Dynamical View of the Positions of Key Side Chains in Protein-Protein Recognition

S. Roy Kimura, Richard C. Brower, Sandor Vajda, and Carlos J. Camacho
Biomedical Engineering Department, Boston University, 44 Cummington Street, Boston, Massachusetts 02215 USA

ABSTRACT When a complex is constructed from the separately determined rigid structures of a receptor and its ligand, some key side chains are usually in wrong positions. These distortions of the interface yield an apparent loss in affinity and would unfavorably affect the kinetics of association. It is generally assumed that the interacting proteins should drive the appropriate conformational changes, leading to their complementarity, but this hypothesis does not explain their fast association rates. However, nanosecond explicit solvent molecular dynamics simulations of misfolded surface side chains from the independently solved structures of barstar, bovine pancreatic trypsin inhibitor, and lysozyme show that even before any receptor-ligand interaction, key side chains frequently visit the rotamer conformations seen in the complex. We show that these simple structural motifs can reconcile most of the binding affinity required for a rapid and highly specific association process. Side chains amenable to induced fit are also identified. These results corroborate that solvent-side chain interactions play a critical role in the recognition process. Our findings are also supported by crystallographic data.

INTRODUCTION

The interface between two proteins in a complex is generally as tightly packed as the protein interior (LoConte et al., 1999). Direct interactions and the removal of water from the interface provides the attractive contributions to the binding free energy that more than compensate for the loss of translational, rotational, and side chain entropy upon association. In addition, for protein pairs such as barnase and barstar, the highly specific electrostatic interactions provide long-range steering effects, resulting in an association rate that is higher than it would be without such attractive forces (Schreiber and Fersht, 1996). We have recently shown that desolvation also provides specific attractive interactions that, albeit shorter-range and weaker than the electrostatic steering, can also increase association rates by several orders of magnitude (Camacho et al., 1999, 2000b).

The tightly packed interface and its thermodynamic and kinetic consequence are all lost when a complex is constructed from the separately determined structures of two proteins. Indeed, more often than not one finds some interfacial structural motifs in "wrong" positions (Koshland, 1958, 1963, 1994; Jorgensen, 1991), resulting in steric clashes and unfavorable electrostatic interactions, even for high resolution x-ray structures and for proteins whose backbone remains practically invariant in the process of binding (Jackson et al., 1998; Vakser et al., 1999; Camacho et al., 2000a). These findings emphasize that upon binding, the protein interactions should lead to some degree of induced fit (Koshland, 1994; Jorgensen, 1991), resulting in the tightly packed interface. Induced fit theory as introduced by Koshland (1958) explains the origin of functional specificity and describes how a substrate changes the structure of an enzyme to bring its catalytic groups into the proper alignment, whereas a nonsubstrate does not. However, the notion of induced fit is also frequently used in a more general sense as the origin of binding specificity, i.e., the collection of conformational changes resulting in optimal interactions when two molecules come in contact (Jorgensen, 1991).

Though induced fit is necessary for forming the well-packed interface seen in protein-protein association, it cannot fully explain important kinetic requirements of the binding process. For instance, if side chains are in wrong conformations before the two molecules contact each other, then the long and medium-range forces due to electrostatics and desolvation that normally bring the two molecules together can be completely eliminated (see below). Furthermore, if these misfolded side chains should rearrange in order to restore the proteins attraction, then these changes must occur within the time scale of an encounter between the receptor and ligand. The lack of nonspecific protein aggregation constrains this time scale to a few nanoseconds (Northrup and Erickson, 1992; Camacho et al., 2000b). On the other hand, even a small rotation of a medium-sized side chain in the protein interior can take as much as 1 s (Creighton, 1993). In this context neither the traditional lock and key (Fischer, 1894) nor the induced fit (Jorgensen, 1991) model seems to provide a rationalization for the attraction between proteins required for a fast and specific binding process. In particular, it becomes an open problem what positions the key side chains occupy before and during the process of protein-protein association.

We use a novel molecular dynamics (MD) algorithm to analyze the coupling between water molecules and flexible surface side chains in three protein ligands. MD has been used to analyze the short time scale behavior of small molecules, mainly water, around typically rigid macromolecules (Brunne et al., 1993; Steinbach and Brooks, 1993;

Received for publication 27 July 2000 and in final form 20 November 2000.
Address reprint requests to Carlos J. Camacho, Biomedical Engineering Department, Boston University, 44 Cummington St., Boston, Massachusetts 02215. Tel.: 617-353-4842; Fax: 617-353-4814; E-mail: ccamacho@bu.edu.

0006-3495/01/02/635/08 $2.00

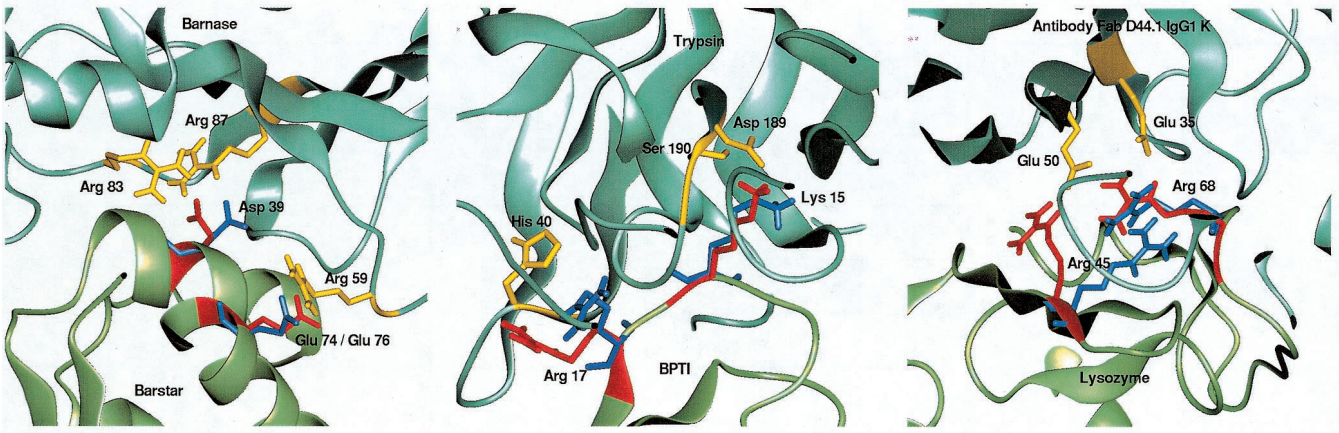

FIGURE 1 Receptor/ligand interfaces in the complex structure with the simulated side chains. In blue are the side chains found in the monomer (unbound) ligand structure. In red are side chains found in the cocrystallized (bound) complex. In yellow are nearby charged residues on the cocrystallized receptor. *Left*: the barnase-barstar complex. Note that Glu76 of 1bta corresponds to Glu74 of 1brs. *Middle*: the trypsin-BPTI complex. *Right*: the Fab arm of antibody IgG1 D44.1, κ, complexed with hen egg-white lysozyme.

Makarov et al., 1998). However, it has also been observed experimentally that water-protein interactions influence protein dynamics (Zanotti et al., 1999). Recognizing that for many protein-protein complexes, surface side chains are the basis for the detailed chemistry that gives rise to their fast and highly specific association, we studied the effects of explicit solvent on the conformational dynamics of surface side chains from the independently solved (unbound) structures of barstar, bovine pancreatic trypsin inhibitor (BPTI), and lysozyme in the absence of the receptor using MD simulations.

This paper revisits the principle, proposed nearly a hundred years ago by Langley (1907), that a ligand is attracted to its receptor by a specific pre-existing affinity between the molecules. In particular, we considered barstar, BPTI, and lysozyme because they show little or no affinity for their receptors when calculated using the x-ray or nuclear magnetic resonance (NMR) structures of the unbound ligands. We focused on the side chains found at the binding interface of the complex structures. Our simulations show that by properly solvating these side chains, the rotamers sampled are frequently those seen in the complex. We also checked that the use of these rotamers restores the expected affinity between the molecules. This confirms that for such side chains, a localized lock-and-key-like complementarity exists before any interdigitation of the monomers. In contrast, side chains close to the perimeter of the interface, though important for the stability of the complex, are generally found in nonspecific rotamers, suggesting that their final conformations are induced by the presence of the receptor.

MATERIALS AND METHODS

Protein complexes

We chose to study representative systems from each of the three major classes of protein-protein complexes for which structures are known. These are the barnase-barstar complex (PDB code 1brs) from the RNase-inhibitor family, the trypsin-BPTI (2ptc) and the trypsin-kallikrein (2kai) complexes from the protease-inhibitor family, and hen egg-white lysozyme bound to two different antibodies, Fab D44.1 (IgG1, κ) (1mlc) and HyHEL5 (2hfl) from the antigen-antibody family (Fig. 1). The structures of these receptors and ligands have also been solved independently and, hence, have been analyzed in numerous docking studies (Jackson et al., 1998; Camacho et al., 2000a; Vakser and Aflalo, 1994; Vakser et al., 1999).

MD simulations were carried out on the unbound protein ligands alone. We focused our attention on two charged side chains on each ligand, selected on the basis of their involvement in interactions with the receptor. The first of the side chains in each structure is fully buried in the complex ("interfacial" in Table 1), whereas the second side chain is always partially exposed to the solvent ("peripheral"). In order to improve the efficiency of the simulations, all other side chains were kept fixed in their unbound conformations (see Table 1 and Fig. 1). We have checked that our results and conclusions did not significantly change when we allowed neighboring side chains to move (see MD protocol). As shown below, the conformations of the selected side chains were substantially affected by the water-ligand interactions. In contrast, the interfacial side chains on the unbound receptors in our examples (e.g., 1bao, 2ptn, 1mlb) differed by <1 Å root mean square deviation (RMSD) between their free and bound states.

Surface of active polarons: semi-explicit solvation model

The MD simulations presented in this study were carried out using a semi-explicit solvent model formulated and developed in previous papers (Brower and Kimura, 1998; Kimura et al., 2000). Briefly, the method involves using a layer of explicit water surrounding the solute and a set of surface charges that are placed directly on the oxygen atoms of the explicit water molecules. The cooperative effects of these charges recreate the polarization of the external bulk not included in the simulation. Here, the

TABLE 1 Side chains studied by MD simulation

	Side chain	
Ligand (PDB)	Interfacial	Peripheral
1bta	asp39	glu76
6pti	lys15	arg17
1lza	arg68	arg45

external water is modeled as a dielectric continuum with $\epsilon_1 = 78.0$. The surface charges are updated continually throughout the simulation using a discretized version of the self-consistent relation,

$$\sigma(\xi) = -\frac{\epsilon_i - \epsilon_0}{\epsilon_1 + \epsilon_0} \sum_k \frac{q_k \hat{\mathbf{n}}(\xi) \cdot (\xi - \mathbf{r}_k)}{2\pi|\xi - \mathbf{r}_k|^3} - \frac{\epsilon_1 - \epsilon_0}{\epsilon_1 + \epsilon_0} \int_{\partial v} \frac{\sigma(\xi')\hat{\mathbf{n}}(\xi) \cdot (\xi - \xi')}{2\pi|\xi - \xi'|^3} dA(\xi') \quad (1)$$

This expression is the same as that used in the boundary element method (Zauhar and Morgan, 1985, 1990; Purisima and Nilar, 1995) and is derived from Poisson's equation and appropriate boundary conditions (Kimura et al., 2000). In the equation, ϵ_0 is the dielectric constant of the vacuum, r_k and q_k are the coordinates and magnitudes of the explicit charges of the system (including both solute and explicit water), ξ are the positions of the oxygens of the explicit water molecules, $\hat{n}$ are the outward normal vectors on these oxygens, and the integration is performed numerically around a closed boundary that is approximated by the oxygen positions of the layer of water surrounding the solute of interest. The first term in this equation accounts for the potential due to the explicit charges in the system, and the second comes from the potential due to the surface charge distribution on the boundary. We refer to these water molecules with their extra charge as a Surface of Active Polarons (SOAP). In addition to the extra charges, short-range forces and thermal fluctuations are applied to the SOAP particles to model hydrostatic pressure and transmission of energy from the bulk. The model has been shown to reproduce solvation energies of 9 ions and 14 polar and charged amino acids with high accuracy (correlation coefficients with experiment 1.000 and 0.995, respectively; Kimura et al., 2000).

Initial conditions

The structures for barstar, BPTI, and lysozyme were obtained from the Protein Data Bank (PDB codes 1bta, 6pti, and 1lza) and modeled using the OPLS/AMBER united atom force field (Jorgensen and Tirado-Rives, 1988) supplied with the TINKER (Pappu et al., 1998) v3.6 distribution. The molecules were solvated with a 9.0-Å layer of explicit TIP3P (Jorgensen et al., 1983) water and minimized until the gradient of potential reached 5.0 Kcal/mol/Å with all solute atoms except the side chains of interest fixed. A very brief initial MD simulation was carried out (4000 steps for 1bta, 3000 steps for 6pti, and 5000 steps for 1lza using a 2-fs time step) to obtain the initial surface charges on the explicit water. Data from these preliminary simulations were not used in our analysis for this paper.

MD protocol

All MD simulations were carried out using the TINKER package (Pappu et al., 1998), which was modified to include the semi-explicit solvation model described above and also a newly parallelized force loop. The modified package was compiled and executed on an Origin 2000 (SGI, Mountain View, CA). The complete coordinate set was saved every 100 fs during each simulation for off-line processing. Water bond lengths were constrained using the rattle algorithm (Andersen, 1983), and all solute atoms other than the two side chains of interest were rigidly constrained to the original coordinates of the unbound ligand PDB structures. Coordinates were updated using the velocity Verlet algorithm with a 2-fs time step. For the electrostatics, a 20 Å distance cutoff was used with cubic spline potential smoothing to avoid potential discontinuities.

To check that fixing neighboring residues did not adversely affect our results, we repeated the lysozyme simulation for 500 ps, allowing the residues Arg45, Asn46, Thr47, Asp48 and Gly67, Arg68, Thr69, Pro70 to move while keeping everything else fixed (not shown). Conformational sampling of the two side chains of interest for this system, Arg45 and Arg68, showed slower convergence but very similar trends, as when the neighboring residues were fixed.

Grid potential

We optimized the simulations by precalculating the effects of fixed atoms under a plane located 6.0 Å below the lowest atom of the simulated side chains. The electrostatic potential was computed at points sampled on a cubic lattice with 0.5 Å spacing within an approximately $60 \times 50 \times 60$ Å box that enclosed the part of the molecule above the plane. This potential was computed using both the solute atoms and the solvent particles with their extra charge, which was calculated and saved during the short initial simulation. This ensured that the effects of the dielectric boundary at the far end of the molecule was approximated in the grid potential as a static reaction field due to the charges of the system. The resulting grid potential was used in all MD simulations using linear interpolation to approximate the continuously varying potential between grid points. In addition, a reflective barrier was placed at the plane separating the implicit and explicit parts of the molecule to prevent water from escaping below the plane. This simplification improves the efficiency of the simulations by approximately a factor of 9.

Evaluation of binding free energy

As in a recent analysis by Lee et al. (2000), we used an effective free energy function to estimate energies for conformations extracted from our MD simulation to compare with that of the crystal complex. Specifically, we computed the binding energies for the cocrystallized protein structures, the complex formed by the native receptor and unbound (independently solved) ligand, as well as for solvated side chain conformations identified in our MD simulations.

The binding free energy is calculated by the expression (Vajda et al., 1994; Novotny et al., 1989)

$$\Delta G_{\text{bind}}^{\text{calc}} = E_{\text{coul}} + \Delta G_{\text{des}} + \Delta G_{\text{rot-trans}}, \quad (2)$$

where E_{coul} denotes the direct electrostatic energy between receptor and ligand, ΔG_{des} corresponds to the desolvation energy including side chain entropy loss, and $\Delta G_{\text{rot-trans}}$ accounts for translational, rotational, vibrational, and cratic effects. E_{coul} is calculated using a distance-dependent dielectric (McCammon et al., 1979) equal to $4r$, enforcing a minimum atom-to-atom distance separation equal to the sum of their corresponding van der Waals radii to avoid artificial overlaps. ΔG_{des} is estimated using an atomic contact energy (ACE; Zhang et al., 1997). For the most part, $\Delta G_{\text{rot-trans}}$ accounts for the loss of rotational and translational degrees of freedom upon binding. This entropic barrier opposing protein association must be surmounted in order for proteins to bind. $\Delta G_{\text{rot-trans}}$ is typically assumed to be a weak function of the size and shape of the interacting proteins (Vajda et al., 1994; Novotny et al., 1989). Here, we assume $\Delta G_{\text{rot-trans}} = 5.0$ Kcal/mol, a value that fits our data well.

Eq. 2 will be used to understand how the side chain conformations affect the binding free energy. In all calculations the receptor is kept in its bound structure, but the conformation of the ligand varies. For example, $\Delta G_{\text{bind}}^{\text{calc}}(\text{ul})$ and $\Delta G_{\text{bind}}^{\text{calc}}(\text{bl})$ will denote the calculated free energies of the bound (b) and the unbound (u) conformations of the ligand (l). Furthermore, within the rigid-body approximation, the apparent loss of affinity ΔG_{loss} for the unbound ligand is estimated as

$$\Delta G_{\text{loss}} = \Delta G_{\text{bind}}^{\text{calc}}(\text{ul}) - \Delta G_{\text{bind}}^{\text{calc}}(\text{bl}). \quad (3)$$

The same equation (Eq. 2) for estimating changes in binding free energy will be used when key side chains on the ligands are replaced by conformations identified through MD simulation (see Results and Table 3).

RESULTS

Free energy calculations using x-ray structures

Table 2 shows a list of the calculated and observed binding free energies, ΔG^{calc}_{bind} and ΔG^{exp}_{bind}, respectively. We also list ΔG_{loss} defined in the methods, and ΔG^{fix}_{loss}, which denotes the loss of affinity of the unbound ligand when the two selected side chains are replaced by their conformations found in the crystal structure of the complex. The agreement between ΔG^{calc}_{bind} and ΔG^{exp}_{bind} is good, with the single exception of the lysozyme-HyHel5 complex (3hfl/1lza), which other groups have also found hard to model (Novotny et al., 1989). Including the water molecules found at the interface of the antibody-antigen complexes yields an improved estimate of $\Delta G^{calc}_{bind} = -17$ Kcal/mol for 3hfl/1lza (unpublished). This, together with previous validation of ACE (Zhang et al., 1997), gives credence to our estimates of ΔG_{loss}.

Because the side chains selected for our study are charged and form salt bridges in the complex, the binding affinity is particularly sensitive to their position. We find that the loss of affinity amounts to >75% of the total binding free energy of the complex. This loss is significantly reduced when the two selected side chains are in the rotamer conformation found in the complex, while all the other side chains remain in their unbound conformations (see ΔG^{fix}_{loss} in Table 2). It is important to emphasize that error bars on ΔG_{loss} do not change the main observation that misfolded side chains can significantly affect the receptor/ligand binding affinity, both for the complex and for near-native structures (Camacho et al., 2000a).

Dynamics

From the analysis of the nanosecond MD simulations of the solvated ligands shown in Fig. 2, we conclude that the side chains fully buried in the complex structure (Asp39 for 1bta, Lys15 for 6pti, Arg68 for 1lza) are frequently in a state within 1.5 Å RMSD of the side chain conformation in the complex structure. In contrast, the side chains that are partially exposed in the complex (Glu76 for 1bta, Arg17 for 6pti, Arg45 for 1lza) do not sample their native-like complex conformation. To quantify the sampling of the different rotamers, each side chain conformation was clustered

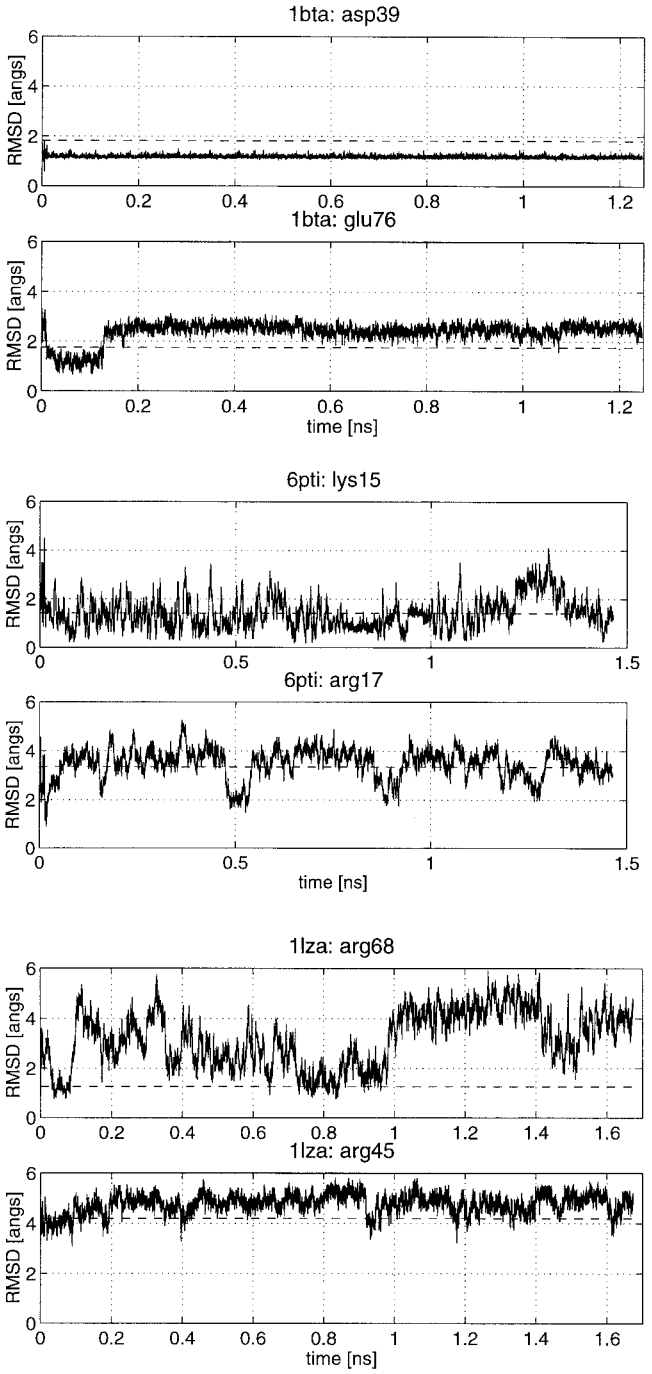

FIGURE 2 Heavy atom RMSD of each side chain studied relative to the corresponding side chains found in the complex. RMSD values are plotted every 50 steps during the MD simulation. The *dashed line* shows the RMSD of the initial unbound monomer structure.

TABLE 2 Binding free energies

Complex/ligand code	$\Delta G^{expt}_{bind}(bl)$	$\Delta G^{calc}_{bind}(bl)$	ΔG_{loss}	ΔG^{fix}_{loss}
		Kcal/mol		
1brs/1bta	−18.9	−18.2	17.6	1.4
2ptc/6pti	−18.1	−17.0	13.2	0.3
2kai/6pti	−12.4	−14.0	11.1	0.3
1mlc/1lza	−9.7	−11.1	16.9	1.0
3hfl/1lza	−14.5	−24.0	8.3	5.5

* Schreiber and Fersht, 1993; Krystek et al., 1993; Schwarz et al., 1995.

around Dunbrack's rotamer library (Dunbrack and Cohen, 1997) and ranked in Table 3.

Buried side chains move to positions favorable for protein-protein recognition

Strikingly, Asp39 of 1bta (Fig. 2, *top*) very rapidly moves and settles into a conformation close to that of the complex (approximately 1.2 Å RMSD), in the absence of the interacting receptor. The simulations show only one dominant cluster (Table 3) containing the native-like rotamer (no. 4) of Asp 39 in the complex crystal structure (1brs). This side chain cluster alone, in the unbound ligand PDB structure, yields a substantial improvement in ligand/bound-receptor interaction energy, described by the mean gain $\langle \Delta G_{gain}^{Asp39} \rangle = \langle \Delta G_{loss}^{Asp39} - \Delta G_{loss} \rangle = -7.1$ Kcal/mol (see Table 2), when compared to the binding affinity of the intact unbound ligand and the bound receptor. In this expression $\langle \Delta G \rangle$ denotes averaging the free energy over the cluster.

TABLE 3 Clusters of side chain conformations*

Ligand/Side chain	Rotamer no.	% share	⟨RMSD⟩	$\langle \Delta G_{gain}^{obs.rot.} \rangle^{\dagger}$
1bta				
Asp39	**4,5,6**	**99.9**	**1.20 (±0.05)**	**−7.07 (±0.96)**
	7,8,9	0.1	1.62 (±0.19)	−0.76 (±2.00)
Glu76	25,27	89.6	2.70 (±0.18)	0.96 (±0.14)
	22,23,24	3.9	1.24 (±0.15)	−2.62 (±0.89)
6pti				
Lys15	70,71,72	34.5	1.26 (±0.23)	−0.81 (±2.84)
	67,68,69	**22.4**	**0.66 (±0.15)**	**−8.05 (±2.27)**
	10,11,12	12.8	1.93 (±0.24)	2.87 (±1.74)
	64,65,66	8.0	1.28 (±0.31)	−2.27 (±4.52)
	55,56,57	5.6	1.37 (±0.13)	−4.04 (±1.12)
	58,59,60	3.9	1.10 (±0.12)	−5.71 (±1.06)
Arg17	67,68,69	30.5	3.85 (±0.21)	−0.68 (±2.38)
	71	14.8	3.33 (±0.23)	−4.71 (±1.77)
	55,57	11.4	3.70 (±0.17)	−4.57 (±1.04)
	72	8.1	3.89 (±0.18)	−4.30 (±2.05)
	10,11	7.9	3.63 (±0.31)	−5.80 (±1.18)
	43,44,45	7.6	2.66 (±0.23)	−2.26 (±1.43)
1lza				
Arg68	31,32,33	13.2	4.21 (±0.24)	13.79 (±4.40)
	79,80,81	8.2	4.12 (±0.27)	12.02 (±3.36)
	28,29,30	7.2	4.17 (±0.22)	10.10 (±2.38)
	16,18	**6.7**	**1.73 (±0.27)**	**11.01 (±5.04)**
	11	**3.6**	**2.04 (±0.20)**	**−1.64 (±2.24)**
	13,14,15	**3.3**	**1.07 (±0.12)**	**0.25 (±2.41)**
	38	**2.9**	**1.57 (±0.31)**	**1.24 (±4.72)**
	17	**1.7**	**2.00 (±0.29)**	**4.08 (±4.43)**
Arg45	4,5,6	64.0	4.92 (±0.21)	−6.04 (±1.97)
	22,23,24	10.0	4.01 (±0.29)	−7.56 (±2.12)
	1,2	9.0	4.76 (±0.15)	−3.68 (±1.75)
	16,17,18	3.1	4.67 (±0.10)	−4.24 (±1.72)
	14,15	2.6	3.99 (±0.16)	−7.85 (±1.25)

* Snapshots were taken every 50 steps during the MD simulations and were binned according to their nearest rotamer tabulated in Dunbrack″s rotamer library (Dunbrack and Cohen, 1997). The bins were further grouped if they shared all but the last χ angle, and were within 0.5 Å RMSD of each other. Rotamers are numbered according to Dunbrack. Bold type corresponds to native-like rotamers.

† $\langle \Delta G_{gain}^{obs,rot} \rangle = \langle \Delta G_{loss}^{obs,rot} - \Delta G_{loss} \rangle$ is the average free energy gain with respect to the unbound structure when the ligand side chain is replaced by the observed rotamers in the cluster.

Although we see more fluctuations for Lys15 of BPTI, this side chain also frequently visits essentially the conformation in the complex at 0.7 Å RMSD. Out of all the possible conformations, the rotamers in the unbound and bound structures rank first and second in their frequency of sampling. Consistent with the interaction energy in the complex, the cluster of structures closer to the native side chain have an energy improvement of $\langle \Delta G_{gain}^{Lys15} \rangle = -8.1$ Kcal/mol.

For lysozyme, the unbound crystal structure shows Arg68 and Arg45 in electrostatically favorable pockets on the protein surface. Despite its large fluctuations between 1 and 6 Å RMSD, upon solvation Arg68 repeatedly returns to a state very close to the complex (rotamer no. 14) near 1.0 Å RMSD and spends as much as 20% of the time in conformations with low (<2.0 Å) RMSD values. Examination of these clusters reveals that they are structurally very close to the bound conformation, and a small adjustment of about 1 Å (perhaps caused by the approaching antibody) is all that is required to bring this side chain to its final bound position. Although this side chain has an initial unbound monomer conformation that is already close to the complex (Fig. 1), the fact that it returns to these low-RMSD conformations suggests the existence of an attractive well in that region that competes favorably with the solvation entropy. Because the unbound rotamer and the rotamer found in the native complex state are very similar, there is no significant energy improvement for this side chain in its solvated conformation.

Partially solvent exposed side chains are typically found in nonspecific rotamer conformations

For Glu76 of 1bta, there is a dominant cluster somewhat close to the rotamer found in the unbound structure (rotamer 27), and only a very brief visit to a rotamer close to the one found in the bound structure (rotamer 22).

The largest cluster for Arg17 of BPTI is near the rotamer in the initial unbound structure; we do not see sampling of the native complex rotamer. Nevertheless, we find substantial energy improvements for many of the nonspecific highly ranked clusters, suggesting that the rotamer in the unbound crystal is in a particularly unfavorable conformation.

Finally, Arg45 in 1lza appears to make only rare visits to a 3.5 Å RMSD baseline, and otherwise stays further away from the initial monomer structure, near 5.0 Å RMSD. Arg45 neither includes the bound (rotamer 33) or unbound (rotamer 59) rotamers in its frequently visited clusters and

yet we see improvement in energies for the top two clusters of about $\langle \Delta G_{gain}^{Arg45} \rangle = -7$ Kcal/mol. These clusters are dominated by solvation effects, suggesting that the rotamer found in the unbound structure is unfavorable in solution. The main observation here is that in the original unbound structure, Arg45 completely blocks the binding site, preventing any possible affinity between the monomers. Our MD simulations suggests that in solution, this side chain moves away from the binding area, allowing the proper complementary surfaces to interact.

DISCUSSION

The goal of this study was to investigate the conformational dynamics of side chains on protein surfaces that play an important role in complex formation (for an animated visualization see http://engpub1.bu.edu/~srk/keys.html). The hypothesis is that key side chains must already be in positions suitable for binding in the absence of their receptors to allow for the recognition process to take place in a biologically feasible time scale, and to yield enough intermolecular affinity to overcome the entropic barrier to binding. Among the side chains we studied, the three that are located deep within the binding interface indeed exhibit key-like behavior. For Asp39 of barstar, Lys15 of BPTI, and Arg68 of lysozyme, we find highly populated clusters of conformations that are within 1.5 to 2 Å RMSD from the structures found in the complex. This sampling occurs within the nanosecond time scale, which also corresponds to the typical lifetime of an encounter complex (Northrup and Erickson, 1992; Camacho et al., 2000b). Thus, we conclude that in all likelihood when ligands interact with their receptors their key side chains are in their "right" bound-like conformations. Table 2 shows that the interactions of the appropriately oriented side chains with their substrate significantly enhances the binding affinity between monomers. The highly specific nature of the interactions between these *key* residues and their receptors also suggest that mutations of these side chains should likely result in a decrease of the association rate. The latter is consistent with the observation (Castro and Anderson, 1996) that the Lys-15-Ala mutation on BPTI leads to a 200-fold decrease of the on rate of the BPTI-trypsin association. From a thermodynamic point of view, we have also checked that an alanine mutation on any of the side chains considered here results in an overall decrease of the binding affinity of the complex.

For the other three side chains on the periphery of the binding interface, we find highly populated nonspecific conformations that differ from the initial unbound structures. In the original unbound ligands, these side chains are often found interfering or even blocking the binding site; we find that solvation effects remove these chains from the binding area to positions accessible to the final complexed state. Our simulations suggest that these side chains act more as "latches" that hold the molecules together rather than as keys fitting snugly into the receptor. Consistent with

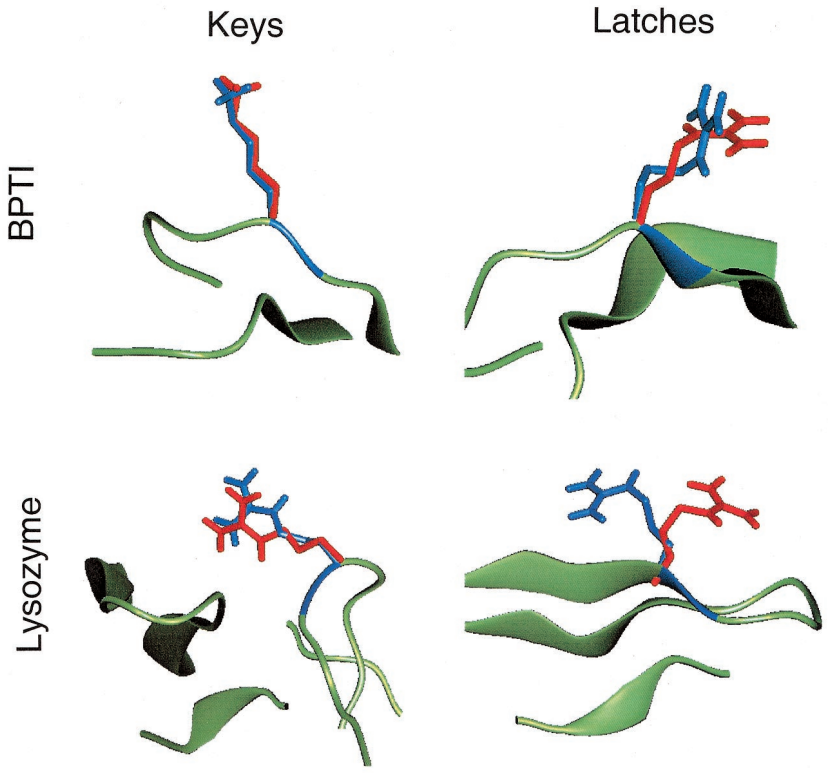

FIGURE 3 Interfacial side chain conformations of BPTI and lysozyme when bound to two different receptors. For BPTI (*top row*), in blue and red are side chain conformations found in the bound complex with trypsin (2ptc) and kallikrein (2kai), respectively. For lysozyme (*bottom row*), in blue and red are the conformations found in the complex with antibodies D44.1 (1 mlc) and HyHel5 (3hfl), respectively.

an induced fit mechanism, these latches should be open before docking and only fasten on to their salt-bridge partners late in the binding process. An implication of these results is that these salt bridges should mostly yield a decrease in the off-rate as opposed to an increase in the association rate. The latter is consistent with the mutagenesis experiments of Schreiber and Fersht (1996) where the mutation of Glu76 to alanine in barstar resulted in a very modest decrease of the association rate with barnase. Similarly, Castro and Anderson (1996) have shown that the mutation of the latch residue Arg17 in BPTI to alanine again leaves the on rate almost unchanged, whereas the off rate is larger than for the wild-type.

Crystallographic data shown in Fig. 3 provide further support for the generality of our findings. Namely, the key-like feature of the buried side chains is confirmed by the fact that BPTI complexed with kallikrein has the same Lys15 rotamer as BPTI with trypsin. The same is found for Arg68 in lysozyme complexed with antibodies D44.1 and HyHEL5. The latter is true despite the fact that these complexes have very different binding free energies and association rate constants. On the other hand, the induced fit mechanism suggested for the nonspecific partially exposed side chains would suggest that their final complex conformations should not necessarily be the same for different complexes. This is indeed the case for the bound rotamer of Arg17 in BPTI complexed with trypsin and kallikrein, and for Arg45 in lysozyme complexed with antibodies D44.1 and HyHEL5 which, as shown in Fig. 3, are found in very different rotamers.

The binding mechanism that emerges from our analysis is one in which the chemical affinity between receptors and their ligands is crucial for a fast recognition process. This affinity is enhanced when key side chains, most important for binding, are properly oriented near their conformations seen in the complex, allowing for the rapid and highly specific interdigitation of the molecules. This mechanism is consistent with the thermodynamic maps (Camacho et al., 1999) of electrostatic and desolvation free energies of receptor/ligand systems, which clearly show a funnel-like shape around a properly aligned precursor state. Furthermore, Brownian dynamic simulations (Camacho et al., 2000b) have confirmed that these molecules could efficiently avoid a lengthy interdigitation process over large contact areas by locking into a well-defined diffusion accessible precursor structure. Partially solvent exposed side chains do not appear to play this key-like role but they contribute to the stability of the complex.

An interesting practical application of our finding is to use the identification of conformational preferences of surface side chains in combination with standard docking algorithms to predict the structure of the complex. Challenges that remain for this type of procedure include the reduction of the number of side chain conformation candidates to a reasonably small number and the inherent computational cost of atomistic MD or Monte Carlo simulation using an accurate solvent model. The benefits gained from this approach are evident, as it decouples the docking procedure into separate side chain and rigid-body searches, thereby reducing the space of possibilities to a manageable size.

In summary, we find that important information can be learned from the characteristic distribution of rotamers resulting from the coupled dynamics of flexible surface side chains and its hydration layer. Our findings suggest that the unique environment of the side chains consisting of neighboring atoms and the solvent, which is ultimately encoded in the primary structure, holds the keys for protein binding. Complex-forming proteins appear to have interfacial side chains that frequently sample favorable conformations in order to meet the stringent kinetic requirements for binding.

This research has been supported by grants DBI-9904834 and GER-9452651 from the National Science Foundation and 1RO1 GM 61867–01 from the National Institute of Health.

REFERENCES

Andersen, H. 1983. Rattle: a velocity version of the shake algorithm for molecular dynamics calculations. *J. Comp. Phys.* 52:24–34.

Brower, R., and S. Kimura. 1998. Semi-explicit bag model for protein solvation. *In* Computer Simulations of Biomolecular Systems: Theoretical and Experimental Applications, vol. 3. van Gunsteren, W., P. Weiner, and A. Wilkinson, editors. ESCOM, Leiden. 223–243.

Brunne, R., E. Liepinsh, G. Otting, K. Wuthrich, and W. van Gunsteren. 1993. Hydration of proteins: a comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J. Mol. Biol.* 231:1040–1048.

Camacho, C., D. Gatchell, S. Kimura, and S. Vajda. 2000a. Scoring docked conformations generated by rigid body protein-protein docking. *Proteins Struct. Funct. Genet.* 55:525–537.

Camacho, C., S. Kimura, C. DeLisi, and S. Vajda. 2000b. Kinetics of desolvation mediated protein binding. *Biophys. J.* 78:1094–1105.

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

Castro, M., and S. Anderson. 1996. Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to β-trypsin and α-chymotrypsin. *Biochemistry.* 35:11435–11446.

Creighton, T. 1993. Proteins: Structures and Molecular Properties. W. H. Freeman and Company, New York, 2nd edition.

Dunbrack, R., and F. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6:1661–1681.

Fischer, E. 1894. Einfluss der configuration auf die wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* 27:2985–2993.

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

Jorgensen, W. 1991. Rusting of the lock and key model for protein-ligand binding. *Science.* 254:954–955.

Jorgensen, W., J. Chandrasekhar, and J. Madura. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Jorgensen, W., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins: energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

Kimura, S., R. Brower, C. Zhang, and M. Sugimori. 2000. Surface of active polarons: a semi-explicit solvation method for biomolecular dynamics. *J. Chem. Phys.* 112:7723–7734.

Koshland, D. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–104.

Koshland, D. 1963. Correlation of structure and function in enzyme action. *Science.* 142:1533.

Koshland, D. 1994. Key-lock theory and induced fit theory. *Angew. Chem. Int. Ed. Engl.* 33:2375–2378.

Krystek, S., T. Stouch, and J. Novotny. 1993. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J. Mol. Biol.* 248: 661–679.

Langley, J. 1907. On the contraction of muscle, chiefly in relation to the presence of "receptive" substances. *J. Physiol. (Lond.)* 36:347–384.

Lee, M., Y. Duan, and P. Kollman. 2000. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins Struct. Func. Genet.* 39:309–316.

LoConte, L., C. Chothia, and J. Janin. 1999. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* 285:2177–2198.

Makarov, V., M. Feig, B. Andrews, and B. Pettitt. 1998. Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophys. J.* 75:150–158.

McCammon, J., P. Wolynes, and M. Karplus. 1979. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry.* 18:927–942.

Northrup, S., and H. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulations. *Proc. Natl. Acad. Sci. USA.* 89:3338–3342.

Novotny, J., R. Bruccoleri, and F. Saul. 1989. On the attribution of binding energy in antigen-antibody complexes mcpc 603, d1.3 and hyhel-5. *Biochemistry.* 28:4735–4749.

Pappu, R., R. Hart, and J. Ponder. 1998. Analysis and application of potential energy smoothing and search methods for global optimization. *J. Phys. Chem.* 102:9725–9742.

Purisima, E., and S. Nilar. 1995. A simple yet accurate boundary-element method for continuum dielectric calculations. *J. Comp. Chem.* 16: 681–689.

Schreiber, G., and A. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry.* 32: 5145–5150.

Schreiber, G., and A. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3:427–431.

Schwarz, F., D. Tello, F. Goldbaum, R. Mariuzza, and R. Poljak. 1995. Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies. *Eur. J. Biochem.* 228:388–394.

Steinbach, P., and B. Brooks. 1993. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 90:9135–9139.

Vajda, S., Z. Weng, R. Rosenfeld, and C. DeLisi. 1994. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry.* 33:13977–13988.

Vakser, I., and C. Aflalo. 1994. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins Struct. Funct. Genet.* 20:320–329.

Vakser, I., O. Matar, and C. Lam. 1999. A systematic study of low-resolution recognition in protein-protein complexes. *Proc. Natl. Acad. Sci. USA.* 96:8477–8482.

Zanotti, J., M. Bellisent-Funel, and J. Parello. 1999. Hydration-coupled dynamics in proteins studied by neutron scattering and nmr: the case of the typical ef-hand calcium-binding parvalbumin. *Biophys. J.* 76: 2390–2411.

Zauhar, R., and R. Morgan. 1985. A new method for computing the macromolecular electric potential. *J. Mol. Biol.* 186:815–820.

Zauhar, R., and R. Morgan. 1990. Computing the electric-potential of biomolecules: application of a new method of molecular-surface triangulation. *J. Comp. Chem.* 11:603–622.

Zhang, C., G. Vasmatzis, J. Cornette, and C. DeLisi. 1997. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* 267:707–726.