Protein–protein docking with a reduced protein model accounting for side-chain flexibility

MARTIN ZACHARIAS

Computational Biology, School of Engineering and Science, International University Bremen, 28759 Bremen, Germany

Abstract

A protein–protein docking approach has been developed based on a reduced protein representation with up to three pseudo atoms per amino acid residue. Docking is performed by energy minimization in rotational and translational degrees of freedom. The reduced protein representation allows an efficient search for docking minima on the protein surfaces within. During docking, an effective energy function between pseudo atoms has been used based on amino acid size and physico-chemical character. Energy minimization of protein test complexes in the reduced representation results in geometries close to experiment with backbone root mean square deviations (RMSDs) of ~1 to 3 Å for the mobile protein partner from the experimental geometry. For most test cases, the energy-minimized experimental structure scores among the top five energy minima in systematic docking studies when using both partners in their bound conformations. To account for side-chain conformational changes in case of using unbound protein conformations, a multicopy approach has been used to select the most favorable side-chain conformation during the docking process. The multicopy approach significantly improves the docking performance, using unbound (apo) binding partners without a significant increase in computer time. For most docking test systems using unbound partners, and without accounting for any information about the known binding geometry, a solution within ~2 to 3.5 Å RMSD of the full mobile partner from the experimental geometry was found among the 40 top-scoring complexes. The approach could be extended to include protein loop flexibility, and might also be useful for docking of modeled protein structures.

Keywords: Protein–protein interaction; protein docking; protein interaction geometry; docking minimization; protein modeling; biomolecular simulation

Many biological processes involve protein–protein interactions. A variety of experimental and computational techniques can be used to identify possible protein binding partners of a given protein. At the same time, the number of experimentally determined protein structures is rapidly increasing. Furthermore, for many more known protein sequences, relatively accurate three-dimensional protein models can be built based on sequence similarity to a known protein structure. However, compared with the available experimental protein structures, the number of experimentally solved protein–protein complexes is still small. Experimental determination of protein–protein complexes is in many cases more difficult than is determination of the isolated components. This is further complicated by the fact that many proteins can interact with several partners, requiring the determination of possible combinations of these protein–protein interactions. The prediction of putative protein–protein interaction geometries by using computational docking methods is therefore of increasing importance to obtain accurate models of protein–protein complexes. The possibility to model the structure of many proteins based on sequence similarity to a known protein structure and the long-term goal to use such models in protein–protein dock-

ing efforts requires algorithms that tolerate errors in loop and side-chain placements of modeled protein structures.

Most current protein–protein docking methods use rigid protein partners to identify protein-binding geometries with optimal surface and electrostatic complementarity. Several methods have been developed to efficiently search the entire six-dimensional rotational and translational space of one protein partner or reduce the task by precalculating putative binding regions on both protein partners before docking (for review, see Halperin et al. 2002; Smith and Sternberg 2002). The search for optimal geometric complementarity can be performed by matching the position of surface spheres and surface normals (Shoichet and Kuntz 1991), application of real space (Palma et al. 2000), or Fourier correlation techniques (Katchalski-Katzir et al. 1992; Meyer et al. 1996; Gabb et al. 1997; Vakser 1997; Ritchie and Kemp 2000; Chen and Weng 2002) to identify an optimal overlap between complementary surface layers of the protein partners.

Other algorithms use matching of surface cubes (Jiang and Kim 1991) or geometric hashing, that is, the identification and matching of convex and concave protein surface regions (Helmer-Citterich and Tramontano 1994; Norel et al. 1994, 1999; Fischer et al. 1995; Lenhof 1996; Ausiello et al. 1997), to find possible binding geometries. Genetic algorithms (Gardiner et al. 2001), Brownian dynamics simulations (Gabdoulline and Wade 2001), and combinations of Brownian dynamics and energy minimization (Fernandez-Recio et al. 2002) have also been applied to tackle the protein docking problem.

To smooth the protein–protein interface, reduced protein models down to one pseudo atom per amino acid (located at the side-chain center) have been used in docking studies by Cherfils et al. (1991, 1994). The search for discrete orientations and translations of both binding partners has been performed either systematically or randomly with a Monte Carlo approach, and each docked complex was evaluated by the amount of buried surface area to determine the degree of surface complementarity (Cherfils et al. 1991, 1994; Cherfils and Janin 1993; Janin 1995). In case of the Fourier correlation techniques, various discrete orientations of one protein relative to the other are used to scan translational space for optimal geometric complementarity. To limit the computational demand, often relatively coarse Euler angle rotation grids of ~15 to 30 degrees are necessary for a systematic docking run, still resulting in many possible docked complexes (Smith and Sternberg 2002).

In a second step, a subset of the generated favorable protein–protein complex geometries are translated into atomic resolution structures and usually filtered based on electrostatics, a knowledge-based residue–residue potential (Moont et al. 1999; Glaser et al. 2001) and/or biochemical information (Cherfils et al. 1994; Gabb et al. 1997; Smith and Sternberg 2002). However, rotation of a protein partner bound to a second protein by only a few degrees can lead to

steric overlap between the two proteins and may interfere with the ability to identify the observed binding position as a favorable site. The problem can be partially overcome by using a very smooth or soft protein surface (Tovchigrechko et al. 2002) and neglect or pruning of larger and flexible surface side-chains (Lorber and Shoichet 1998; Lorber et al. 2002). Usually, only after generation of rigidly docked protein complexes, conformational flexibility of side-chains, and possibly of the protein backbone, is introduced (Jackson et al. 1998; Camacho and Vajda 2001; Fernandez-Recio et al. 2002; Lorber et al. 2002; for review, see Smith and Sternberg 2002). A shortcoming of this scheme is the possibility to overlook realistic docking geometries in the first rigid docking search due to the neglect of possible conformational rearrangements necessary for favorable protein–protein association.

An alternative docking procedure is to use energy minimization in the rotational and translational degrees of freedom of one protein partner to generate docked protein complexes. An advantage of such an approach is that any residual unfavorable interactions caused by the discrete sampling of rotational degrees of freedom (used by some docking methods) that can interfere with a realistic complex evaluation is removed. In addition, instead of filtering large numbers of complexes after docking, the evaluation of docked complexes based on surface complementarity, electrostatics, knowledge-based potentials, or other biochemical information can be taken into account during docking minimization. However, at atomic resolution, and in particular if one takes conformational flexibility into account energy, minimization is very time-consuming and might not be useful for systematic docking efforts. In addition, at atomic resolution, a realistic evaluation of docked complexes can be difficult due to the relatively "hard" Lennard-Jones (LJ) interatomic potentials and due to partial charges on each atom that require a well-balanced calculation of pair-wise Coulomb and electrostatic solvation contributions (Zacharias et al. 1994; Jackson and Sternberg 1995; Jackson et al. 1998; Smith and Sternberg 2002). A recent study, however, that combines preselection of docking geometries generated with a pseudo Brownian dynamics simulation (with rigid protein partners) followed by energy minimization using a limited number of start geometries, shows promising results (Fernandez-Recio et al. 2002).

A reduced protein model with pseudo atoms that represent groups of atoms allows a much faster calculation of protein–protein interactions. At the same time, the number of distinct docking minima is much smaller compared with an atomic detail representation of the protein surfaces. In the present study, a reduced protein model with up to three pseudo atoms per amino acid residue has been used. The resolution is higher than in previous docking studies with reduced models of one pseudo atom per residue (Cherfils et al. 1991, 1994), and results in energy-minimized complex

geometries close to the experimental geometry. Interactions between pseudo atoms are calculated by using a soft LJ-type potential with parameters that reflect the hydrophilic or hydrophobic character of the side-chain. The flexibility of critical side-chains at the protein surface has been treated by using several side-chain copies that represent different rotametric states of the side-chains. To select the best fitting copy, the approach allows to switch between different side-chain copies during the docking minimization. Iteration of this process for each putative docking geometry allows a simultaneous optimization of docking geometry and optimal fitting side-chain geometries. The application of the approach to several protein pairs in unbound conformations shows promising results.

Results

Energy minimization of protein–protein complexes

Energy minimization of several protein–protein complexes in the pseudo atom representation (parameters are given in Table 1) in translational and rotational degrees of freedom of the ligand protein yielded energy-minimum geometries that stay close to experimental start geometries (Table 2; Figs. 1, 2). The root mean square deviation (RMSD) of C_α atoms between experimental start structure and energy-minimized structures was <1.5 Å for the full complexes and was, in most cases, between 1.0–2.0 Å when considering only the mobile ligand protein (in two cases, however, the deviation was ~3 Å for the mobile partner; see Table 2). Energy minimization and docking at atomic resolution yield deviations of similar magnitude between calculated and experimental structures (Fernandez-Recio et al. 2002). It should be noted that such small deviation could not be achieved with a lower-resolution representation of the protein, for example, by using only one pseudo atom per residue (data not shown).

Docking from multiple start positions and orientations

The density of surface start positions and orientations was systematically varied to find out how many starting positions and orientations are necessary to identify energy minima close to the experimental geometry in a systematic docking effort. No information about the experimentally known docking geometry was taken into account during the search. For the trypsin/BPTI (bovine pancreatic trypsin inhibitor) test case, the number of energy minima within 10 energy units of the lowest-energy complex geometry has been plotted as the number of start configurations (Fig. 3). A density of ~35,000 start configurations corresponds to about one start position every 4 to 5 Å of the receptor surface and ~260 different ligand orientations at each starting position. Although the number of new low energy

Table 1. *Lennard-Jones (8/6) parameters*

Atom	R_i (Å)	A_i ($[RT]^{0.5}$)
ALL 1	2.00	1.0
ALA 2	1.90	1.0
ARG 2	1.95	2.0
ARG 3	1.90	0.6
ASN 2	1.90	0.6
ASP 2	1.90	0.5
CYS 2	2.00	2.0
GLN 2	2.00	1.5
GLN 3	1.90	0.6
GLU 2	2.00	1.5
GLU 3	1.90	0.5
HIS 2	2.00	1.5
HIS 3	1.90	0.5
ILE 2	2.20	3.0
LEU 2	2.20	3.0
LYS 2	2.00	2.0
LYS 3	1.90	0.6
MET 2	2.00	2.0
MET 3	2.00	2.9
PHE 2	2.20	2.0
PHE 3	2.20	3.0
PRO 2	1.90	1.0
SER 2	1.90	0.6
THR 2	1.90	0.6
TRP 2	2.00	1.5
TRP 3	2.20	2.6
TYR 2	2.20	2.0
TYR 3	2.20	2.6
VAL 2	2.00	2.5

Pair-wise interactions between pseudo atoms i and j at a distance r_{ij} are calculated from $[E_{ij} = B_{ij}/(r_{ij})^8 - C_{ij}/(r_{ij})^6]$ with $B_{ij} = A_i A_j (R_i + R_j)^8$ and $C_{ij} = A_i A_j (R_i + R_j)^6$. The A_i's are in units $(RT)^{0.5}$, with R and T being the gas constant and temperature, respectively. ALL 1 corresponds to the first (backbone) pseudo atom of each amino acid.

minima still increases beyond 60,000 start configurations, the increase is much smaller than in the range <40,000 start structures. Energy minimizations of ~35,000 start geometries for this test case took ~3.5 h CPU time on a 1.3 GHz Athlon PC. At this density, the energy-minimized experimental complex structure (i.e., same energy minimum as the one obtained by starting minimization from the experimental complex structure) was found >30 times. A similar start configuration density was used for all subsequent docking simulations.

Test runs on all complexes shown in Table 2 using the protein partners in their bound conformations (i.e., the conformation in the protein–protein complex) yielded the energy-minimized experimental geometry at least three times and in most cases more than 10 times (Table 2). This result indicates that the above density of start configurations is sufficient to obtain with high probability an energy minimum close to experiment among the docking solutions.

In addition, the systematic docking approach applied to a set of protein pairs in their bound conformations yielded, in

Table 2. *Systematic docking of bound protein structures*

Complex name	PDB code	Energy (RT)	Rmsd (Å)	Rank$_{EMxray}$	N$_{EMxray}$
β-Trypsin/BPTI	2PTC	−21.6	0.45/1.5	1	39
Subtilisin BPN/inhibitor	2SIC	−28.5	0.65/2.2	2	22
Kallikrein/PTI	2KAI	−27.5	0.55/1.7	1	31
α-Chymotrypsin/ovomucoid 3[d]	1CHO	−21.4	0.61/1.8	2	26
α-Chymotrypsin/eglin C	1ACB	−25.2	1.12/3.4	1	7
Uracil-DNA glycosylase/inhib.	1UGH	−29.7	0.41/1.1	1	4
1gG1 D44.1 Fab-frag/lysozyme	1MLC	−14.4	0.71/2.1	94	6
Human growth hormone/receptor	3HHR	−25.4	0.70/1.2	12	3
Barnase/barstar	1BRS	−20.3	0.60/1.9	1	5
Acetylcholinesterase/fasciculin II	1FSS	−23.3	0.30/1.4	4	3
Trypsin/soybean inhibitor	1AVW	−25.1	1.10/3.1	1	5
Thermithase/eglin C	2TEC	−19.2	0.8/1.8	2	32
Subtilisin N/chymotrypsin-inhibitor	2SNI	−25.1	0.5/1.5	1	17
Trypsinogen/BPTI	1TGS	−25.2	0.3/0.7	1	4
α-Chymotrypsin/BPTI	1CGI	−33.3	0.35/1.0	1	5
Subtilisin Carlsberg/inhibitor	1CSE	−18.3	0.5/1.2	4	15
Thrombin mutant/BPTI	1BTH	−26.3	0.45/1.5	1	16
Trypsin/amyloid-β-inhibitor	1BRC	−20.9	0.68/2.2	1	35

Column 2 gives the energy of the docking minimum that comes closest to the experimental geometry when using protein structures in their bound conformation taken from the complex structures indicated in column 1 (PDB indicates Protein Data Bank; RT, gas constant and temperature). The complex geometry EMxray indicates the same complex geometry (same minimum) as observed when starting minimization from the experimental complex structure. The root mean square deviation (RMSD, backbone) is given for the whole complex/mobile ligand protein with respect to the experimental complex structure. The fifth and sixth columns indicate the ranking of the geometry that comes closest to the experimental structure and the number of times it was found in a systematic docking run, respectively.

most cases, the energy minimum close to the experimental complex structure as lowest-energy structure or at least among a small number of low-energy complexes (Table 2; Fig. 2). However, a notable exception is the D44.1-Fab fragment/lysozyme complex (an antibody antigen complex). The minimum energy geometry closest to the X-ray structure (PDB code 1MLC) scored significantly worse than several (unrealistic) alternative predicted complexes (Table
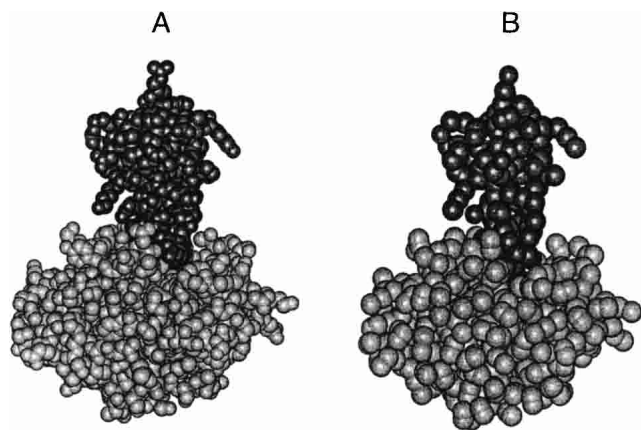

Figure 1. Comparison of the trypsin (light gray)-BPTI (dark gray) complex (PDB code 2PTC) at atomic resolution (*A*; each van der Waals sphere represents a heavy atom) and in the reduced model (*B*; each sphere represents a pseudo atom in the reduced protein model).

2). In this case, the receptor protein consists of two domains with a very large concave interdomain interface. Most of the low-energy complexes show the lysozyme molecule docked at this interface. A second reason for the poor performance in this case might be that in the Fab/lysozyme case, more polar residues are located at the protein–protein interface (Lo Conte et al. 1999), indicating that electrostatic interactions play a more dominant role than hydrophobic interactions. It should be noted that restriction of the start configurations to a region close to the known lysozyme binding region (which is common practice in many docking approaches) selects a position close to experiment as among the low-energy binding positions (data not shown). The result, however, also shows the need to further improve the scoring function or to use different parameter sets for different classes of protein complexes. Although the present set of test complexes includes a variety of protein complexes with known structure antibody–antigen complexes were excluded from further analysis due to the poor performance of the scoring function, even in case of using "bound" structures during docking (see above).

Effect of side-chain conformational changes on docking results

To test the sensitivity of the docking results on the side-chain placement, the conformation of extended side-chains
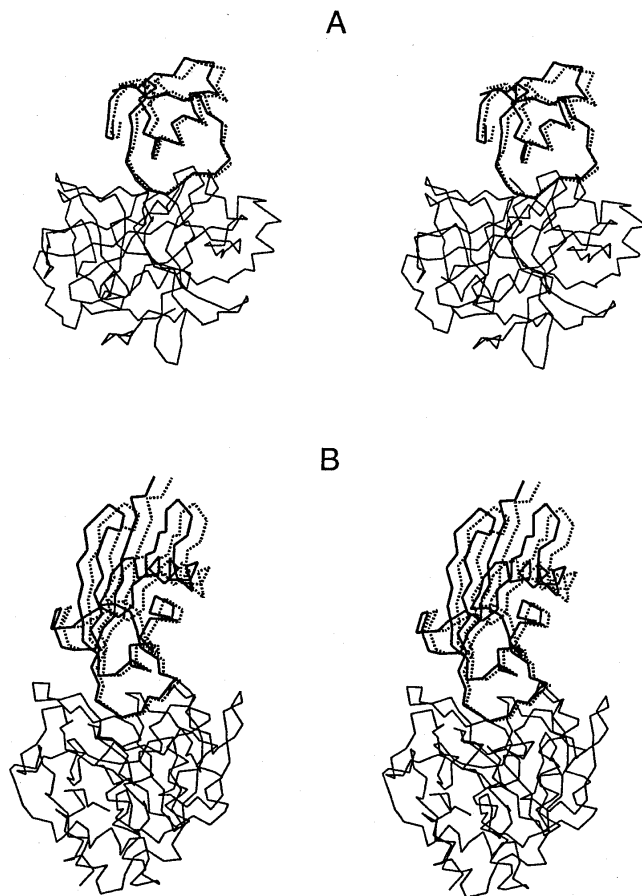
Figure 2. Comparison of energy-minimized (Cα-backbone of mobile protein partner in bold) and experimental protein complexes (dashed line indicates ligand protein backbone; continuous line, receptor protein backbone). (*A*) Stereo view of the subtilisin/inhibitor complex (PDB entry 2SIC); (*B*) Stereo view of the α-chymotrypsin/ovomucoid complex (PDB entry 1CHO). Energy minimization has been performed in translational and rotational degrees of freedom of the ligand proteins by using the reduced protein model and starting from the corresponding experimental structures.

located at the interface between ligand and receptor protein was slightly distorted. In case of BPTI, the Lys_{15}, Arg_{17}, and Arg_{39} side-chains contact the trypsin receptor protein upon complex formation. By randomly rotating side-chain torsion angles of these three residues, four variants with increasing conformational deviation from the bound BPTI structure were generated (Table 3). The average deviation over all side-chain atoms of the three critical residues was in the range of 1.0 to 2.8 Å. These models were translated into the reduced model and used as ligands for systematic docking with the same protocol as for the undistorted bound receptor and ligand structures (see above). As indicated in Table 3, models with a side-chain deviation of 1.0 and 1.5 Å with respect to the original bound BPTI structure still gave a docked complex close to experiment as the top-scoring solution. A side-chain deviation of >2 Å, however, resulted in an increased deviation of the complex that is

closest to experiment (3.1 Å) and decreased rank of this complex (24) with respect to alternative complex geometries.

Docking of unbound structures

For a subset of complexes given in Table 2, structures of both binding partners in the unbound states are available. For these cases, docking simulations were performed by using unbound structures (Table 4). However, only complexes were included for which docking of bound structures yielded top-scoring complexes (among the five top-ranking complexes) in good agreement with experiment.
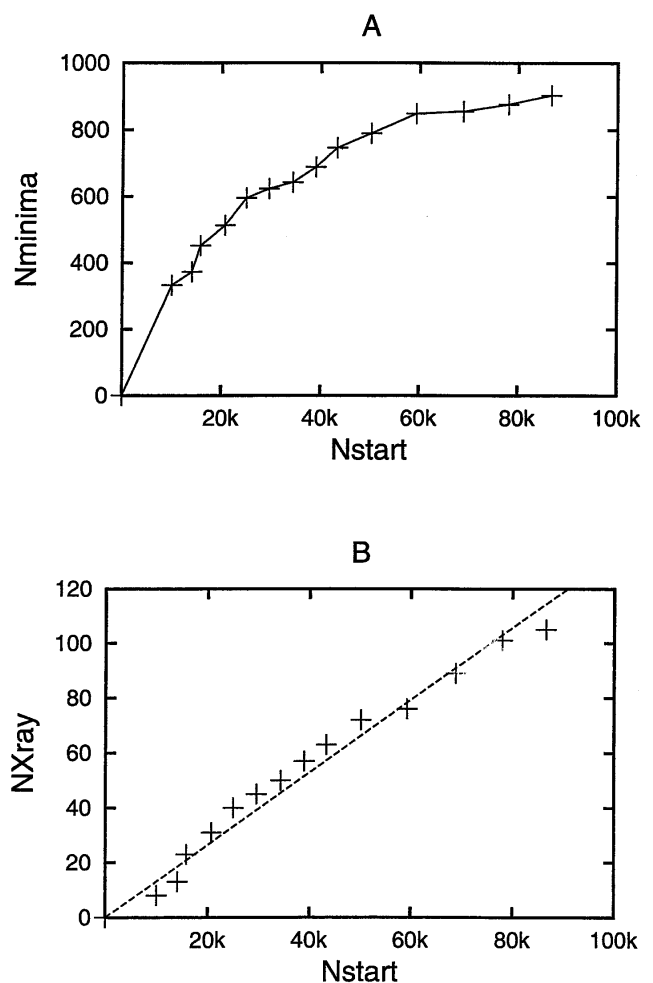

Figure 3. (*A*) Number of distinct energy-minimized complexes (Nminima, within 10 energy units of the lowest-energy complex) versus the number of start configurations (Nstart) for systematic docking studies on the trypsin/BPTI (PDB entry 2PTC) system. Two energy minima are considered as distinct minima if the root mean square deviation of the mobile protein partner is >0.2 Å relative to other energy minima. (*B*) Number of docking trials that yielded the energy-minimized X-ray structure (Nxray) versus the number of docking start configurations spread over the surface of the receptor (Nstart).
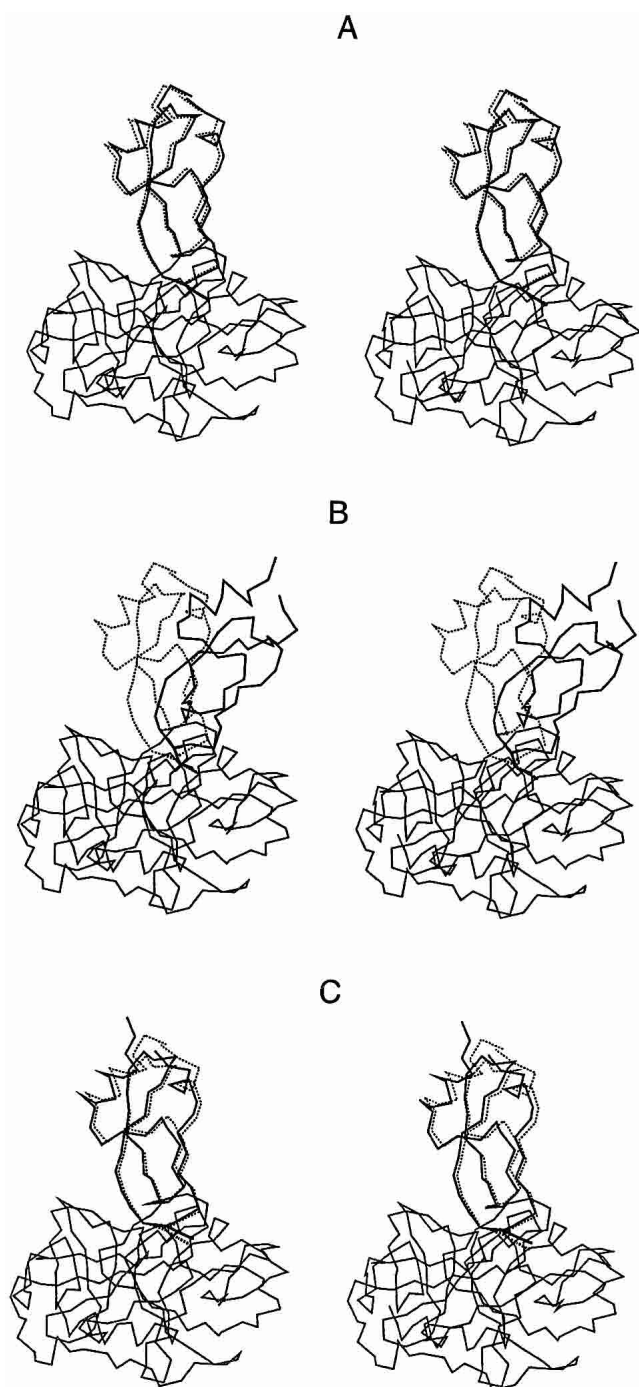
Figure 4. Comparison (stereo view) of energy minimization of the trypsin/BPTI complex using the bound (*A*; PDB code 2PTC), unbound (*B*; PDB codes 2PTN + 4PTI), and unbound conformations including side-chain copies on surface-exposed side-chains for BPTI (*C*), starting from the same placements as found in the experimental complex structure. The Cα-backbone of the mobile BPTI ligand protein after energy minimization is shown in bold (including the Lys$_{15}$ side chain at the docking interface). The backbone of BPTI in the experimental complex is shown as a dashed line (continuous thin line indicates backbone of trypsin in the experimental complex structure).

Table 3. *Effect of side-chain conformational changes on docking results*

Ligand docked to β-trypsin (2PTC)	RMSD of side-chains 15, 17, 39	Rank/RMSD (Å)
BPTI (as in 2PTC)	0.0	1/1.5
Variant 1	1.0	1/1.8
Variant 2	1.5	2/2.0
Variant 3	1.8	6/2.9
Variant 4	2.2	24/3.1

Side-chain conformations of three residues (Lys$_{15}$, Arg$_{17}$, Arg$_{39}$) at the β-trypsin/BPT1 interface were randomly distorted (variants 1–4) to different degrees (column 2) and used in systematic docking efforts (ranking/RMSD, respectively, of the solution that comes closest to experiment is given in column 3). RMSD indicates root mean square deviation.

Even for docking protein partners in their unbound conformations, the obtained energy minima are, in several cases, reasonably close to the experimental complex structures and still score among the top-ranking docked complexes. Examples are α-chymotrypsin in complex with ovomucoid three-dimensional domain, the uracil DNA glycosidase/inhibitor complex, the trypsin/soybean inhibitor complex, the trypsinogen/BPTI case, and the subtilisin Carlsberg/inhibitor complex (see Table 4). Although the conformation of the unbound protein structures can deviate slightly from the conformation in the protein complex, this appears to be tolerated due to the reduced protein representation and soft interaction functions. The results for some other complexes are still acceptable, in which complexes in reasonable agreement with experiment (overall RMSD, ~3 Å and 5 to 8 Å for the ligand protein alone), were found among the 20 top-ranking energy minima (e.g., the barnase/barstar docking case).

It might be of interest to note that in some docking studies, the RMSD of atoms within ~10 Å of the protein–protein binding site has been reported as a criterion for successful docking (Gabb et al. 1997; Chen and Weng 2002). With this measure, the interface backbone deviation in case of the uracil DNA glycosidase/inhibitor system is 1.3 Å (versus 3.2 Å if one considers only the mobile ligand protein; see Table 4) for the top-ranked complex using unbound partners compared with the experimental interface geometry. This is close to the RMSD given for the whole complex (first number in the RMSD column in Table 4) and well within an interface deviation of up to 2.5 Å considered as successful docking (Gabb et al. 1997; Chen and Weng 2002). The RMSD reported for the whole complex in the present study might be more appropriate for comparing the present results with other docking studies that report the RMSD of interface residues of the docked complex relative the experimental geometry.

However, for several cases no low-energy minima close to the experimental complex geometry were found and/or those with the smallest deviation from experiment show a

Table 4. *Docking of unbound protein structures*

Complex	Protein Data Bank codes of partners	Energy (RT)	RMSD (Å)	Rank I	Rank II
β-Trypsin/BPTI	2PTN + 4PTI	−11.2/−19.8	1.4/3.5	765/~15k	275/~6k
Subtilisin-BPN/inhibitor	1SUP + 3SSI	−17.3/−18.5	2.9/9.9	7/~13k	4/~5k
Kallikrein/BPTI	2PKA + 6PTI	−12.6/−17.1	1.5/3.4	167/~12k	68/~3k
α-Chymotrypsin/ovomucoid 3^d	5CHA + 2OVO	−19.2/−21.8	1.2/2.8	9/~12k	3/~3k
U-DNA glycosylase/inhibitor	1AKZ + 1UGI	−23.5/−23.5	1.3/3.2	1/~14k	1/~4k
Barnase/barstar	1A2P + 1A19	−17.3/−19.3	2.2/5.8	10/~13k	5/~3.5k
Acetylcholinesterase/fasciculin	2ACE + 1FSC	−17.1/−22.2	0.7/2.4	68/~24k	22/~6k
Trypsin/soybean inhibitor	2PTN + 1BA7	−19.5/−26.4	1.1/4.9	17/~18k	5/~5.5k
Thermithase/eglin C	1THM + 2TEC	−17.0/−18.3	1.1/3.2	10/~17k	6/~4k
SubtilisinN/chymotryp.-inhibitor	1SUP + 2CI2	−9.7/−18.6	1.6/3.8	1943/~15k	671/~3.5k
Trypsinogen/BPTI	2PTN + 1HPT	−16.2/−20.3	1.3/1.8	17/~3k	14/~3k
α-Chymotrypsin/BPTI	1CHG + 1HPT	−15.8/−20.6	1.7/3.4	154/~14k	34/~3.5k
Subtilisin Carlsberg/inhibitor	1SCD + 1ACB	−16.3/−19.4	1.5/2.7	10/~17k	6/~4k
Thrombin mutant/BPTI	2HNT + 6PTI	−15.7/−19.4	3.1/7.8	17/~10k	12/~3k
Trypsin/amyloid-β-inhibitor	1BRA + 1AAP	−19.1/−23.0	4.4/9.2	27/~14k	21/~3k

Column 1 gives name and Protein Data Bank entry of the protein structures used for docking. The energy of the complex that came closest to experiment/lowest energy complex, respectively, is given in column 3 (R and T indicates gas constant and temperature, respectively). The root mean square deviation (RMSD) of the complex/mobile ligand protein (closest to experiment), respectively, is given in column 4. The last two columns indicate the ranking with respect to two criteria (fifth column, two docked complexes are counted as nonequivalent if the RMSD of the mobile ligand or the distance from the center of the receptor differs by <0.2 Å; sixth column, same but using a 0.5 Å limit; the approximate number of energy minima is also given in units of 1000).

poor scoring (Table 4). Comparison of bound and unbound structures indicates that one major reason for the poorer performance of the approach in case of unbound structures is the presence of a few critical surface side-chains that change conformation upon complex formation. As an example, the effect of using the unbound BPTI structure on docking energy minimization to trypsin has been illustrated in Figure 4. It leads to a reduced ranking and to a complex geometry that significantly deviates from the energy-minimized bound structure (Table 4). Conformational changes in particular of side-chains important for mediating protein–protein interactions is a major difficulty that limits the applicability of many docking approaches (Halperin et al. 2002). For docking with the present reduced protein model, this is mainly due to residues with two pseudo side-chain atoms.

Docking with side-chain copies

Amino acid side-chains preferably adopt conformations in folded proteins and also in protein complexes that correspond to distinct rotameric states of the side-chain dihedral angles (Tuffery et al. 1991; Dunbrack and Karplus 1993; Dunbrack 2002). Based on this observation, a possible way to treat side-chain flexibility is to use several discrete (fixed) side-chain copies during docking and to select the lowest-energy copy for each critical side-chain.

In the present approach, several side-chain copies are present during the docking minimization (in rotational and translational degrees of freedom), and the selected side-chain copy can change during the minimization procedure. It was shown above that the reduced representation tolerates side-chain misplacements by up to ~1.5 to 2 Å during docking (Table 3). In particular, changes in the 'higher' dihedral angles beyond χ_1 and χ_2 (for long side-chains such as Lys or Arg) lead only to small cartesian coordinate changes of <2 Å. As a consequence, the reduced side-chain description allows to represent the conformational side-chain space by fewer copies than necessary at atomic resolution.

For an efficient selection of reasonable start structures for docking with multiple side-chain copies, a systematic docking run was first performed, in which solvent exposed side-chains (with two side-chain pseudo atoms) were pruned of their second side-chain pseudo atom. This leads to several docked complexes that come relatively close to within ~3 to 5 Å of the experimental geometry and scoring at least among the top 1000 complexes. This subset of the docked complexes served as start structures for docking with several side-chain copies (the ~1000 start structures can be minimized in a few minutes of workstation time).

Because both docked complexes with a realistic and with an unrealistic geometry can profit from additional side-chain copies, the influence of the number of side-chain copies on the docking results was first tested for the trypsin/BPTI system (Table 5). For the test, increasing numbers of side-chain copies were placed onto three residues (Lys_{15}, Arg_{17}, Arg_{39}) located at the trypsin/BPTI interface in the experimental complex structure. Inspection of Table 5 in-

Table 5. *Docking of trypsin/BPTI (2PTN + 4PTI) with side-chain copies on As 15, 17, and 39*

No. of side-chain copies	Rank/RMSD (Å)
1	765/3.5
2	7/1.5
3	9/1.5
4–5	13/1.5

The number of side-chain copies of residues Lys_{15}, Arg_{17}, and Arg_{39} of BPTI present during docking is indicated in column 1. In case of one copy, only the conformation in unbound structure was present. For larger copy numbers, the conformation as found in the bound complex (two-copy case) and additional possible rotamer states have been included. RMSD indicates root mean square deviation.

dicates that addition of side-chain copies with a conformation as found in the bound structure of BPTI to the free BPTI structure significantly improves the docking results. A docking placement close to the experimental geometry was identified among the 10 top-scoring complexes. This is a very significant improvement compared with docking using the free BPTI structure during docking (see Table 4). A further increase of the copy number still identifies a ligand protein placement close to experiment among the top-scoring complexes. However, because alternative complex geometries also profit from the availability of multiple side-chain copies, the specificity of the approach begins to degrade beyond more than five copies on the critical positions. It should be noted that for simplicity, each side-chain copy has the same weight (no energetic weighting of each rotamer). Inclusion of such a rotamer-dependent penalty in future studies may further enhance the specificity of the approach.

The approach was also applied to those protein complexes that performed poorly during docking with unbound structures. In these cases all surface-exposed side-chains were represented by up to five or six side-chain copies. The copies correspond to reduced representations of favorable rotamers with respect to the χ_1 (Phe, Tyr, His) and χ_2 (for Arg, Lys, Gln, Glu, Met, and Trp) dihedral angles of the corresponding side-chain. Other side-chain torsions were kept in extended trans-configuration. No information about the side-chain conformations in the experimentally known complex structure was included during side-chain copy generation. The use of several side-chain copies significantly improved the results of docking simulations using unbound structures (compare Tables 4, 6). For all cases shown in Table 6, a significantly better agreement of docked complexes with experiment was found, and the ranking of this complex was in all cases also improved compared with docking with unbound structures and no side-chain copies (Table 6; Fig. 4). For the kallikrein/BPTI, the subtilisin/chymotrypsin inhibitor, and the α-chymotrypsin/BPTI case, the complex that showed closest agreement with the experimental structure still scored only at positions 30 to 70 (or 10 to 40 for the less stringent distinction of minima). The second part of Table 6 indicates that this is to a large part due to conformational changes in the receptor protein not considered in the present study. To test the influence of the receptor protein structure, docking studies were performed with the unbound ligand protein structures, including side-chain copies and the corresponding receptor proteins in the bound conformations (Table 6, lower part). In all cases, the complex geometry that comes closest to experiment scored among the top 10 docking minima. This demonstrates that conformational changes in the receptor protein during complex formation are mainly responsible for the nonoptimal scoring of the above-mentioned three cases. It should be noted that it is also possible to use the side-chain copy

Table 6. *Docking using unbound structures with side-chain copies*

Complex (binding to unbound receptor)	Protein Data Bank code	Energy (RT)	RMSD (Å)	Rank I	Rank II
β-Trypsin/BPTI	2PTN + 4PTI	−19.6 (−18.9)/−22.3	1.9/2.7 (0.6/1.4)	11(24)/~10k	8(13)/~3.0k
Subtilisin/BPN/Inhibitor	1SUP + 3SSI	−20.9 (−17.3)/−22.9	2.6/5.7 (1.6/3.7)	2(82)/~10k	2(38)/~4.5k
Kallikrein/BPTI	2PKA + 6PTI	−13.0−16.7	1.4/2.4	39/~9k	16/~2.6k
SubtiN/chymotryp.-I.	1SUP + 2CI2	−17.3/−21.2	2.1/3.7	69/~10k	39/~2.9k
α-Chymotrypsin/BPTI	1CHG + 1HPT	−18.2/−21.5	1.8/3.0	32/~10k	12/~2.7k
Trypsin/amyloid-β-I.	1BRA + 1AAP	−23.2/−23.5	2.0/3.1	5/~10k	5/~2.5k
Complex (binding to bound receptor)					
β-Trypsin/BPTI	2PTC + 4PTI	−21.6/−24.3	0.8/1.8	2/~10k	2/~3.0k
Subtilisin/BPN/Inhibitor	1SIC + 3SSI	−21.5/−22.7	1.6/3.3	7/~10k	5/~4.5k
Kallikrein/BPTI	2KAI + 6PTI	−14.7/−16.1	1.7/3.0	8/~9k	5/~2.4k
SubtiN/chymotryp.-I.	1SNI + 2CI2	−19.4/−25.0	1.6/2.7	21/~10k	9/~3.0k
α-Chymotrypsin/BPTI	1CGI + 1HPT	−21.3/−26.3	1.9/3.0	6/~10k	3/~2.6k
Trypsin/amyloid-β-I.	1BRA + 1AAP	−22.4/−22.4	1.9/3.1	1/~10k	1/~2.6k

The upper part of the table summarizes the results of systematic docking studies with both ligand and receptor proteins in unbound conformations and including side-chain copies for the ligand proteins. Results given in the lower part of the Table have been obtained from docking to the receptor protein in the bound conformation (ligand in the unbound conformation including side chain copies). See also legend of Table 4.

approach for the receptor protein, and this possibility will be tested in future studies.

Discussion

In contrast to many existing docking approaches that use discrete orientations of protein partners to search for an optimal translational fit, in the present docking approach, energy minimization in rotational and translational space of one partner has been performed. Small discrete rotations of one protein partner relative to the other may affect the energetic ranking of a docked complex, even after translational readjustment by correlation methods. Docking by energy minimization avoids such effects. A reduced protein representation during energy minimization enhances the efficiency of the approach, because it allows much faster minimization than at atomic resolution and, at the same time, greatly reduces the number of energy minima on the surface of the protein partners. The results of the present study show that with a reduced model that represents amino acids by up to three pseudo-atoms, it is possible to obtain energy minima in very reasonable agreement with the corresponding experimental complex structures. The deviation is not much larger than are deviations that are obtained in energy minimization studies at atomic resolution (see Fernandez-Recio et al. 2002). This may allow a direct translation of the energy-minimized complexes into atomic resolution models (by superimposing the atomic resolution structures onto the reduced models), resulting in structures with reasonable stereochemistry that are useful start structures for energy minimization at atomic resolution.

In the initial stage of most present docking methods, docked complexes are collected that show reasonable geometric and/or electrostatic complementarity. In a second stage, these solutions are ranked according to an energy function, for example, an atomic resolution force field or a knowledge-based residue-residue potential (for review, see Smith and Sternberg 2002). In the current approach, a residue–residue potential has been used that ranks not only the amount of surface complementarity but also the hydrophilic and/or hydrophobic character of the contacting protein regions during the docking process. The energy function was sufficient to rank docking geometries close to experiment as top-ranking complexes in systematic docking simulations using protein partners in their bound conformation. However, association can lead to conformational changes in protein partners (Betts et al. 1999; Najmanovich et al. 2000). Test calculations showed that the reduced protein representation allows for some tolerance with regard to the exact conformation of side-chains at the protein–protein interface during docking. This leads also to reasonable performance in docking studies using protein partners in their unbound conformations for cases in which conformational changes upon complex formation are small. However, in several

other cases the docking results were unsatisfactory when using unbound structures, mostly due to the conformation of some critical side-chains. In these cases, the use of several side-chain copies present during docking and selection of the best-fitting copy during EM significantly enhances the performance of the approach also for docking of unbound structures.

Discrete side-chain copies at atomic resolution have been used in order to predict side-chain placements in protein homology modeling (see Tuffery et al. 1991) and also in protein docking simulations for preselected docked complexes (Jackson et al. 1998; Lorber and Shoichet 1998; Lorber et al. 2002). Most related to the present approach is the method by Jackson et al. (1998), which introduces a mean field calculated from possible rotameric states of side-chains located at the protein–protein interface. This mean field is used to further refine the complex geometry in translational and rotational coordinates. However, the method is limited to a small number of putative complex starting geometries because each refinement at atomic resolution requires several minutes of workstation time (Jackson et al. 1998). As a consequence, a small number of putative complexes reasonably close to the experimental geometry must either be known prior to application of the search for optimal side-chain rotamers or the initial docking result (with incorrect side-chain geometry) must be close to the correct (experimental) geometry.

In a recent approach by Lorber et al. (2002), the rigid part of the ligand protein (without flexible side-chains) is first docked in various orientations (many million discrete orientations) into the (known) receptor binding site. For each ligand orientation, the best-fitting combination of side-chain conformations at the interface is selected. This approach shows very promising results on a number of test systems (Lorber et al. 2002). However, it is expected that the optimal side-chain geometry depends on the correct initial positioning of the ligand protein at the receptor binding site, which may not be always possible when docking with unbound structures.

In the present method, side-chain copy selection is coupled to the energy minimization in translational and rotational degrees of freedom of the mobile protein partner in an iterative way. For each docking attempt, the final docked complex contains the lowest-energy (best fitting) side-chain copies for the given complex geometry. It is important to note that the present method tackles the problem of side-chain flexibility but is expected to fail in cases in which large-scale conformational changes occur during the protein–protein association process.

A relatively simple scoring function for energy minimization and evaluation of docked complexes has been used that mainly distinguishes between hydrophobic and hydrophilic side-chains, penalizes contacts between charged side-chains, and weakly attracts pairs of side-chains with oppo-

site charges. This function performed reasonably well in case of docking bound structures, but also for most docking applications with unbound structures in case of a side-chain geometry close to the geometry in the experimental protein–protein complex. However, for certain types of protein–protein complexes, in particular the antibody/antigen test case (D44.1 Fab fragment/lysozyme complex), the performance of the scoring function was unsatisfactory. Although the current simple energy function includes solvation effects indirectly (due to the neglect of attractive interactions between hydrophilic side-chains or hydrophilic and hydrophobic side-chains), the scoring scheme may not be accurate enough to account for the balance between electrostatic solvation and Coulomb interaction that could make a more dominant contribution in the D44.1 Fab fragment/lysozyme case compared with other complexes (Jackson 1999).

It should be emphasised that the primary focus of the current study was to present a rapid docking approach that approximately accounts for side-chain conformational flexibility and to demonstrate its performance. There are several ways to improve the scoring function used in the current approach in future studies. Recent work by Janin and co-workers (Lo Conte et al. 1999; Chakrabarti and Janin 2002) indicates characteristic patterns of amino acids at and around a protein–protein interaction interface. Such information could be incorporated into the docking scoring scheme. Another possibility is to use different sets of parameters for different types of protein–protein complexes.

However, even with the present scoring function, the multistart energy minimization docking approach with side-chain copies on the ligand protein yields complex geometries in reasonable agreement with experiment among the top 20 docked complexes, in most cases using unbound ligand and receptor protein structures. In cases, however, in which global protein conformational changes or side-chain conformational changes on the receptor side play a dominant role, the score of complex geometries close to experiment becomes worse. Most existing docking methods identify complexes close to experiment among several 100 top-docking solutions and may require extensive filtering of many structures to reduce the list of possible solutions (Gabb et al. 1997; Camacho et al. 2000; Chen and Weng 2002). Subsequent refinement steps, including Monte Carlo searches and minimization at atomic resolution, can then be applied to further refine the docking solutions (Jackson et al. 1998; Camacho et al. 1999, 2001). A recent approach by Fernandez-Recio et al. (2002) identified complex geometries close to experiment among a few top-scoring complexes when using unbound protein structures. In this approach, a first initial screen is performed using Brownian dynamic simulations with rigid protein partners, followed by an energy minimization phase and a final refinement step that allows also side-chain motion at the protein–protein interface. All steps are performed at atomic resolution, and

the search is restricted to only part of the receptor region. The initial screen using rigid unbound structures may, however, produce many false-positive complex geometries that need to be further evaluated in the subsequent optimization and refinement steps, which could be computationally expensive (a time of 7 to 20 min per structure has been reported by Fernandez-Recio et al. 2002).

A possible advantage of the present approach is that docking optimisation and selection of a best fitting side-chain copy are performed simultaneously/iteratively. In addition, the use of a reduced protein model allows the search on the complete surface of the receptor protein in reasonable computer time of ~4 to 8 h (single processor time), which is comparable or faster than Fourier correlation methods or real-space correlation techniques (Gabb et al. 1997; Palma et al. 2000).

It should also be noted that in many cases, some knowledge about putative protein–protein interaction regions exists prior to docking. It is expected that restriction of the number of start geometries will further improve results in terms of prediction accuracy and speed. Another possible way to improve the performance of the approach is to combine it with a smarter generation of docking start geometries. In geometric hashing methods, surface patches on both protein partners that fit to each other are matched by using an appropriate rotational and translational transformation of one protein relative to the other (Lenhof 1996; Norel et al. 1999). These approaches can identify very rapidly (within minutes) putative protein–protein docking complexes (Norel et al. 1999). However, in case of unbound protein conformations, the resulting complexes that come closest to the experiment may not be ranked at the top of the possible solutions and may deviate from the experimental complex geometry (Norel et al. 1999). By using a set of only 1000 possible geometries as start structures for docking, minimization with several side-chain copies could considerably enhance the speed of the current approach. It is also possible to combine the present side-chain copy selection scheme with Brownian dynamics simulations to study protein association (Gabdoulline and Wade 2001; Fernandez-Recio et al. 2002). In such simulations, the protein is usually treated as one rigid hydrodynamic particle undergoing rotational and translational motion during time steps in the ns to μs regime. It is reasonable to assume that during such time intervals, a side-chain conformational adjustment toward an optimal fitting side-chain rotamer can occur, which would justify to separate rotamer selection from the overall translational and rotational motion of the protein. Another possible extension of the present docking approach is the use of several conformational copies for loop regions of proteins and selection of the best fitting copy during docking. This possibility might be important in docking applications that involve protein models that have, for example, been generated based on sequence similarity to a known protein struc-

ture. Such protein models can have errors, especially in loop regions, or the loop might be presented by several alternative loop conformations.

Conclusions

A protein–protein docking approach based on energy minimization with a reduced protein model and accounting approximately for side-chain flexibility with a side-chain copy approach has been presented. The method results in docking minima that stay close to the experimental start structure when using protein structures in their bound conformation. In case of docking with unbound protein structures, ranking of complexes close to the experimental geometry among the top 40 highest-ranking complexes was obtained for several test cases. A side-chain copy approach was effective for complexes in which side-chain conformational changes occur during complex formation. The docking approach could be an efficient alternative to existing docking methods and can be easily extended to treat the flexibility of protein loop regions in a similar way as side-chain mobility.

Materials and methods

Coordinates for protein complex structures and for isolated protein partners were obtained from the PDB (Bernstein et al. 1977). For the docking simulations, the protein partners were first translated into a reduced pseudo atom representation. Each amino acid was represented by one pseudo atom located at the position of the amino acid $C\alpha$ atom and up to two pseudo atoms for each side-chain (except for Gly). The side-chains of Ala, Ser, Thr, Val, Leu, Ile, Asn, Asp, and Pro were represented by one pseudo atom located at the center of geometry of all side-chain heavy atoms. In case of larger amino acids (Arg, Lys, Glu, Gln, His, Met, Phe, Tyr, Trp), a representation by two pseudo atoms appeared to be necessary in order to get good agreement between energy-minimized complexes and corresponding experimental complex geometries. This description is similar to reduced models used in low-resolution protein folding studies (Wallqvist and Ullner 1994; Gibbs et al. 2001).

For the two pseudo atom representation, the first pseudo atom was located exactly in the middle between C_β and C_γ of each side-chain, and the second pseudo atom was calculated as the center of geometry of all other side-chain atoms. A soft LJ-type potential between pseudo atom pair i,j at distance rij was used to describe the effective interaction $[(E(rij) = Bij/(rij)8 - Cij/(rij)6]$, with Bij and Cij indicating repulsive and attractive LJ parameters, respectively (see Table 1). To keep the model as simple as possible, only a few types of repulsive and attractive LJ-parameters were used that represent approximately the size and the physicochemical character of the amino acid side-chains (Table 1). With these parameters, the effective amino acid interaction energies at the LJ minima for hydrophobic side-chains (Phe, Met, Ile, Leu, Val, Cys, Trp, Tyr, Ala) give energies between −1.0 and −0.3 RT units (R indicates gas constant; T, room temperature), similar to effective pair-wise interactions derived from an analysis of folded proteins by Miyazawa and Jernigan (1999). The interaction between polar side-chains and polar with nonpolar side-chains is close to zero at contact distance. Between charged side-chains

(charge located on the second side-chain atom), electrostatic interactions were calculated by using a distance-dependent dielectric constant $\varepsilon = 15r$, which penalizes contacts between side-chains of the same charge and weakly attracts side-chains of opposite charge.

For systematic docking studies, one of the proteins (the smaller protein, called the ligand protein) was used as probe and placed at various positions on the surface of the second fixed (receptor) protein. A probe radius was chosen that was slightly larger than the maximum distance of any atom from the ligand center. At each starting position on the receptor protein, various initial ligand protein orientations were generated. For a systematic docking run, typically ~260 ligand protein orientations for each starting position were used (corresponding to a ~30- to 60-degree spacing for each Euler rotation angle, a finer sampling for the Euler angle φ [30-degree steps] was used in case of Euler angle θ ~90 degrees compared with θ ~0 or 180 degrees). For each start configuration, several energy minimizations with respect to translational and rotational degrees of freedom of the ligand protein were performed. During the first minimization, a harmonic restrain between the center of the fixed protein and the closest $C\alpha$ pseudo atom of the ligand protein was applied. The force constant for the harmonic restrain ranged from k = 0.005 to k = 0.01 kcal/mole/$Å^2$. This first minimization served to generate a close contact between the two proteins. For the subsequent energy minimizations, the ligand protein was free to move to the closest energy minimum. To achieve high computational efficiency, a short cut-off radius of 7.0 Å was used for the final energy minimizations. Because a non-bonded pair list was generated only at the beginning of each minimization, several consecutive minimizations were needed to achieve complete convergence. Typically, six consecutive energy minimizations (three with the above harmonic distance restrain, and three without) were performed. A Quasi-Newton minimizer that gradually builds up an approximation to the second derivative of the energy function with respect to translational and rotational variables was used for all energy minimizations, which allowed convergence down very small residual energy changes per step ($<10^{-5}$ energy units per step) in <40 energy minimization steps for most start configurations. The approach has been implemented in a docking program called ATTRACT.

The docking approach allows the inclusion of several side-chain copies during energy minimization for those residues that are represented by three pseudo atoms. After a preset number of minimization steps, the interaction energy of each copy of a given residue with the other protein is calculated, followed by selection of the copy with the lowest interaction energy and restart of the energy minimization. This procedure allows a switch between side-chain copies toward the best-fitting copy during several consecutive docking minimizations. To generate side-chain copies for the reduced model, several favorable rotamers of the selected residue were first generated at atomic resolution. Variation of rotamers was restricted to the first dihedral angle (χ_1) in case of Phe, Tyr, and His because variation of the second side-chain dihedral angle (χ_2) results in the same (or very similar placement in case of His) of the second side-chain pseudo atom. For the extended side-chains of Arg, Gln, Glu, Lys, Met, and Trp, variation of rotamers was restricted to the first two side-chain dihedral angles (χ_1 and χ_2). To get an impression of the conformational difference between rotamers in the reduced model description, it is useful to look at the distance of the second pseudo atom for two different rotamers. Taking Lys as an example rotation of the χ_1 side-chain torsion angle by 120 degrees (e.g., a g+ to-t or g+ to-g transition of the torsion angle) results in a second pseudo atom position that has a distance of ~5 Å from the corresponding position in the initial conformation. In case of varying χ_2 by 120 degrees, this

leads to a second pseudo atom distance of ~2.5 Å. Variation of other side-chain dihedrals (beyond χ_2) leads to much smaller changes in the reduced representation of the side-chains (distance between second pseudo atoms for different rotamers was <1.5 Å), which affected docking results much less than changes in χ_1 and χ_2 (see Results). Rotamer generation was further restricted to favorable states described by Dunbrack and Karplus (1993) and to those states that do not generate steric conflicts with other parts of the protein. For simplicity, all generated side-chain copies were assumed to be isoenergetic. So the influence of the intrinsic energy of each side-chain rotamer has been assumed to be small and has been neglected. Because the number of additional interactions that need to be calculated due to the side-chain copies is small compared with the total number of interacting pairs, inclusion of a few additional copies did not markedly increase the docking computer time.

Acknowledgments

References

Ausiello, G., Cesareni, G., and Helmer-Citterich, M. 1997. A new docking procedure applied to reconstruction of protein tertiary structure. *Proteins* **28:** 556–567.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E., Brice, M.D., Rodgers, J.R., Kennard, O., Shianouchi, T., and Tasumi, M. 1977. The PDB: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112:** 535–542.

Betts, M.J. and Sternberg, M.J.E. 1999. An analysis of conformational changes on protein–protein association: Implications for predictive docking. *Protein Eng.* **12:** 271–283.

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

Camacho, J.C., Gatchell, D.W., Kimura, S.R., and Vajda, S. 2000. Scoring docked conformations generated by rigid-body protein–protein docking. *Proteins* **40:** 525–537.

Camacho, J.C. and Vajda, S. 2001. Protein docking along smooth association pathways. *Proc. Natl. Acad. Sci.* **98:** 10636–10641.

Chakrabarti, P. and Janin, J. 2002. Dissecting protein–protein recognition sites. *Proteins* **47:** 334–343.

Chen, R. and Weng, Z. 2002. Docking unbound proteins using shape complementarity, desolvation and electrostatics. *Proteins* **47:** 281–294.

Cherfils, J. and Janin, J. 1993. Protein docking algorithms: Simulating molecular recognition. *Curr. Opin. Struct. Biol.* **3:** 265–269.

Cherfils, J., Duquerroy, S., and Janin, J. 1991. Protein-protein recognition analysed by docking simulation. *Proteins* **11:** 271–280.

Cherfils, J., Bizebard, T., Knossow, M., and Janin, J. 1994. Rigid-body docking with mutant constraints of influenza hemagglutinin with antibody HC19. *Proteins* **18:** 8–18.

Dunbrack, R.L. 2002. Rotamer libraries in 21st century. *Curr. Opin. Struct. Biol.* **12:** 431–440.

Dunbrack, R.L. and Karplus, M. 1993. Backbone-dependent rotamer library for proteins: Application to side chain prediction. *J. Mol. Biol.* **230:** 543–574.

Fernandez-Recio, J., Totrov, M., and Abagyan, R. 2002. Soft protein–protein docking in internal coordinates. *Protein Sci.* **11:** 280–291.

Fischer, D., Lin, S.L., Wolfson, H., and Nussinov, R. 1995. A geometry based suite of molecular docking processes. *J. Mol. Biol.* **248:** 459–477.

Gabb, H.A., Jackson, R.M., and Sternberg, M.J.E. 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272:** 106–120.

Gabdoulline, R.R. and Wade, R.C. 2001. Protein-protein association: Investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol.* **306:** 1139–1155.

Gardiner, E.J., Willet, P., and Artymiuk, P.J. 2001. Protein docking using a genetic algorithm. *Proteins* **44:** 44–56.

Gibbs, N., Clarke, A.R., and Sessions, R.B. 2001. Ab initio protein structure prediction using physicochemical potentials and a simplified off-lattice model. *Proteins* **43:** 186–202.

Glaser, F., Steinberg, D.M., Vakser, I.A., and Ben-Tal, N. 2001. Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins* **43:** 89–102.

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

Helmer-Citterich, M. and Tramontano, A. 1994. Puzzle: A new method for automated protein docking based on surface shape complementarity. *J. Mol. Biol.* **235:** 1021–1031.

Jackson, R. and Sternberg, M.J.E. 1995. A continuum model for protein/protein interactions: Application to the docking problem. ITLJ. Mol. Biol. **250:** 258–275.

Jackson, R., Gabb, H.A., and Sternberg, M.J.E. 1998. Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *J. Mol. Biol.* **276:** 265–285.

Jackson, R.M. 1999. Comparison of protein–protein interactions in serine protease inhibitors and antibody-antigen complexes: Implications for the protein docking problem. *Protein Sci.* **8:** 603–613.

Janin, J. 1995. Protein–protein recognition. *Prog. Biophys. Mol. Biol.* **64:** 145–166.

Jiang, F. and Kim, S. 1991. Soft docking: Matching of molecular surface cubes. *J. Mol. Biol.* **219:** 79–102.

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesen, A.A., Aflalo, C., and Wodak, S.J. 1992. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci.* **89:** 2195–2199.

Lenhof, H.-P. 1996. An algorithm for the protein docking problem. In *Bioinformatics: From nucleic acids and proteins to cell metabolism* (eds. D. Schomburg and U. Lessel), pp. 125–139. GBF monographs 18th edition, Braunschweig, Germany.

Lo Conte, L., Chotia, C., and Janin, J. 1999. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285:** 2177–2198.

Lorber, D.M. and Shoichet, B.K. 1998. Flexible ligand docking using conformational ensembles. *Protein Sci.* **7:** 938–950.

Lorber, D.M., Udo, M.K., and Shoichet, B.K. 2002. Protein–protein docking with multiple residue conformations and residue substitutions. *Protein Sci.* **11:** 1393–1408.

Meyer, M., Wilson, P., and Schomburg, D. 1996. Hydrogen bonding and molecular surface complementarity as a basis for protein docking. *J. Mol. Biol.* **264:** 199–210.

Miyazawa, S. and Jernigan, R.L. 1999. Self-consistent estimation of inter-residue protein contact energies based on an equilibrium mixture approximation of residues. *Proteins* **34:** 49–68.

Moont, G., Gabb, H.A., and Sternberg, M.J.E. 1999. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* **35:** 364–373.

Najmanovich, R., Kuttner, J., Sobolev, V., and Edelman, M. 2000. Side-chain flexibility in proteins upon ligand binding. *Proteins* **39:** 261–268.

Norel, R., Lin, S.L., Wolfson, H., and Nussinov, R. 1994. Shape complementarity at protein–protein interfaces. *Biopolymers* **34:** 933–940.

Norel, R., Petry, D., Lin, S.L., Wolfson, H., and Nussinov, R. 1999. Examination of shape complementarity in docking of unbound proteins. *Proteins* **35:** 403–419.

Palma, P.N., Krippahl, L., Wampler, J.E., and Moura, J.J. G. 2000. Bigger: A new (soft) docking algorithm for predicting protein interactions. *Proteins* **39:** 372–384.

Ritchie, D.W. and Kemp, J.L. 2000. Protein docking using spherical polar Fourier correlations. *Proteins* **47:** 334–343.

Shoichet, B.K. and Kuntz, I.D. 1991. Protein docking and complementarity. *J. Mol. Biol.* **221:** 327–346.

Smith, G.R. and Sternberg, M.J.E. 2002. Prediction of protein–protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12:** 28–35.

Tovichigrechko, A., Wells, C.A., and Vakser, I.A. 2002. Docking protein models. *Protein Sci.* **11:** 1888–1896.

Tuffery, P., Etchebest, C., Hazout, S., and Lavery, R. 1991. A new approach to the rapid determination of protein side chain conformations. *J. Biomol. Struct. Dyn.* **8:** 1267–1289.

Vakser, I.A. 1997. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins Suppl.* **1:** 226–230

Wallqvist, A. and Ullner, M. 1994. A simplified amino acid potential for use in structure predictions of proteins. *Proteins* **18:** 267–280.

Zacharias, M., Luty, B.A., Davis, M.E., and McCammon, J.A. 1994. Combined conformational search and finite-difference Poisson-Boltzmann approach for flexible docking: Application to an operator mutation in the λ repressor-operator complex. *J. Mol. Biol.* **238:** 455–465.