

Hierarchical modeling of protein interactions

Mateusz Kurcinski · Andrzej Kolinski

Received: 15 November 2006 / Accepted: 18 January 2007 / Published online: 13 February 2007
© Springer-Verlag 2007

Abstract A novel approach to hierarchical peptide–protein and protein–protein docking is described and evaluated. Modeling procedure starts from a reduced space representation of proteins and peptides. Polypeptide chains are represented by strings of alpha-carbon beads restricted to a fine-mesh cubic lattice. Side chains are represented by up to two centers of interactions, corresponding to beta-carbons and the centers of mass of the remaining portions of the side groups, respectively. Additional pseudoatoms are located in the centers of the virtual bonds connecting consecutive alpha carbons. These pseudoatoms support a model of main-chain hydrogen bonds. Docking starts from a collection of random configurations of modeled molecules. Interacting molecules are flexible; however, higher accuracy models are obtained when the conformational freedom of one (the larger one) of the assembling molecules is limited by a set of weak distance restraints extracted from the experimental (or theoretically predicted) structures. Sampling is done by means of Replica Exchange Monte Carlo method. Afterwards, the set of obtained structures is subject to a hierarchical clustering. Then, the centroids of the resulting clusters are used as scaffolds for the reconstruction of the atomic details. Finally, the all-atom models are energy minimized and scored using classical tools of molecular mechanics. The method is tested on a set of macromolecular assemblies consisting of proteins and peptides. It is demonstrated that the proposed approach to the flexible docking could be successfully applied to prediction of protein–peptide and protein–protein

interactions. The obtained models are almost always qualitatively correct, although usually of relatively low (or moderate) resolution. In spite of this limitation, the proposed method opens new possibilities of computational studies of macromolecular recognition and mechanisms of assembly of macromolecular complexes.

Keywords Protein modeling · Flexible docking · Macromolecular recognition · Protein assemblies · Monte Carlo

Introduction

Tens of thousands of experimental structures of proteins are known already. Although this is only a small fraction (range of 0.1%) of the number of known protein sequences, it is safe to assume that the known structures are representative for at least half of all existing protein structures. Indeed, using advanced contemporary methods of comparative modeling, it is possible to build reasonable quality theoretical molecular models for more than 50% of newly sequenced proteins [1]. This opens enormous opportunities for theoretical interpretation of protein biological functions at the molecular level. Modeling of enzymatic activity, mechanisms of molecular signaling, protein associations (including autocatalytic changes of native structures, as the ones responsible for some neurodegenerative diseases), and other processes has recently become possible. At the same time, computer studies of biomacromolecular processes face numerous technical difficulties, due to the high level of molecular complexity of the systems under consideration.

One of the most important and challenging tasks of computational molecular biology is prediction of interactions between biologically important molecules. Docking of

M. Kurcinski · A. Kolinski (✉)
Faculty of Chemistry,
Warsaw University,
ul. Pasteura 1,
02-093 Warsaw, Poland
e-mail: kolinski@chem.uw.edu.pl

smaller molecules into receptors and protein–protein docking are essential for computer-aided, rational drug design, understanding metabolic pathways and theoretical interpretations of molecular basis of neurodegenerative diseases, etc. In general, the term “docking” means finding the lowest free energy conformation(s) of a ligand–receptor complex [2]. The traditional approach assumes knowledge of the structure of the macromolecular components [3, 4]. Thus, in the case of small molecule docking, its location in respect to the receptor and its internal conformation, compatible with the complex structure and intra- and intermolecular interactions, need to be calculated. In the cases of protein–protein docking, the two structures need to be connected, i.e., the interface(s) of the complex has to be found. So called “flexible docking” is usually limited to an optimization of the ligand structure, relaxation of the involved side chains, and (sometimes) very limited readjustment of portions of the main chain of proteins [5, 6]. This kind of approach sometimes works very well. However, in general, the assumptions of the traditional docking approaches could be non-physical. Structures of components of an assembly are always a function of the interactions in the entire complex. Frequently, the same molecule (macromolecule) in isolation has a qualitatively different structure from that observed in various complexes. There are numerous examples where, upon a ligand docking, the structure of the receptor changes dramatically. The same is true for protein–protein association.

In this work, we propose a new approach, which is the first step towards a fully flexible docking. At this moment, the new method is limited to docking of peptides to proteins and protein–protein docking only, with some restrictions imposed on the range of conformational changes of the macromolecular components, induced by the assembly process. The approach is hierarchical—a multiscale molecular modeling is employed. Here, examples of two-molecule complexes are discussed and, although the algorithms are general, larger than two-molecule systems could be treated in the same way. The modeling process starts from a random mutual orientation of the molecules under consideration and a random conformation of one of them (the smaller one or the one randomly selected if the two components are of the same molecular mass). Simulation of the assembly process is done using the multichain version of our reduced-space protein modeling tool CABS [7]. CABS (CA, Cb, and Side chain) employs a united atom representation of protein conformations. A residue in a model polypeptide chain is represented by up to four interaction centers corresponding to the alpha carbon, the beta carbon, the center of mass of the remaining portion of the side chain (where applicable) and pseudoatoms located in the center of the virtual bonds connecting the alpha carbons. Conformational space of this model is sampled by

means of multicopy Monte Carlo method (REMC, Replica Exchange Monte Carlo) versions of which are also known as the Simulated Tempering MC [8]. Subsequently, multiple structures resulting from the REMC simulations with the CABS model are subject to clustering, rebuilding the atomic details, all-atom optimization/minimization, and scoring of the models. Finally, the model structures are compared with the known crystallographic structures. Most of the modeled assemblies were very close to the native ones, showing the predicting power of the new method.

Methods

Figure 1 shows a flow-chart of the flexible docking procedure employed in this work. The simulations have a

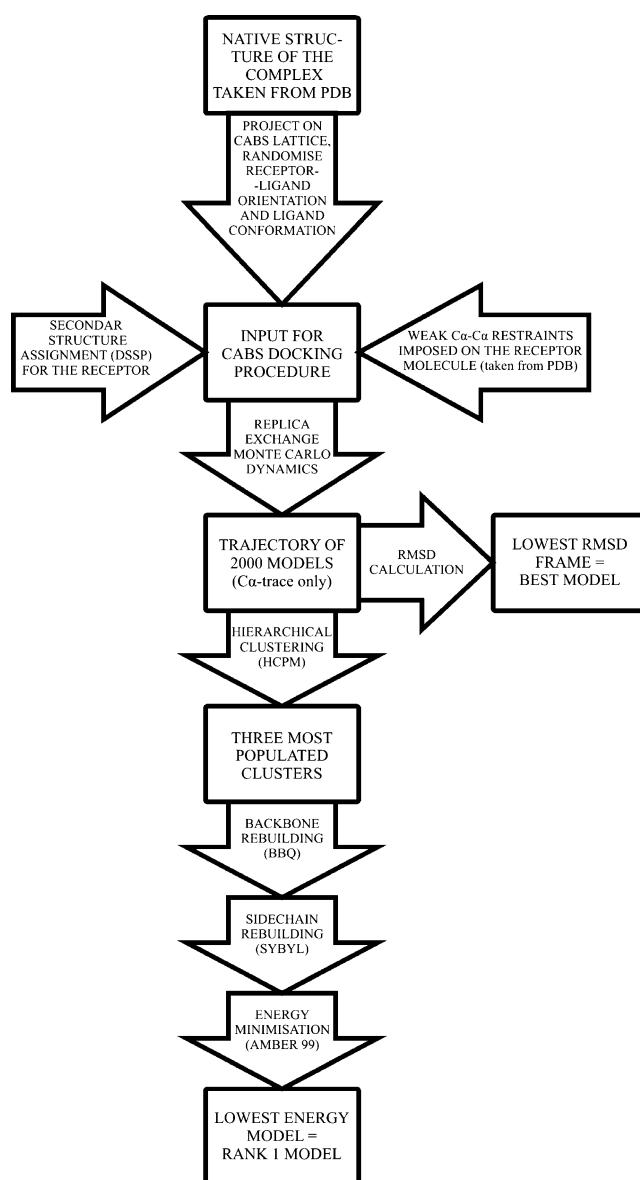


Fig. 1 Flow chart of multiscale assembly of protein complexes

multiscale character. Search of the conformational space is done by means of REMC simulations with CABS reduced model of polypeptides. Refinement and final selection of molecular models is done on the all-atom level using an AMBER 99 [9] force field. Details of the proposed methodology are given below.

CABS model and simulation method

CABS is a simplified model of polypeptides. A detailed description of this model and its force field has recently been published [7]. The model has been extensively tested during the CASP6 (Critical Assessment of Protein Structure Prediction) community-wide experiment [10]. The results of CASP6 have proved that CABS was one of the two best protein folding algorithms, especially when applied to more difficult cases of blind structure prediction [11, 12]. Other applications of CABS, as structure prediction based on sparse NMR data, modeling of folding dynamics and thermodynamics, etc., have also been described recently [13–15]. Here, for reader convenience, we provide just a concise outline of the CABS features.

A single residue in the CABS model is represented by up to four interaction centers corresponding to alpha carbons, beta carbons, the center of mass of the remaining portion of the side chain (where applicable) and a pseudoatom located in the center of the virtual Ca-Ca bond, corresponding roughly to the center of mass of the peptide bond. The positions of alpha carbons are restricted to the underlying cubic lattice, with the mesh size equal to 0.61 Å. The lengths of the Ca-Ca virtual bonds are allowed to oscillate around their physical value of 3.78 Å. These pseudobonds belong to a set of 800 lattice vectors. Thus, possible lattice artifacts can be safely avoided. The fluctuating bond length facilitates very fast sampling and a better packing of the model side groups. Lattice-restricted Ca-trace provides a convenient reference frame for the definition of the (off-lattice) coordinates of the remaining united atoms. Two-rotamers approximation of the side groups' mobility is assumed. Geometry of the model amino acids is derived from statistical analysis of high resolution protein structures. Lattice-based reference frame of the model chain facilitates very rapid and straightforward calculations of conformational transitions.

The interaction scheme of CABS consists of several heuristic potentials derived from the statistical analysis of the structural regularities seen in the folded proteins. The short range potentials have the form of energy histograms as functions of the distances between i -th and $i+2$, $i+3$, and $i+4$ th alpha carbons, the amino acid composition and the chirality of the corresponding fragments. The long range potentials consist of hard core repulsions between the Ca, C β united atoms and peptide bond pseudoatoms, and soft

(finite) repulsions between the side chains. Attractive parts of the side chain pairwise potentials have a form of square-well potentials. The strength of the side chain interactions is context-dependent and is different for different orientations of the interacting groups and for different types of the local geometry of the main chain. Additionally, there is a model of the main chain hydrogen bonds defined as a set of geometrical criteria for a configuration of the interacting fragments of the main chain that is consistent with the presence of the main chain hydrogen bonds.

Intermolecular interactions and intramolecular interactions were treated in the same way. This is certainly an approximation. There are, however, two reasons partially justifying this approximation. First, a derivation of separate statistical potentials for the interfaces is more difficult due to the limited set of the crystallographic data available. Second, as it is shown in this work, the assumption of the same potentials works very well, indicating that the possible differences in these interactions are probably not that significant. Nevertheless, this issue needs to be further investigated.

Conformational updating of the CABS chains consists of a random sequence of one residue, two, three and four residues local micromodifications. For instance, the one residue update requires a small random displacement of a randomly selected alpha carbon and subsequent rearrangements of the three involved side chains. Occasionally, a small displacement of a larger fragment of a chain (up to 22 residues) is attempted, allowing for a rigid-body component of the simulated motion. The sequence (types of motion and their placement along the polypeptide chain) is controlled by a random algorithm. As a result, the entire chain can relax its internal coordinates and move in space. Consequently, the ligand is allowed to sample the full spectrum of its internal coordinates and all possible locations in respect to the receptor. The receptor mobility is limited to a local relaxation (backbone and side chains) by a set of weak harmonic restraints imposed onto a small fraction of the Ca-Ca distances in the receptor (five restraints per residue). The strength of these restraints was sufficiently low to allow 2–4 Å relaxation of the receptor starting structure. The last limitation (receptor's restraints) can be omitted, although it would lead to significantly poorer quality of the resulting models. In future work, it is planned to identify rigid and flexible regions of the receptor, thereby allowing docking with large scale structural rearrangements of the all assembling components. The details of Monte Carlo dynamics employed here are described in detail elsewhere [7]. Thanks to the lattice-type representation of the model chains, the computation of the attempted micromodifications is extremely simple and requires only a few references to randomly selected elements of large tables describing precomputed, allowed conformational transitions. Thus, the

simulations with CABS are much faster than would be the case with an otherwise similarly reduced model in the continuous space.

Sampling of the CABS model is done by means of a version of the Replica Exchange Monte Carlo method. The starting conformations were constructed as follows. First, the receptor (or one of the members of a homodimer structure) has been projected onto the CABS representation using the native structure from PDB as a template. Then, the second molecule was added in a random conformation and in a random position in respect to the first molecule. Each replica had a different random configuration. During the simulations, both molecules were mobile, although the structure of the first molecule (receptor) was controlled by a set of weak restraints, taken from the native structure of the assembly. The strength of these restraints was high enough to keep the receptor structure in a region of globally native-like conformations, though allowing for significant readjustments of the assembly interface. Additionally, a weak harmonic force has been turn-on, when the two molecules separated on distances precluding any chain contacts. This was done for a purely technical reason, namely to avoid useless sampling of both molecules in separation. At distances allowing even sparse contacts of two simulated molecules the harmonic force has been turn-off.

The results of single simulations were stored as pseudotrajectories containing 2,000 snapshots of the system Ca-traces. Depending on the size of studied assemblies simulations took a couple of hours (for the smallest assembly) to a couple of days (for the largest one studied here) on a single LINUX box.

Clustering

Simulation trajectories were subject to a hierarchical clustering procedure using the HCPM software [16] (<http://www.biocomp.chem.uw.edu.pl/HCPM>). The clustering was stopped when about 50% of the structures were assigned to the growing clusters. The structures closest to the clusters' centroids of the three biggest clusters were selected for further processing. For the purpose of an additional evaluation of the subsequent model-ranking procedure, one more structure (the fourth one), which was the closest to the native (RMSD-coordinate root-mean-square deviation for the alpha traces after the best superimposition with the crystallographic structure) in the entire trajectory of 2,000 snapshots, was also selected.

All-atom refinement and model selection

Selected reduced structures from the Monte Carlo simulations (centroids of the top three clusters and the best structure observed in a simulation) were used as scaffolds

for the all-atom building. The recently developed BBQ (Backbone Building from Quadrilaterals) program was first applied to the reconstruction of the all-backbone atoms [17] (<http://www.biocomp.chem.uw.edu.pl/BBQ>). The BBQ algorithm employs a large set of high resolution protein fragments which are superimposed on the Ca-trace. The algorithm is very fast, accurate and tolerates the inaccurate distances between alpha carbons from CABS simulations very well. In the next step, approximate positions of the side chains are calculated using a large library of side chain rotamers and the main chain coordinates as the reference frame [18].

The crude all-atom models were finally refined during a relatively short minimization with an AMBER force field using TRIPOS software [9]. The solvent was treated in the implicit fashion using the Generalized Born solvation model. It has been proved that the energy from the all-atom minimization correlates better with the RMSD than does the CABS energy. Thus, the multiscale approach improves the accuracy of the model selection. A similar effect was also observed by Baker [19] for a combination of the Rosetta reduced models with a refinement on the atomic level.

Test set of macromolecular assemblies

A set of 11 protein–peptide and protein–protein assemblies was selected for the test predictions using the method described above. Although relatively small, the test set seemed to be representative. It contained heterodimers and homodimers, the size of the component of the assemblies ranging from 5 to 197 residues, and the participating proteins representing various structural classes (alpha, beta and alpha/beta). An additional criterion of the selection was high resolution of the crystallographic structures, which could be of some importance for a proper evaluation of the accuracy of predictions. Table 1 contains a brief description of the studied systems.

Results and discussion

Using the pipeline described in [Methods](#) and outlined in the flow-chart given in [Fig. 1](#), molecular models for the 11 test systems ([Table 1](#)) were generated and evaluated. The reduced conformational space modeling with the CABS is the core element of this pipeline. The CABS modeling started from a number of random configurations of the two molecules–replicas in the REMC simulations. The conformations of the first molecule (receptor, or an arbitrarily selected molecule from a homodimer) were weakly restrained to the near-native region of its conformational space. The second molecule was fully mobile in a neighborhood of the first one and internally flexible during

Table 1 Summary of the test set of macromolecular assemblies

PDB code	Complex	Receptor class	Receptor length	Ligand length
1A2X	Troponin C – Troponin I	α	158	31
1CKA	CRK SH3 Domain – Proline rich peptide	β	57	9
1KLQ	MAD2 Protein – MAD2 binding peptide	α/β	197	10
1KYF	Alpha adaptin C – Growth factor substrate	β	117	5
1NTV	Disabled PTB domain – ApoER2 peptide	β	152	10
1OGT	HLA antigen – Vasoactive polypeptide	α/β	179	9
1OKH	Crambin-like homodimer	α/β	46	46
1RPR	ROP homodimer	α	63	63
1VWA	Streptavidin – Peptide containing HPQ	β	121	7
2BBM	Calmodulin – Myosin light chain	α	148	26
2ZTA	GCN4 leucine zipper	α	31	31

the simulations. Totals of 2,000 snapshots were extracted from each simulation, at equal intervals of the clock of the simulation algorithm.

After clustering, the centroids of the top three clusters were subject to the all-atom reconstruction and refinement,

and the lowest energy (AMBER 99) model was reported as the RANK 1 model in Table 2. In order to also evaluate the proposed strategy for the selection of models the best CABS structure observed in the simulations was all-atom reconstructed and energy minimized in exactly the same

Table 2 Summary of the modeling results

PDB code	Model	AMBER99 energy	Cluster size	RMSD			Fraction of native contacts
				Total	Receptor	Ligand	
1A2X	RANK 1	-3,086.64	957	1.76	1.02	3.83	0.66
	BEST	-2,562.66	1	1.24	0.95	2.24	0.70
1CKA	RANK 1	-852.46	489	2.69	0.77	7.87	0.57
	BEST	-852.71	186	1.60	0.50	4.77	0.64
1KLQ	RANK 1	-968.02	157	1.56	1.04	5.58	0.75
	BEST	-1,001.23	712	1.30	0.97	4.10	0.74
1KYF	RANK 1	-1,931.22	21	1.87	1.07	7.80	0.57
	BEST	-1,573.20	1	1.37	1.13	4.05	0.52
1NTV	RANK 1	-2,456.88	117	2.20	0.99	8.17	0.58
	BEST	-2,337.50	362	1.09	0.87	2.77	0.65
1OGT	RANK 1	-3,261.38	19	1.92	0.55	8.65	0.30
	BEST	-3,612.17	1	1.18	0.47	5.08	0.51
1OKH	RANK 1	-958.41	720	3.63	1.12	9.27	0.48
	BEST	-437.67	720	2.14	1.29	4.01	0.59
1RPR	RANK 1	-2,050.36	542	5.21	1.68	10.93	0.40
	BEST	-2,321.46	542	4.95	1.21	10.57	0.45
1VWA	RANK 1	-1,729.21	28	2.95	1.32	11.54	0.30
	BEST	-2,339.11	897	1.59	1.20	4.70	0.38
2BBM	RANK 1	-3,030.26	640	2.30	0.91	5.61	0.80
	BEST	-2,872.88	1	1.06	0.55	2.49	0.78
2ZTA	RANK 1	-1,400.24	687	1.69	1.18	2.69	0.51
	BEST	-1,148.31	3	1.43	1.35	2.66	0.58

RANK 1 the lowest energy (AMBER 99) all-atom model of a cluster centroid from the CABS simulations

BEST: the all-atom model of the best structure from the set of 2000 CABS models

Total coordinate RMSD of the entire complex after the best superimposition with the crystallographic structure

Receptor coordinate RMSD for the receptor (or the first chain in a homodimer) after the best superimposition with the crystallographic structure of the receptor alone

Ligand coordinate RMSD for the ligand (or the second chain in the homodimer) after the best superimposition of the receptor (or the first chain of the homodimer)

Fraction of native contacts fractions of native contacts observed in the model structures on their protein-protein (or protein-peptide) interfaces. After Vajda [20] the “contacts” are defined for pairs of alpha-carbon atoms using 10Å cut-off

fashion as was done for the clusters' centroids. These are reported as the BEST models in Table 2.

Various measures of the accuracy of the models could be used. In Table 2, we reported the coordinate root-mean-square deviation (RMSD) for the entire complex after the best superimposition (the column with the header Total), the RMSD for the restrained part of the system (Receptor), the RMSD for the entirely free-moving part of the system after the best superimposition of the restrained part of the assembly (Ligand), and the fraction of the native contacts observed in the complex interface.

Analysis of the modeling results summarized in Table 2 leads to a number of interesting observations. First of all, it can be noted that in all cases the resulting models are qualitatively correct (see also Figs. 2, 3, 4 and 5), as evidenced by RMSD values for entire complexes (ranging from 1.06 Å for the 2BBM, Calmodulin-Myosin light chain complex, to 5.21 Å in the worst case of 1RPR, ROP homodimer, with a typical value below 2 Å for most cases) and significant fractions of recovered native contacts in the model interfaces (typically above 50%). It should be

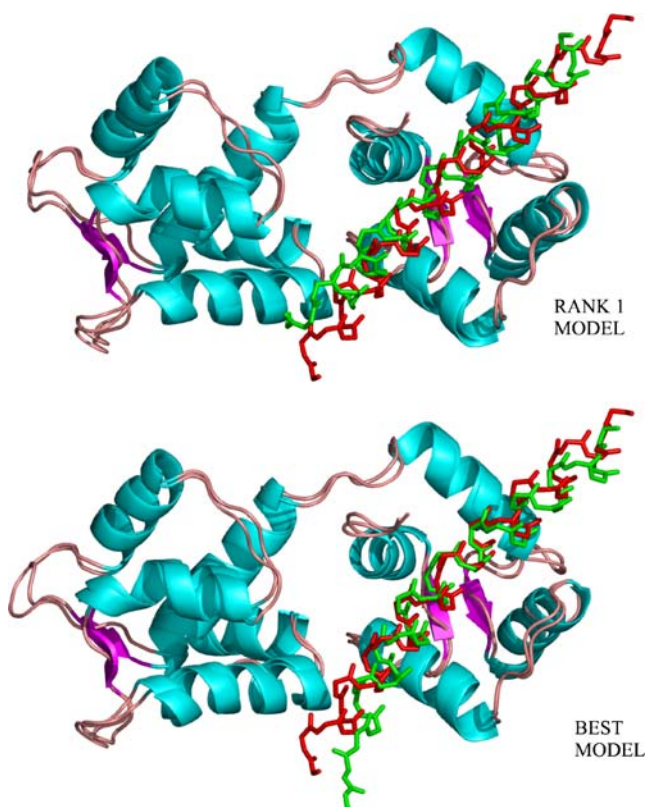


Fig. 2 Superimposition of the predicted and native structures of 1A2X complex. The receptor is shown as ribbons. Covalent structure for heavy atoms is shown for the peptide ligand. *Rank 1 model* is the lowest energy (according to the AMBER force-field) model after refinement of the clusters' centroids. The *best model* is the refined all-atom model built on the best (according to the RMSD deviation from the crystallographic structure) CABS structure found in the trajectory of 2000 snapshots. Ligand native structure is shown in red, modeled in green

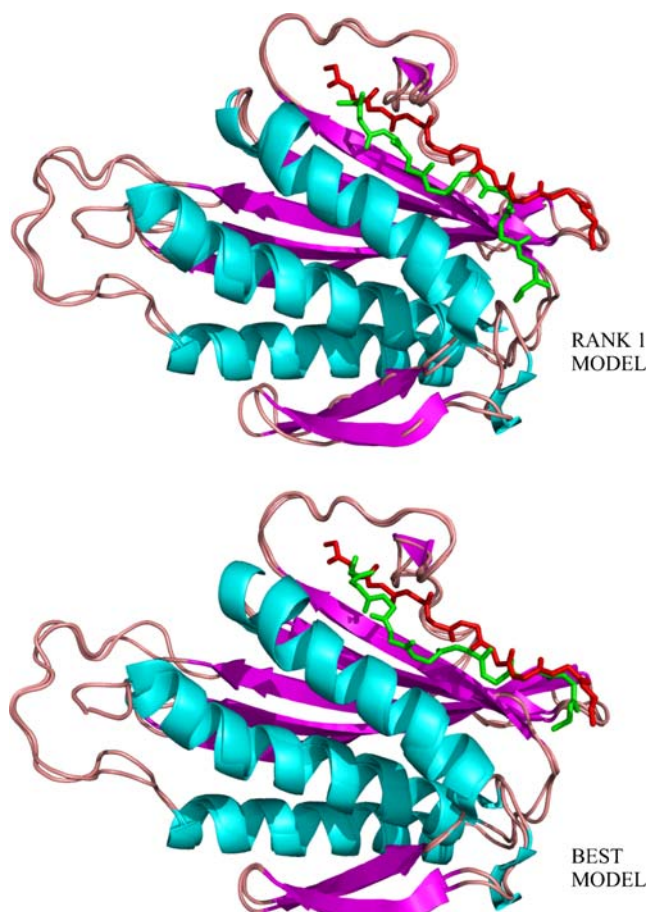


Fig. 3 Superimposition of the predicted and native structures of 1KLQ complex. The receptor is shown as ribbons. Covalent structure for heavy atoms is shown for the peptide ligand. See also the legend to Fig. 2

pointed out that the employed definition of contacts [20] is very sensitive to small structural changes and therefore the value of 50% is highly significant and reflects qualitatively correct association interface. Even in the worst case of the ROP homodimer, the high value of RMSD (5.21 Å) does not mean a qualitatively wrong prediction. Figure 5 shows that the two molecules are in a proper mutual orientation and that the high value of RMSD is caused by misfolding of small terminal fragments of one of the monomers.

Very small values of the first molecules' RMSD (typically below 1 Å) indicate that the CABS model, even with relatively weak distance restraints, facilitates a very accurate representation of protein conformations, in spite of its reduced representation. Indeed, during the CASP6 experiment several comparative models built by means of the template restrained CABS simulations were of an experimental quality (in the range of 1.5 Å RMSD from the native). In some cases, the models were better than the used templates [10] (<http://www.predictioncenter.org/casp6>). On the other hand, such good quality of the restrained parts of the complexes studied in this work does

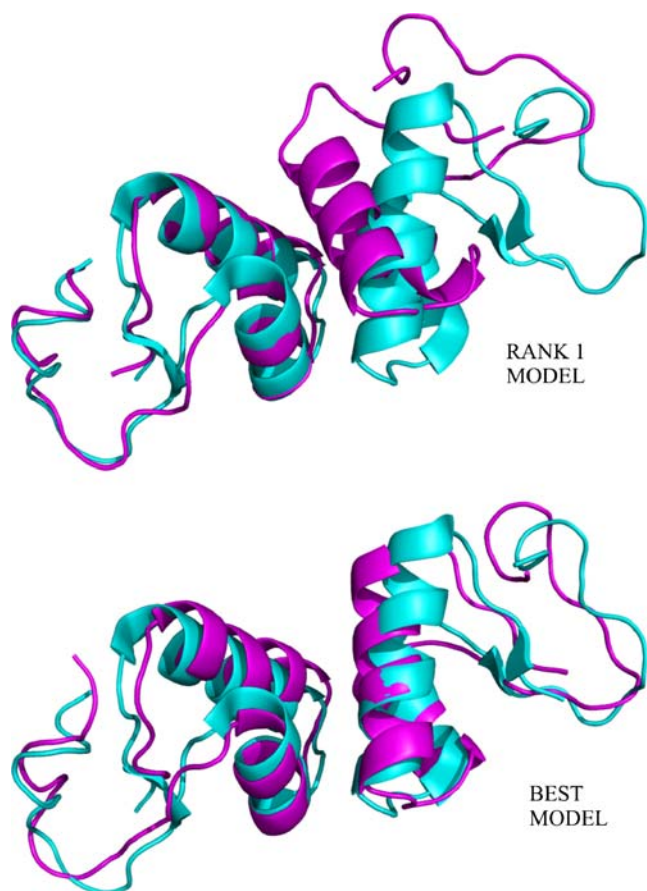


Fig. 4 Superimposition of the predicted and native structures of 1OKH homodimer. Native structure in cyan and predicted in magenta ribbons, respectively

not prove that a completely de novo macromolecular assembly using the CABS model is feasible. This problem is beyond the scope of this work. Previous work has shown that for not too large systems an unrestrained assembly is possible, although the accuracy of the resulting models would be lower. Incomplete templates for multimers (as also demonstrated during CASP6) are often sufficient for a successful modeling by CABS.

The stringiest criterion of the model quality is the RMSD of the ligand (or the second molecule of a dimer) after the best superimposition of the receptor structure (Ligand). Here, the deviations from the crystallographic structure are of the largest magnitude (ranging from 2 to 11 Å). Nevertheless, in all cases the native docking interface was at least partially recovered during modeling. Small structural inaccuracies of the internal conformations of the ligands are the main sources of the RMSD errors (see Figs. 2, 3, 4 and 5). Good examples of such a situation provide the models of ROP homodimers (see Fig. 5).

Finally, the proposed procedure for the selection of the best models needs a comment. The all-atom refinement certainly leads on average to a better selection than that based on ranking according to the CABS energy or/and

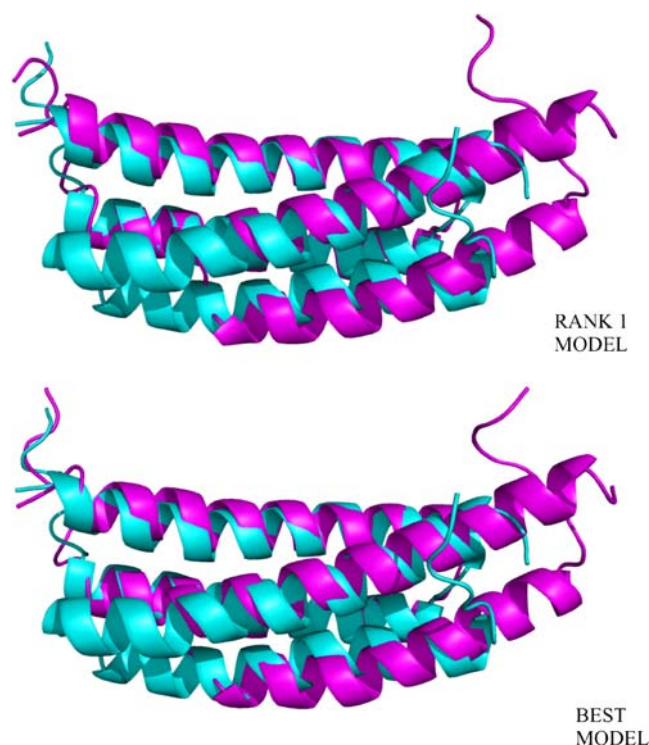


Fig. 5 Superimposition of the predicted and native structures of 1RPR homodimer. Native structure in cyan and predicted in magenta ribbons, respectively

based on the clusters' size after the CABS simulations (data for the CABS-only modeling not given). The proposed model selection procedure almost always (the case of 1VWA is an exception) leads to reasonable quality models. At the same time (as demonstrated in Table 2) better models could always be found in the simulation trajectories. Thus, there is quite a lot of room for an improvement of the proposed modeling method just by a design of more dependable model scoring and selection procedures. It is evident (Table 2) that the AMBER energy after minimization is not always the lowest for the best structures. Other force fields and more elaborated all-atom refinement protocols need to be investigated. Model selection is always an important component of all approaches to reduced-space modeling of proteins. It has been given significant attention during the CASP experiments [10], and it has been concluded that a fully satisfying solution to the problem remains to be found. The most promising possibilities for an improvement of this aspect of proteins and macromolecular assembly modeling include: analysis of a larger number of clusters, modification of the clustering criteria (not only RMSD but also the overlap ratio of the side chain contact maps, etc.) and more extensive all-atom optimization protocols (Molecular Dynamics relaxations after the minimization procedures). All these possibilities are now being carefully investigated. The problem of the protein folding “end-game” (model refinement, scoring and selec-

tion) is being intensively pursued in leading laboratories [19], with a promise of success in the near future. Hopefully, the multiscale approach proposed here will contribute to the solution of this problem.

Conclusions

A hierarchical multiscale method of semi-automated modeling of macromolecular assemblies composed of protein and peptides is described and evaluated. The modeling pipe-line starts from Monte Carlo simulation of the assembly process using the CABS reduced model of polypeptides. The crude models are subsequently subject to clustering, all-atom reconstruction and energy minimization with an AMBER 99 force field. It has been shown that the method generates qualitatively correct structures, in many cases of a near-experimental quality. The model selection procedure is capable of picking models that are much better than average from a large set of CABS structures. While acceptable, the model selection is far from being perfect—better models could always be found in the trajectories from the CABS simulations. Thus, the problem of model scoring and selection needs further study. Poor correlation of geometric accuracy with model energy seems to be a common deficiency of all reduced-space approaches to protein folding [11, 19].

Finally, it should be mentioned that the performance of the proposed method does not deteriorate significantly when, instead of the experimental structures, the models built by means of comparative modeling are used as the source of the restraints for the receptor. Also, “suspicious” parts of the receptors could be safely made fully flexible. Work in progress is directed towards improvement of the model scoring, and an extension of the method onto non-peptide ligands and onto interactions with nucleic acids.

Acknowledgement This work was partially supported by Polish Ministry of Scientific Research and Information Technology (PBZ-KBN-088/P04/2003).

References

1. Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Sali A (2000) *Nat Struct Biol* 7:986–990
2. Schneidman-Duhovny D, Nussinov R, Wolfson HJ (2004) *Curr Med Chem* 11:91–107
3. Gabb HA, Jackson RM, Sternberg MJ (1997) *J Mol Biol* 272:106–120
4. Mandell JG, Roberts VA, Pique ME, Kotlovyy V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF (2001) *Protein Eng* 14: 105–113
5. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) *J Comput Aided Mol Des* 15:411–428
6. Claussen H, Buning C, Rarey M, Lengauer T (2001) *J Mol Biol* 308:377–395
7. Kolinski A (2004) *Acta Biochim Pol* 51:349–371
8. Swendsen RH, Wang JS (1986) *Phys Rev Lett* 57:2607–2609
9. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) *J Comput Chem* 24:1999–2012
10. Kolinski A, Bujnicki JM (2005) *Proteins* 61:84–90
11. Vincent JJ, Tai CH, Sathyanarayana BK, Lee B (2005) *Proteins* 61:67–83
12. Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A (2005) *Proteins* 61:27–45
13. Kmiecik S, Kurcinski M, Rutkowska A, Gront D, Kolinski A (2006) *Acta Biochim Pol* 53:131–144
14. Kurcinski M, Kolinski A (2006) *J Steroid Biochem* (in press)
15. Plewczynska D, Kolinski A (2005) *Macromol Theor Simul* 14:444–451
16. Gront D, Kolinski A (2005) *Bioinformatics* 21:3179–3180
17. Gront D, Kmiecik S, Kolinski A (2006) *J Comput Chem* (in press)
18. Claessens M, Van Cutsem E, Lasters I, Wodak S (1989) *Protein Eng* 2:335–345
19. Bradley P, Misura KM, Baker D (2005) *Science* 309:1868–1871
20. Vajda S, Camacho CJ (2004) *Trends Biotechnol* 22:110–116