The Impact of Protein Flexibility on Protein–Protein Docking

Lutz P. Ehrlich,[†] Michael Nilges,[‡] and Rebecca C. Wade[*]
European Molecular Biology Laboratory and EML Research, Heidelberg, Germany

ABSTRACT **Accounting for protein flexibility in protein–protein docking algorithms is challenging, and most algorithms therefore treat proteins as rigid bodies or permit side-chain motion only. While the consequences are obvious when there are large conformational changes upon binding, the situation is less clear for the modest conformational changes that occur upon formation of most protein–protein complexes. We have therefore studied the impact of local protein flexibility on protein–protein association by means of rigid body and torsion angle dynamics simulation. The binding of barnase and barstar was chosen as a model system for this study, because the complexation of these 2 proteins is well-characterized experimentally, and the conformational changes accompanying binding are modest. On the side-chain level, we show that the orientation of particular residues at the interface (so-called hotspot residues) have a crucial influence on the way contacts are established during docking from short protein separations of approximately 5 Å. However, side-chain torsion angle dynamics simulations did not result in satisfactory docking of the proteins when using the unbound protein structures. This can be explained by our observations that, on the backbone level, even small (2 Å) local loop deformations affect the dynamics of contact formation upon docking. Complementary shape-based docking calculations confirm this result, which indicates that both side-chain and backbone levels of flexibility influence short-range protein–protein association and should be treated simultaneously for atomic–detail computational docking of proteins. Proteins 2005;58:126–133.**

Key words: protein docking; protein association; induced fit; molecular dynamics; protein flexibility; torsion angle dynamics; side-chain conformation

INTRODUCTION

The association of proteins plays a vital role in important biological processes such as transcription, signal transduction, and immune response. A profound understanding of the association process is required in order to design drugs that interfere with disease-related protein interactions. A variety of computational methods is available to predict the complex of 2 proteins, given their respective unbound conformations as obtained from crystallographic (X-ray) or NMR structure determination experiments.[1] These methods range from scanning all relative orientations by means of convolution-based techniques at low[2,3] and high[4] resolution or spherical polar Fourier correlations[5] via approaches utilizing pattern recognition techniques[6] to minimization of detailed energy functions.[7–9]

The majority of contemporary computational methods for protein–protein docking share the approximation that proteins can be considered rigid at the backbone level, while allowing for limited flexibility at the side-chain level. Underlying this approximation is the so-called "lock-and-key" paradigm, on the basis of which it is assumed that the association of 2 proteins is governed by their geometric complementarity, and little importance is attributed to structural changes of the backbone upon binding. This assumption is clearly a shortcoming for proteins that exhibit large-scale conformational changes upon binding. However, it is usually considered reasonable when adjustments in backbone conformation upon binding are small, especially as rigid body protein models are often parameterized to implicitly allow for some intermolecular penetration. In the case of small backbone adjustments upon binding, the assumption of backbone rigidity is generally justified for the diffusional association of 2 proteins over the relatively large length and timescales at which association is dominated by long-range electrostatic and hydrodynamic interactions.[10] However, its applicability to association from short protein separations is less clear. The transition from a diffusional encounter complex[11] to a fully bound complex is governed by short-range (van der Waals and hydrophobic) interactions, as well as electrostatic and hydrogen-bonding interactions. Here, local structural changes, such as rotation of surface side-chains taking 10^{-11}–10^{-10}s,[12] alter the microscopic environment on the same timescale as the final docking together of the pro-

Grant sponsor: Klaus Tschira Foundation.

*Correspondence to: Rebecca C. Wade, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany. E-mail: rebecca.wade@eml-r.villa-bosch.de

[†]Current address: Zetvisions AG, Im Breitspiel 21, D-69126 Heidelberg, Germany.

[‡]Current address: Unité de Bioinformatique Structurale, Institut Pasteur, 25-28 rue du Dr. Roux, F-75015 Paris, France.

Received 31 March 2004; Accepted 21 June 2004

Published online 28 October 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20272

teins. Large-scale structural changes, which generally occur on longer timescales, are less relevant to the final short-range association step, as they mostly precede it.[13] Consequently, here, we aim to assess the degree to which local side-chain and backbone flexibility influences the final, short-range stage of the association process of 2 proteins. The approach taken is designed to provide results relevant to understanding the physical process of protein–protein association and to the construction of algorithms for generating docked protein–protein complexes.

In this work, we study the effects of flexibility on computational docking of the extracellular ribonuclease barnase (bn) and its intracellular inhibitor barstar (bs). Structures of both the bound and unbound conformations have been solved experimentally to high resolution[14–16] (see Fig. 1). Their binding thermodynamics have been studied extensively by spectroscopic and calorimetric approaches.[18–20] The binding of bn and bs has been shown to be fast and result in a high affinity complex. At pH 8, the association rate constant of $k_{on} = 6.0 \times 10^8\ M^{-1}\ s^{-1}$ and dissociation rate constant of $k_{off} = 8 \times 10^{-6}\ s^{-1}$ result in a dissociation constant of $K_{diss} = 1.3 \times 10^{-14}\ M$.[18] This dissociation constant is equivalent to a binding free energy of $\Delta G = -18.9$ kcal/mol. Fast association is achieved by favorable electrostatic interactions between the proteins, which enhance their association rate.[19] Association rates for the wild-type proteins and a series of mutants could be reproduced by Brownian dynamics computer simulations,[13,21] in which the proteins are treated as rigid bodies with interactions dependent on shape and electrostatics, as represented by the Poisson–Boltzmann continuum solvent model.

Both bn and bs exhibit only limited structural changes upon binding. For bn, the C_α root-mean-square deviation (RMSD) between bound and unbound structures is 0.5 Å for the entire protein backbone. The binding loop around R59 in bn exhibits local deviations of up to 2 Å between bound and unbound conformations (see Fig. 1). For bs, the overall C_α RMSD is 0.5 Å, and there are no regions of larger deviation.[16] Therefore, bn and bs provide an experimentally well-characterized example system for investigating the effects of modest conformational changes on protein–protein docking.

To investigate the effects of modest conformational changes on protein–protein docking, we performed atomic-detail simulations of the association of bn and bs, starting with the protein structures separated by a short distance (5 Å). This separation distance is comparable to the deviations from the fully bound complex found in diffusional encounter complexes generated by Brownian dynamics simulation.[11] During the simulations described here, the degree of protein flexibility was varied. First, the proteins were treated as rigid, then full side-chain flexibility was introduced, and this was complemented by backbone deformations. We supplemented these simulations with shape-based rigid-body docking with experimentally determined structures and with backbone-deformed structures. These docking computations allow us to assess the effects of different contributions to conformational flexibility on the docking process.

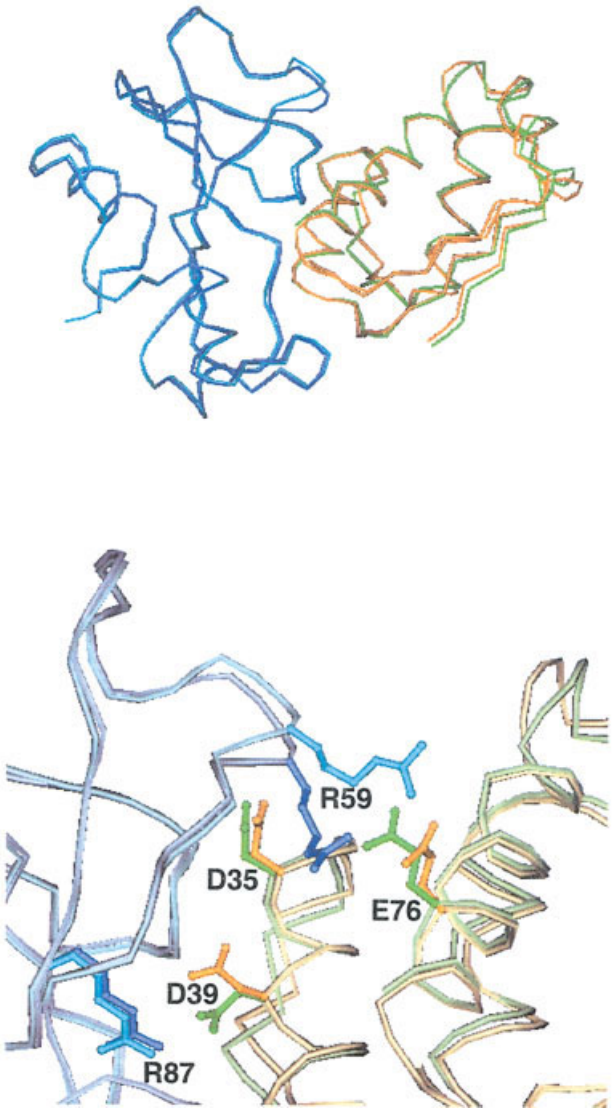

Fig. 1. Top: Ribbon structure of the complex of barnase (**left**) and barstar (**right**). The unbound (light blue and orange) conformations are superimposed onto the bound (dark blue and green) conformations in the complex determined crystallographically.[14–16] Bottom: A close-up view shows the location of the interfacial hotspot residue pairs studied. The figure was prepared using Molscript.[17]

MATERIAL AND METHODS

Initial Protein Structures

The following experimentally determined protein structures from the Protein Data Bank (PDB)[22] were used for this study:

Unbound barnase: PDB code: 1bnj, chain A, residues 3–110. This structure was determined by X-ray crystallography at a resolution of 2.1 Å by Mauguen et al.[14]

Unbound barstar: PDB code: 1bta. This structure contains the minimized average of 30 NMR structure models solved by Lubienski et al.[15] The average RMSD of these 30

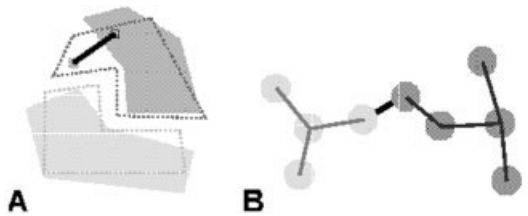

Fig. 2. Schematic diagrams of the definition of the docking metrics. (**A**) C_α RMSD of barstar interface residue; the deviation of a residue in the docked (shaded) barstar from its position in the bound crystal structure (dashed outline) is indicated. (**B**) Percentage of residue contacts correct (a contact is defined by a distance less than 1 Å between any of the atoms in the residues from the different proteins). See text for details.

structures from the mean structure is 0.42 ± 0.07 Å for the backbone atoms and 0.90 ± 0.07 Å for all nonhydrogen atoms.

Bound barnase/barstar complex: PDB code: 1brs, barnase chain A, barstar chain D. This structure was determined by X-ray crystallography at a resolution of 2.0 Å. The unit cell contains 3 bound complexes. The structure was solved for a mutant of barstar with cysteines 40 and 82 replaced by alanine.

Bound Structure Preparation Protocol

The bound structure of the complex of bn and bs with missing side-chains and polar hydrogen atoms added was provided by Dr. Razif Gabdoulline and had been prepared as described.[21]

Nonpolar hydrogen atoms were added to the proteins using the X-PLOR HBUILD command.[23] The hydrogen-bonding network was then reoptimized as follows: Ordered water molecules from the PDB file that were within 5 Å of any protein atom of chain A (bn) or chain D (bs) were selected and retained, using Quanta.[24] To generate hydrogen coordinates for the water and protein molecules, the ADDHYD procedure implemented in WHATIF[25] was used. This procedure generates hydrogen coordinates using an algorithm to optimize the hydrogen-bonding network.[26] During this procedure, barnase-H18 and barstar-H17 were assigned to the tautomer with $N_{\epsilon 2}$ protonated. The complex of bn and bs with the surrounding water molecules was then subjected to 100 steps of Powell energy minimization using X-PLOR.[23] The CHARMM22 all-atom force field[27] in which the Coulombic energy function was modified by a switching function with switching distances r_{on} = 7 Å and r_{off} = 11 Å, respectively,[23] was used. A cutoff of 12 Å was used for the construction of the nonbonded pair list.

Application of the preparation protocol led to a small C_α RMSD of 0.15 Å for both bn and bs. After energy minimization, the water molecules were discarded and the proteins were separated to generate starting structures for simulations and docking.

Unbound Structure Preparation Protocol

Unbound barnase: Residues 3–110 of chain A from PDB file 1bnj were extracted. The missing atoms of barnase-E60 were generated by mutating the residue to alanine and back to glutamic acid in WHATIF.[25] N- and C-terminii were generated with the program package InsightII.[28] The side-chain of barnase-H18 was flipped manually by 180° using Quanta to match the conformation found in the bound structure. Hydrogen atoms were added using the ADDHYD procedure[26] in WHATIF. The structure was energy-minimized using the same protocol as that used for the bound conformation. The RMSD of the C_α atoms after minimization was 0.17 Å.

Unbound barstar: For consistency with the bound structure, the C(40,82) → A mutations were generated in the 1bta structure, using WHATIF.[25] Hydrogen atoms were generated using WHATIF, as described above. The protonation of barstar-H17 was manually changed in order to match the protonation state of this residue in the bound state. Then energy minimization was performed exactly as for barnase. Preparation according to this protocol led to a C_α RMSD of 0.42 Å.

Barnase loop deformation protocol: In the unbound conformation of bn, we deformed the bn loop region, comprising residues 58–63, to its bound conformation by setting all backbone atom positions to those found in the bound structure after superposing unbound and bound conformations. After fixing the positions of the backbone atoms, the rest of the structure was relaxed by extensive energy minimization (1000 steps of steepest descent followed by 500 steps of conjugate gradient minimization using the AMBER94 forcefield as implemented in the Molecular Modelling Toolkit[29]).

Simulation Procedures

For all simulations, the initial relative orientation of bn and bs was obtained by pulling the respective bound structures apart by 5 Å along the line connecting the centers of mass. Depending on the particular simulation setup, different protein conformations were superposed onto the separated proteins. No relative rotation was applied to any of the conformations.

Simulations were performed using the program X-PLOR[23] in rigid body and torsion angle dynamics modes. Rigid body dynamics is implemented with an algorithm developed by Head-Gordon and Brooks.[30] Both proteins are considered completely rigid thoughout the simulation. Hence, the association process can be described by the time evolution of 6 degrees of freedom. The forces and torques acting on the rigid protein units are derived from the atomic molecular mechanics forces.

The torsion angle dynamics algorithm describes the time evolution of chosen internal coordinates.[31] In our work, all side-chain torsion angles that are not part of an aromatic ring were considered dynamic variables. The backbone was held rigid during all simulations. Simulations were performed using the all-atom CHARMM22 parameterization[27] in vacuo. To treat dielectric solvent screening, a distance-dependent dielectric of $\epsilon = 4R$ was chosen, with R being the separation distance of 2 interacting charges in angstroms. A switching function with 7(10) Å for the on(off) radius was used; a cutoff of 11 Å was

applied to the nonbonded pair list generation. Interaction with the solvent was mimicked by performing all simulations in a Langevin bath at 300 K. To enhance orientational sampling, a reduced friction coefficient of $2ps^{-1}$ was assigned to all atoms accessible to a 1.4-Å-radius probe in the isolated protein starting conformation prior to the simulation. Buried atoms were not subjected to any hydrodynamic interaction.

Before starting the simulations, steric clashes were removed. This was done by first performing 300 steps of "debumping" minimization with a simplistic force field consisting of only van der Waals interactions, with atomic radii scaled down by a factor of 0.9. Then an additional 200 steps of Powell minimization were performed with the full force field. This procedure was omitted for the rigid body simulations starting from the bound barnase/barstar conformations. Simulations were performed for 100 ps.

Shape-Based Docking

Shape-based docking was performed with the FTDOCK program package of Gabb et al.[4] In this method, all relative orientations of 2 rigid proteins are scanned by means of an efficient algorithm using fast Fourier transforms. Candidate docking orientations are filtered with respect to geometric and electrostatic complementarity.[4] Three different sets of starting structures for the 2 proteins were used: both proteins in unbound conformations, both proteins in bound conformations, and both proteins in unbound conformations with the backbone of the bn loop region (residues 58–63) deformed to the bound conformation as described above.

Parameters for the 2-stage FTDOCK procedure were set to the default values described in the FTDOCK manual.[32] During the first stage (coarse-grain docking), the 3 highest scoring candidate complexes were stored for each translational search. An angular granularity of 20° was used. An electrostatic filter removed unfeasible candidates. During the second stage (fine-grain docking), an angular granularity of 3° was used, while restraining the rotational range during refinement to 9°. A surface thickness of 1.2 Å was used. The best 100 candidate complexes were retained and scored with respect to the contacts and the bs interface RMSD metric (see below).

Analysis of Docking Progress

To assess docking progress, 2 metrics were used. These are the same as those introduced in the protein docking section of the CASP2[33] contest. The first metric [see Fig. 2(A)] is the RMSD of those C_α atoms of bs that are within 8 Å of any bn atom in the bound complex. To measure this metric for an instantaneous complex, the current bn C_α trace was superposed onto its bound conformation; the bs interface C_α RMSD was then determined with respect to the respective bound conformation.

The second metric measures the percentage of correctly established residue contacts. Two residues are considered to be in contact when the van der Waals sphere of any atom of 1 residue is within 1 Å of any atom belonging to the other residue [see Fig. 2(B)]. In the experimentally determined bound complex, bn and bs establish 45 such residue contacts.

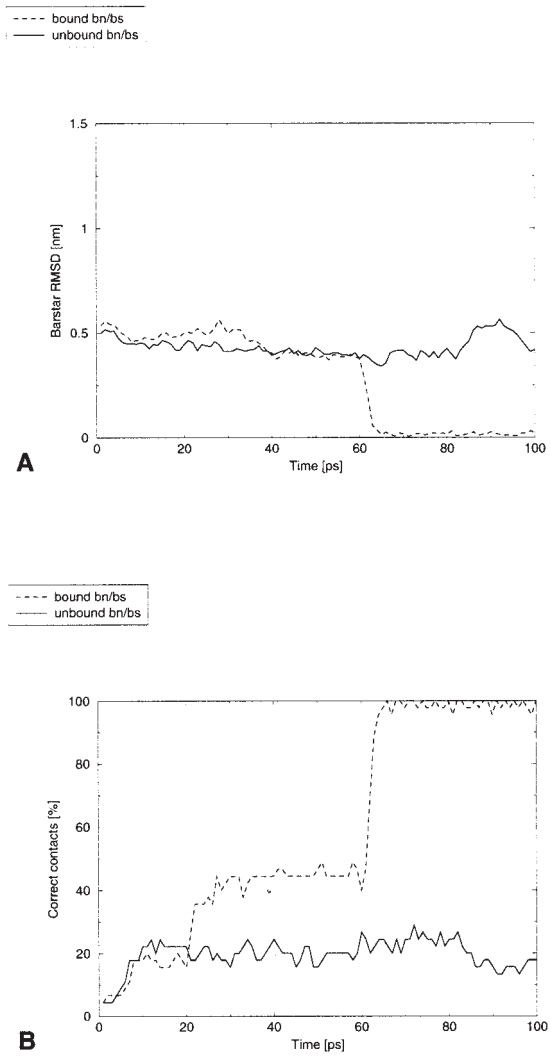

Fig. 3. Rigid body dynamics simulations. The simulations start with the bound (dashed lines) or unbound (solid lines) conformations of barnase and barstar separated by 5 Å. The plots show progress toward formation of the bound complex during the simulations, as given by 2 metrics (see Materials and Methods section for definition). (**A**) Barstar interface C_α RMSD. (**B**) Percentage of correct residue contacts.

RESULTS AND DISCUSSION

Structural Differences Prohibit Rigid Body Docking

To establish a baseline from which to assess how flexibility affects docking from short separations, rigid body simulations were performed first with the bound, and then with the unbound, conformations of bn and bs. Simulations were started with bn and bs simply separated 5 Å along their line of centers in the crystal structure of the complex. As is shown in Figure 3, under the conditions of the simulations, association to the correct structure of the complex occurs on timescales of approximately 50 ps for the bound conformations of the proteins. For the unbound

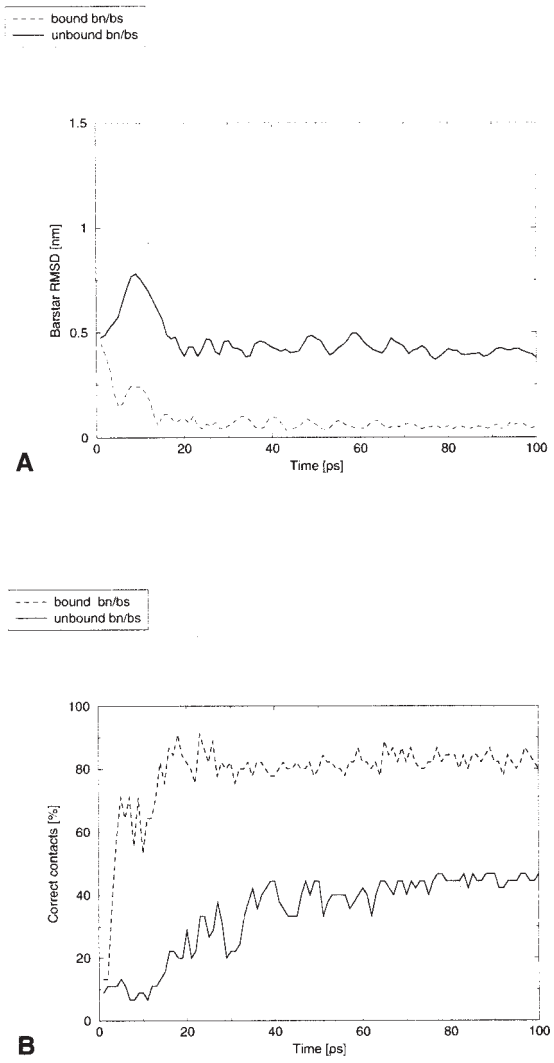

Fig. 4. Side-chain torsion angle dynamics simulations. The simulations start with the bound (dashed lines) or unbound (solid lines) conformations of barnase and barstar separated by 5 Å. The plots show progress toward formation of the bound complex during the simulations as given by 2 metrics (see Materials and Methods section for definition). (**A**) Barstar interface C_α RMSD. (**B**) Percentage of correct residue contacts.

protein conformations, on the other hand, a protein complex is not established during the 100-ps simulation time. The docking metrics have approximately constant values over the last 50 ps of the simulation, providing no indication that association would occur if the simulation were to be run longer. These results show the necessity of accounting for minor structural changes in atomically detailed docking.

Incorporating Side-Chain Flexibility Improves Docking

Figure 4 shows the results for the side-chain torsion angle dynamics simulations. Inclusion of side-chain flexibility leads to significant improvement in contact formation upon association starting from the separated unbound conformations, with nearly half of the correct residue contacts formed after 40 ps. On the other hand, the simulation starting from the correct bound protein conformations only leads to formation of 80% of the correct residue contacts. The reduced number of correct residue contacts compared to the rigid body simulations can be attributed to the different rotameric states of the interface side-chains that are acquired during the simulation due to thermal fluctuations. However, contact formation occurs much faster (within 20 ps for the bound conformations) than in the rigid body dynamics simulation. This can be explained by the "soft" side-chains making way for, instead of hindering, the approach of the 2 proteins.

The Conformations of the Hotspot Residue Side-Chains Influence Protein Association

In order to limit the extent to which flexibility is treated at an atomic level during computational docking, it is vital to know which interfacial residues are of special importance to the docking process. A compilation of mutagenesis experiments by Bogan and Thorn[34] indicates that certain (so-called *hotspot*) residues are far more important than others for binding affinity. We performed torsion angle dynamics simulations to check what role these hotspot residues play in the association process. These simulations used different starting conditions, in which the rotameric states of the respective residues were set to those found in the bound conformation, while keeping the rest of the protein in the respective unbound conformation and vice versa. Three pairs of hotspot residues, involving the 5 residues labeled in Figure 1 were, one at a time, switched to different rotamers and simulations run. Four of these 5 residues result in large changes in binding free energy when mutated to alanine (bs-D39A, 7.7 kcal/mol; bn-R87A, 5.5 kcal/mol; bn-R59A, 5.2 kcal/mol; bs-D35A, 4.5 kcal/mol; bs-E76A, 1.3 kcal/mol). A comparison of these results to the control simulations is displayed in Figure 5.

Figure 5(A) shows that, for simulations with the bound conformations, starting any of the 3 hotspot residue pairs in their unbound conformers reduced the percentage of correct contacts in the initial part of the simulations but had little effect on the results in the last 70 ps, when, for all 4 simulations (a control and those with the 3 rotated hotspot rotamers), the percentage of correct residue contacts was about 80%. The torsion angle dynamics had overcome the initial structural differences, and the hotspot residues mostly moved to rotamers closer to their bound rotamers.

For simulations with the unbound protein structures, Figure 5(B) shows that changing the rotamers of just 2 residues to their bound rotamers at the beginning of the simulations can have a big effect, resulting in faster association in the initial 10 ps, and in 2 of the 3 cases, much closer association, with about 70% correct residue contacts formed after 100 ps. In the cases with improved association, the hotspot residues retained their bound rotamers, whereas in the case with weaker association, the bn-R59 side-chain moved away from the bound (and also the unbound) rotamer conformation during the simulation. The difference in backbone conformation at R59

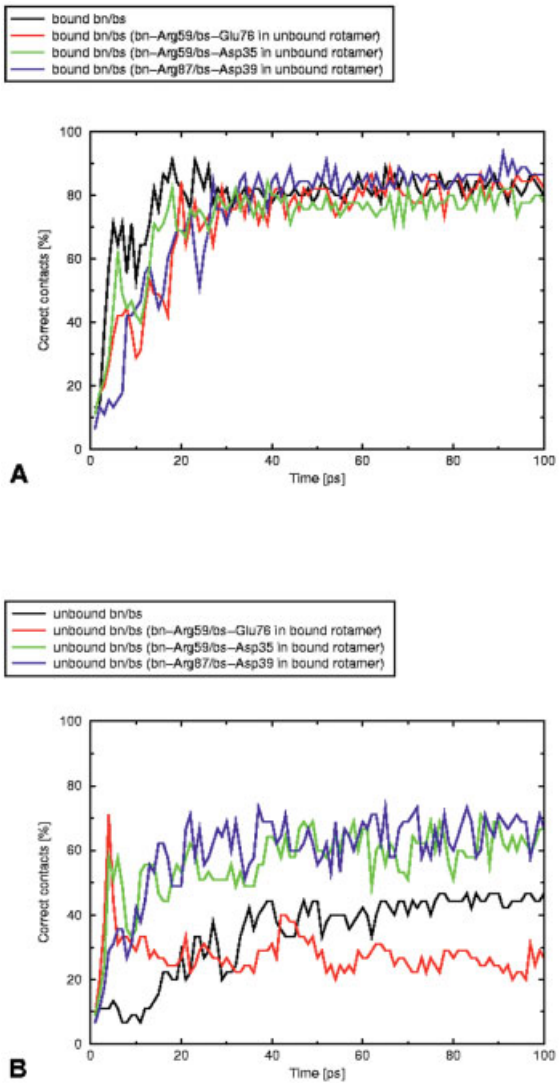

Fig. 5. Side-chain torsion angle dynamics simulations with hotspot residues assigned to different rotamers at the beginning of the simulations. The plots show progress toward formation of the bound complex during the simulations, as given by the percentage of correct residue contacts formed. (**A**) Bound conformations of the proteins, with specified side-chain pairs set to rotamers observed in the unbound conformations. (**B**) Unbound conformations of proteins, with specified side-chain pairs set to rotamers observed in the bound conformations.

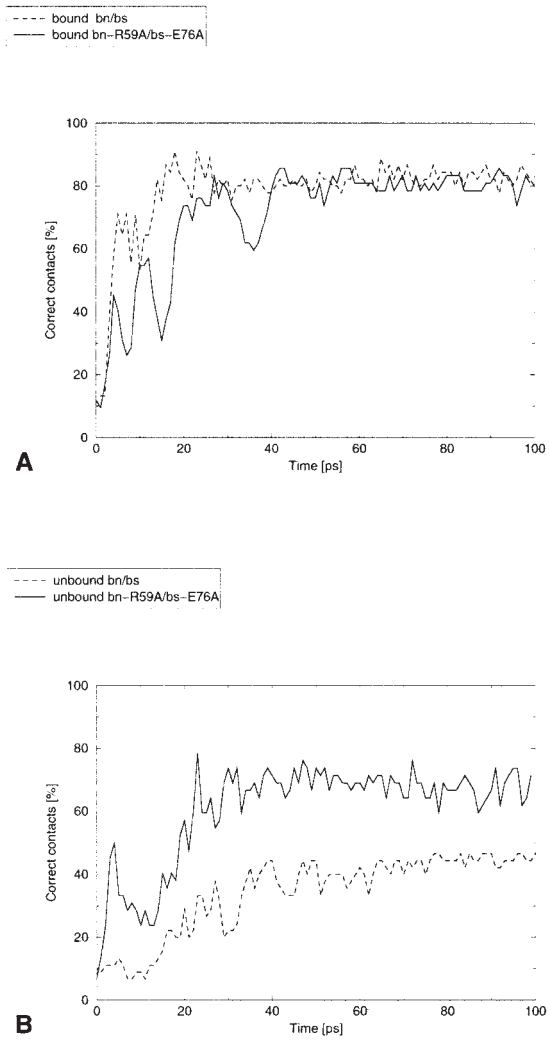

Fig. 6. Side-chain torsion angle dynamics simulations for wild-type proteins and for the bn-R59A and bs-E76A mutants. The plots show progress toward formation of the bound complex during the simulations, as given by the percentage of correct residue contacts formed. Bound (**A**) and unbound (**B**) conformations of the proteins are shown.

between bound and unbound states affected salt-link formation to R59 even though the side-chain was initially placed in the bound rotamer.

To further probe the impact of these hotspot residues, additional simulations with the respective alanine mutants were performed. The results are shown in Figure 6 for the bn-R59–bs-E76 hotspot residue pair; the other hotspot residue pairs lead to similar results, although with less pronounced differences in complex formation metrics between the wild-type and mutant proteins (data not shown). Mutation of these 2 residues to alanine in the bound protein structures slowed the association process but nevertheless resulted in about 80% of correct contacts being formed in the latter half of the simulations [see Fig. 6(A)]. Mutation of bn-R59 and bs-E76 to alanine in the unbound protein structures notably improved association compared to the wild-type proteins, resulting in about 70% of correct residue contacts being formed. This improvement is due to removal of the obstruction caused by the unbound rotamers of bn-R59 and bs-E76.

To summarize these findings, the correct conformation of hotspot residues is vital to establishing the proper contacts in the initial docking phase. These residues play an active role in contact formation. Consequently, one should account for their flexibility explicitly in protocols for protein–protein docking at atomic resolution.

Minor Backbone Deformations Influence Protein Docking

To elucidate how minor backbone deformations affect protein–protein association, the following torsion angle

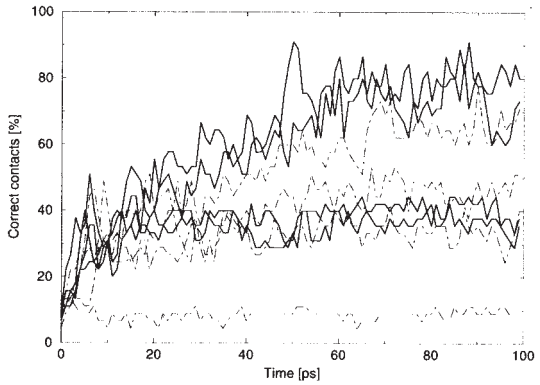

Fig. 7. Side-chain torsion angle dynamics simulations with deformed barnase-R59 loop region. The plot shows progress toward formation of the bound complex during the simulations, as given by the percentage of correct residue contacts formed. The starting protein structure was in the unbound conformation except that the backbone of the bn-R59 loop was deformed to the bound conformation. Dashed lines indicate 4 different simulations with bn-R59 and bs-E76 initially set to their respective unbound side-chain rotamers; solid lines represent 4 simulations with these two residues initially set to their respective bound side-chain rotamers. Within each set of 4 simulations, the initial random number seed differs.

dynamics simulations were performed. The starting structures were the unbound protein conformations with a deformation introduced into bn in its reactive loop, spanning residues 58–63, to match the conformation found in the bound structure. Two different setups were used for the simulations: The residues bn-R59 (located in the deformed loop) and bs-E76 were initially set to either their bound or their unbound rotamers.

As shown in Figure 7, multiple torsion angle dynamics simulations do not indicate a clear-cut separation in association behavior between the two sets of starting structures. It is notable, though, that the average contact metric for the simulations with the deformed loop and bound rotameric states for the hotspot residue pair improves by 20% compared to the torsion angle dynamics simulation starting from the unbound bn and bs structures. In addition, the simulation performing best with this setup performs as well as the simulation with entirely bound conformations (see Fig. 5).

These findings indicate that a local change in the backbone contributes to the formation of the correct contacts. To confirm this hypothesis, we performed additional shape-based docking with the FTDOCK package,[4] using 3 different sets of starting structures. Apart from using the bound and unbound conformations of the proteins, the unbound conformation of bs was combined with the unbound but locally deformed (at the 58–63 loop) conformation of bn.

As can be seen from Table I, the quality of the best candidate with respect to the residue contact and RMSD metrics is improved by using the deformed bn conformation instead of the unbound one. This is an indication that even minor backbone changes affect docking on an atomic level. Table II shows the rankings of the best candidates out of 100 kept with respect to the other 3 metrics studied. These data provide further insight into the shortcomings of the rigid shape-based docking approach when dealing with unbound structures. From the bound via the deformed to the unbound conformation, the ranking of the best candidates with respect to the internal shape-based FTDOCK score deteriorates. Qualitatively, this can also be seen from the FTDOCK scores of the respective best candidates in Table I.

TABLE I. Best-Candidate Complexes Generated Using FTDOCK

Conformation	Best in Contacts [%]	Best RMSD [Å]	Best FTDOCK Score
Unbound	13.3	5.8	205
Deformed	22.2	4.7	195
Bound	51.1	2.4	202

For 3 different protein conformations (see text for details), the 100 best candidates generated by FTDOCK were retained and scored with respect to the 2 docking metrics and the FTDOCK score. The best complexes according to each metric are listed.

TABLE II. Ranking of the Best FTDOCK Candidate Complexes

		Best in Contacts	Best RMSD	Best FTDOCK
Unbound	Contacts ranking	1	3	N/A[a]
	RMSD ranking	2	1	42
	FTDOCK ranking	91	53	1
Deformed	Contacts ranking	1	1	73
	RMSD ranking	1	1	53
	FTDOCK ranking	96	96	1
Bound	Contacts ranking	1	1	3
	RMSD ranking	2	1	3
	FTDOCK ranking	16	3	1

[a]Candidate was one of 80 out of 100 structures with 0% correct contacts.
See text for details.

CONCLUSIONS

We report results of rigid body and torsion angle dynamics simulations to investigate the impact of local protein flexibility at the side-chain and main-chain level on protein–protein docking. Comparing simulations starting with the bound and unbound conformations of the proteins, we find that treatment of side-chain flexibility is needed, if the docking of proteins from their native unbound structures is to be described at an atomic level. Furthermore, certain hotspot contact residue pairs, which make important contributions to the binding energetics, seem to contribute actively to contact formation; treatment of their conformational flexibility is vital in atomic docking simulations.

Treatment of side-chain flexibility alone was, however, not sufficient to achieve satisfactory docking starting from unbound protein structures in side-chain torsion angle dynamics simulations. We find that minor loop deformations in the binding interface influence contact formation upon docking significantly. In particular, the proper backbone conformation of the bn interface loop, comprising residues 58–63, and

the correct (bound) rotameric states for the hotspot residue pair bn-R59–bs-E76 are shown to result in improved contact formation, comparable to that in simulations starting with the bound conformations. The impact of minor backbone deformation was confirmed by rigid body shape-based docking simulations using FTDOCK. Small backbone motions are particularly important when they are coupled with side-chain motions, such as when a small peptide chain motion is necessary to make a side-chain rotamer sterically accessible.

From our study, we find that minor backbone deformations cannot be neglected in protein–protein docking procedures. They may be partly treated implicitly by, for example, using low-resolution rigid body docking grids[2] or by scaling van der Waals interactions to make docking "soft."[8,35] A more complete treatment can be achieved by coupling rigid body docking with subsequent refinement by molecular dynamics (MD) simulation. This was recently shown in MD simulations of the association of bn and bs from 5 Å separation using unbound structures and an implicit solvent model.[36] Docked complexes with approximately 70% of native contacts were achieved, but the quality of the complexes, including hydrogen-bonding patterns of hotspot residues, was dependent on the implicit solvent model used; the best results were obtained with the neutralized, polarized, surface area dependent (NPSA) implicit solvent model.[36]

We conclude that simulation protocols for protein–protein docking at an atomic level of detail aimed at presenting a realistic picture of protein–protein binding events should include simultaneous treatment of side-chain and main-chain flexibility, as even small backbone conformational changes can have as much impact on protein–protein association as changes in side-chain rotameric states.

ACKNOWLEDGMENTS

We thank Jens Linge for critical reading of the manuscript and Juergen Schlitter, Shoshana Wodak, and Razif R. Gabdoulline for helpful discussions.

REFERENCES

1. Ehrlich L, Wade RC. Protein–protein docking. Rev Comp Chem 2001;17:61–97.
2. Vakser IA, Aflalo C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. Proteins 1994;20:320–329.
3. Vakser IA. Protein docking for low-resolution structures. Protein Eng 1995;8:371–377.
4. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 1997;272:106–120.
5. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178–194.
6. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins 1998;32:159–174.
7. Abagyan R, Totrov M, Kuznetsov D. ICM—a new method for protein modelling and design: Applications to docking and structure prediction from the distorted native conformation. J Comput Chem 1994;15:488–506.
8. Camacho CJ, Vajda S. Protein docking along smooth association pathways. Proc Natl Acad Sci USA 2001;98:10636–10641.
9. Zacharias M. Protein–protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci 2003;12:1271–1282.
10. Gabdoulline RR, Wade RC. Biomolecular diffusional association. Curr Opin Struct Biol 2002;12:204–213.
11. Gabdoulline RR, Wade RC. On the protein–protein diffusional encounter complex. J Mol Recogn 1999;12:226–234.
12. McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. New York: Cambridge University Press; 1987.
13. Gabdoulline RR, Wade RC. Protein–protein association: investigation of factors influencing association rates by Brownian dynamics simulations. J Mol Biol 2001;306:1139–1155.
14. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C,, Jack A. Molecular structure of a new family of ribonucleases. Nature 1982;297:162–164.
15. Lubienski MJ, Bycroft M, Freund SM, Fersht AR. Three-dimensional solution structure and ^{13}C assignments of barstar using nuclear magnetic resonance spectroscopy. Biochemistry 1994;33:8866–8877.
16. Buckle AM, Schreiber G,, Fersht AR. Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-Å resolution. Biochemistry 1994;33:8878–8889.
17. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
18. Schreiber G, Fersht AR. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry 1993;32:5145–5150.
19. Schreiber G, Fersht AR. Rapid, electrostatically assisted association of proteins. Nat Struct Biol 1996;3:427–431.
20. Frisch C, Schreiber G, Johnson CM, Fersht AR. Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation. J Mol Biol 1997;267:696–706.
21. Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. Biophys J 1997;72:1917–1929.
22. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
23. Brünger AT. X-PLOR: a system for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1992.
24. Accelrys. San Diego, CA: Quanta. 1992.
25. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56, 29.
26. Hooft RW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 1996;26:363–376.
27. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher I, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies. J Phys Chem B 1998;102:3586–3616.
28. Accelrys. San Diego, CA: InsightII. 1999.
29. Hinsen K. The Molecular Modelling Toolkit: a new approach to molecular simulations. J Comput Chem 2000;21:79–85.
30. Head-Gordon T, Brooks CL III. Virtual rigid body dynamics. Biopolymers 1991;31:77–100.
31. Rice LM, Brünger AT. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins 1994;19:277–290.
32. Gabb HA. FTDOCK manual. London: Biomolecular Modelling Laboratory, Imperial Cancer Research Fund; 1997.
33. Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, Olson A, Duncan B, Rao M, Jackson R, Sternberg M, James MN. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to beta-lactamase. Nat Struct Biol 1996;3:233–239.
34. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
35. Fernandez-Recio J, Totrov M,, Abagyan R. Soft protein–protein docking in internal coordinates. Protein Sci 2002;11:280–291.
36. Wang T, Wade RC. Implicit solvent models for flexible protein–protein docking by molecular dynamics simulation. Proteins 2003;50:158–169.