WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking

Andreas May and Martin Zacharias*

School of Engineering and Science, Jacobs University Bremen, D-28759 Bremen, Germany

ABSTRACT

Protein–protein association can frequently involve significant backbone conformational changes of the protein partners. A computationally rapid method has been developed that allows to approximately account for global conformational changes during systematic protein–protein docking starting from many thousands of start configurations. The approach employs precalculated collective degrees of freedom as additional variables during protein–protein docking minimization. The global collective degrees of freedom are obtained from normal mode analysis using a Gaussian network model for the protein. Systematic docking searches were performed on 10 test systems that differed in the degree of conformational change associated with complex formation and in the degree of overlap between observed conformational changes and precalculated flexible degrees of freedom. The results indicate that in case of docking searches that minimize the influence of local side chain conformational changes inclusion of global flexibility can significantly improve the agreement of the near-native docking solutions with the corresponding experimental structures. For docking of unbound protein partners in several cases an improved ranking of near native docking solutions was observed. This was achieved at a very modest (∼2-fold) increase of computational demands compared to rigid docking. For several test cases the number of docking solutions close to experiment was also significantly enhanced upon inclusion of soft collective degrees of freedom. This result indicates that inclusion of global flexibility can facilitate in silico protein–protein association such that a greater number of different start configurations results in favorable complex formation.

Proteins 2008; 70:794–809.

Key words: *conformational change; collective degrees of freedom; Gaussian network modes; molecular docking; protein–protein interaction.*

INTRODUCTION

Protein–protein interactions play a major role in many biological processes. The experimental structure determination of protein–protein complexes has so far only been achieved for a small fraction of complexes important for biological function. Besides of systematic structure determination efforts the prediction of protein–protein complex geometries is important to obtain at least structural models of protein–protein interactions. To achieve a realistic prediction, protein–protein docking approaches require both, a realistic description of protein–protein interactions during docking simulations and the accurate inclusion of possible conformational changes of the binding partners during the binding process.[1–5] In addition, the observed conformational changes during complex formation can also serve important biological functions. These changes can involve local rearrangements of amino acid side chains or loops at the interface of protein–protein complexes but also global motions that involve collective changes of the protein backbone stucture.[3–6] In principle, molecular dynamics (MD) simulation techniques or multi-start energy minimizations (EMs) are capable of sampling the protein conformational space treating each degree of freedom at atomic resolution explicitly.[1,7,8] However, prediction of a realistic protein–protein docking geometry including full flexibility of the binding partners is a global optimization problem that involves an enormous number of degrees of freedom (∼3n in Cartesian variables if n is the number of atoms).

MD simulations including full protein flexibility are only applicable in cases where the binding placement is approximately known and if experimental data can be included to restrict the search for putative protein–protein docking geometries.[4,9–11] As a consequence, most of the time-efficient systematic protein–protein docking approaches make use of the rigid body approximation, leading to a tremendous dimensional reduction (from ∼3n to 6).[12–22] These

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat.

Grant sponsor: DFG (Deutsche Forschungsgemeinschaft); Grant number: Za153-5.

*Correspondence to: Martin Zacharias, School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany.
E-mail: m.zacharias@jacobs-university.de

Received 14 November 2006; Revised 30 March 2007; Accepted 10 April 2007
Published online 29 August 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21579

approaches perform quite well in case of docking "bound" structures, meaning protein structures that have been extracted from experimentally determined protein–protein complex structures.[2] Because of conformational changes upon complex formation the performance on "unbound" structures (determined in the absence of the partner) is often much poorer.[1–3] Limited flexibility can be introduced implicitly by using a soft interaction potential to allow for some steric mismatch at the interface.[18,21] Explicit flexibility can be introduced on the level of side chains using a discrete set of side chain rotamers for a selected number of residues.[8,9,23] Typically, this is done in a postsampling step,[9,24–26] on preselected docked complexes or in more demanding approaches also during the docking search itself.[21,27,28] Inclusion of side chain flexibility can significantly enhance the prediction performance.[27,28] However, in many cases side chain flexibility alone is still not sufficient to model the conformational changes that occur upon protein–protein complex formation.[1,4,5] Even comparably small-scale backbone rearrangements cannot be neglected in some cases, and should preferably be modeled explicitly. For example, receptor backbone conformational changes of less than 1 Å can significantly affect the scoring of a docked complex.[2] It is also possible to perform docking searches using conformational ensembles of (rigid) protein structures instead of a single structure (extension of the rigid docking approach).[29–34] This is in accordance with the idea of pre-existing ensembles.[30] However, it can dramatically increase the computational demand and also creates many more possible solutions that must be appropriately scored. Apart from that, global backbone flexibility has so far been addressed via placement of hinges[35,36] that requires knowledge or prediction of possible hinge-regions.

In case of localized changes of the backbone structure, for example protein loop segments, it is possible to include approximate loop flexibility efficiently during the docking search employing a meanfield based on sterically allowed loop copy conformations.[37,38] However, the efficiency of this method depends significantly on the size of the loop region and number of necessary loop copies relative to the whole protein structure. Efficient representation of global backbone flexibility during docking that involves collective motions of large protein segments still remains a challenge. Zacharias and Sklenar[39] suggested the inclusion of soft harmonic modes as additional variables during docking of biomolecule-ligand complexes. The soft collective degrees of freedom were obtained by normal-mode analysis with respect to the atomic coordinates at an energy minimum of the receptor structure. The method was applied to DNA in complex with a minor groove binding ligand.[39] Use of such soft modes as additional variables allows to rapidly relax the receptor structure during docking and to estimate the receptor deformation energy avoiding costly calculation of the internal receptor energy at every docking minimization step. The deformation of the receptor in the soft degrees of freedom can be treated implicitly using a penalty potential to limit deformations. An additional advantage compared to docking methods that employ ensembles of rigid receptor structures is that the receptor conformation can change continuously during docking in the precalculated soft degrees of freedom and has therefore a greater capacity for induced fit adaptation. A drawback of the method is the computationally expensive precalculation of harmonic modes of large receptor molecules which requires very extensive EM. The EM is often performed in the absence of solvent and can lead to large deviations from a realistic receptor geometry and the calculated soft harmonic modes may in turn not correspond to realistic soft degrees of freedom. It is also possible to extract soft principal components of motion (also termed essential dynamics) obtained under more realistic conditions from a MD simulation including surrounding waters and ions.[40] A subset of such soft modes has also been used as additional docking variables (termed principal component relaxation).[41] Application to a receptor ligand system resulted in significantly improved docking performance compared to docking to a rigid 'unbound' receptor structure at a modest increase in computational demand.[41] However, the accuracy of the resulting docked complexes depends on how well the soft precalculated collective modes overlap with experimentally observed conformational changes upon complex formation. Principal components of motion extracted from MD simulations can significantly depend on the simulation length and the calculation is also computational demanding and grows rapidly with the size of the protein. A recent MD analysis of many proteins involved in protein–protein interactions indicated little overlap of extracted soft degrees of freedom for the isolated proteins with conformational changes in the protein partners that have been observed experimentally upon complex formation.[42]

It has been shown that soft normal modes obtained from Gaussian network models (GNMs) of proteins frequently overlap with experimentally observed global conformational changes.[43–47] In a Gaussian network model the experimental structure serves as a reference (energy minimum) structure and the mobility of a residue or protein segment depends on the local density and number of short range contacts.[43,44] Compared to normal mode analysis using a molecular mechanics force field or essential dynamics (MD) analysis the calculation of modes based on GNMs (GNM) does not require a costly EM or MD simulation and the reference structure is identical to the experimental protein structure. Normal modes extracted from GNMs have recently been used to approximately account for global flexibility during docking simulations[5,48,49] either for predeformation[49] or as additional flexible variables.[5,48]

Applications to protein–protein and a protein–ligand complexes as well as a protein-DNA complexes showed promising results.[5,48,49] However, these docking simulations have been performed on selected examples only starting exactly at or from positions close to the experimentally known ligand binding geometries. It is therefore not clear if inclusion of soft collective modes can also improve the performance of systematic docking simulations starting from many possible positions and orientations of the protein partners without prior knowledge of the binding geometry. In the present study systematic docking simulations have been performed on several docking test cases that include significant global conformational changes upon complex formation. Docking minimization in translational and rotational degrees of freedom and up to five precalculated soft collective degrees of freedom of one protein partner was used starting from tens of thousands of start geometries. For comparison protein docking cases that involve no significant changes during complex formation or for which the conformational change showed little or no overlap with soft GNM modes were also included. The results indicate that approximate inclusion of global flexibility using precalculated soft modes can improve the docking performance in systematic searches at a modest increase of computational costs. In docking experiments with one of the partners (ligand) in the bound conformation the deviation of the complex closest to experiment was for many cases significantly smaller upon inclusion of soft modes during docking compared to rigid docking. For docking simulations employing unbound protein conformations an improved ranking of near-native solutions was observed in particular for those examples with a significant overlap between soft-modes and experimentally observed conformational changes. In addition, we found in several cases an increase of the number of minimized docked complexes within a preset deviation range from experiment in case of flexible versus rigid docking. This indicates that inclusion of global soft mode flexibility can facilitate the accessibility of a "binding funnel" for the associating proteins.

METHODOLOGY

Force field and docking

Systematic docking searches were performed using the program attract[21] based on a reduced protein model. Each residue is represented by up to three pseudo atoms (one for the backbone and up to two for each side chain) allowing to approximately model the dual character of some amino acid side chains (e.g. hydrophobic and hydrophilic parts of a side chain). The energy function consisted of pair-wise soft Lennard–Jones type functions and an electrostatic interaction term with a distance dependent dielectric constant [$\varepsilon(r) = 15\ r$] for the interaction of charged residues.

$$V_{\text{rigid}} = \sum \left(\frac{R_{AB}}{r_{ij}}\right)^8 - \left(\frac{R_{AB}}{r_{ij}}\right)^6 + \frac{q_i q_j}{\varepsilon(\sigma_{ik})\sigma_{ik}} \qquad (1)$$

Rigid protein–protein docking was performed by EM in translational and rotational degrees of freedom of one protein (ligand) with respect to the second protein (receptor). The details of the model and its performance on bound and unbound docking cases have been published.[21] The original parameters for the reduced model resulted in energy minimized docking complexes when starting from the experimental complexes that deviated on average by ~1.5–2 Å from experiment (Rmsd of the mobile ligand after superposition of the receptor structures: Rmsd_{lig}). In the present study we are interested to test if inclusion of global flexibility can improve the scoring and predicted geometry of docked complexes in systematic docking simulations. To minimize the influence of the force field, the pseudo atom radii in the energy function were slightly reparameterized to optimize the ligand placement in the energy minimized test (bound) complexes (Table I) such that the Rmsd_{lig} for the complexes closest to experiment (in case of bound structures) was on average ~0.5 Å (Table Is and IIs in supplementary material).

Inclusion of global flexibility

In case of docking including global flexibility either only the receptor structure or both receptor and ligand protein structures were allowed for relaxation (deformation) along precalculated soft collective degrees of freedom.[39,41] The soft collective degrees of freedom corresponded to eigenvectors of the protein calculated using an approximate normal-mode analysis related to GNM as described by Hinsen.[45] The normal modes were calculated with respect to the protein backbone (Cα atoms) only based on a pair-wise distance dependent energy function:

$$V(R_1, \ldots, R_N) = \sum_{C\alpha-\text{pairs}} V_{ij}\ (R_i - R_j), \qquad (2)$$

with the pair-wise term:

$$V_{ij} = k(R_{ij}^{(0)})(|r| - R_{ij}^{(0)})^2, \qquad (3)$$

where $R_{ij}^{(0)}$ is the pair's equilibrium distance. The force constant k is distance-dependent:

$$k(r) = \lambda \cdot \exp\left(-\frac{|r|^2}{r_0^2}\right) \qquad (4)$$

DOI 10.1002/prot

Table I
Protein–Protein Docking Test Cases

No.	System name	Complex	Receptor	Ligand
1.	RNAse/Inhibitor (receptor)	1DFJ[50]	2BNH[51]	9RSA[52]
2.	Xylanase (receptor)/Inhibitor (CAPRI[53] Target 18)	1T6G[54]	1UKR[55]	1T6E[56]
3.	Origin recognition complex subunit 1 (receptor)/Regulatory protein SIR1 (CAPRI[53] Target 21)	1ZHI	1M4Z[57]	1Z1A[58]
4.	HIV-1 neutralizing antibody 50.1 (receptor)/v3 Loop peptide antigen	1GGI[59]	1GGC[60]	–
5.	HIV-Gag CA C-term. Dom. (receptor)/Inhibitor	2BUO[61]	1A43[62]	–
6.	Barnase (receptor)/Barstar	1AY7[63]	1RGH[64]	1A19[65]
7.	Chymotrypsin (receptor)/Eglin C	1ACB[66]	2CGA[67]	1EGL[68]
8.	Kallikrein A (receptor)/Trypsin Inhibitor	2KAI[69]	2PKA[70]	1EAW[71]
9.	Adrenoxin Reductase (receptor)/Adrenoxin	1E6E[72]	1E1N[73]	1CJE[74]
10.	Importin Beta (receptor)/Ran GTPase	1IBR[75]	1F59[76]	1QG4[77]

such that small distances are strongly restrained and long distances are weakly controlled. The constant λ controls the overall flexibility. Harmonic modes with respect to the above energy function can be obtained by diagonalization of the second derivative matrix of the energy function as described by Hinsen.[45] In test calculations it was found that setting $r_0 = 4$ Å gave the best overlap between the softest nontrivial modes and the conformational difference between apo and holo structures for all test systems. In the present study, normal modes are exclusively employed for modeling global low-frequency backbone movements. Side chains were allowed to move as rigid bodies attached to the corresponding backbone pseudo atom (means the whole side chain moves as a rigid body). This separation allows in principle an independent efficient treatment of side chain mobility (e.g. in terms of discrete rotameric states)[21] on top of the global backbone motion. During docking including global flexibility the protein structure was minimized in up to five softest modes obtained from the approximate normal mode calculations (in addition to the six translational + orientational degrees of freedom of the ligand protein). Following earlier experience the deformability of the protein in the soft modes[5,41] was limited by a fourth order function in the deformation along each mode. The total energy of the system is,

$$V_{\text{flex}} = V_{\text{rigid}} + \lambda \sum_{m=1}^{M} \kappa_m \, (R^0 - R_m)^4 \qquad (5)$$

The parameter λ is a scaling factor that was set to 6.0 RT (R, gas constant, T, temperature) units. R^0 is the coordinate set describing the experimental structure of the protein, $\|R^0 - R_m\|$ is the conformational change of the protein in mode m. The force constant κ_m is the square of the eigenvalue corresponding to mode m. M is the total number of modes employed. The form of the penalty allows significant conformational freedom of the protein close to the reference structure but strong penalty for larger scale deformations. Approximate normal modes were typically calculated for the unbound protein structures. However, to adjust the flexibility in the soft modes we also performed docking simulations on the complex structures in the bound form. The scaling factor λ which was identical for all systems was chosen such that the deviation from experiment when using bound structures was small (<1 Å) but still allowing for sufficient flexibility for receptor relaxation when starting from the apo receptor conformation (conformational backbone changes of up to ~2–3 Å, see Results and Discussion section).

Test systems

For each test system (Table I) unbound (apo) and bound (holo) structures were available at the protein data bank (PDB).[78] Except for System 6 each of the test systems involved a significant backbone conformational change during complex formation for one protein partner. In the current study this partner is defined as the receptor structure (in most but not all cases this is also the larger protein partner structure). Two systems (2 and 3) were also target structures of the CAPRI challenge (Critical Assessment of Predicted Interaction).[53] System 6 was chosen as it is usually considered as fairly rigid to assess if already slight explicit backbone changes result in an increase of prediction accuracy. Two systems not showing any overlap between mode space and conformational change were selected to investigate the inclusion of precalculated soft modes (that point in the incorrect direction) on docking performance (Systems 7 and 8). The receptor structure was represented by the unbound (apo) conformation. Since we are interested in an evaluation of the benefits of including global flexibility during docking

both bound (holo) and unbound (apo) conformations of the second partner (ligand) were used during docking. Ligand protein conformations in the bound and unbound form were employed to check the interference of the docking performance due to a nonoptimal conformation of the ligand protein.

To further separate effects because of local structural differences (e.g. side chain conformational changes) between bound and unbound receptor structures from influences due to global backbone changes, we constructed "hybrid" receptor structures with the side chain conformations nearly as observed in the bound structure on the unbound backbone structure. This was achieved by targeted minimization of the bound receptor structure using the backbone of the unbound structure as a target reference structure and employing the Amber 8 program suite.[79] It is important to note that since the side chains are free to move this results in energy minimized "bound" side chain conformations on an "unbound" protein backbone structure. The energy minimized side chain conformations can still slightly differ from the "bound" conformation by up to ~0.5 Å. Prior to docking all proteins were translated into the reduced protein representation.[21]

Docking minimization

The individual docking runs were performed by EM in a 6 + n-dimensional space (n is the number of modes + 6 rotational and translational degrees of freedom, illustrated in Fig. 1). Starting positions on the surface of the receptor protein during systematic docking runs were obtained by using the ligand protein as probe rolled over the receptor surface. The probe radius was set to the radius of the ligand protein +5 Å, so that initial ligand orientations do not overlap with the receptor protein. To ensure a dense sampling up to 1000 starting positions and 300 different orientations per start point (~300,000 starting structures) were generated. Two docking protocols (1 and 2) were used. In the first protocol the complete receptor surface was covered with docking minimization start points. Resulting rigidly docked structures (minima) were used as start structures for flexible docking including the soft precalculated modes (up to 5 modes). For the second protocol only a region in the vicinity of the known binding site (start positions 30 Å from the binding site) was considered and docking included all flexible soft degrees of freedom from the beginning of the docking simulations. To allow direct comparison between small and large proteins and between the two protocols the final evaluation of docked complexes was restricted to the same distance range from the known binding site (all docked complexes with ligand protein distances within 10 Å from the experimental placement). As an option a shake routine[80] was employed to adjust the bonded geometry of the receptor after deformation in the soft modes. In addition to flexible docking in the approximate mode space, rigid docking to predeformed structures was also applied. The predeformed unbound structures represent the best possible approximation of the corresponding bound structures based on deformation of the unbound structure in a set of precalculated soft modes (for the unbound structure).

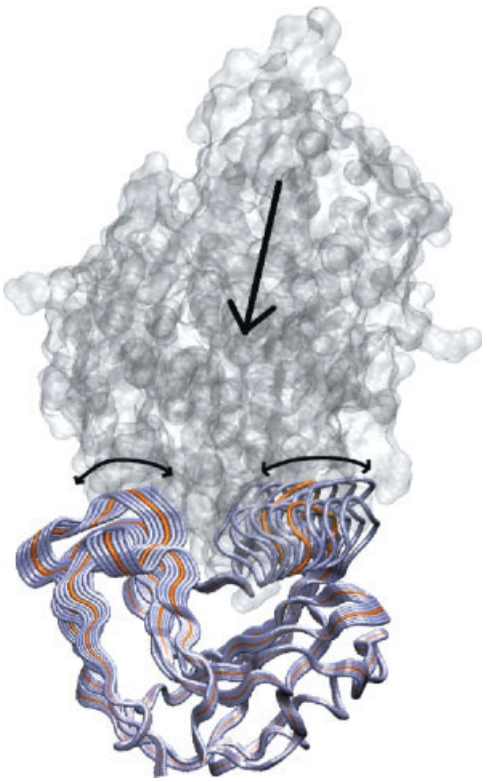


Figure 1

Illustration of protein–protein docking including global collective receptor degrees of freedom for System 2: Xylanase (tube representation) and TAXI inhibitor (surface representation). A docking minimization consists of orientational and translational optimization (solid arrow) of the ligand protein coupled to minimization of the recepetor protein in a subset of soft collective degrees of freedom obtained from normal mode analysis (superposition of deformed structures in the softest Xylanase mode, also indicated by two double arrows). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The docking solutions obtained from each systematic search were ranked according to the interaction energy score calculated at the final docking minimization (with a 7 Å cutoff). In case that several nearly identical solutions were generated only the copy with the lowest energy was kept, all the others were discarded. Nearly identical solutions are those that differ by less than 1 RT unit in interaction score, in less than 1 Å ligand-receptor center-to-center distance and in less than 1 Å of the $Rmsd_{lig}$. Images of docked complexes were generated using VMD.[81]

DOI 10.1002/prot

RESULTS AND DISCUSSION

Overlap between precalculated soft modes and conformational differences of apo and holo protein structures

The current flexible protein–protein docking approach includes soft collective degrees of freedom of protein structures as additional variables during docking minimization (illustrated in Fig. 1). It is desirable that these soft collective degrees of freedom relate to realistic global degrees of freedom of the protein structures. Ideally, observed conformational differences between unbound and bound protein structures should overlap with precalculated soft normal modes calculated for the unbound protein structure. It has already been shown in a number of studies that soft modes obtained from GNMs can show a significant overlap with experimentally observed conformational changes including differences between apo and ligand-bound protein structures.[46,47] Since the eigenvectors (modes) from the GNM form an orthogonal set each mode contributes independently and the best possible approximation can be obtained as a sum of projections of the conformational difference on each mode (scalar product between conformational difference and each mode). As a result one can calculate the best possible deformation (smallest Rmsd from bound structure) of the deformed apo form from the holo structure for a given number of included modes (Fig. 2). The conformational changes of receptor protein structures for Systems 1–5 and 10 showed the most significant overlap. In these cases it is possible with only the five softest modes to account for ~20–75% of the backbone conformational difference between unbound and bound receptor structures. Contributions beyond mode 5 are much smaller than for the first five modes (Fig. 2). During the docking simulations the number of soft modes was therefore limited to the first five softest modes. It is important to note that in several cases the overlap of precalculated soft modes with conformational changes of the region involved in protein–protein binding is even larger than the overlap with the conformational change of the complete protein. For example, in case of the receptor of System 2 (Xylanase) the softest modes overlap very well with the conformational change near the inhibitor binding site but the bound and unbound forms of the enzyme structure differ in a loop region (far away from the inhibitor binding site) that cannot be well approximated by the softest GNM modes. In case of Systems 7 and 8, apo and holo receptor structures also differ significantly in the backbone conformation. However, in these two cases the precalculated soft modes (for the apo form) show little overlap with the experimentally observed conformational difference. The experimentally observed conformational changes corresponded to large "localized" changes of loop regions that cannot be modeled accurately by a few precalculated soft modes. This agrees with studies by Petrone and Pande[82] on several cases where experimentally observed structural changes corresponded to large localized changes that could not be approximated using a small set of soft modes from a GNM. However, inclusion of these test cases is important since for a realistic protein–protein docking simulation it is neither known in advance if the proteins undergo conformational changes during association nor if a conformational change overlaps with a precalculated GNM mode. For the present ligand protein structures the backbone conformational differences between bound and unbound was in all cases smaller then 1 Å and the overlap of this difference with the 5 softest modes (for the

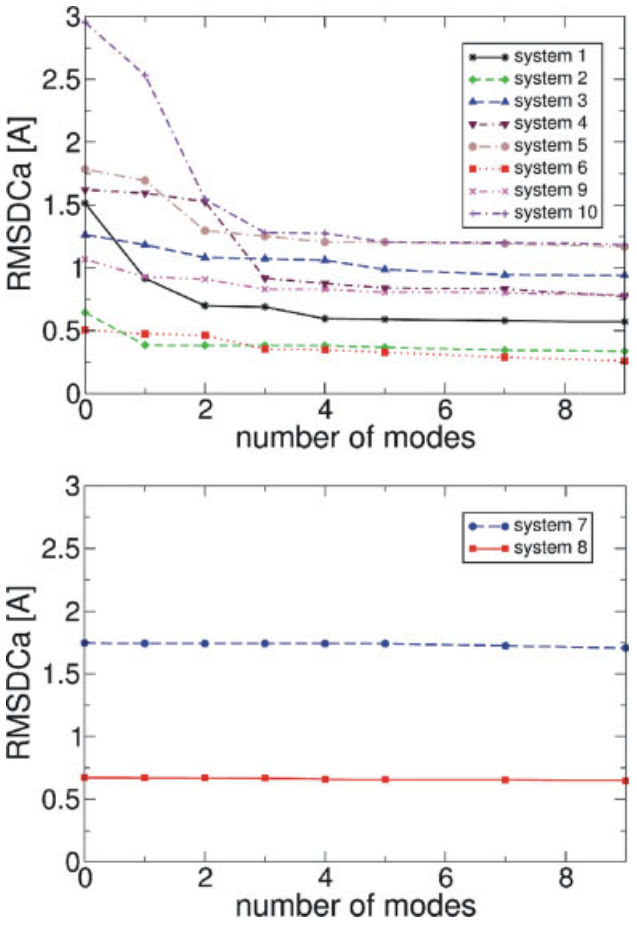


Figure 2

Projection of the conformational difference between bound and unbound receptor structures on a subset of soft normal modes obtained for the receptor in the unbound form. The normal modes were obtained from a Gaussian network model (see Methodology section). The residual conformational difference between optimally deformed structures and number of included modes has been plotted. The upper panel includes all systems for which at least 20% of the conformational difference can be approximated by a linear combination of the softest (5) normal modes. For Systems 7 and 8 (lower panel) less than 5% of the conformational difference can be approximated by the softest five normal modes. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ligand) ranged from 5 to 23% (corresponding to between 0.03 and 0.23 Å in terms of backbone Rmsd, not shown). It is interesting to note that in all present protein complexes that involved a significant conformational change of one partner protein (the receptor) the eigenvalues of the softest modes for this protein partner were smaller than for the partner that underwent little or no conformational change. Hence, the mobility in the softest modes (given by the eigenvalues) might be a useful indicator for the probability of the protein partner to undergo conformational changes upon binding.

Flexible docking of bound structures starting from the known complex structure

Since the ultimate goal of protein–protein docking simulations is to reproduce the "bound" complex structure starting from the apo-conformations of the partner structures it is important to check the performance of the docking approach on the experimentally known bound structures. In this case one starts from ideal protein conformations (no deviations from the structure in the complex). The docking calculation on bound structures sets a lower bound with respect to the best possible agreement with experiment for docking simulations using the unbound protein structures. Using the current force field version very good agreement between energy minimized complex structures and experiment were observed for all complexes with an average Rmsd of the ligand protein after superpositioning of the receptors ($Rmsd_{lig}$) of ~0.5 Å (largest $Rmsd_{lig}$ = 1.6 Å, supplementary material, Table IIs). Inclusion of an increasing number of precalculated soft modes allowed the receptor to deform in the modes to further improve the interaction with the ligand. This has the undesired side effect that even during docking of bound structures the flexible receptor structure can deform away from the bound reference conformation. Although the deformation in the receptor was on average only ~0.5 Å (upon inclusion of up to five softest modes) it in turn also affected the energy minimized placement of the ligand protein. Inclusion of up to five modes resulted in an increased deviation from the experimental placement to on average ~1.3 Å (supplementary material, Table IIs). These deviations were observed for ideal protein conformations. The result gives an estimate on a possible lower bound for the best possible results when using unbound protein structures.

Docking starting from the known binding geometry using apo receptor structures

The effect of including global flexibility during docking was first tested during single docking minimizations starting from the protein placements observed in the known complex. As receptor input conformations either the unbound structures or the hybrid structures (apo backbone and holo side chain conformations, see Methodology section) were employed. The second protein (ligand protein) was rigid and in the bound conformation. Docking to receptor hybrid structures was included because conformational changes of side chains at the protein interface can also strongly affect the ligand protein placement after docking minimization. Docking to hybrid structures can help to separate effects because of global conformational changes from those due to side chain conformational changes. However, one should keep in mind that the side chain conformations attached to the apo backbone structures may not be in perfect agreement with the side chain conformations in the bound form since the structure generation involved an EM step (see Methods). For all systems that showed a significant overlap between conformational changes upon complex formation (1–5) also a significant improvement of the Ligand Rmsd ($Rmsd_{Lig}$) after flexible docking compared to rigid docking was observed (Table II). Surprisingly, in some cases rigid docking to the hybrid structures resulted in energy minima farer away from experiment than using the unbound receptor structures. This shows that the effect of incorrect backbone structures and of incorrect interface side chains is not necessarily additive. It is also important to note that depending on the degree of initial steric overlap, EM starting from the experimental ligand protein placement does not necessarily result in the docking minimum closest to the experimental structure. This may also explain why in some cases the change in ligand placement was nonmonotonic with respect to the number of flexible modes included during docking. In these cases the sterical overlap at the beginning of the minimization that creates large initial forces may have caused the ligand to move towards a docking minimum that is not the closest to the experimental placement. Overall the improvement was in most cases slightly more dramatic for docking to the hybrid receptor structures compared to the "full" apo-receptor conformations (Table II). In several cases especially those with a significant overlap between precalculated soft modes and difference between bound and unbound receptor conformation not only an improvement of the ligand placement but also an improvement in receptor structures towards a smaller deviation from the bound form was observed. As expected, for the Barnase/Barstar system (System 6), only a slight improvement was observed when side-chains were in the holo conformation. In case of side chains in the apo conformation the receptor structure slightly deformed in the incorrect direction during docking including receptor flexibility (slight increase in receptor Rmsd from experiment). As a side effect, this led to an improvement of the ligand placement from 3.1 Å (rigid docking) down to 2.3 Å (inclusion of five receptor modes). Generally, the observed improvement is not directly proportional to the degree of receptor deformability, that is number of modes employed. For both the antibody-peptide system (4) and the GAG-CCA-inhibitor

DOI 10.1002/prot

Table II
Influence of Soft Mode Flexibility on Docking Using Unbound Receptor Structures

Syst.	Side chain conformation	Number of included soft modes during docking minimization					
		0	1	2	3	4	5
1	apo	5.5 (1.5)	3.3 (0.9)	5.9 (0.9)	3.4 (1.3)	3.6 (1.1)	3.5 (1.1)
	holo	3.9 (1.5)	3.1 (1.0)	3.7 (1.0)	1.3 (1.2)	1.6 (1.4)	2.5 (1.1)
2	apo	10.6 (0.6)	2.6 (0.4)	2.4 (0.5)	3.0 (0.5)	2.8 (0.6)	2.9 (0.6)
	holo	15.1 (0.6)	2.3 (0.4)	1.5 (0.5)	1.7 (0.6)	1.2 (0.6)	1.2 (0.6)
3	apo	6.0 (1.3)	4.6 (1.2)	2.8 (1.1)	3.5 (1.1)	3.5 (1.1)	3.6 (1.2)
	holo	8.8 (1.3)	1.9 (1.2)	3.2 (1.1)	3.6 (1.1)	3.7 (1.1)	6.7 (1.2)
4	apo	6.5 (1.6)	6.9 (1.6)	6.2 (1.5)	6.0 (2.0)	4.7 (1.3)	4.7 (1.3)
	holo	4.8 (1.6)	4.8 (1.6)	4.6 (1.7)	4.6 (2.0)	4.5 (2.0)	4.5 (2.0)
5	apo	11.6 (1.8)	12.6 (1.9)	4.6 (1.7)	5.3 (3.1)	5.1 (3.3)	3.4 (2.9)
	holo	4.7 (1.8)	6.0 (1.7)	7.5 (2.1)	10.3 (1.7)	17.1 (1.7)	4.0 (1.7)
6	apo	3.1 (0.5)	2.5 (0.7)	2.4 (0.7)	2.6 (0.8)	2.4 (0.8)	2.3 (0.7)
	holo	0.6 (0.5)	0.8 (0.5)	0.6 (0.5)	0.5 (0.4)	0.5 (0.4)	0.6 (0.5)
7	apo	4.0 (1.7)	2.0 (1.8)	2.0 (1.8)	2.0 (1.8)	2.3 (1.8)	2.2 (1.8)
	holo	8.4 (1.7)	8.7 (1.7)	8.7 (1.7)	8.9 (1.7)	9.5 (1.7)	9.3 (1.7)
8	apo	12.5 (0.7)	15.8 (0.7)	10.6 (0.7)	18.4 (0.7)	7.1 (0.7)	7.1 (0.7)
	holo	3.5 (0.7)	3.6 (0.7)	3.3 (0.7)	3.3 (0.7)	3.6 (0.8)	3.5 (0.7)

Single docking minimization was started from the ligand protein position as found in the experimental complex structure but replacing the bound receptor structures by the corresponding unbound (apo) structures or "hybrid" receptor structures. The hybrid structures contain side chain coordinates close to the conformations in the bound receptor structure (holo, see Methods).The protein backbone Rmsd (in Å) is given for the docked ligand protein and receptor protein (in brackets) after docking minimization including between 0 and 5 soft receptor modes.

system (5) only a modest improvement in ligand placement and receptor deformation was obtained upon inclusion of receptor flexibility during docking. For System 5 using holo side chains the ligand placements showed significant variation. It is likely that the initial ligand-receptor overlap created large initial forces that, depending on the number of variables, resulted in docking minima not necessarily corresponding to the minima closest to experiment. Interestingly, for some systems with little or no overlap of observed conformational changes and precalculated modes an improvement of the ligand protein placement was observed when using the apo receptor structure (not in case of using the hybrid structures, Table II). A possible explanation could be a small compensatory motion of the backbone to outbalance the effect of incorrect side chains in case of docking using the apo-receptor structures.

Systematic multi start docking searches using bound ligand protein structures

The docking simulations starting from the known protein binding sites demonstrated that inclusion of precalculated soft modes has the potential to significantly improve the docking results with only a modest increase of computer time of factor ~2 compared to rigid docking. Note, that a single docking minimization (of a medium sized protein complex) to a converged docking geometry takes typically ~0.1 s. The modest increase of the computational demand for approximate inclusion of global flexibility offers the possibility to efficiently account for global flexibility also during systematic docking runs that require the efficient minimization of tens or even hundreds of thousands of starting geometries. It was tested using several different protocols as described in the Methods section and the legends of Tables III and IV. Briefly, systematic docking runs were performed starting either from positions that covered the complete receptor surface (Protocol 1) or from start positions not too far from the known binding site (Protocol 2). Note, that the second protocol does not correspond to an unrealistic docking example. For cases of two interacting proteins where structures have been determined experimentally for each partner there is often also biochemical/mutagenesis data available that allows one to restrict the binding site to a region on the protein surface of at least one partner. Flexibility was either included from the beginning of the systematic search (Protocol 2) or only after refining rigid docking results (Protocol 1). To keep the results comparable we restricted in all cases the solutions that were further analyzed to those within 10 Å of the known binding site (see Methodology section for details). Tables III and IV represent the results for all tested docking protocols applied to System 1 whereas Figures 3 and 4 show the results for a subset of protocols that had the largest impact on the docking results for all other systems.

In case of using apo backbone and apo side chains for the System 1 receptor the $Rmsd_{Lig}$ of solutions closest to experiment improved only in case of using Protocol 2 (first part of Table III). The best possible solutions with an Rmsd of ~3.3 Å in case of flexible docking agree with the results on single docking simulations on this system (Table II). Even in case of docking to a best possible

Table III
Systematic Docking: Rmsd of Docking Solutions from Experiment for System 1

Rec.					Number of included soft modes during docking minimization					
bb	*sc*	Pr.	Shake	Restr.	0	1	2	3	4	5
A	a	1			4.0 (1.5)	4.4 (1.2)	4.6 (1.1)	5.0 (1.2)	4.9 (1.3)	5.0 (1.4)
A	a	1	x		4.0 (1.5)	4.4 (1.2)	4.6 (1.1)	4.9 (1.2)	4.9 (1.3)	5.0 (1.4)
A	a	1		x	4.0 (1.5)	4.4 (1.2)	4.5 (1.2)	4.9 (1.2)	4.9 (1.3)	5.0 (1.4)
A	a	1	x	x	4.0 (1.5)	4.4 (1.2)	4.5 (1.2)	5.0 (1.2)	4.9 (1.3)	5.0 (1.4)
A	a	2	x	x	4.0 (1.5)	3.3 (0.9)	3.5 (0.8)	3.4 (1.3)	3.5 (1.0)	3.5 (1.0)
P	a	2		x	4.0 (1.5)	3.4 (0.9)	3.3 (0.7)	3.3 (0.7)	3.6 (0.6)	3.7 (0.6)
A	a	2		x	4.0 (1.5)	4.5 (1.2)	3.7 (1.4)	3.7 (1.1)	3.0 (1.1)	3.2 (1.1)
A	h	1			3.3 (1.5)	3.1 (1.1)	3.4 (1.1)	3.1 (1.1)	3.2 (1.2)	3.3 (1.2)
A	h	1	x		3.4 (1.5)	3.1 (1.1)	3.4 (1.1)	3.1 (1.1)	3.2 (1.2)	3.3 (1.2)
A	h	1		x	3.4 (1.5)	3.1 (1.1)	3.4 (1.1)	3.0 (1.1)	1.2 (1.4)	3.3 (1.2)
A	h	1	x	x	3.3 (1.5)	3.1 (1.1)	3.4 (1.1)	3.1 (1.1)	3.2 (1.2)	3.3 (1.2)
A	h	2	x	x	3.3 (1.5)	2.4 (0.9)	0.9 (1.2)	1.3 (1.2)	1.6 (1.4)	1.2 (1.4)
P	h	2		x	3.3 (1.5)	0.9 (0.9)	2.0 (0.7)	1.1 (0.7)	1.9 (0.6)	1.6 (0.6)
A	h	2		x	3.4 (1.5)	3.1 (1.1)	3.4 (1.1)	3.0 (1.1)	1.2 (1.4)	1.2 (1.2)

Systematic docking minimization was performed for system 1 using either the receptor with the backbone (bb) in apo (A) or best possible predeformed (P) structure (that can be achieved within the included soft modes). The receptor side chain (sc) conformations were those of the unbound form (a in column 2: apo) or those of the bound receptor structure (h: hybrid). The inclusion of a shake routine to adjust the bonded geometry after relaxation in soft modes as well as the inclusion of restraints to push the partners together at the beginning of each docking run are indicated in columns 3 and 4, respectively.

predeformed receptor towards the bound form only a modest improvement in the $Rmsd_{lig}$ was seen (best solution ~3.3 Å). However, for basically all conditions during systematic docking a significant improvement of the receptor conformation towards a closer agreement with experiment was obtained (Table V).

In case of using the "hybrid" receptor structures (apo backbone and holo side chains; second part of Table III) application of Protocol 1 gave an overall similar result as in case of using apo-receptor structures. For Protocol 2 the ligand protein placement of the solution closest to experiment improved significantly reaching 1–1.5 Å upon inclusion of 1–5 soft receptor modes. In general the inclusion of a restraint to push the protein partners together during an initial minimization followed by free docking minimization had only a modest effect. Similar, the application of a shake routine to readjust small deformations in the bonded geometry during flexible docking had a minor effect on the docking result.

In case of docking using apo receptor structures the inclusion of global flexibility had an influence on the ranking of the solutions closest to experiment for System 1

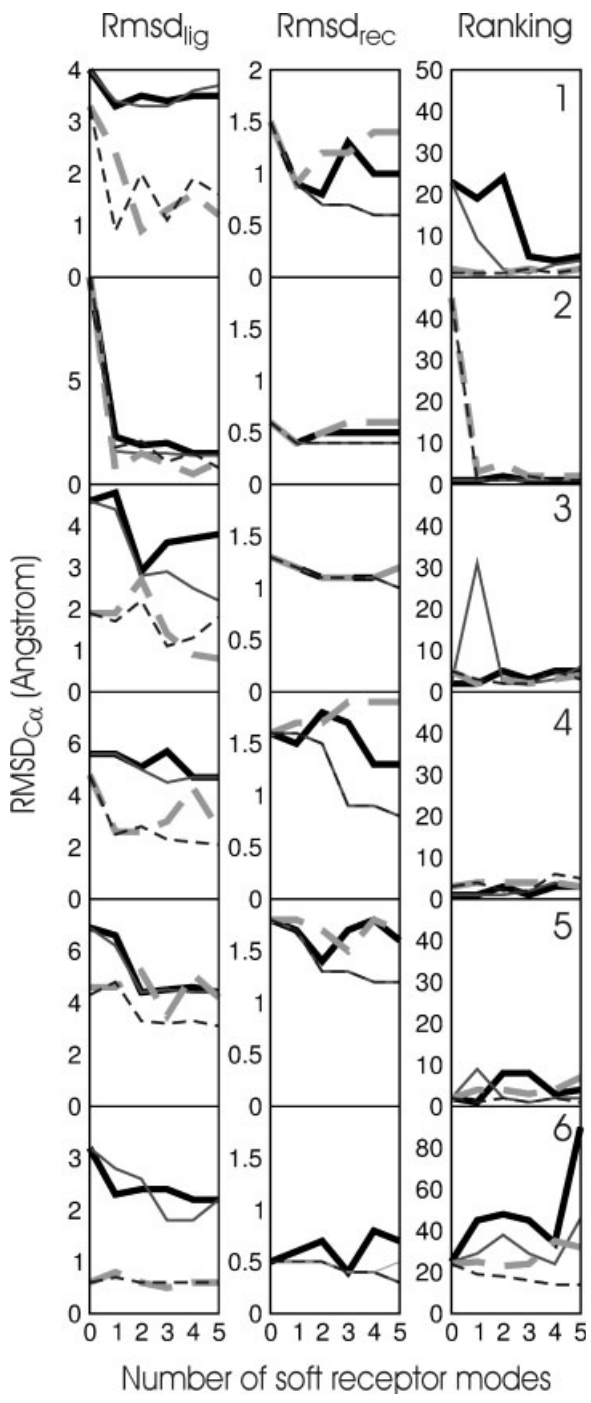
Table IV
Systematic Docking: Ranking of Docking Solutions for System 1 Closest to Experiment

Rec.					Number of included soft modes during docking minimization					
Bb	*sc*	Pr.	Shake	Restr.	0	1	2	3	4	5
A	a	1			21 (−12.0)	21 (−13.4)	14 (−14.1)	16 (−14.5)	15 (−14.9)	10 (−15.6)
A	a	1	X		21 (−12.0)	27 (−12.6)	14 (−14.1)	15 (−14.5)	13 (−14.9)	10 (−15.7)
A	a	1		x	21 (−12.0)	14 (−13.4)	12 (−14.1)	20 (−13.7)	12 (−14.9)	5 (−15.7)
A	a	1	X	x	21 (−12.0)	13 (−13.4)	11 (−14.1)	16 (−14.5)	15 (−14.9)	10 (−15.6)
A	a	2	X	x	23 (−12.0)	19 (−12.9)	24 (−13.1)	5 (−16.8)	4 (−14.9)	5 (−14.9)
P	a	2		x	23 (−12.0)	9 (−13.3)	2 (−15.2)	1 (−14.9)	3 (−14.3)	4 (−14.2)
A	a	2		x	78 (−12.0)	34 (−14.5)	13 (−14.9)	1 (−19.3)	2 (−19.2)	2 (−19.3)
A	h	1			2 (−17.0)	1 (−18.2)	3 (−18.8)	1 (−21.5)	1 (−21.5)	1 (−21.7)
A	h	1	x		2 (−17.0)	1 (−18.2)	2 (−18.8)	1 (−21.5)	1 (−21.5)	1 (−21.8)
A	h	1		x	2 (−17.0)	2 (−18.2)	2 (−18.8)	1 (−21.5)	2 (−20.2)	1 (−21.7)
A	h	1	x	x	2 (−17.0)	3 (−18.2)	2 (−18.8)	1 (−21.5)	1 (−21.5)	1 (−21.7)
A	h	2	x	x	2 (−17.3)	1 (−18.3)	1 (−19.1)	2 (−20.4)	1 (−20.9)	2 (−22.3)
P	h	2		x	1 (−17.3)	1 (−17.3)	1 (−18.8)	2 (−16.2)	1 (−18.2)	2 (−16.8)
A	h	2		x	3 (−17.1)	2 (−18.8)	1 (−19.1)	1 (−21.2)	22 (−11.7)	1 (−23.3)

Columns 1–5 are the same as in Table IV (see legend of Table IV). Columns 6–11 give the ranking of the docking result in closest agreement with experiment (numbers in brackets give the corresponding final energy score). Only docked ligand protein placements within 10 Å of the experimental position (center-to-center distance) were considered.

DOI 10.1002/prot

(Table IV). The most dramatic improvement was observed for using Protocol 2. The ranking of the solution closest to experiment improved from position 23–78 (rigid docking) to rank 2–5 in case of using five flexible receptor modes during docking. For rigid docking of System 1 with a hybrid receptor structure the ranking of the solution closest to experiment was already quite good and improved only slightly in a few cases upon inclusion of soft modes (Table IV).



The various docking protocols were also tested on the other test systems. Similar to the results obtained for System 1 small changes in the docking protocol (e.g. inclusion of Shake) had only a minor effect on the docking results and did not change the general trend. Therefore, Figures 3 and 4 show only the results for a few variations of the best docking protocol (2) including options for side chain structures and receptor structures predeformed in subsets of soft modes. A general trend that can be extracted from Figures 3 and 4 is that the deviation of the docking solution closest to experiment improved for most systems and for different docking protocols upon inclusion of soft modes during docking. The improvement correlates with the results concerning the overlap of precalculated modes and the conformational difference between apo and bound form of the receptor structures (Fig. 2). For example, for Systems 1–6 and 10 improvement in the Rmsdof the ligand protein and in several cases also of the receptor was observed upon inclusion of receptor flexibility (Figs. 3–5). The largest improvements were seen for Systems 2, 3, and 10 (also illustrated in Fig. 5). A smaller effect was seen for the Systems 1, 4, and 5. Note that cases 4 and 5 involve relatively small ligand proteins that may provide less interactions between protein partners to drive conformational changes compared to the cases with larger ligand-proteins. In several cases the improvement reached similar results as seen for docking to predeformed receptor structures indicating that the docking minimization conditions are close to optimal. It is important to stress that the deflection in each mode from the apo conformation, when docking against nonpredeformed structures, is solely determined by the minimization routine and no structural information from the bound receptor conformation has been employed. For System 8 (no overlap between observed receptor change and soft modes) no improvement of both ligand protein placement and receptor backbone conformation was obtained (Fig. 4). In contrast, for System 7 an improvement

Figure 3

Systematic docking using bound ligand and apo-receptor structures accounting for global receptor flexibility in up to 5 softest normal modes. Results of systematic docking searches for Systems 1–6 (indicated in the right panels) are shown using the receptor protein with both backbone and side chains in the apo form (bold line) or apo-backbone but side chains in a conformation close to the bound form (bold dashed line). Systematic docking involved in each case minimization of up to 60,000 start configurations. For comparison the results for rigid docking to receptor structures optimally pre-deformed in the softest modes is also shown (continuous thin line: side chains in apo form; dashed line: side chains close to conformation in the bound form). The first column indicates the ligand protein backbone Rmsd (Rmsd$_{lig}$) with respect to the experimental structure (after best superposition of the receptor structure) of the solution closest to experiment. The second column (Rmsd$_{rec}$) gives the Rmsd (Cα) of the receptor structure with respect to experiment of the docking solution closest to experiment. The ranking of the near-native docking solution is given in the third column of plots. All systematic docking searches were performed according to Protocol 2 (inclusion of receptor flexibility during all docking runs, see Methodology section for details).

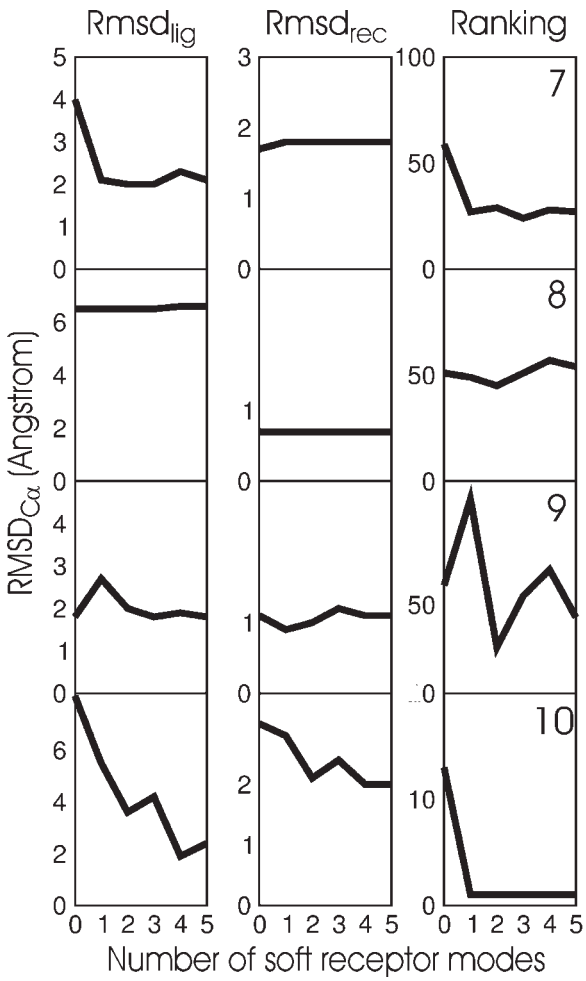


Figure 4

Systematic flexible docking of ligand proteins for Systems 7–10 (right panels) to apo receptor structures. Systematic docking minimization following Protocol 2 was performed using up to five softest modes as additional variables to account for receptor flexibility. Bound ligand protein structures and apo receptor structures were used. The columns of plots indicate Rmsd$_{lig}$, Rmsd$_{rec}$ and ranking, respectively, of the docking solution closest to experiment.

of ligand protein placement at the cost of a slightly deformed receptor structure was observed. In this case normal mode analysis correctly predicted a mobile loop that undergoes a conformational change upon ligand protein binding but the direction of loop movement did not overlap with the predicted mode direction. However, this "incorrect" movement still allowed for a better access to the ligand protein binding site and an improvement of Rmsd$_{lig}$. Inclusion of soft receptor modes for systems with small overlap of the soft modes and the observed receptor backbone changes (e.g. Systems 6 and 9) resulted in slight or no improvement of ligand protein placement, but also a slightly increased deviation of the receptor structure from experiment (Figs. 3 and 4).

Inspection of Figures 3 and 4 indicates that for most systems the ranking of the generated solutions closest to experiment is already quite good in case of rigid docking and consequently only a marginal or no improvement could be observed in most cases. However, for test Systems 1, 7, and 10 when using the receptor in the apo-conformation a significantly improved ranking of the near-native docking minimum was observed. In some cases[4–6] the ranking of solutions closest to experiment showed some fluctuation upon variation of the number of soft modes included during docking. One possible reason is that in these cases alternative docking geometries profit more from inclusion of some of the soft modes than near native docking solutions.

Normal modes facilitate the discovery of binding funnels

The systematic docking simulations indicate that inclusion of receptor flexibility improved the ligand protein docking placement in several cases dramatically but the influence on the ranking of the near-native solutions was insignificant. However, for several cases the systematic docking searches indicated that inclusion of soft flexible receptor modes can increase the number of docking solutions near the native binding geometry (Figs. 6 and 7). Similar trends were found for using apo or hybrid receptor structures (Fig. 6). In particular for Systems 1 and 2 the number of docking solutions with Rmsd$_{lig}$ $<$ 6–8 Å from experiment is larger for docking including receptor flexibility than in case of rigid docking (~30% for System 1). For System 2 no solutions with Rmsd$_{lig}$ $<$ 10 Å were found without receptor flexibility. The effect is also illustrated in Figure 6 for System 2. The three most accurate solutions in the rigid case are distributed around the

Table V

Systematic Docking Using Unbound Protein Partner Structures

	Rmsd$_{Lig}$ [Å]			Rank		
No.	Rigid	Semi-flexible	Fully-flexible	Rigid	Semi-flexible	Fully-flexible
1	4.8	4.8	4.4	2	2	13
2	9.7	5.0	4.6	14	1	2
3	6.3	7.7	7.2	36	4	11
6	4.3	3.8	3.9	111	105	164
7	4.1	7.2	5.7	99	69	111
8	8.7	7.7	5.5	64	114	97
9	2.0	2.7	3.7	19	17	16
10	9.5	9.8	10.5	42	19	24

Systematic docking minimizations (protocol 2) were performed for systems 1–10 with both partners in the apo-conformation. Both partner structures were either rigid or using the 5 softest receptor modes (semi-flexible) or including the 5 softest ligand and receptor protein modes as additional variables during docking minimization. Rmsd$_{Lig}$ and Rank indicate the ligand protein backbone Rmsd with respect to the experimental structure (after best superposition of the receptor structure) and the ranking, respectively, of the solution closest to experiment.

DOI 10.1002/prot

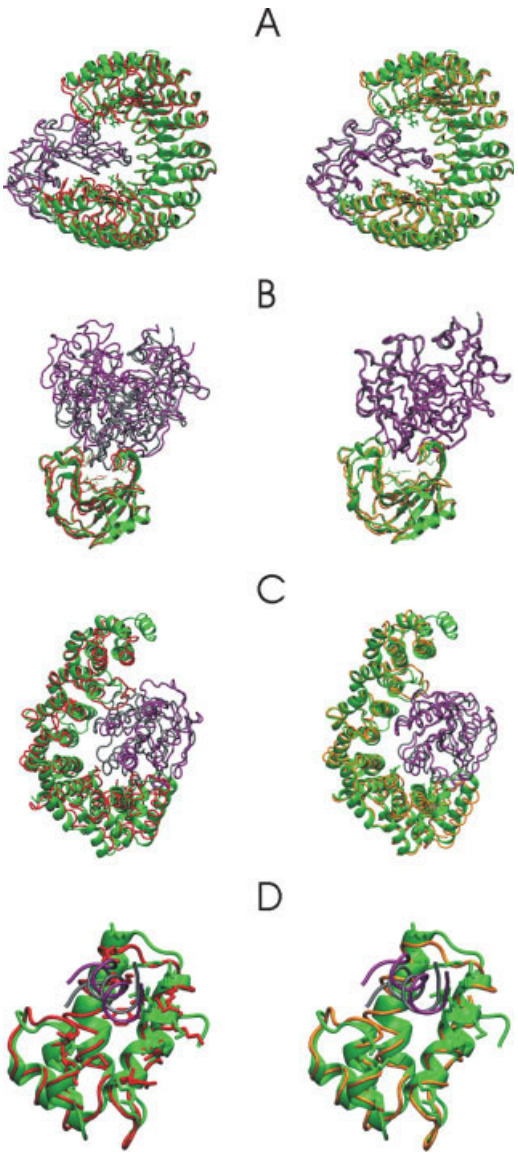


Figure 5

*Comparison of rigidly docked complexes (left panels) and docked complexes including receptor flexibility (right panels) with smallest deviation from experiment. Unbound (apo) receptor structures are shown as red tupes (left panels) and bound receptor structures in cartoon representation (green). Mobile side chains at the interface are also included. In the right panels the receptor structure minimized in five soft receptor modes during docking is shown as orange tubes. The ligand proteins at the experimental binding positions are indicated in grey (tube representation) and the docked ligand proteins as purple tubes. (**A**) System 1: RNAseA/inhibitor (receptor) complex, (**B**) System 2: Xylanase (receptor)/Taxi inhibitor complex (**C**) System 10: Importin Beta (receptor)/RanGTPase complex (**D**) System 5: HIV-Gag CA-C-terminal domain (receptor)/inhibitor complex. PDB entries of the docking cases are given in Table I.*

native binding position with an $Rmsd_{lig} > 10$ Å. In case of including precalculated soft modes during docking the three solutions closest to experiment converged to similar low energy structures with significantly reduced $Rmsd_{lig}$ (~3 Å). The inclusion of soft global modes appears to increase the chance of convergence towards a near native solution starting from the same docking start sites.

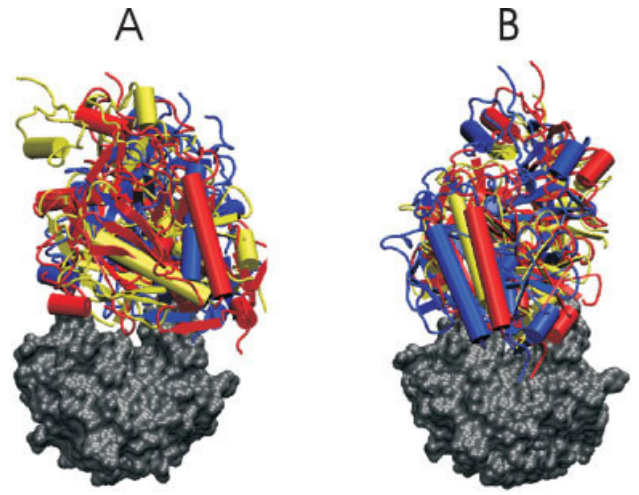


Figure 6

*Superposition of the three docking solutions closest to experiment for systematic docking runs on System 2 using a rigid receptor structure (**A**) and including the five softest receptor modes (**B**). The three docked ligand proteins are depicted in cartoon representation (coloured in yellow, red and blue, respectively). The receptor (bound) structure is shown as surface representation (grey). The three docking solutions closest to experiment obtained in case of including global receptor flexibility are much closer to experiment and show smaller pair-wise deviation compared to rigid docking (compare placement of a long helix of the ligand protein in panel B vs. A).*

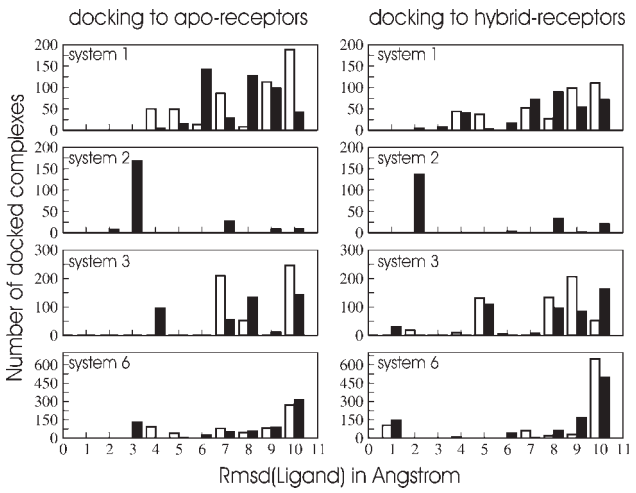


Figure 7

Number of docking solutions versus $Rmsd_{Lig}$ for systematic rigid docking (open bars) and flexible docking (filled bars, including the five softest receptor modes at all stages of docking minimization). Identical starting geometries for rigid and flexible docking were used (see Methodology section for details).

However, the effect was smaller for the systems with little overlap between precalculated soft modes and observed conformational difference between apo and bound structures (e.g. System 6, Fig. 7).

Systematic docking searches using unbound receptor and ligand structures

In addition to docking simulations employing the ligand protein in the bound form systematic docking searches with both partners in the unbound form were also performed. The docking searches included either both partners in the rigid apo-form, including global flexibility of the receptor structure (five softest modes, termed semi-flexible docking) or global flexibility of both partners (inclusion of the five softest modes for both protein partners). For the systematic searches we employed Protocol 2 (flexibility included at all docking stages). The deviation of the ligand placement closest to experiment improved in only three cases upon inclusion of soft-mode flexibility of the receptor or of both partner proteins (System 2, Table V). This contrasts to the results obtained using the ligand protein structures in the bound conformation (with an optimal local side chain structure at the protein–protein interface). In these cases a more significant improvement of the ligand placement closest to experiment was observed and also for more systems (Figs. 3 and 4). The result shows that an incorrect side chain conformation at the protein–protein interface can interfere with a realistic docking prediction of the position and orientation of the ligand protein even if soft-mode flexibility for the backbone is included. However, although the inclusion of the soft modes improved the $Rmsd_{lig}$ of near native docking solutions only in a few cases to some degree it improved the ranking of near-native solutions in several cases. Especially for systems with large overlap between soft modes and conformational difference between bound and unbound receptor structure (e.g. Systems 2, 3, and 10) a significant improvement of the ranking of the near native solution was found upon inclusion of global receptor flexibility (Table V). The ranking did not improve further upon inclusion of global ligand protein flexibility (fully-flexible case). In case with little overlap of the soft receptor modes with the experimentally observed conformational change (Systems 6–8) and also for System 1 the ranking for the fully-flexible docking became worse than for rigid docking. Again a likely reason is that conformational differences of side chains between unbound and bound states of the partner proteins interfere with the conformational adjustment in global modes. Future studies will explore the possibility to model side chain movement upon complex formation explicitly in addition to the inclusion of global soft-mode flexibility.

CONCLUSIONS

The inclusion of precalculated soft collective degrees of freedom during docking as additional variables offers a computationally efficient approach to approximately account for global flexibility during docking. In particular, the simple and rapid calculation of soft collective modes of proteins based on GNM allows to rapidly extract flexible collective modes prior to a docking simulation. Several previous studies indicate that such soft modes show considerable overlap with experimentally observed conformational changes associated with complex formation.[46,47,83] In agreement with previous work we also found significant overlap of precalculated soft GNMs with experimentally observed conformational changes for several systems. In previous work the inclusion of such soft modes during docking had only been applied during refinement of predocked complexes starting from placements close or exactly at the known binding geometry of the protein partners.[48,83] Purpose of the present study was to systematically test the effect of including soft modes obtained from a Gaussian network type model during docking minimization starting from thousands of start configurations and including several protein complexes. The systematic docking approach was first tested using a setup that was intended to reduce the influence of local conformational changes such as side chain changes on docking performance. This was achieved by employing one of the partners (the ligand protein) in its bound conformation and in a subset of cases also to use an apo receptor structure with side chains adjusted to the conformation in the bound (holo) structure. As a trend the inclusion of the flexible modes in several of these cases significantly reduced the deviation of the docking minima closest to experiment compared to rigid docking at a modest increase of computer time (~2-fold). In addition, the frequency of convergence to solutions close to the native binding geometry was considerably larger in several cases upon inclusion of global flexibility during docking. As a consequence it would be possible to start from fewer initial docking start geometries. The relative ranking of the solutions close to experiment improved for some docking test cases but the improvement was less dramatic than the improvement of the final docking geometry. The ranking of docking solutions depends on sterical fit but also on the performance of the scoring function. For several of the present docking runs when using the ligand in the bound form a quite good ranking of the near-native solutions was already found in case of rigid docking with little or no possibility of ranking improvement upon inclusion of soft mode flexibility. In addition, the approach was tested in a more realistic docking setup using both partner structures in the apo-form. In this case on average little or no improvement of ligand placement in the docking minima closest to experiment compared to rigid docking

DOI 10.1002/prot

was observed. However, in several cases especially those where backbone conformational changes showed significant overlap with the precalculated global soft modes a significant improvement of the ranking of near-native docking solutions was observed. As expected, no or only a small improvement of the docking performance was observed for control cases that did not show any significant conformational change upon complex formation or for those with little overlap of the changes with the precalculated soft modes. Together the docking results with apo partner structures and the docking experiments to reduce the influence of side chain conformational changes indicate the potential of the approach to improve both the ranking of near-native solutions and the agreement with experiment (if the side chains are close to the structure in the bound form). However, this requires a simultaneous efficient adjustment of side chain conformational changes during docking. Future research efforts will focus on combining adjustment of global conformational changes using soft precalculated soft modes with rapid adjustment of side chain conformations at the protein–protein interface. The present flexible protein–protein docking method might also be useful in case of docking employing protein model structures. Efficient inclusion of structural flexibility during docking may also improve docking results in these cases to cope with structural inaccuracies of the global structure of such models. Other applications include efficient inclusion of receptor flexibility during docking of drug-like molecules to protein target structures for rapid prescreening of putative ligands.

ACKNOWLEDGMENTS

This work was performed using computational resources of the CLAMV (Computational Laboratories for Analysis, Modeling and Visualization) at Jacobs University.

REFERENCES

1. Ehrlich LP, Wade RC. Protein–protein docking. Rev Comput Chem 2001;17:1761–1797.
2. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
3. Smith GR, Sternberg MJ. Prediction of protein–protein interactions by docking methods. Curr Opin Struct Biol 2002;12:28–35.
4. Bonvin AA. Flexible protein–protein docking. Curr Opin Struct Biol 2006;16:194–200.
5. May A, Zacharias M. Accounting for protein deformability during protein–protein and protein-ligand docking. Biochim Biophys Acta 2005;1754:225–231.
6. Teodoro ML, Phillips GN, Kavraki LE.Molecular docking: a problem with thousands of degrees of freedom. Proceedings 2001 IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea, May 2001. pp 960–965.
7. Adock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 2006;106:1589–1615.
8. Fernandez-Recio J, Trotov M, Abagyan R. ICM-DISCO docking by global optimization with fully flexible side chains. Proteins 2003;52:113–117.
9. Schueler-Furman O, Wang C, Baker D. Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with improved side-chain flexibility. Proteins 2005;187–194.
10. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–1737.
11. van Dijk AD, Kaptein R, Boelens R, Bonvin AM. Combining NMR relaxation with chemical shift perturbation data to drive protein–protein docking. J Biomol NMR 2006;34:237–244.
12. Jiang F, Kim S. Soft docking: matching of molecular surface cubes. J Mol Biol 1991;219:79–102.
13. Shoichet BK, Kuntz ID. Protein docking and complementarity. J Mol Biol 1991;221:327–346.
14. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 1992;89:2195–2199.
15. Norel R, Lin SL, Wolfson H, Nussinov R. Shape complementarity at protein–protein interfaces. Biopolymers 1994;34:933–940.
16. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
17. Vakser IA. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. Proteins 1997;1:226–230.
18. Palma PN, Krippahl L, Wampler JE, Moura JJ. Bigger: a new (soft) docking algorithm for predicting protein interactions. Proteins 2000;39:372–384.
19. Richie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178–194.
20. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation and electrostatics. Proteins 2002;47:281–294.
21. Zacharias M. Protein–protein docking with a reduced protein model accounting for side-chain flexibility. Proteins Sci 2003;12:1271–1282.
22. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins 2003;53:693–707.
23. Leach AR. Ligand docking to proteins with discrete side-chain Flexibility. J Mol Biol 1994;235:345–356.
24. Wang C, Schueler-Furman O, Baker D. Improved side chain modeling for protein–protein docking. Protein Sci 2005;14:1328–1339.
25. Jackson RM, Gabb HA, Sternberg MJE. Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. J Mol Biol 1998;276:265–285.
26. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci USA 2001;98:10636–10641.
27. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
28. Abagyan R, Totrov M, Kuznetsov D. ICM—a new method for protein modeling and design: applications to docking and structure prediction from distorted native conformation. J Comp Chem 1994;15:488–506.
29. Knegtel RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.
30. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol Pharmacol 2000;57:213–218.
31. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
32. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS- Automated docking to multiple target structures: incorporation of pro-

tein mobility and structural water heterogeneity in AutoDock. Proteins 2002;6:34–40.
33. Carlson HA, Masukawa KM, McCammon JA. Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 1999;103:10213–10219.
34. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc 2002;124:5632–5633.
35. Sandak B, Nussinov R, Wolfson HJ. A method for biomolecular structural recognition and docking allowing conformational flexibility. J Comput Biol 1998;5:631–654.
36. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins: Struct Funct Genet 1998;32:159–174.
37. Bastard K, Thureau A, Lavery R, Prévost C. Docking macromolecules with flexible segments. J Comput Chem 2003;24:1910–1920.
38. Bastard K, Prévost C, Zacharias M. Accounting for loop flexibility during protein–protein docking. Proteins: Struct Funct Bioinf 2006; 62:956–969.
39. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: applications to DNA minor groove ligand complex. J Comput Chem 1999;20:287–300.
40. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
41. Zacharias M. Rapid protein-ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. Proteins: Struct Funct Genet 2004; 54:759–767.
42. Smith GR, Sternberg MJE, Bates PA. The relationship between the flexibility of proteins and their conformational states on forming protein–protein complexes with an application to protein–protein docking. J Mol Biol 2005;347:1077–1101.
43. Tirion MM. Large amplitude elastic motions in proteins from a single-parameter atomic analysis. Phys Rev Lett 1996;77:1905–1908.
44. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
45. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins 1998;33:417–429.
46. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng 2001;14:1–6.
47. Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proc Natl Acad Sci USA 2005;102:18908–18913.
48. Lindahl E, Delarue M. Refinement of docked protein-ligand and protein-DNA structures using low-frequency normal mode amplitude optimization. Nucleic Acids Res 2005;33:4496–4506.
49. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640.
50. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 1995;374: 183–186.
51. Kobe B, Deisenhofer J. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. J Mol Biol 1996;264:1028–1043.
52. Nachman J, Miller M, Gilliland GL, Carty R, Pincus M, Wlodawer A. Crystal structure of two covalent nucleoside derivatives of ribonuclease A. Biochemistry 1990;29:928–937.
53. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: A critical assessment of predicted interactions. Proteins 2003;52:2–9.
54. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A. Structural basis for inhibition of *Aspergillus niger* xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.
55. Krengel U, Dijkstra BW. Three-dimensional structure of Endo-1,4-β-xylanase I from *Aspergillus niger*: molecular basis for its low pH optimum. J Mol Biol 1996;263:70–78.
56. Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns, A. Structural basis for inhibition of *Aspergillus niger* xylanase by triticum aestivum xylanase inhibitor-I. J Biol Chem 2004;279:36022–36028.
57. Zhang Z, Hayahashi MK, Merkel O, Stillman B, Xu MR. Structure and function of the BAH containing domain of ORC1P in epigenetic silencing. EMBO J 2002;21:4600–4612.
58. Hou Z, Bernstein DA, Fox CA, Keck JL. Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc Natl Acad Sci USA 2005;102:8489–8494.
59. Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA. Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. Proc Natl Acad Sci USA 1993;90:6325–6329.
60. Stanfield RL, Takimoto-Kamimura M, Rini JM, Profy AT, Wilson IA. Major antigen-induced domain rearrangements in an antibody. Structure 1993;1:83–93.
61. Ternois F, Sticht J, Duquerroy S, Krausslich HG, Rey FA. The HIV-1 capsid protein C-terminal domain in Compl a virus assembly inhibitor. Nat Struct Mol Biol 2005;12:678–706.
62. Worthylake DK, Wang H, Yoo S, Sundquist WI, Hill CP. Structures of the HIV-1 capsid protein dimerization domain at 26 A resolution. Acta Crystallogr D Biol Crystallogr 1999;55:85–92.
63. Sevcik J, Urbanikova L, Dauter Z, Wilson KS. Recognition of RNase Sa by the inhibitor barstar: structure of the complex at 1.7 A resolution. Acta Crystallogr D Biol Crystallogr 1998;54:954–963.
64. Sevcik J, Dauter Z, Lamzin VS, Wilson KS, Ribonuclease from Streptomyces aureofaciens at atomic resolution. Acta Crystallogr D Biol Crystallogr 1996;52:327–344.
65. Ratnaparkhi GS, Ramachandran S, Udgaonkar JB, Varadarajan R. Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. Biochemistry 1998; 37:6958–6966.
66. Frigerio F, Coda A, Pugliese L, Lionetti C, Menegatti E, Amiconi G, Schnebli HP, Ascenzi P, Bolognesi M. Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0 A resolution. J Mol Biol 1992;225:107–123.
67. Wang D, Bode W, Huber R. Bovine chymotrypsinogen A X-ray crystal structure analysis and refinement of a new crystal form at 1.8 A resolution. J Mol Biol 1985;18:5595–624.
68. Hyberts SG, Goldberg MS, Havel TF, Wagner G. The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. Protein Sci 1992;1:736–751.
69. Chen Z, Bode W. Refined 2.5 A X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. Crystallization Patterson search, structure determination, refinement, structure and comparison with its components and with the bovine trypsin-pancreatic trypsin inhibitor complex. J Mol Biol 1983;164:283–311.
70. Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. Refined 2 A X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase. Crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. J Mol Biol 1983;164:237–282.
71. Friedrich R, Fuentes-Prior P, Ong E, Coombs G, Oehler M, Hunter R, Pierson D, Gonzalez R, Huber R, Bode W, Madison EL. Catalytic domain structures of Mt-Sp1/Matriptase, a matrix-degrading transmembrane serine proteinase. J Biol Chem 2001;277:2160–2173.
72. Muller JJ, Lapko A, Bourenkov G, Ruckpaul K, Heinemann U. Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. J Biol Chem 2001; 276:2786–2789.

 DOI 10.1002/prot

73. Ziegler GA, Schulz GE. Crystal structures of adrenodoxin reductase in complex with NADP+ and NADPH suggesting a mechanism for the electron transfer of an enzyme family. Biochemistry 2000;39: 10986–10995.
74. Pikuleva IA, Tesh K, Waterman MR, Kim Y. The tertiary structure of full-length bovine adrenodoxin suggests functional dimers. Arch Biochem Biophys 2000;373:44–55.
75. Vetter IR, Arndt A, Kutay U, Gorlich D, Wittinghofer A. Structural view of the Ran-Importin beta interaction at 2.3 A resolution. Cell 1999;97:635–646.
76. Bayliss R, Littlewood T, Stewart M. Structural basis for the interaction between FxFG nucleoporin repeats and importin-β in nuclear trafficking. Cell 2000;102:99–108.
77. Kent HM, Moore MS, Quimby BB, Baker AM, McCoy AJ, Murphy GA, Corbett AH, Stewart M. Engineered mutants in the switch II loop of Ran define the contribution made by key residues to the interaction with nuclear transport factor 2 (NTF2) and the role of this interaction in nuclear protein import. J Mol Biol 1999;289:565–577.
78. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computerbased archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
79. Case D, Pearlman DA, Caldwell JW, Cheatham TE, III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner PK, Kollman PA. Amber 8. University of California, San Francisco, 2003.
80. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
81. Humphrey W, Dahlke A, Schulten K. VMD—visual molecular dynamics. J Mol Graph 1996;14:33–38.
82. Petrone P, Pande VS. Can conformational change be described by only a few normal modes? Biophys J 2006;90:1583–1593.
83. Bahar I, Rader EJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.