Restricted Mobility of Conserved Residues in Protein-Protein Interfaces in Molecular Simulations

Osman N. Yogurtcu,* S. Bora Erdemli,* Ruth Nussinov,[†‡] Metin Turkay,* and Ozlem Keskin*

*Koc University Center for Computational Biology and Bioinformatics and College of Engineering Rumelifeneri Yolu, Istanbul, Turkey; [†]Basic Research Program, SAIC-Frederick, Inc. Center for Cancer Research Nanobiology Program, The National Cancer Institute at Frederick, Frederick, Maryland; and [‡]Sackler Institute of Molecular Medicine Department of Human Genetics and Molecular Medicine Sackler School of Medicine Tel Aviv University, Tel Aviv, Israel

ABSTRACT Conserved residues in protein-protein interfaces correlate with residue hot-spots. To obtain insight into their roles, we have studied their mobility. We have performed 39 explicit solvent simulations of 15 complexes and their monomers, with the interfaces varying in size, shape, and function. The dynamic behavior of conserved residues in unbound monomers illustrates significantly lower flexibility as compared to their environment, suggesting that already before binding they are constrained in a boundlike configuration. To understand this behavior, we have analyzed the inter- and intrachain hydrogen-bond residence-time in the interfaces. We find that conserved residues are not involved significantly in hydrogen bonds across the interface as compared to nonconserved. However, the monomer simulations reveal that conserved residues contribute dominantly to hydrogen-bond formation before binding. Packing of conserved residues across the trajectories is significantly higher before and after the binding, rationalizing their lower mobility. Backbone torsional angle distributions show that conserved residues assume restricted regions of space and the most visited conformations in the bound and unbound trajectories are similar, suggesting that conserved residues are preorganized. Combined with previous studies, we conclude that conserved residues, hot spots, anchor, and interface-buried residues may be similar residues, fulfilling similar roles.

INTRODUCTION

Proteins function cooperatively. The sites through which they associate are believed to contribute to the recognition and binding of proteins by providing specific chemical and physical properties (1–4). Protein interfaces have been analyzed with respect to size, shape, hydrophobicity, amino-acid propensity, segmentation, secondary structure and complementarity (3,5); function, and the cellular pathways in which they are found (6,7). Some characteristic differences have been observed between various types of interfaces: homodimers are better packed than other types of interfaces, and the changes in the accessible surface areas upon complexation for the homodimers are known to be larger than for heterocomplexes; interfaces of obligate complexes are more hydrophobic; transient complexes, especially enzyme-inhibitors, have polar surface patches and are more hydrophilic. In both transient and obligatory interfaces, most of the interfaces are relatively planar and accessible, whereas homodimers and enzyme-inhibitor interfaces are not as planar as other interfaces. In addition to these static physicochemical differences, flexibility is a key factor that allows optimization of the interfaces for better packing and electrostatic complementarity.

The binding free energy of protein-protein association is unevenly distributed across the interfaces (1,8) with some regions and individual amino acids contributing dominantly (1,9,10). These hot-spot residues have been defined as residues adding to the free energy of binding >2 kcal/mol. The characteristics of hot spots have been discussed previously (1,3,8–10). Kinetic analyses of mutagenesis experiments provide clues to the role played by individual residues in protein binding. Replacement of the hot spots may enhance or hinder protein recognition. Experimental and computational methods have been developed for predicting interface hot spots. Experimentally, alanine scanning measures the effect of mutating interface residues on the stability of the binding. Computationally, Kortemme and Baker (11) developed a model which includes hydrogen bonds, implicit solvent and packing interactions and ignores changes in backbone conformation or effects on the dynamic interface. Hu et al. and Ma et al. carried out structural comparisons of 11 interface families (12,13), observing that structurally conserved residues strongly correlate with the experimental hot spots, consistent with evolutionarily conserved residues being critical for the function and stability of the complexes (8,13–16). Keskin et al. observed that hot spots from both interface sides cluster in densely packed hot regions (9).

It has long been recognized that formation of a complex between two proteins, and between proteins and DNA, RNA or small molecules may lead to an increase in configurational entropy, which is related to an increase in the flexibility of the system. As proposed by Steinberg and Scheraga (17) more than 40 years ago, this increase may compensate for the loss

Submitted June 12, 2007, and accepted for publication November 7, 2007.

Osman N. Yogurtcu and S. Bora Erdemli contributed equally to this article.

Address reprint requests to O. Keskin, Tel.: 90-212-338-1538; E-mail: okeskin@ku.edu.tr; or R. Nussinov, Tel.: 301-846-5579; E-mail: ruthn@ncifcrf.gov.

Editor: Kathleen B. Hall.

0006-3495/08/05/3475/11 $2.00

doi: 10.1529/biophysj.107.114835

of translational and rotational entropy upon association (18–20). Rajamani et al. (21) have proposed that the mechanism for molecular recognition requires one of the interacting proteins, usually the smaller of the two, to anchor a specific side chain in a structurally constrained binding groove of the other protein, providing a steric constraint that assists in stabilizing a nativelike bound intermediate. They performed 11 molecular dynamics (MD) simulations and suggested that one or a few key anchor residues frequently visit their bound state and that these residues are critical in early recognition. Kimura et al. (22) explained that specific side chains act as ready-made recognition motifs by having nativelike bound conformations before an association with the receptor. Recently, energetic hot-spot residues have been observed to frequently locate themselves in complemented pockets (23). Further, these pockets preexist binding. In 16 out of the 18 analyzed complexes, the root mean-squared deviations (RMSD) of the atoms lining these pockets between the bound and unbound states are as small as 0.9 Å, implying that the unbound and bound forms of the pockets are very similar. Smith et al. (24) studied the extent to which the conformational fluctuations of proteins in solution reflect the conformational changes that they undergo when they form binary protein-protein complexes. They observed that some residues in the unbound state that get buried in the core of the interfaces upon binding are less mobile, and during the simulations assume conformations similar to those in the bound state. These articles are very valuable in understanding the mechanisms of protein-protein interactions.

Here we study protein associations focusing on the role of conserved residues in interfaces (9,12,13) before and after association. The complexed and monomeric forms of 15 protein complexes are analyzed using molecular dynamics simulations, leading to 39 simulations. Additional analytical methods have been applied to study the dynamic behavior of these complexes as well as a set of 11 protein complexes obtained from Protein-Protein Docking Benchmark 2.0 (25). Similar to Kimura et al. (22) and Smith et al. (24), we target certain residues. While they analyzed interface anchor and core residues, respectively, we focus on evolutionarily conserved residues. The complexes are diverse with respect to size, shape, and type. We chose three types: homodimers, enzyme-inhibitors, and antibody-antigens. We have investigated whether conserved (C) residues are restricted in their mobility as compared to nonconserved (NC) residues in the interfaces of these complexes. We observed that side-chain flexibilities of C residues are significantly lower than those of NC residues. The lower mobility of C residues suggests that these residues have favorable potential energy despite their low entropy. Across the interfaces, hydrogen-bonding ability does not differ between C and NC residues when solvent molecules are considered. On the other hand, C residues contribute dominantly in H-bonding within the monomers before binding, making them less mobile. The accessible surface areas (ASAs) of the C residues are significantly lower than of NC residues in all complex types. Therefore, once complexation occurs these residues change their role in stabilization from forming hydrogen bonds to forming highly dense networks of Van der Waals interactions. The backbone torsional angle distributions of the C and NC residues indicate that the conformational space sampled by C residues is much more restricted in the complex simulations. The bound and unbound simulation trajectories indicate that the most visited conformations are similar. Modular interface architectures (26) and densely packed hot regions (9) have been shown to exist across the interfaces. Hence, similarly constrained dynamic behavior of C residues and interface regions suggest a preorganization of the C residue populated modules already in the unbound state.

MATERIALS AND METHODS

A schematic representation of the procedure followed in this study is illustrated in Fig. 1. Molecular dynamics simulations of 15 complexes and their individual monomers are performed to analyze the role of conserved residues in the binding regions. We mainly focus on the flexibility of side chains in the binding region before and after the formation of the complexes.

Complexes analyzed

The selection of the complexes aims to generate an ensemble of protein interfaces with diverse size, shape, and type from our recently compiled dataset of nonredundant protein-protein interfaces (27). We further filtered the proteins that have no homologs so that we will be able to obtain conservation scores for the proteins in our set. The complexes are classified into three groups with six homodimers, five enzyme-inhibitors, and four antibody-antigens. The three-dimensional structures of the complexes were obtained from the Brookhaven Protein Data Bank (PDB) (28). These are mainly dimers except for the three antibody-antigens (PDB IDs 1vfb, 1ddh, and 1qnz) that are trimers in the PDB. The list of protein complexes and the chain IDs that form the complexes are given in Table 1. The last column gives the total number of residues in the two chains of the complexes. The prosthetic groups are removed before the simulations. In addition, some of the complexes have

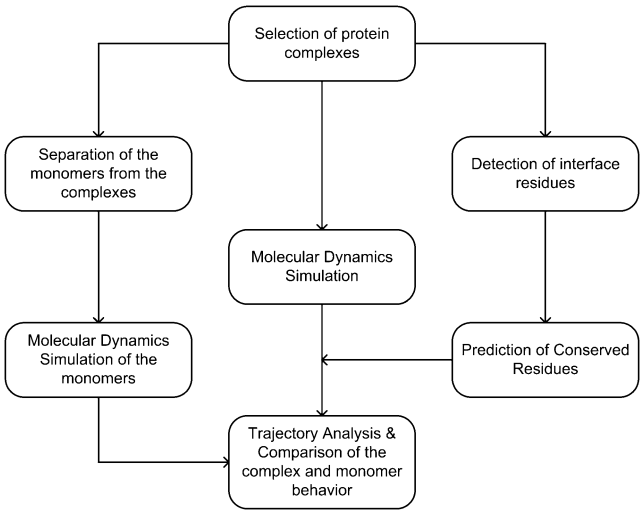

FIGURE 1 A flow chart of the procedure.

TABLE 1 Protein complexes

PDB ID	Chain IDs	Description	Complex type	Interface size	Number of conserved residues	Number of residues
1bft	A B	Nuclear factor	Homodimer	33	16	202
1bjr	E I	Hydrolase, lactoferrin	Enzyme-inhibitor	37	8	289
1bzd	A B	Binding protein	Homodimer	44	20	254
1cd0	A B	Immunoglobulin	Homodimer	44	14	222
1cjq	A B	Ribonuclease, peptide	Enzyme-inhibitor	34	13	116
1cqk	A B	Immune system	Homodimer	54	6	202
1ddh	A P	Histocompatibility, antigen	Antibody-antigen	45	4	284
1fcc	A C	Immunoglobulin, antigen	Antibody-antigen	35	4	262
1mr8	A B	Metal transport	Homodimer	67	12	186
1qnz	H P	Anti-HIV antibody complex	Antibody-antigen	28	1	137
1sbw	A I	Hydrolase, inhibitor	Enzyme-inhibitor	41	8	238
1ugh	E I	Glycosylase, inhibitor	Enzyme-inhibitor	64	15	305
1vfb	B C	Antibody, lysozyme	Antibody-antigen	24	3	245
1x2i	A B	Helicase	Homodimer	55	2	134
2sni	E I	Subtilisin, inhibitor	Enzyme-inhibitor	40	8	358
Total				645	134	3434

The first column lists the PDB IDs of the protein-protein complex. The second column gives the chains which are involved in the interfaces. The third and fourth columns give the descriptions of the complexes and the group to which they belong. The fifth and sixth columns correspond to the numbers of interface residues and conserved residues located in these interfaces. The last column gives the number of residues in the two-chain complexes. Note that ''1cjq'' is classified as an enzyme-inhibitor complex although it is actually an enzyme-peptide complex.

disulphide bridges. The list of the removed ions and disulphide bridges is given in the Table S5. Two of these complexes are nonbiological (crystal contact) complexes according to NOXCLASS (noxclass.bioinf.mpi-inf.mpg.de). The PDB IDs for the crystal complexes are 1bzd and 1bft. We further analyzed 11 complexes from Protein-Protein Docking Benchmark 2.0 (25) as listed in Table S1. The list includes five rigid-body, three medium-difficulty, and three difficult cases.

Determination of Interface residues

NACCESS (http://wolf.bms.umist.ac.uk/naccess/) was used to calculate the accessible surface areas (4) (ASAs). Residues losing 1 $Å^2$ of ASA after complexation were taken as interface residues (8). The interface between chains A and B of human λ-6 light chain dimer (pdb id: 1cd0) is highlighted in Fig. 2 *A*. The interface residues are colored cyan; the two chains are colored yellow and gray.

Detection of conserved residues

C residues are mapped using CONSURF (Ver. 3.0) (29). The program identifies hot spots by their conservation during evolution. The continuous conservation scores calculated by CONSURF are discretized into nine bins for every protein, such that the evolutionary conservation of a residue increases with this score. CONSURF score should be highest among the interface residues (CONSURF scores 8 or 9), complex residue ASA should be <62 $Å^2$, and the change in ASA between the bound and unbound forms should be >5 $Å^2$ (8). These additional filters are applied to assure that conserved residues act as energy hot spots (8). For the antibody-antigen complexes, we ease the criterion on conservation scores (residues with CONSURF score larger than five were taken). We have used the Bayesian method with default parameters to extract the conservation scores for each interface residue. Fig. 2 *B* displays the positions of the C residues in human λ-6 light chain dimer (pdb id: 1cd0) as red and purple spheres. These residues cluster in two regions consistent with the hot-region proposition (9).

Molecular dynamics

Molecular dynamics (MD) simulations are performed for the 15 complexes and their individual monomers separately leading to 39 simulations (six of the monomers are peptides and do not have stable conformations on their own). The monomers are extracted from the complexed forms. Some of these monomers are available in unbound forms in the PDB. When we exclude antibodies (since an antibody structure will always be in a complex formed by a light and heavy chain), two crystal packing complexes, and peptides, 24 proteins are searched in the PDB to see if their monomers exist. We found that 16 exist in other entries: nine exist in a monomeric structure in the PDB; seven exist in other complex structures and no data are found for the eight monomers (three are homodimers leading to six monomers). The root mean-squared deviation (RMSD) values of the monomers in the unbound form and the complexes are given in Table S6. Considering all heavy atoms, the largest RMSD difference (1.73 Å) is between chain C of 1vfb (monomer structure separated from the complex) and 132l (native monomeric structure). Fig. S1 illustrates the structures of two proteins in bound and unbound forms. These results show that the cases discussed here do not show large conformational changes upon complexation. Since the RMSD values do not show large

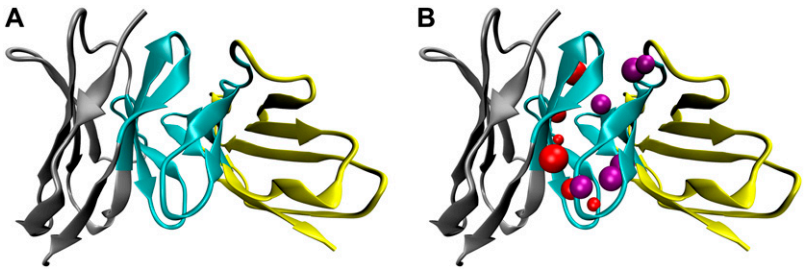

FIGURE 2 (*A*) Ribbon diagram of a complex formed by two proteins (*gray* and *yellow chains*) that represents the structure of an immunoglobulin, human λ-6 light chain dimer (pdb ID: 1cd0). The interface region between the two monomers is shown in cyan. (*B*) The C residues in the yellow chain are shown as purple spheres, whereas the ones in the gray chain are red spheres. The residue numbers of these C residues are Q38, S42, Y49, P55, Y87, and Y91.

conformational changes, we carried out all MD simulations with the structures separated from the complex forms.

The simulations were performed with NAMD (30) with CHARMM27 (31) force-field parameters for 6 ns. The protein complexes were solvated using the TIP3 water molecules in the VMD (32) package. The total number of atoms and water molecules, and the size of the boxes in each of the protein system, are given in the Table S7. Particle-mesh Ewald (33) was applied in the simulations. We used the VMD Autoionize, adding sodium and chlorine ions, to neutralize the system. NVT ensemble and periodic boundary conditions with a rectangular box were applied in the simulations. The temperature in the simulations was kept constant at 300 K by using Langevin dynamics. Initial equilibration was done for 10,000 steps, followed by 6-ns runs. The time step was 2 fs. The first 1-ns runs were further discarded to assure that the data collected are after equilibration. Trajectories were sampled at 40-ps intervals. The simulations were carried out in a Linux-based cluster from a Racksaver cluster and each node has two 3.06 GHz Intel Pentium Xeon processors and Beowulf Cluster with nodes having Intel Pentium 4 2.4 GHz processors.

Gaussian network model (GNM)

Gaussian network model (GNM) is a simple method to study the equilibrium fluctuations of proteins (34). This model assumes that the protein in the folded state is equivalent to a three-dimensional elastic network. Therefore, the protein structure is modeled as a chain of N beads (residues) connected by $N-1$ springs. The beads are subject to harmonic potentials from all neighboring beads within a cutoff distance ($R_c = 7.3$ Å, in this study) regardless of backbone connections. The topology of the network is recorded in an $N \times N$ Kirchoff matrix, Γ, where the off-diagonal elements are -1 if the nodes are within a cutoff distance, R_c, and zero otherwise (34–36). The diagonal elements represent the coordination number of each residue. Assigning a uniform spring constant, γ, the cross-correlations between the fluctuations ΔR_i and ΔR_j of residues i and j are evaluated as

$$\langle \Delta R_i \cdot \Delta R_j \rangle = 3\, k_B T/\gamma [\Gamma^{-1}]_{ij}, \quad (1)$$

where k_B is the Boltzmann constant, T is the absolute temperature, and $[\Gamma^{-1}]_{ij}$ is the ij^{th} element of the inverse of Γ (34,35). Setting j equal to i in the above equation, we obtain the mean-square fluctuations of residue i, $\langle(\Delta R_i^2)\rangle$.

The equilibrium dynamics of the structure results from the superposition of $N-1$ nonzero modes found by the eigenvalue decomposition of Γ. The elements of the k^{th} eigenvector, u_k, describe the displacements of the residues along the k^{th} mode coordinate associated with its frequency (eigenvalue, λ_k) where $1 \le k \le N-1$. The contribution of the k^{th} mode to the mean-square fluctuations of residue i is

$$\langle(\Delta R_i^2)\rangle_k = (3\, k_B T/\gamma)\,(1/\lambda_k)\,(u_k u_k^T)_{ii}, \quad (2)$$

where $(u_k u_k^T)_{ii}$ designates the i^{th} diagonal element of the Γ matrix. The first three dominant modes ($1 \le k \le 3$) are used in this study to represent the extent of fluctuations of the residues.

The GNM is applied to the 15 complexes in our list and 11 complexes in the benchmark complexes. The detailed results are provided in the Supplementary Material.

Propensities of conserved residues

Propensities are calculated for C residues for each of the three complex types according to

$$P_i^* = \frac{\dfrac{n_i^*}{N_i^*}}{\dfrac{n}{N}}, \quad (3)$$

where n_i^* is the number of residue type i C residues in the chains, N_i^* is the number of residue type i in the chains, n is the total number of C residues, and N is the total number of residues in the chain belonging to a particular complex type.

H-bond definitions

The hydrogen bonds formed by interface residues (between the two complementary proteins) were monitored using default cutoff distance (3.5 Å) and bond angle (30°) formed between donor and acceptor atoms using VMD (32).

RESULTS AND DISCUSSION

Interface residues and conserved residues

C residues in protein interfaces are identified as detailed in the Materials and Methods. Table 1 lists the protein complexes: the first two columns give the PDB IDs and the chain IDs, the last three columns list the corresponding number of interface residues and C residues and the total residue numbers on both chains of the complexes, respectively. A total of 134 C residues are found among 645 interface residues as listed. The ranges of interface residues and C residues per complex are between 24–67 and 1–20 residues, respectively. We observe no correlation between the interface sizes and the number of C residues. Fig. 2 *A* displays the interface between chains A and B of human λ-6 light chain dimer (1cd0) as an example. Fig. 2 *B* shows the positions of C residues in the interface. These C residues are shown as spheres whereas the rest of the protein complex is in a cartoon representation.

Analysis of interface residues and conserved residues: Is the mobility distributed homogeneously in the interfaces?

The results of 5-ns MD simulations are analyzed for each of the complexes and their constituent binary proteins. Our goal is to determine whether there is a difference in the dynamic behavior of C residues and NC residues before and after complex formation; and if there exists one, possible reasons. We have assessed the flexibility of the residues in the interfaces during the simulations. First, the average RMSD of each residue in the interface is calculated over the entire simulation time. Following Rajamani et al.'s strategy (21), before calculating the residue side-chain RMSD values, all heavy backbone atoms (N, C^{α}, C, O) of the interface residues are aligned with the initial structure at the beginning of the simulations to avoid systematic errors caused by translational motions. Side-chain RMSD values are obtained (both from the complex and the monomeric simulations) by comparing each frame during the simulations with the structure at the beginning of the simulations after the equilibration step. We analyzed 15 complexes and their monomers. For 11 out of the 15, the superimpositions of backbone atoms display low RMSD values between their initial ($t = 1$ ns) and the final ($t = 6$ ns) structures during the simulations (RMSD values <3.86 Å), although for four other complexes (1cd0, 1mr8, 1cqk, and 1ugh), the RMSD values are higher. These interfaces display

breathing motions with RMSD values between 7 and 10 Å (data not shown).

Side-chain RMSD values provide a measure of the residue movement during the simulation. We expect that the C residues would have less displacement from their initial coordinates compared to NC residues in the interfaces, as Kimura et al. suggested for anchor residues (22). To test this hypothesis, we calculated the RMSD values of C and NC residues for the homodimers, antibody-antigen, and enzyme-inhibitor complex-types, separately. Table 2 summarizes the average side-chain RMSD values of C and NC residues for the 15 protein interfaces. The first and second columns list the PDB IDs of the complexes and the chains involved in forming the interfaces. The third and fourth columns in this table give the average RMSD values for C and NC residues in the interfaces of the complexes, respectively. Some complexes accommodate C residues only in one side of their interfaces. Therefore, a single row for a PDB ID suggests that the complex does not have C residues at both interface sides (for example, 1vfb has C residues only in its C chain). The antibody-antigen and enzyme-inhibitor complexes also do not have C residues in their antigen and inhibitor chains. Table 2 suggests that the side-chain flexibility is different for C and NC residues.

TABLE 2 Overall averages of RMSD of conserved residues and nonconserved residues for all complexes

		Complex simulations		Monomer simulations	
PDB ID	Chain	C	NC	C	NC
Homodimers					
1bft	A	2.00	1.65	3.46	2.95
	B	1.97	2.11	2.39	2.70
1bzd	A	1.96	1.56	2.41	1.90
	B	1.87	1.66	2.11	1.93
1cd0	A	2.92	3.42	2.33	2.27
	B	3.28	3.61	2.71	3.44
1cqk	A	1.70	2.58	2.02	2.74
	B	1.61	2.48	2.34	2.97
1mr8	A	2.07	2.41	2.48	2.33
	B	2.10	2.37	3.01	3.38
1x2i	A	1.18	1.76	2.72	4.22
	B	1.65	1.87	3.64	3.65
Average		2.06	2.56	2.66	3.12
SD		0.66	0.61	0.47	0.63
Antibody-antigen					
1ddh	A	1.60	1.92	2.26	1.92
	P	—	—	—	—
1fcc	A	1.41	1.6	1.87	2.42
	C	—	—	—	—
1qnz	H	1.36	1.67	1.41	2.20
	P	—	—	—	—
1vfb	B	—	—	—	—
	C	1.09	2.02	2.38	2.99
Average		1.36	1.80	1.98	2.39
SD		0.18	0.17	0.38	0.39
Enzyme-inhibitor					
1bjr	E	1.61	1.91	1.68	1.93
	I	—	—	—	—
1sbw	A	2.48	2.66	2.24	2.41
	I	—	—	—	—
1ugh	E	2.04	2.78	2.78	3.36
	I	—	—	—	—
2sni	E	1.56	2.04	1.99	2.70
	I	—	—	—	—
1cjq	A	—	—	—	—
	B	3.81	4.24	1.95	2.79
Average		2.30	2.73	2.13	2.64
SD		0.83	0.83	0.37	0.47

C, conserved residue; *NC*, nonconserved residue. The results of the complex and monomer simulations are given.

We have further investigated the results of the simulations from the individually separated monomers of the complexes. Only the proteins that have C residues on their interfaces are analyzed, and peptides and inhibitors which are devoid of C residues are not investigated. Overall, as expected the residues in unbound proteins are more mobile compared to the bound cases. When the flexibility differences are inspected for the component monomeric proteins as well as the complexes, the differences are pronounced in all three types, antibody-antigen, enzyme-inhibitor, and homodimer complexes. Fig. 3 *A* summarizes the RMSD results for the complexes and monomers studied. The first and last three data points correspond to complex and monomer simulations. Solid circles and shaded squares are for C and NC residues, respectively, with corresponding error bars. This figure clearly shows that there is a difference between the mobility of the two types of residues. The average RMSD of C and NC residues in enzyme-inhibitors are 2.30 and 2.73 Å. The corresponding RMSDs are 2.06 Å and 2.56 Å for homodimers. Antibody-antigens have a flexibility difference between their C and NC residues of 1.36 Å vs. 1.80 Å for the complex simulations. The values are 2.66 Å vs. 3.12 Å for homodimers, 1.98 Å vs. 2.39 Å for antibodies, and 2.13 Å vs. 2.64 Å for enzymes (Table 2, the last two columns). We should note that two crystal contact (nonbiological) complexes (PDB IDS 1bft and 1bzd) are also investigated but not included in the analyses. No such flexibility difference is observed there between C and NC residues. The C residues in these complexes are as mobile as the NC residues. This further validates that not all C residues on protein surfaces act as those in the interfaces.

These results show that evolutionarily conserved residues are less flexible before the binding, consistent with the proposition of anchor and core residues of Smith et al. (24) and Rajamani et al. (21). The low flexibility of some residues may be a result of steric constraints which may further help in stabilizing nativelike bound intermediates in the unbound state, facilitating the binding of the two partner proteins. Here, we further show that these residues continue to be restricted after complex formation. Therefore, these residues can be seen as the signatures of the binding pockets. The binding pockets are more constrained as compared to the other regions of the interfaces (21). Our observations from the extensive simulations extend the current findings: binding pockets are more constrained, and the C residues tend to cluster at the binding

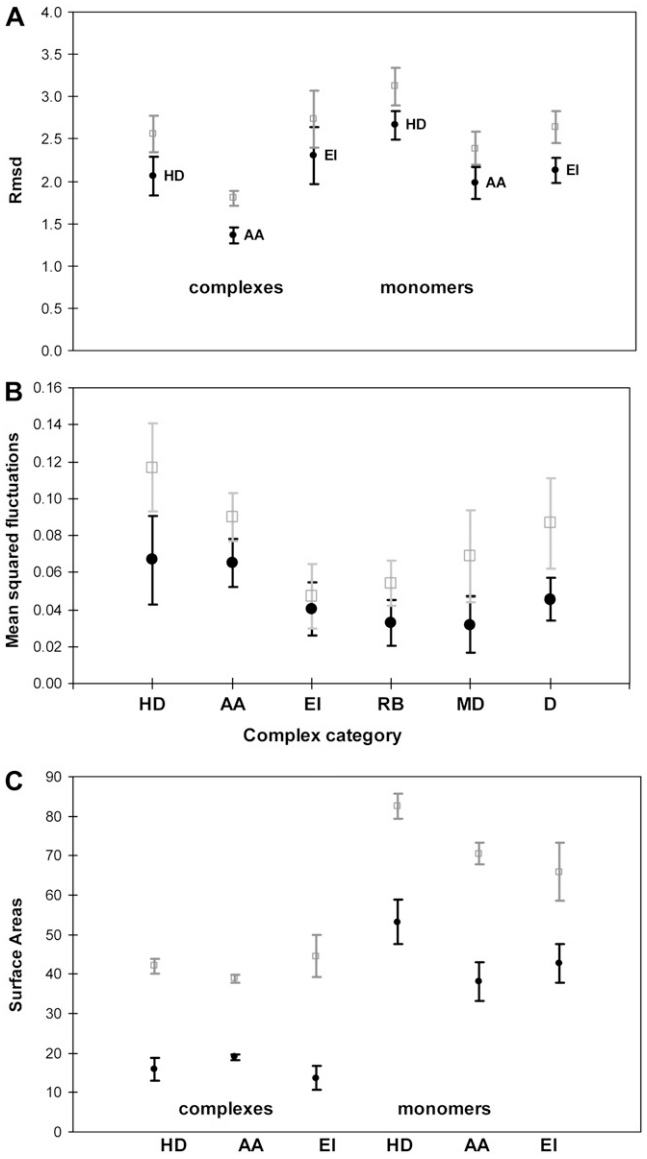

FIGURE 3 Averages of root mean-squared deviations, root mean-squared fluctuations, and surface area of C and NC-residues calculated along the trajectories. (*A*) Differences of averages of the residue RMSD values. The first and last three points correspond to the complexes and monomers. The shaded squares and solid circles are for NC and C residues, respectively. (*B*) Differences between the C and NC RMSFs for homodimers (*HD*); antibodies (*AA*); enzymes (*EI*); rigid-body (*RB*); medium-difficulty (*M*); and difficult (*D*) cases, respectively. (*C*) Differences between the surface areas of C and NC residues for HD, AA, and EI complexes. The first and last three points correspond to complex and monomers, respectively.

pockets of the interfaces. This suggests that nature attempts to minimize the entropy penalty during protein-protein association. If the binding were to occur through a flexible region, upon binding the entropy of the new system would dramatically decrease; on the other hand, for a rigid region, the entropy change would not be as high and the penalty would be optimized. These also suggest that at least in the cases studied here, entropy is a major component in binding.

Similarly, the dynamic characteristics of protein complexes are analyzed using the GNM (34,35). GNM was applied to all complexes in Table 1 as well as the 11 complexes and their monomers obtained from Protein-Protein Docking Benchmark 2.0 (25). The list of the latter proteins is given in Table S1. The results for both sets of complexes consistently show that the conserved residues move significantly less compared to the rest of the interface residues both before and after complexation. A summary of the GNM results is given in Fig. 3 *B* for the monomers. The first three data points correspond to the averages of homodimers (HD), antibodies (AA), and enzymes (EI), respectively. The last three points are for the rigid-body (RB), medium-difficulty (M), and difficult (D) cases from the benchmark, correspondingly. Solid circles and shaded squares represent the C and NC residues. Error bars are added to the figures to guide the eye. As seen in the figure, in all cases the C residues display less flexibility compared to NC residues, consistent with MD simulation results. The detailed results for individual complexes are given in Tables S2 and S3. Homodimers exhibited average normalized root mean-squared fluctuation values of 0.034 and 0.044 for C and NC residues (in the complex), respectively. Similarly, antibodies have 0.070 and 0.072, respectively. And enzymes are observed to have 0.019 and 0.026 for C and NC residues (Table S2). Thus, the docking benchmark was studied for three cases: rigid-body, medium-difficulty, and difficult cases; and in all the cases, C residues showed less fluctuations compared to the NC residues (Table S3).

Fig. 4 *A* shows the positions of the conserved residues in Uracil-DNA Glycosylase in complex with a peptide inhibitor (PDB ID: 1ugh) (37). These C residues are shown in red whereas the rest of the protein complex is yellow. Glycosylases cleave the glycosidic bonds in DNA. The inhibitor in the complex mimics the structure of DNA so that the enzyme could not catalyze hydrolysis. There are 18 residues on Uracil-DNA Glycosylase that are important in binding to the inhibitor. Of these 18 residues, 11 have been defined as conserved by our method. From Table 2, we observe that overall the C residues are less mobile than NC residues (RMSD values 2.04 Å and 2.78 Å, respectively). Val^{274} is the least mobile NC residue with RMSD value 0.74 Å (data not shown). The rigidity of this residue suggests an immobilization effect of a nearby cluster of C residues, Ser^{270}, Pro^{271}, Leu^{272}, and Ser^{273}. The C residues Gln^{144}, His^{148}, and Ser^{169} of chain E form strong hydrogen-bond pairs with Leu^{23}, Ser^{21}, and Glu^{20} of inhibitor (I) chain, respectively. The active site residue Leu^{272}, which is also a C residue, makes hydrophobic interactions with the residues Val^{32} and Met^{56} of the inhibitor chain (Fig. 4 *B*).

Does hydrogen bonding play a role in the lower flexibility of conserved residues?

To understand the origin of the lower flexibility of C residues in the interfaces, we investigate the hydrogen-bond forming

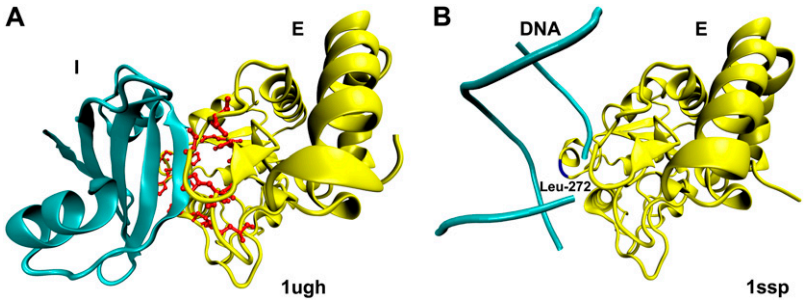

FIGURE 4 (*A*) Ribbon diagram of an enzyme-inhibitor complex (pdb ID: 1ugh) that represents the structure of Uracil-DNA Glycosylase (*yellow*) in complex with a peptide inhibitor (*cyan*). The C residues (Q144, D145, Y147, H148, Q152, P167, P168, S169, A214, S216, H268, S270, P271, L272, and S273) are shown in red. (*B*) Ribbon diagram of Uracil-DNA Glycosylase in complex with a short DNA fragment (pdb ID: 1ssp). Protruding hydrophobic active site residue Leu-272 is shown in blue.

ability of the C and NC residues across the interfaces by analyzing the H-bond residence times in the interfaces. The number of hydrogen bonds formed by each residue in the interface is averaged over all frames during the simulations. These numbers were further normalized by the number of C and NC residues in the respective interface. To get the C residue H-bond enrichment, we divide the contribution of C residues by NC residues: (Enrichment $= H_C/H_{NC}$, where H_C and H_{NC} are the H-bonding capability of the C and NC residues). For the complex simulations (Table S4), the average enrichments are >1, which might indicate that hot-spots are favored to form H-bonds. However, no consistent trends are observed in individual proteins. Thus, there is no significant difference between the C and NC residues in the H-bond forming ability after complex formation. On the other hand, when water molecules are considered, H-bonds formed by conserved residues are consistently underutilized. This is consistent with the work of Keskin et al. over a much larger dataset of two-chain interfaces (9). For example, histocompatibility/antigen complex (1ddh) has four computationally identified hot-spots as shown in Fig. 5 *A*. When we examine the hydrogen bonds formed between the antibody and antigen, we observe that out of eight hydrogen bonds only one is formed by a C residue, Tyr^{159} (Fig. 5 *B*).

Table 3 shows the results for H-bonding enrichments in the separated monomers. The first two columns give the PDB and chain ID of the interfaces. The third and fourth columns list the values when water molecules are included and excluded, respectively, in the analysis. This table clearly shows that in all cases, C residues are less involved in H-bonding when both protein and water atoms are considered. On the other hand, when only H-bonds between the protein atoms are counted, more H-bonds involving C residues are observed (enrichments >1). These results confirm that conserved residues are largely protected from water molecules and have a stabilizing effect in forming the intraprotein H-bonding network. In summary, C residues are shielded from water molecules, thus do not form a significant number of H-bonds with the water molecules in the interfaces. Although the averages are not very dissimilar, the difference between the number of H-bonds in the presence and absence of water is statistically significant (a p-value of 0.001 is obtained after a t-test with an α-value of 0.05). Levy et al. (38) emphasize the role of water, proposing that water molecules actively participate in molecular recognition. They mediate the interactions between binding partners and contribute to either enthalpic or entropic stabilization. Here, we observe that water molecules make a significant difference in H-bonds.

These results suggest that

1. In the complexes, the C residues do not show increased H-bond formation across the interfaces; and that
2. They contribute in intrachain protein-atoms (with water exclusion) H-bonds in the monomer simulations; but on the other hand,
3. Once the complex is formed, their contribution is insignificant.

In agreement with these results, it is known that hot-spot residues are usually large (8), and regardless of the interface type are often aromatic (12). According to the O-ring proposition

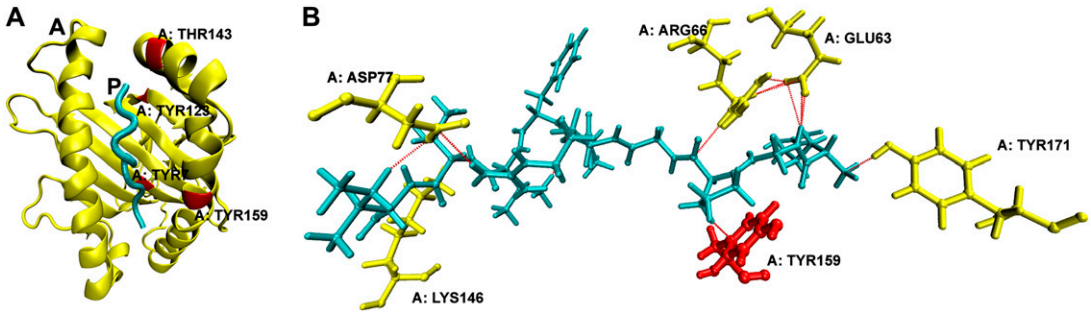

FIGURE 5 (*A*) Ribbon diagram of an antibody-antigen complex (pdb ID: 1ddh) that represents the structure of MHC Class I: Antibody Heavy Chain (*yellow*) in complex with Human Immunodeficiency Virus Envelope Glycoprotein 120 (gp120) antigen (*cyan*). The C residues (Y7, Y123, T143, and Y159) are shown in red. (*B*) The hydrogen bonds formed between the antigen and the antibody are shown in orange and the C residues are in red.

TABLE 3 H-bond formation enrichments in monomers

		Monomers	
PDB ID	Chain	With water	Without water
Homodimers			
1cd0	A	0.99	1.89
	B	1.33	2.31
1cqk	A	0.91	1.73
	B	0.90	1.38
1mr8	A	1.23	1.01
	B	1.26	1.21
1x2i	A	0.40	0.96
	B	0.46	0.72
Average		0.94	1.40
SD		0.35	0.54
Antibodies			
1ddh	A	0.93	0.95
	P		
1fcc	A	0.96	1.44
	C		
1qnz	H	0.61	0.79
tek	P		
1vfb	B		
	C	1.26	1.55
Average		0.94	1.18
SD		0.27	0.37
Enzymes			
1bjr	E	0.79	0.68
	I		
1sbw	A	0.71	1.73
	I		
1ugh	E	0.66	0.89
	I		
2sni	E	0.97	1.00
	I		
1cjq	A		
	B	1.05	1.35
Average		0.84	1.13
SD		0.17	0.41

(8), the hot spots are shielded from bulk solvent. The dielectric constant around the hot spots decreases, strengthening the electrostatic effect of the hot spots and their environment.

Surface areas (SAs) of the interfaces and hot spots

It is logical to consider that packed residues will be restricted in their flexibility. Since we observe that C residues are less flexible, we also expect them to be buried. We have performed an analysis on solvent accessibility versus evolutionary information on the list of proteins studied in this work. Hence, we address the relevant question of whether the C residues in the interfaces of these complexes are, on average, more buried. The results show that there is a high correlation between being buried and being conserved when all protein residues are considered. However, this correlation reduces when considering only interface residues. For enzymes, 57 of the 67 (85%) conserved residues are buried (have SA <40 Å^2); for antibodies, 14 of the 19 (74%) conserved residues are buried; as for homodimers, 65 of the 84 (77%) conserved residues are buried. Similar analyses indicate that for enzymes, 57 of the 85 (67%) buried residues are conserved. Fourteen of the 54 (26%) buried residues are conserved for antibodies, and 65 of the 183 (36%) buried residues are conserved. Similarly, we computed the percentages of the nonconserved residues to be buried: we obtained 58%, 70%, and 44% for enzymes, antibodies and homodimers, respectively. These numbers suggest that the majority of the conserved residues are buried for all complex types. On the other hand, there is no clear trend for NC residues.

The interface sizes vary between 24 and 67 residues and the total accessible surface areas of the interfaces range between 795 Å^2 and 2428 Å^2. The interface SAs are calculated as the cumulative sum of the interface residue SAs in the complex. As expected, there is a linear correlation (with a correlation coefficient of 0.83) between the interface size and the surface areas of the interfaces. Comparison of the flexibilities (i.e., the average side-chain RMSD of each residue) and the SAs of the residues give a weak correlation for C residues (a correlation of 0.24). NC residues get more flexible as their exposure to the solvent increases (correlation coefficient of 0.50). Since C residues are more buried, there is no correlation between their flexibility and exposure to the

TABLE 4 Molecular dynamics trajectory cluster sizes of the proteins in the complex and of the monomer conformations

	Cluster sizes		RMSD values between the most observed interface clusters of
PDBID	Monomer	Complex	monomer and complex forms (Å)
1bft_A	18	6	1.32
1bft_B	14	9	1.58
1bjr_E	15	6	2.04
1bzd_A	14	7	1.78
1bzd_B	5	5	1.57
1cd0_A	9	8	2.02
1cd0_B	5	12	1.73
1cjq_B	7	20	1.83
1cqk_A	10	7	1.63
1cqk_B	10	6	1.53
1ddh_A	11	14	2.21
1fcc_A	16	17	1.78
1mr8_A	14	13	2.23
1mr8_B	19	11	2.51
1qnz_H	10	4	2.41
1sbw_A	15	6	2.46
1ugh_E	17	8	2.76
1vfb_C	14	9	1.90
1x2i_A	19	10	4.09
1x2i_B	18	11	3.27
2sni_E	14	8	1.33
Mean			2.09
SD			0.67

solvent. The same holds in the monomeric proteins. The SA values for C residues are smaller compared to NC residues.

Fig. 3 *C* illustrates the difference between SAs of C and NC residues. The average normalized accessible surface areas of C and NC residues are presented for all complex types. The open squares are for NC residues and the solid circles are for C residues. In all cases, C residues have lower SAs than the NC residues. To understand the significance of the difference between C and NC residues, paired *t*-test is applied on the C and NC residue SAs. A p-value of 0.001 was found well under an α-value of 0.05. Therefore, the difference is statistically significant for all types of interfaces. Enzyme-inhibitor complexes show more distinct differences between their C and NC residue ASA values. This might be due to the difference in interface size in these cases. Thus we conclude that highly packed C residues may contribute to H-bonds within the protein; upon complexation they are significantly involved to form tight Van der Waals interactions across the interface.

Do the unbound proteins assume bound conformations before binding?

We structurally clustered the conformations from the MD simulation trajectories of the monomer and complex forms of the interfaces and compared the clusters of the bound and unbound conformations by atomistic cross-RMSD values. To cluster the interfaces along time, we used the program *g_cluster*, which is available in the GROMACS MD-simulation package (39). A clustering cutoff radius of 1 Å is used to obtain the clusters. The number of clusters varies between 4 and 20 (as seen in Table 4). Fig. 6 *A* shows the trajectory-clusters formed by the monomer and the complex of chain B of 1bzd. We selected the most highly recurring interface clusters in the last 4 ns of the simulations for every complex and conformation type (in complex or monomer form). Fig. 6 *B* illustrates the most frequently observed clusters of 1bzd. The RMSD value between these conformations is only 1.57 Å. The most frequently observed clusters in the monomer and complex simulation trajectories of the proteins are compared in Table 4. The last column of Table 4 shows the cross RMSD values. The mean of the RMSD values for 21 chains is 2.09 Å. Therefore, it is reasonable to propose that molecular interfaces are preorganized in the bound conformation before the binding event.

We have further investigated the backbone torsional angle distributions of the C and NC residues in the interfaces along the trajectories. The backbone torsional angles of each snapshot along the trajectory are plotted in two histograms (one for ϕ- and one for ψ-angles). The results reveal that usually the C residue backbone conformations differ from the NC residues except Ala, Pro, and Thr. Fig. 7, *A–D*, shows the histograms of Ile and His. The left panel is for C and the right for NC residues. Fig. 7, *A* and *B*, shows the distributions of ϕ-and ψ-angles for Ile, respectively. The ϕ-angles assume a distribution between $-100°$ and $-40°$ for the C residues and

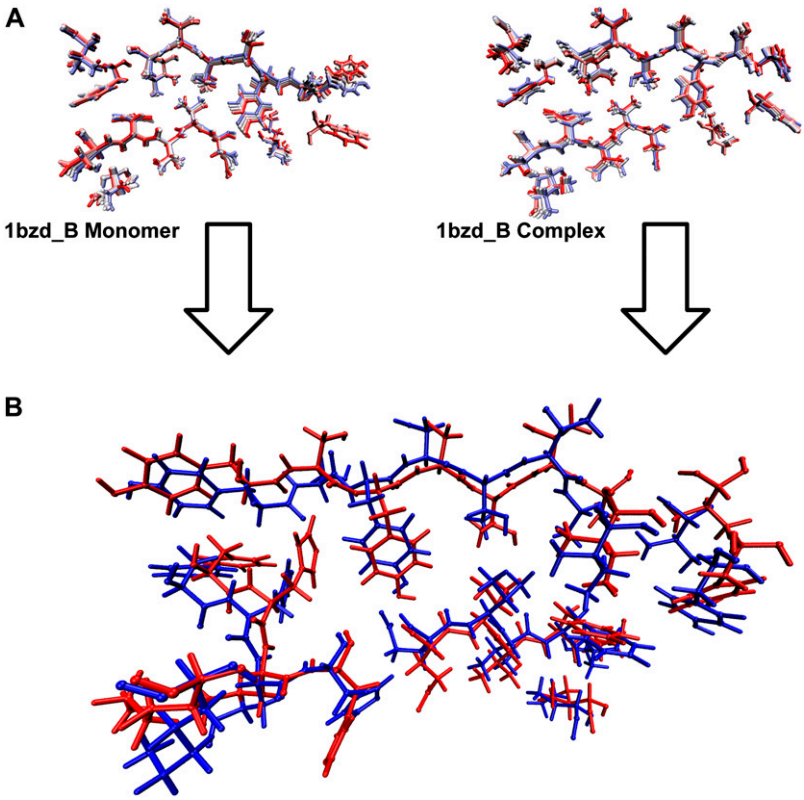

FIGURE 6 (*A*) The clusters from the MD simulation trajectory for the interface of chain B of 1bzd are shown, for the monomer and the complex form. There are five clusters in each. (*B*) The monomer conformation which is observed most frequently is in blue and its complex counterpart is shown in red. The RMSD value between these conformations is 1.57 Å.

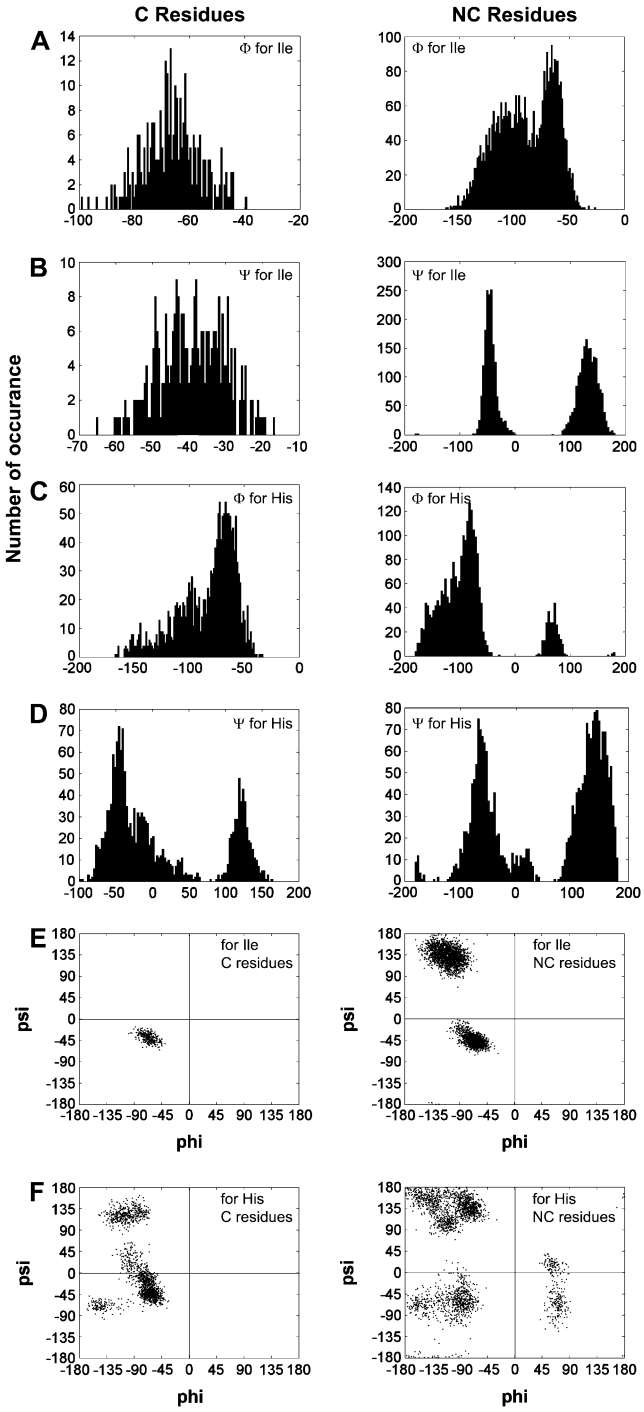

FIGURE 7 Backbone torsional angle distributions of the C and NC residues. The left and right figures are for C and NC residues, respectively. (*A* and *B*) The ϕ- and ψ-angle distribution for ILE; (*C* and *D*) ϕ- and ψ-angle distribution for His; and (*E* and *F*) Ramachandran maps for Ile and His, respectively.

$-160°$ and $-40°$ for the NC residues. Similarly, there are two peaks for NC residues in the ψ-distribution ($\sim -40°$ and 130°), but only one peak is observed for the C residues ($\sim -40°$). Fig. 7, *C* and *D*, displays similar histograms for His. We have further investigated the Ramachandran maps of C and NC residues. Fig. 7, *E* and *F*, displays the maps for Ile and His. It is clearly seen that C residues visit restricted regions of the map, whereas NC residues can assume a higher number of isomeric states.

CONCLUSIONS

Here we study the mobility of conserved residues in protein-protein interfaces with the goal of obtaining insight into their role in complex formation. Conserved residues have been shown to strongly correlate with residue hot spots. We carry out 39 explicit solvent molecular dynamics simulations of 15 protein-protein complexes and their monomers. We observe that throughout the simulations of the monomers the mobility of conserved residues in the interfaces is low, and the clustered conformations are remarkably similar to those observed in the simulations of the complexes. We further observe that the backbone torsional angle distributions differ significantly for conserved and nonconserved residues. Conserved residues are found to visit restricted regions during the simulations. The residence times of across-the-interface H-bonds formed by conserved residues are not significantly higher than those formed by other interface residues; however, there is an intramolecular H-bond enrichment when the water molecules are not considered. Conserved residues are observed to have higher burial, even during the simulations of the monomeric state. We conclude that conserved residues are preorganized in a bound-state-like conformation. Their lower mobility optimizes the entropic penalty term in the complex formation. Overall, the monomer and complex simulations show that conserved residues in interfaces are less mobile as compared to the rest of the interface residues. This may suggest that specific side chains in an interface are structurally constrained both before and after the binding takes place. Considering the results obtained by Rajamani et al. (21), Smith et al. (24), and Keskin et al. (9) on key anchor residues and on residues buried in interfaces, respectively, leads us to conclude that these three types reflect similar residues: anchor residues are conserved residue hot spots and they tend to be buried in interfaces.

SUPPLEMENTARY MATERIAL

To view all of the supplemental files associated with this article, visit www.biophysj.org.

We thank Drs. Hui-Hsu (Gavin) Tsai and Chung-Jung Tsai for assistance.

O.K. has been granted with Turkish Academy of Sciences Young Investigator Program (TUBA-GEBIP). This project has been funded in whole or in part with TÜBİTAK (research grant No. 104T504) and Federal funds from the National Cancer Institute, National Institutes of Health, under contract No. N01-CO-12400.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported (in part)

by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research.

REFERENCES

1. Clackson, T., and J. A. Wells. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science*. 267:383–386.
2. Young, L., R. L. Jernigan, and D. G. Covell. 1994. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci.* 3:717–729.
3. Janin, J., and C. Chothia. 1990. The structure of protein-protein recognition sites. *J. Biol. Chem.* 265:16027–16030.
4. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
5. Jones, S., and J. M. Thornton. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA.* 93:13–20.
6. Brun, C., F. Chevenet, D. Martin, J. Wojcik, A. Guenoche, and B. Jacq. 2003. Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network. *Genome Biol.* 5:R6.
7. Jansen, R., and M. Gerstein. 2004. Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction. *Curr. Opin. Microbiol.* 7:535–545.
8. Bogan, A. A., and K. S. Thorn. 1998. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280:1–9.
9. Keskin, O., B. Ma, and R. Nussinov. 2005. Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* 345:1281–1294.
10. Ma, B., H. J. Wolfson, and R. Nussinov. 2001. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr. Opin. Struct. Biol.* 11:364–369.
11. Kortemme, T., D. E. Kim, and D. Baker. 2004. Computational alanine scanning of protein-protein interfaces. *Sci. STKE.* 2004:pl2.
12. Hu, Z., B. Ma, H. Wolfson, and R. Nussinov. 2000. Conservation of polar residues as hot spots at protein interfaces. *Proteins.* 39:331–342.
13. Ma, B., T. Elkayam, H. Wolfson, and R. Nussinov. 2003. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. USA.* 100:5772–5777.
14. Bell, R. E., and N. Ben-Tal. 2003. In silico identification of functional protein interfaces. *Comp. Funct. Genomics.* 4:420–423.
15. Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. *Science.* 296:750–752.
16. Valdar, W. S., and J. M. Thornton. 2001. Conservation helps to identify biologically relevant crystal contacts. *J. Mol. Biol.* 313:399–416.
17. Steinberg, I. Z., and H. A. Scheraga. 1963. Entropy changes accompanying association reactions of proteins. *J. Biol. Chem.* 238:172–181.
18. Arumugam, S., G. Gao, B. L. Patton, V. Semenchenko, K. Brew, and S. R. Van Doren. 2003. Increased backbone mobility in β-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. *J. Mol. Biol.* 327:719–734.
19. Gohlke, H., L. A. Kuhn, and D. A. Case. 2004. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins.* 56:322–337.
20. Zidek, L., M. V. Novotny, and M. J. Stone. 1999. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* 6:1118–1121.
21. Rajamani, D., S. Thiel, S. Vajda, and C. J. Camacho. 2004. Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. USA.* 101:11287–11292.
22. Kimura, S. R., R. C. Brower, S. Vajda, and C. J. Camacho. 2001. Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.* 80:635–642.
23. Li, X., O. Keskin, B. Ma, R. Nussinov, and J. Liang. 2004. Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J. Mol. Biol.* 344:781–795.
24. Smith, G. R., M. J. Sternberg, and P. A. Bates. 2005. 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.* 347:1077–1101.
25. Mintseris, J., K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, and Z. P. Weng. 2005. Protein-Protein Docking Benchmark 2.0: an update. *Proteins.* 60:214–216.
26. Reichmann, D., O. Rahat, S. Albeck, R. Meged, O. Dym, and G. Schreiber. 2005. The modular architecture of protein-protein binding interfaces. *Proc. Natl. Acad. Sci. USA.* 102:57–62.
27. Keskin, O., C. J. Tsai, H. Wolfson, and R. Nussinov. 2004. A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications. *Protein Sci.* 13:1043–1055.
28. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
29. Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal. 2003. CONSURF: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics.* 19:163–164.
30. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.
31. MacKerell, A. D., N. Banavali, and N. Foloppe. 2000. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers.* 56:257–265.
32. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD—visual molecular dynamics. *J. Mol. Graph.* 14:33–38.
33. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
34. Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* 2:173–181.
35. Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* 79:3090–3093.
36. Keskin, O., R. L. Jernigan, and I. Bahar. 2000. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.* 78:2093–2106.
37. Mol, C. D., A. S. Arvai, R. J. Sanderson, G. Slupphaug, B. Kavli, H. E. Krokan, D. W. Mosbaugh, and J. A. Tainer. 1995. Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell.* 82:701–708.
38. Levy, Y., and J. N. Onuchic. 2006. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* 35: 389–415.
39. Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* 91:43–56.