algorithm (Zacharias 2003; May and Zacharias 2007). The method employs a coarse-grained representation of side chains and the docking procedure consists of two steps: rigid docking and optimization of the resulting poses, allowing for flexibility of the side chains of interface residues. ATTRACT algorithm was also used in docking simulations allowing for backbone flexibility of the loop regions. It has been demonstrated that even limited flexibility improves the docking results for most of the tested cases.

Small conformational changes, induced by docking, could be accommodated to some extent by reduced representations or/and coarse-grained potentials describing the interface residues. This is probably one of the reasons for surprisingly good performance of docking procedures based on just shape complementarity and smoothened details of the surface. Recently it has been shown that smoothed low-resolution representation of the surface residues leads to more consistent shape complementarity (Zhang et al. 2009). Another way to increase recognition specificity of the interfaces could be achieved by use of multibody knowledge-based potentials. For instance, four-body statistical pseudo-potentials proven to be useful in protein–peptide docking, allowing for full flexibility of the peptide moieties (Aita et al. 2010). Before the applications in protein docking, four-body potentials proven to be very effective efficient in scoring protein decoys (Krishnamoorthy and Tropsha 2003; Feng et al. 2010).

In summary, while small, local conformational changes accompanying protein docking are relatively well handled by a variety of docking algorithms, large deformations of components are more difficult to predict. RosettaDock is one of few exceptions, where *de novo* prediction of new (compared to known structures of components) structures is sometimes feasible. The multiscale approach described below, based on CABS modeling software, bootstrapped with all atom molecular mechanics, is a step toward fully flexible docking of proteins and peptides.

## 2.4 Multiscale Flexible Docking with CABS

CABS (the acronym stands for the united atoms representing a residue in a polypeptide chain: CA – alpha carbon of the main chain, CB – beta carbon, and the center of side group). C $\alpha$  trace in CABS is restricted to a high-resolution cubic lattice grid, where the lattice spacing is set to 0.61 Å. C $\alpha$ –C $\alpha$  distances in CABS are allowed to fluctuate near the 3.8 Å. An additional pseudo-atom is located in the center of

the C $\alpha$ –C $\alpha$  bond and supports a model of main-chain hydrogen bonds. The accuracy of a projection of high-resolution protein structure onto the lattice is of about half of the lattice spacing. The coordinates of beta carbons and side chains are not restricted to the lattice and are defined in the reference frame defined by the C $\alpha$  trace. Due to the lattice representation, computations of local conformational transitions of the model chains are extremely fast and, in most cases, they are reduced to straightforward shuffling of integer numbers. Similarly, most of interactions could be computed via simple references to large hashing tables. Such computations with CABS are about two orders of magnitude faster than it would be possible for – otherwise equivalent – continuous space model (Boniecki et al. 2003). It should be pointed out that, due to the fine grid of the lattice representations, the model does not exhibit any lattice artifacts. Actual accuracy of the molecular models generated by CABS is lower than the resolution resulting from the lattice representation. The results of free modeling, when successful, are accurate within a low-resolution range of 2.5–5 Å. Comparative models are more accurate and their accuracy depends on the quality of templates from which the distance restraints between C $\alpha$  atoms are extracted. The best models have an accuracy of about 1 Å. CABS allows for easy implementation of various restraints, not only C $\alpha$ –C $\alpha$  distances from templates but also restraints from sparse experimental data, as chemical shifts, residual dipolar coupling, side chain–side chain contacts from mutagenesis, etc. This opens a convenient framework for the treatment of docking flexibility at various levels. A geometric fidelity of the CABS representation is sufficient for a reasonably accurate all-atom reconstructions (Gront et al. 2007). A good measure of this fidelity is an experiment of projecting the structure onto CABS lattice followed by subsequent reconstruction of atomic details. The reconstruction consists of two stages: the first one is a very fast rebuilding of the main chain and beta carbons, executed by Backbone Building from Quadrilaterals (BBQ) program, the second one could be side-chain fitting via side-chain replacement with rotamer libraries (SCWRL) program by Dunbrack (Canutescu et al. 2003). The accuracy of such a reconstruction cycle is within a range of few tens of angstroms for the main-chain atoms and a range of 1.5–2 Å for the side-chain atoms, depending on the structure. The all-atom structures could be generated at any stage of the docking and scored by force fields other than CABS. Details of the CABS knowledge-based force field (Kolinski 2004) and description of combinations of CABS with all-atom molecular mechanics (Kmieciak and Kolinski 2007, 2008; Kmieciak et al. 2007) could be found in earlier publications.

Sampling protocols of CABS employ various Monte Carlo-based algorithms. When the folding mechanisms are of interest, simple simulated annealing or isothermal Monte Carlo dynamics could be appropriate. Since CABS conformational updating employs various local rearrangements controlled by a pseudo-random mechanisms, the trajectories from such simulations represent solutions of a certain Master Equation of motions and thereby provide a coarse-grained picture of the system dynamics. In this respect, CABS differs from the most of other structure assembly-reduced models, such as Rosetta (Rohl et al. 2004). There are, however, reduced space models enabling similar studies. The continuous models (like united

residues (UNRES), (Ołdziej et al. 2005) allow for Molecular Dynamics simulations and, similarly to CABS, for Monte Carlo dynamics. When just structure prediction is a goal, more effective than simulated annealing are various multicopy MC algorithms. Most docking experiments with CABS proceed according to a combination of Replica Exchange Monte Carlo (REMC) (simulated tempering) (Hukushima and Nemoto 1996) with simulated annealing. During a typical simulation, a large number of replicas (50) are subject to slow annealing of the entire stack.

### ***2.4.1 Treating of Flexibility***

CABS facilitates various levels of docking flexibility. Schematically, different instances of docking could be outlined as follows:

- A. A semi-flexible docking of two or more proteins of known structures.
- B. Docking of a fully flexible (unrestrained folding) protein or peptide on a semi-flexible scaffold of other proteins.
- C. De novo assembly (fully flexible, free folding) of protein (peptide) complex.

Obviously, the success rate and accuracy of the resulting structures decreases from A to C. The unrestricted de novo assembly of a protein complex (C) is now feasible only for relatively small and structurally not too complex proteins. Various symmetry-resulting (as for homo-dimers, trimers, etc.) restraints could be easily implemented (similarly as it was done before for RosettaDock), increasing the docking accuracy significantly.

In semi-flexible docking (A), intra-protein restraints are read from the unbonded structures and the corresponding distances are allowed to fluctuate around their unbonded values. The poses could be generated by the CABS REMC in an unrestricted fashion or the initial poses (structures placed at starting replicas) could be obtained from various fast docking shape-complementarity-based algorithms. To speed up assembly, centers of gravity of the assembled molecules are subject to a weak generic attractive force acting at distances larger than the plausible estimated distance within the complex. In such procedures, trivial flexibility related to different conformations of the interface side chains is approximately accounted for at the stage of the all-atom reconstruction. Flexibility of interacting proteins does need to be treated in a uniform fashion. It is easy to restraint just parts of each molecule, allowing the remaining portions to freely adjust during the docking. Prediction of flexible fragments could be achieved in various ways, including structural comparison of different complexes of proteins of interests, normal modes, or Gaussian Network, analysis of these proteins, etc. A relative strength of restraints could be also included in the input data for the docking.

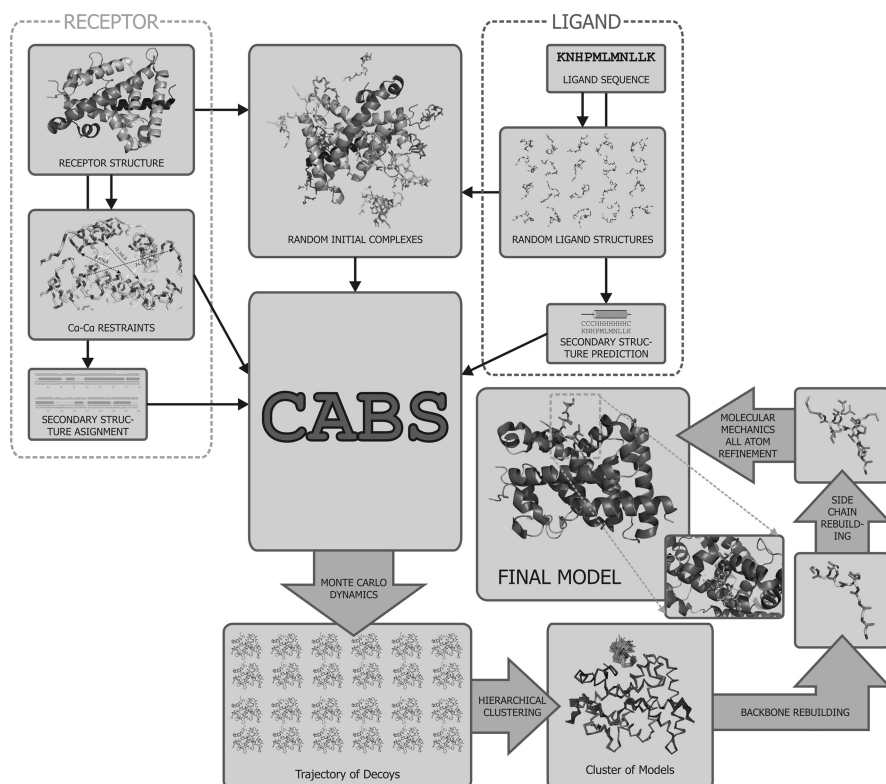
Docking of a fully flexible small protein (or a peptide) to semi-flexible scaffold of other protein (a receptor) is almost always successful, without assuming a priori anything about the pose (except the penalty for large distances between molecules) and internal conformation of the free molecule.

In principle, the interaction between the interface amino acids does need to be the same as the intra-protein interactions. The interactions between the side chains in CABS model are described by statistical pseudo-potentials derived from regularities seen in known structures. These potentials are context dependent (accounting for mutual orientations of the side chains and local conformations of the main chain). Thus, the potentials account in an implicit way for complex multi-body packing effects. Also the averaged solvent effect is encoded in these potentials. Interestingly, potentials derived separately for the interfaces in known complexes do not differ significantly from the potentials derived for monomeric proteins. In the example docking simulations described in this chapter, generic CABS potentials were used.

### ***2.4.2 Example of Peptide Docking to Receptor Protein***

Frequently small peptides act as coactivators for larger proteins. Below we describe a typical example of such a docking experiment (Kurcinski and Kolinski 2007). The receptor protein is the vitamin D receptor (or rather the receptor part of the entire protein). The receptor is treated in a semi-flexible fashion. A large number of  $C\alpha$ – $C\alpha$  intra-molecular distances are extracted from the crystallographic structure of the protein. Additionally, the secondary structure defined according to the define secondary structure of proteins (DSSP) assignment is a part of the simulation input. Assigned secondary structure provides a bias toward the proper short-range geometry and favors the hydrogen-bonding patterns consistent with this secondary structure. The simulation set-up for the receptor is schematically depicted in the left-hand side of the flow chart given in Fig. 2.1. During the simulation, the receptor structure oscillates around its native structure. The initial set of 50 replicas for the REMC simulations are generated by replication of the receptor structure with randomly placed peptide chains near the protein surface. Internal conformations of the peptide and its location in respect to the receptor are both selected in a random fashion. The set-up for the peptide is illustrated in the top right-hand part of the flow chart. The starting replica with the superimposed receptor structure is shown in the top central panel of the flow chart. The main part of the docking simulations is executed by the CABS algorithm. CABS produces a large number of conformations, stored in a pseudo-trajectory read from the lowest-temperature replica. Typically, the CABS output contains some thousands of structures. Single run generates millions of states and requires several hours of a single LINUX computing unit. The structures stored in the pseudo-trajectory are subject to a clustering procedure (hierarchical clustering or *K*-means clustering). In the case of the example illustrated here, there is only one well-defined cluster of solutions, containing majority of the structures. Remaining structures are scattered in apparently random fashion. The main cluster is very dense with nicely superimposed receptor structures. Only the end and some loop residues deviate a little (0.5–1.0 Å) from the mean structure, which is almost identical with the crystallographic structure. The cloud of the peptide structures is also very well defined (bottom, right-hand panel of Fig. 2.1), with





**Fig. 2.1** Flow chart of multiscale hierarchical peptide-protein docking. See the text for details

the mean-square dispersion below 1 Å. The centroid structure from the main cluster provides a scaffold for the all-atom reconstruction (left-hand panels) of the complex. The reconstructed structure is optimized with all-atom force field and rectified in Molecular Dynamics. The final structures obtained in such test docking are of crystallographic resolution.

In several tests of peptide docking to various receptors, the proper pose has always been found to be within the main cluster of solution. For longer peptides (25–30 amino acids), the resulting coordinates of end residues of the peptide were usually of lower accuracy (2–3 Å), although the interface contact maps were always predicted (or rather *postdicted*) with high accuracy.

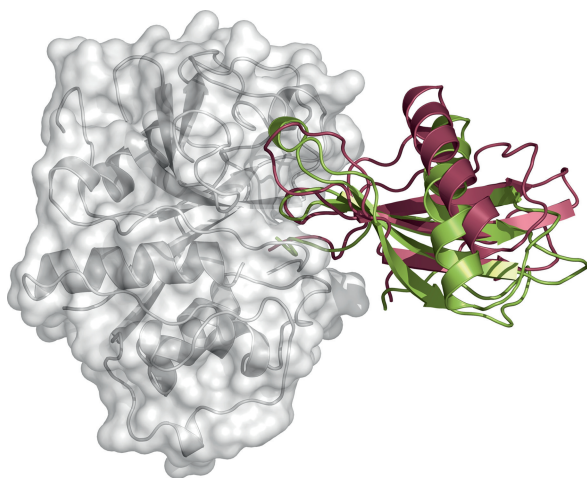
### 2.4.3 Protein–Protein Docking

The methodology described in the previous sections could be used for protein–protein docking where one or two (or more) proteins could be treated in a fully



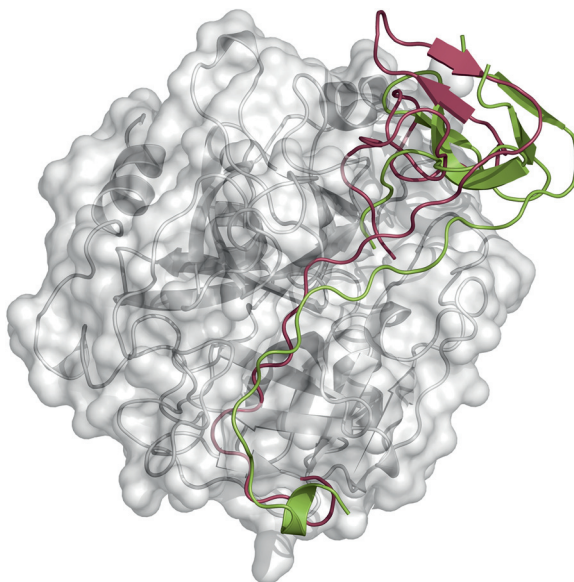
flexible fashion, without assuming anything about their structures within the complex. Good results of fully flexible, unrestricted docking, could be expected only for relatively small and structurally not too complex proteins, of a size of the Rop homo-dimer that consists of two antiparallel long-helical hairpins or the crambin pseudo-dimer. For larger proteins, properly folded complex structures are not always obtained. At the present status of the CABS methodology, the semi-flexible docking simulations are more dependable, where at least parts of the modeled structures are controlled by weaker or stronger intra-molecular restraints derived from non-bonded structures or from different complexes of the proteins of interest.

Two examples of semi-flexible docking results are illustrated in Figs. 2.2 and 2.3. In both cases, the structures of the larger proteins in the complex were strongly restrained to their non-bonded native structures. Docking simulations modified these structures very little (deviation range of 0.3–0.9 Å). Actually, this is very close to the structural differences seen between the individual proteins in the complexes and their unbound structures. The structures of larger protein are shown in gray. The second components of the complexes had higher flexibility, the restraints were much weaker allowing for large fluctuations, ranging within 5–10 Å. In both cases, near-native structures were found in the largest clusters. The resulting poses are qualitatively correct, although the details of the internal structures of these proteins have several errors. For easy comparison, Figs. 2.2 and 2.3 show both experimental (green) and calculated (red) structures. The drawings were done assuming the best superimposition of the larger proteins in the complex. The resulting superimposition of the second proteins illustrates a sum of the errors of pose and the errors of internal coordinates. The sum of these errors (coordinate Root-Mean-Square deviation of the smaller protein after the best superimposition of the larger protein) is 2.6 Å in the first case and 3.8 Å in the second case, respectively. Thus, qualitatively correct poses have been predicted (nothing has been assumed about the mutual orientation



**Fig. 2.2** Structure obtained from docking procedure. The “receptor protein” (PDB code 1ppn) shown in gray, “ligand” protein (PDB code 2oct) shown in *green* in the crystallographic structure (PDB code 1stf), and in *red* for the final model. See the text for details

**Fig. 2.3** Structure obtained from docking procedure. The “receptor protein” (PDB code 2hnt) shown in *gray*, “ligand” protein (PDB code 5hir) shown in *green* in the crystallographic structure (PDB code 4htc), and in *red* for the final model. See the text for details



of the components) although the structural details have been distorted, especially in the second case. Initial 50 replicas for the REMC simulations were generated by FTdock program.

## 2.5 Perspectives

The problem of *in silico* flexible docking, especially in cases where the docking-induced conformational changes are large, is far from being solved. Nevertheless, there are numerous encouraging small steps toward a partial solution of this problem. Multiscale procedures, where flexible docking is performed using various coarse-grained protein models, followed by refinement of the resulting poses by more detailed molecular mechanics seem to be very promising. Here, we described combinations of the CABS-reduced space modeling methodology with all-atom refinements applied to flexible and semi-flexible protein-protein and protein-peptide dockings. The method is now mature enough for large-scale predictions of protein interactomes. In the large-scale applications, it may be necessary to introduce a pre-screening phase employing fast docking procedures based on a shape complementarity.

At present, the described method is limited to proteins and peptides. An extension onto nucleic acids will require development of their coarse-grained representation, consistent with the CABS representation of proteins. Also the treatment of small ligands within such multiscale docking procedures requires significant extensions of the knowledge-based force fields, and this is still an open problem.

Another possibility of applications of the outlined method is related to the assembly mechanisms of protein complexes. CABS sampling techniques and its force field enable meaningful simulations of folding pathways (see Chapter 12). Extension of the method on multimeric assemblies is straightforward, although it will require larger computing resources, due to a higher complexity of the problem.

Finally, we would like to note that the coarse-grained protein models are potentially very interesting in the context of yet different class of docking problem, namely, fitting molecular structures into cryo-electron microscopy (EM) (or similar) low-resolution experimental data (Lindert et al. 2009; Orzechowski and Tama 2008; Jolley et al. 2008). This problem, however, is beyond the scope of this chapter.

## References

- Aita T, Nishigaki K, Husimi Y (2010) Toward the fast blind docking of a peptide to a target protein by using a four-body statistical pseudo-potential. *Comput Biol Chem* 34:53–62
- Aloy P, Russell RB (2004) Ten thousand interactions for the molecular biologist. *Nat Biotechnol* 22:1317–1321
- Anand GS, Law D, Mandell JG, Snead AN, Tsigelny I, Taylor SS, Ten Eyck LF, Komives EA (2003) Identification of the protein kinase A regulatory RI $\alpha$ -catalytic subunit interface by amide H/2H exchange and protein docking. *Proc Natl Acad Sci USA* 100:13264–13269
- Andrusier N, Mashiach E, Nussinov R, Wolfson HJ (2008) Principles of flexible protein–protein docking. *Proteins* 73:271–289
- Boniecki M, Rotkiewicz P, Skolnick J, Kolinski A (2003) Protein fragment reconstruction using various modeling techniques. *J Comput Aid Mol Des* 17:725–738
- Bonvin AM (2006) Flexible protein–protein docking. *Curr Opin Struct Biol* 16:194–200
- Burgoyne NJ, Jackson RM (2006) Predicting protein interaction sites: binding hot-spots in protein–protein and protein–ligand interfaces. *Bioinformatics* 22:1335–1342
- Camacho CJ, Vajda S (2001) Protein docking along smooth association pathways. *Proc Natl Acad Sci USA* 98:10636–10641
- Canutescu AA, Shelenkov AA, Dunbrack RL (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Prot Sci* 12:2001–2014
- Carter P, Lesk VI, Islam SA, Sternberg MJ (2005) Protein–protein docking using 3D-dock in rounds 3, 4, and 5 of CAPRI. *Proteins* 60:281–288
- Cerutti DS, Ten Eyck LF, McCammon JA (2005) Rapid estimation of solvation energy for simulations of protein–protein association. *J Chem Theory Comput* 1:143–152
- Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52:80–87
- Daily MD, Masica D, Sivasubramanian A, Somarouthu S, Gray JJ (2005) CAPRI rounds 3–5 reveal promising successes and future challenges for RosettaDock. *Proteins* 60:181–186
- Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D (2009) Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci USA* 106:18978–18983
- Del Carpio-Muñoz CA, Ichiishi E, Yoshimori A, Yoshikawa T (2002) MIAX: a new paradigm for modeling biomacromolecular interactions and complex formation in condensed phases. *Proteins* 48:696–732
- 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
- Feng Y, Kloczkowski A, Jernigan RL (2010) Potentials ‘R’ Us web-server for protein energy estimations with coarse-grained knowledge-based potentials. *BMC Bioinformatics* 11:92
- Fischer D, Lin SL, Wolfson HL, Nussinov R (1995) A geometry-based suite of molecular docking processes. *J Mol Biol* 248:459–477

- 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
- Gront D, Kmiecik S, Kolinski A (2007) Backbone building from quadrilaterals: a fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. *J Comput Chem* 28:1593–1597
- Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *J Phys Soc Jpn* 65:1604–1608
- 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 L, Gao Y, Mao F, Liu Z, Lai L (2002) Potential of mean force for protein–protein interaction studies. *Proteins* 46:190–196
- Jolley CC, Wells SA, Fromme P, Thorpe MF (2008) Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys J* 94:1613–1621
- Jones S, Thornton JM (1997) Prediction of protein–protein interaction sites using patch analysis. *J Mol Biol* 272:133–143
- 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* 89:12195–12199
- Kmiecik S, Gront D, Kolinski A (2007) Towards the high-resolution protein structure prediction. Fast refinement of reduced models with all-atom force field. *BMC Struct Biol* 7:43
- Kmiecik S, Kolinski A (2007) Characterization of protein-folding pathways by reduced-space modeling. *Proc Natl Acad Sci USA* 104:12330–12335
- Kmiecik S, Kolinski A (2008) Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys J* 94:726–736
- Koehl P (2006) Electrostatics calculations: latest methodological advances. *Curr Opin Struct Biol* 16:142–151
- Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol* 51:349–371
- Kozakov D, Brenke R, Comeau SR, Vajda S (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* 65:392–406
- Krishnamoorthy B, Tropsha A (2003) Development of a four-body statistical pseudo-potential to discriminate native from non-native protein conformations. *Bioinformatics* 19:1540–1548
- Kurcinski M, Kolinski A (2007) Steps towards flexible docking: modeling of three-dimensional structures of the nuclear receptors bound with peptide ligands mimicking co-activators' sequences. *J Steroid Biochem* 103:357–360
- Lichtarge O, Bourne HR, Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 257:342–358
- Lindert S, Staritzbichler R, Wötzel N, Karakaş M, Stewart PL, Meiler J (2009) EM-fold: de novo folding of alpha-helical proteins guided by intermediate-resolution electron microscopy density maps. *Structure* 17:990–1003
- Mandell JG, Roberts VA, Pique ME, Kotlovyy V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 14:105–113
- May A, Zacharias M (2005) Accounting for global protein deformability during protein–protein and protein–ligand docking. *Biochem Biophys Acta* 1754:225–231
- May A, Zacharias M (2007) Protein–protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins* 69:774–780
- Nilges M (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J Mol Biol* 245:645–660
- Orzechowski M, Tama F (2008) Flexible fitting of high-resolution X-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys J* 95:5692–5705

- Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nancias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kaźmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA (2005) Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. *Proc Natl Acad Sci USA* 102:7547–7552
- Res I, Lichtarge O (2005) Character and evolution of protein–protein interfaces. *Phys Biol* 2:S36–S43
- Res I, Mihalek I, Lichtarge O (2005) An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics* 21:2496–2501
- Ritchie DW (2008) Recent progress and future directions in protein–protein docking. *Curr Protein Pept Sci* 9:1–15
- Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66–93
- Salwinski L, Eisenberg D (2003) Computational methods of analysis of protein–protein interactions. *Curr Opin Struct Biol* 13:377–382
- Schueler-Furman O, Wang C, Baker D (2005) Progress in protein–protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins* 60:187–194
- Sheinerman FB, Norel R, Honig B (2000) Electrostatic aspects of protein–protein interactions. *Curr Opin Struct Biol* 10:153–159
- Sternberg MJ, Gabb HA, and Jackson RM (1998) Predictive docking of protein–protein and protein–DNA complexes. *Current Opinion in Structural Biology* 8:250–256
- Tobi D, Bahar I (2006) Optimal design of protein docking potentials: efficiency and limitations. *Proteins* 62:970–981
- Vajda S, Kozakov D (2009) Convergence and combination of methods in protein–protein docking. *Curr Opin Struct Biol* 19:164–170
- Vakser IA, Afllalo C (1994) Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins* 20:320–329
- Vakser IA (1995) Protein docking for low-resolution structures. *Protein Eng* 8:371–377
- Valencia A, Pazos F (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol* 12:368–373
- van Dijk AD, de Vries SJ, Dominguez C, Chen H, Zhou H, Bonvin AM (2005) Data-driven docking: HADDOCK’s adventures in CAPRI. *Proteins* 60:232–238
- Wang C, Schueler-Furman O, Andre I, London N, Fleishman SJ, Bradley P, Qian B, Baker D (2007) RosettaDock in CAPRI rounds 6–12. *Proteins* 69:758–763
- Wang C, Schueler-Furman O, Baker D (2005) Improved side-chain modeling for protein–protein docking. *Prot Sci* 14:1328–1339
- Wodak SJ, Janin J (1978) Computer analysis of protein–protein interaction. *J Mol Biol* 124:323–342
- Zacharias M (2005) ATTRACT: protein–protein docking in CAPRI using a reduced protein model. *Proteins* 60:252–256
- Zacharias M (2003) Protein–protein docking with a reduced protein model accounting for side-chain flexibility. *Prot Sci* 12:1271–1282
- Zhang C, Liu S, Zhu Q, Zhou Y (2005) A knowledge-based energy function for protein–ligand, protein–protein, and protein–DNA complexes. *J Med Chem* 48:2325–2335
- Zhang Q, Sanner M, Olson AJ (2009) Shape complementarity of protein–protein complexes at multiple resolutions. *Proteins* 75:453–467

<http://www.springer.com/978-1-4419-6888-3>

Multiscale Approaches to Protein Modeling

Kolinski, A. (Ed.)

2011, XII, 355 p., Hardcover

ISBN: 978-1-4419-6888-3