



# Side-Chain Conformational Changes upon Protein–Protein Association

Anatoly M. Ruvinsky<sup>1</sup>, Tatsiana Kirys<sup>1,2</sup>, Alexander V. Tuzikov<sup>2</sup>  
and Ilya A. Vakser<sup>1,3\*</sup>

<sup>1</sup>Center for Bioinformatics, The University of Kansas, Lawrence, KS 66047, USA

<sup>2</sup>United Institute of Informatics Problems, National Academy of Sciences, 220012 Minsk, Belarus

<sup>3</sup>Department of Molecular Biosciences, The University of Kansas, Lawrence, KS 66045, USA

Received 26 October 2010;  
received in revised form  
31 January 2011;  
accepted 11 February 2011  
Available online  
25 February 2011

Edited by B. Honig

## Keywords:

protein binding;  
protein recognition;  
structure prediction;  
side-chain rotamers;  
structural flexibility

Conformational changes upon protein–protein association are the key element of the binding mechanism. The study presents a systematic large-scale analysis of such conformational changes in the side chains. The results indicate that short and long side chains have different propensities for the conformational changes. Long side chains with three or more dihedral angles are often subject to large conformational transition. Shorter residues with one or two dihedral angles typically undergo local conformational changes not leading to a conformational transition. A relationship between the local readjustments and the equilibrium fluctuations of a side chain around its unbound conformation is suggested. Most of the side chains undergo larger changes in the dihedral angle most distant from the backbone. The frequencies of the core-to-surface interface transitions of six nonpolar residues and Tyr are larger than the frequencies of the opposite surface-to-core transitions. The binding increases both polar and nonpolar interface areas. However, the increase of the nonpolar area is larger for all considered classes of protein complexes, suggesting that the protein association perturbs the unbound interfaces to increase the hydrophobic contribution to the binding free energy. To test modeling approaches to side-chain flexibility in protein docking, conformational changes in the X-ray set were compared with those in the docking decoy sets. The results lead to a better understanding of the conformational changes in proteins and suggest directions for efficient conformational sampling in docking protocols.

© 2011 Elsevier Ltd. All rights reserved.

## Introduction

Protein structure–function relationships with a focus on conformational changes upon protein–protein association have been the subject of extensive research, including systematic studies on

protein sets<sup>1–10</sup> and specific proteins.<sup>11–16</sup> The theory of such conformational changes has been evolving from the early “lock-and-key” concept,<sup>17</sup> through the induced-fit model<sup>18</sup> and the allosteric model,<sup>19</sup> to the paradigm of the conformational selection and population shift.<sup>20–27</sup> The knowledge and understanding of these conformational changes have been accumulated and implemented in algorithms for predicting the structure of protein complexes, as evidenced by the CAPRI experiment.<sup>28</sup> However, the conformational changes upon the formation of a complex remain one of the greatest challenges for researchers studying protein interactions. A direct

\*Corresponding author. Center for Bioinformatics, The University of Kansas, Lawrence, KS 66047, USA.  
E-mail address: [vakser@ku.edu](mailto:vakser@ku.edu).

Abbreviations used: RASA, relative solvent-accessible surface area; RSD, root-square deviation.

way to tackle this problem is to study the differences between the unbound and bound structures of the same protein<sup>1,2,6</sup> or the differences between the alternative conformations in unbound proteins.<sup>4,5</sup> An encouraging factor is the growth of the PDB,<sup>29</sup> which has been the source for studies of the side-chain conformations in proteins in general.<sup>30–37</sup> In the 1990s, when only a few proteins had both bound and unbound structures known,<sup>38</sup> Betts and Sternberg<sup>2</sup> studied 39 pairs of bound and unbound proteins, with only eight of the complexes having unbound structures of both binding proteins. Recently side-chain transitions were analyzed on a set of 124 protein complexes with known unbound structures.<sup>6</sup> Currently, such sets (called docking benchmark sets because of their primary use in docking validation) contain a significantly larger and growing number of complexes.<sup>39,40</sup> Our DOCKGROUND nonredundant benchmark set<sup>40</sup> used in this study has 233 protein complexes, 99 of them having unbound structures of both binding proteins (134 complexes have the unbound structure for one of the proteins).

Protein structures reveal a rich variety of conformational changes that occur at different scales upon binding.<sup>41</sup> This includes domain motions, local folding–unfolding transitions, transitions between regular secondary structure elements in “chameleon sequences,”<sup>42,43</sup> disorder-to-order transitions,<sup>43,44</sup> and other changes in protein backbone and side chains. Although, in general, the different types of the conformational changes may be interrelated, in this study we focus on the conformational changes in the side chains. This choice is motivated by the fact that the majority of protein complexes in the nonredundant benchmark sets<sup>39,40</sup> have small C $\alpha$  RMSD between bound and unbound structures. Indeed, 71% of the DOCKGROUND set<sup>40</sup> used in this study has C $\alpha$  RMSD < 2 Å for 71% of the complexes. The Benchmark set from Weng's group<sup>39</sup> has interface C $\alpha$  RMSD < 2.2 Å for 84% of complexes. Thus, studying conformational changes in the side chains is important for the development of better protein–protein docking procedures.<sup>45–49</sup> The focus on the side chains also follows the “divide-and-conquer” paradigm: elucidating the side-chain conformational changes first, then proceeding to the backbone flexibility, and eventually to their combination (objective of our future study).

Previous studies related to the side-chain conformational changes analyzed the dynamics of the changes.<sup>50–54</sup> The scale of the conformational change was found to be determined to a significant extent by the residue's surrounding (the environment effect). The effect appears as a decreased number of rotamers in the buried residues in comparison with the surface residues,<sup>50, 51, 52, 53</sup> as a small RMSD between bound and unbound states of pocket side chains,<sup>3</sup> or as reduced fluctuations of the center of

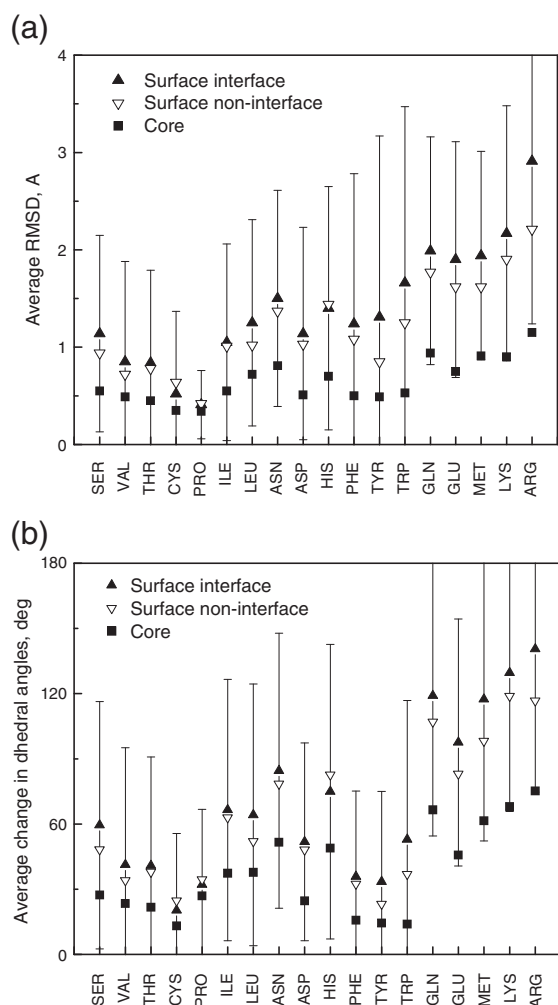
mass of such residues.<sup>55</sup> The side-chain dynamics made it possible to differentiate the roles of the interface residues in binding and develop a concept of anchor and latch residues that show restricted mobility and pass through similar conformations in molecular dynamics trajectories of the bound and unbound states.<sup>50,54</sup> The concept was extended to include conserved residues at protein interfaces with similar properties.<sup>53</sup> The influence of the conformational changes on the binding entropy has been studied.<sup>52</sup> Guharoy *et al.*<sup>6</sup> found that the interface residues undergo more significant conformational changes and often have higher energies than the other surface residues.

Our study presents an analysis of the conformational changes of the core and the surface side chains accompanying noncovalent protein heterodimerization. We show that the mechanism and the scale of the conformational changes depend on the side chain length and the proximity of the dihedral angle to the protein backbone. Long side chains, with three or more dihedral angles, are more often subject to large conformational transitions ( $\sim 120^\circ$  of  $\chi$ -angle change). Shorter residues, with one or two dihedral angles, typically undergo small conformational changes ( $\sim 40^\circ$ ) leading to local readjustments. We suggest that the local readjustments result from the equilibrium fluctuations of the side chain around its unbound conformation. The results show that about one-tenth of the complexes in our study went through the local interface changes only. All other complexes are subject to the interplay of the large conformational transitions and the local readjustments. In most residues, the largest conformational changes occur in the dihedral angle most distant from the backbone. The opposite trend is found in the residues with symmetric aromatic (Phe and Tyr) and charged (Asp and Glu) groups, where the  $\chi$  angle closest to the backbone changes most. The study also reveals the interface conformational changes leading to disorder-to-order transitions and the changes in the residue surface area that result in core-to-surface and surface-to-core transitions.

## Results and Discussion

### Analysis of unbound-to-bound conformational changes in crystal structures

Comparison of the dihedral angles values in bound and unbound residues was performed on the DOCKGROUND docking benchmark set containing the bound and unbound structures of same proteins (see [Methods](#)). The results ([Fig. 1](#)) reveal two trends: (1) generally, the extent of the conformational changes increases with the increase in the number of dihedral angles in the side chain; (2) the extent is



**Fig. 1.** Average conformational change between unbound and bound residues. (a) Change in Cartesian coordinates (RMSD) and (b) change in dihedral angles (RSD; see [Methods](#)). The residues are sorted left to right according to the increasing number of  $\chi$  angles and increasing mass (if the number of  $\chi$  is the same). The SD is shown for the interface residues.

larger for the surface interface residues than for the surface non-interface and the core residues. The relatively smaller conformational changes in the core can be explained by the tight packing. A number of the surface non-interface residues are part of the crystal packing interfaces. The relatively smaller conformational changes in the surface non-interface residues may suggest that the crystal packing interactions, on average, are weaker than interactions across the biological interfaces. However, the exact contribution of the crystal packing effect on the non-interface residues is beyond the scope of this study, which is focused on the analysis of the residues at the biological interfaces.

The results show that Pro, Cys, and His have larger conformational changes at the non-interface

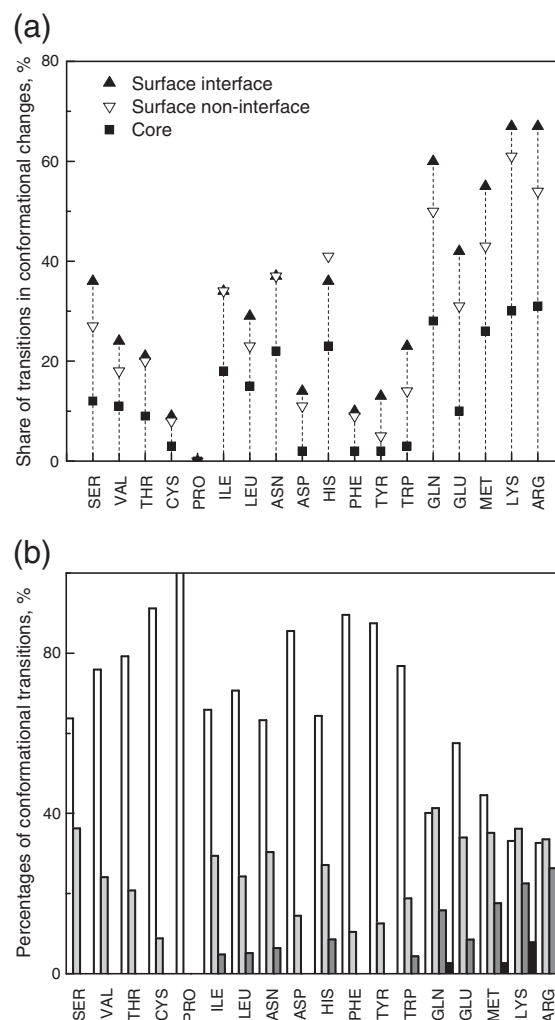
surface than at the interface ([Fig. 1](#)). However, the increase is not statistically significant. The average RMSD of the interface residues with one, two, three, and four dihedral angles is 0.75, 1.22, 1.94, and 2.54 Å, and the average root-square deviation (RSD) of the dihedral angles (see [Methods](#)) is 40.5°, 55.1°, 111.3°, and 135.0°, respectively. Since the dihedral angle tends to cluster near 180°, 60°, and -60° (the *trans*, *gauche*<sup>+</sup>, and *gauche*<sup>-</sup> conformations), one can conclude that the side chains with one or two  $\chi$  angles undergo local conformational changes, whereas the side chains with three or four  $\chi$  angles can undergo a conformational transition between the energy minima. Moreover, since the average RSDs for the long side chains (the ones with more  $\chi$  angles) vary around 120° (the distance between two adjacent energy minima), one can assume that the conformational transitions most likely occur in a single  $\chi$  angle. Other dihedral angles in the long side chains as well as those in the short side chains typically undergo a local readjustment upon binding. It is worth noting that all long side chains are polar, except Met. Among the side chains with two and one  $\chi$  angles, Asn, His, and Ser, which show the largest average dihedral angle RSD in the group, are also polar. Three nonpolar residues, Cys, Pro, and Phe, and polar Tyr have the smallest changes of dihedral angles. Cys and Pro are the least variable in terms of RMSD. One can assume that the difference in the degree of conformational changes of polar and nonpolar residues may result from different packing around these residues. The nonpolar residues have high propensity for the tightly packed protein core, whereas the polar residues often have exposed conformations that loosen their structural surrounding, allowing more space for change.

The test of statistical significance (Supplementary Tables S1–S4) showed that the difference between the average RMSDs as well as the average RSDs of the core and the surface non-interface residues is statistically significant for all amino acids, with *P* values much less than the significance level of 5%. Still, the corresponding average changes in the core and at the non-interface surface in terms of RSD/RMSD are highly correlated. The core–interface and the non-interface surface–interface correlations were also high. The high correlation reflects the larger flexibility of the longer side chains observed in all protein regions ([Fig. 1](#)). The difference between the average RMSDs/RSDs of the core and interface was statistically significant for all residues except Cys, which is one of the most uncommon residues at protein interfaces and can form disulfide bonds restricting the scale of the conformational changes upon binding.<sup>56</sup> The difference between the average RMSDs/RSDs of the surface non-interface and interface was statistically significant for about half of the residues. Among the most conserved interface residues (Trp, Phe, and Met<sup>57</sup>), only Met had a *P*

value <5% for both metrics. Trp had a  $P$  value of 3.8% for the difference between the average RMSDs and a  $P$  value of 5.5% for the difference between the average RSDs. Phe had  $P$  values >5% in both cases. Four of the five charged residues (Arg, Glu, Lys, and Asp) had  $P$  values within the significance level for the difference between the average RMSDs. Arg, Glu, and Lys also had  $P$  values within the significance level for the difference between the average RSDs.

The local readjustments of the short side chains likely occur due to the thermal fluctuations of dihedral angles. Since the thermal fluctuations deviate, on average,  $\pm 20^\circ$  from the equilibrium,<sup>58</sup> one can estimate the average  $\chi$ -angle RSD (Eq. 1) due to the thermal fluctuations as  $\Delta\bar{\chi}^T \approx 40\sqrt{n}$ . For the side chains with one or two  $\chi$  angles ( $n=1$  or 2), we obtain  $\Delta\bar{\chi}_1^T = 40^\circ$  and  $\Delta\bar{\chi}_2^T = 56.6^\circ$ , which are in excellent agreement with the statistically derived average RSDs  $\Delta\chi_1 = 40^\circ$  and  $\Delta\chi_2 = 55.1^\circ$  (see above). The thermal fluctuations likely play the role of a “lock-and-key lubricant,” providing plasticity of interfaces needed for the exact fit upon binding. The average RSDs  $\Delta\chi_3 = 111.3^\circ$  and  $\Delta\chi_4 = 135.0^\circ$  in the longer side chains deviate from the fluctuation-based estimates  $\Delta\bar{\chi}_3^T = 69.3^\circ$  and  $\Delta\bar{\chi}_4^T = 80^\circ$ . The deviations increase with the increase in the number of the  $\chi$  angles from 2 to 4. The increase of the deviations may be explained by the ability of the longer residues to establish interactions across interface earlier than the shorter residues. Indeed, binding proteins “optimize” conformations of the long residues at the early stages of their approach. The short residues get involved at later stages of binding, and thus have less time for conformational sampling. The increase of the deviations also points toward a greater role of the induced fit mechanism in comparison to the “lubricated lock-and-key” mechanism for the longer side chains. The results show that only the “lubricated lock-and-key” mechanism is present in 11% of the complexes in the set. The rest 89% of the complexes are subject to the interplay of both the induced fit and the “lubricated fit” mechanisms. In 66% of the complexes, there are one to four conformational transitions per interface, which on average consists of 18 residues.

The share of the conformational transitions, defined as conformational changes  $\geq 100^\circ$  in one of  $\chi$  angles, among all the conformational changes is larger for the side chains with three or four  $\chi$  than for the side chains with one or two  $\chi$  (Fig. 2). The conformational transitions occur more frequently at the interface than at the nonbinding surface and in the core. This observation is in agreement with the results of Guharoy *et al.*<sup>6</sup> obtained on a smaller set of complexes using different definitions of surface and interface. The data in Fig. 2 show that only the His residue has a higher frequency of the conformational transitions at the nonbinding surface. Thr, Cys, Ile,



**Fig. 2.** The share of conformational transitions. The conformational transitions are defined as those with  $\Delta\chi \geq 100^\circ$  in any of the dihedral angles. (a) Percentage of conformational transitions in all conformational changes between unbound and bound structures. (b) Percentage of simultaneous conformational transitions in one (light gray), two (dark gray), and three (black)  $\chi$  angles for interface residues. The percentage for four  $\chi$  angles is negligible (not shown). The percentage of conformational changes  $< 100^\circ$  in each dihedral angle is shown by open bars.

Asn, and Phe have similar frequencies of the interface and non-interface surface conformational transitions. The probability of the interface transitions  $> 100^\circ$  simultaneously in several dihedral angles decreases significantly with the increase in the number of the dihedral angles. The dihedral angle change associated with the rotation of a symmetric group in Asp, Phe, Tyr, and Glu cannot exceed  $90^\circ$ ; thus, these residues do not undergo conformational transitions simultaneously in all  $\chi$  angles (Fig. 2).

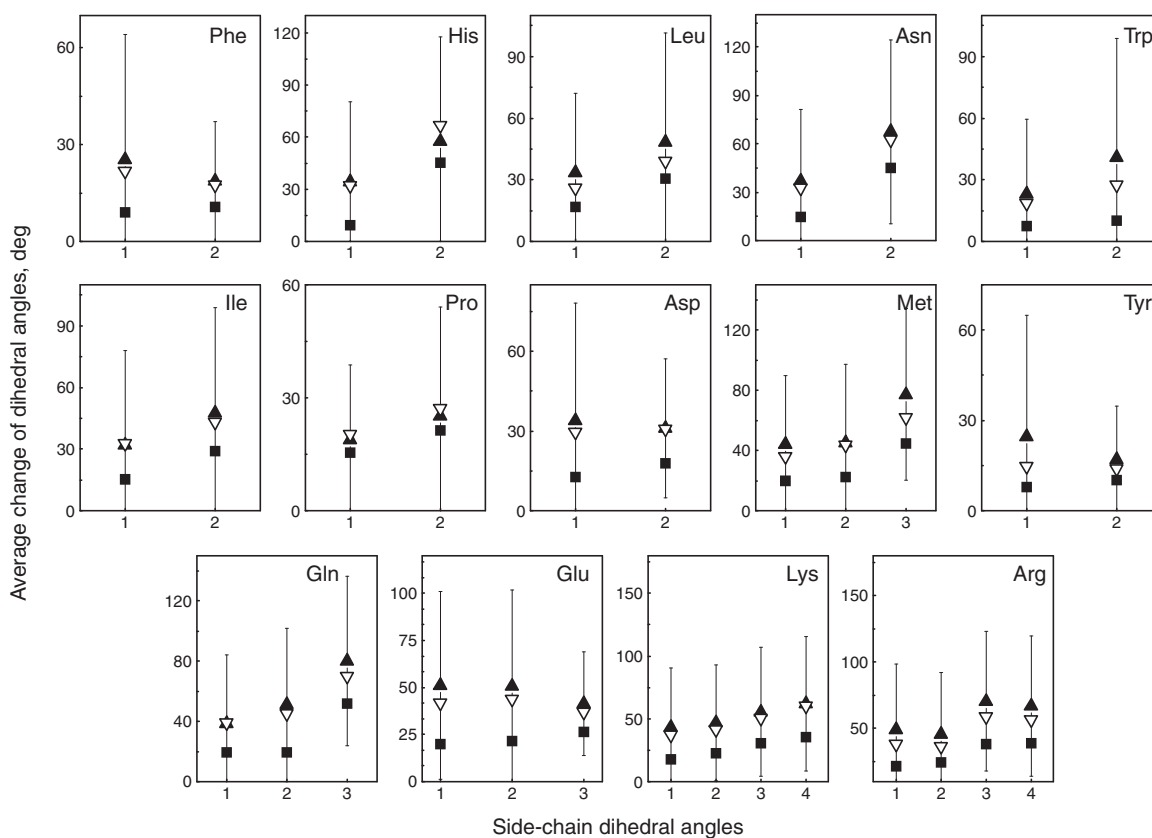
To further detail the picture, we computed the average changes of each  $\chi$  angle in the amino acids (Fig. 3). Six of the nine side chains with two dihedral angles had larger changes of the outer angle ( $\chi_2$ ) in comparison with the near-backbone angle ( $\chi_1$ ). The same trend was observed for all the side chains with three and four  $\chi$  angles, with the exception of Glu, which is slightly more prone to the changes in the first and second  $\chi$  angles. Two amino acid side chains with aromatic groups (Phe and Tyr) and two charged amino acids (Asp and Glu) showed an opposite trend where the outer  $\chi$  changes less than the near-backbone one. This trend is explained partly by the reduced interval of variability of the outer  $\chi$  due to the symmetry of the amino acid's terminal groups.

In agreement with earlier studies,<sup>43,44,59</sup> the results indicate that binding can decrease structural disorder at protein interfaces. Four percent (164 residues) of all interface residues in the set had disorder-to-order transition upon binding. The disordered residues were defined as those with missing coordinates in the crystal structure. Most of the disordered residues (39%) were Ala, Gly, Glu, or Thr. On the other hand, only 11% of the disordered

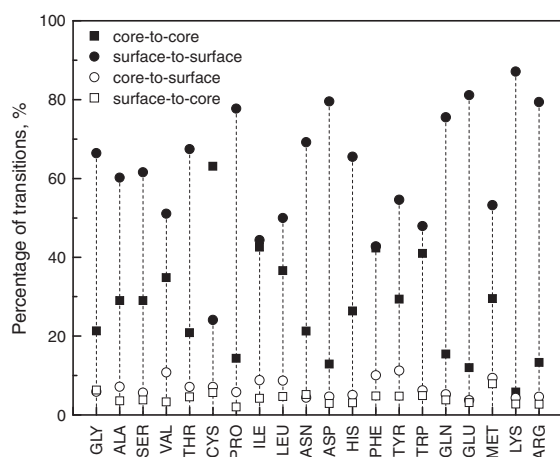
residues were Cys, His, Phe, Ile, Pro, Trp, or Tyr. This observation is in agreement with the classification of amino acids into order-promoting and disorder-promoting ones<sup>44</sup> and correlates well with the ability of the amino acids to fluctuate.<sup>55</sup>

An important conformational aspect of protein association is the changes of the residue surface area upon binding. The rate of the core-to-surface interface transitions (Fig. 4), calculated as a percentage of all transitions, varied from 10–11% (Tyr, Val, and Phe) to 4% (Asn, Glu, and Lys). Examples of the core-to-surface interface transitions are shown in Figs. 5 and 6. The rate of the surface-to-core interface transitions varies from 2% (Pro) to 8% (Met). Interestingly, on average, the rate of the core-to-surface transitions exceeds that of the surface-to-core transitions for all amino acids, except Asn.

To test the statistical significance of the differences between the corresponding rates of the core-to-surface and surface-to-core interface transitions, the jackknife test was applied (Supplementary Tables S5). Each protein complex in the benchmark set was left out, one at a time, and the rest of the complexes were used to calculate the rates of the transitions. Then, the excluded complex was put back in the set,

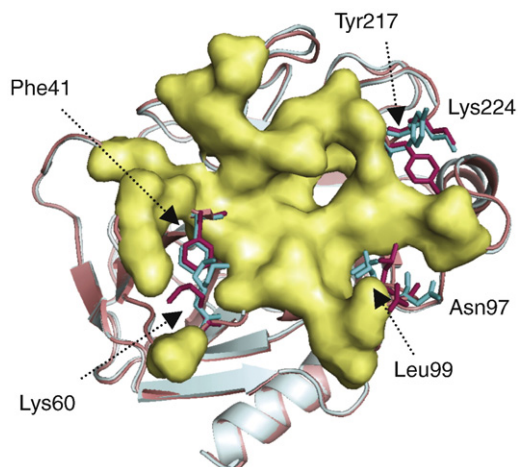


**Fig. 3.** Average dihedral angle change for different structure regions. The change between unbound and bound conformers is shown for the core (+), non-interface surface ("), and interface (#) residues. The SD is shown for the interface residues.



**Fig. 4.** Frequencies of transitions between surface and core at the interface.

and a next complex was left out, etc., continuing through all the complexes. Finally, the mean rates and their standard deviations (SDs) were calculated. The unpaired and paired two-sample *t* tests showed that the differences between the rates of the transitions are statistically significant for all the amino acids, with *P* values close to zero. This is

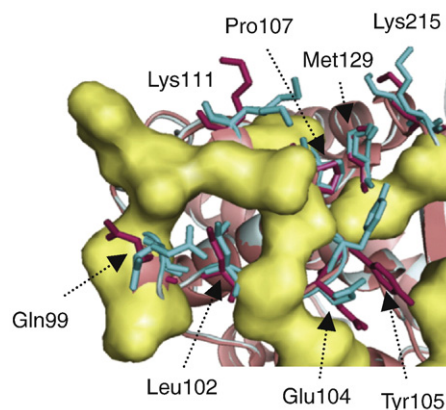


**Fig. 5.** Core-to-surface interface transitions in porcine pancreatic trypsin induced by soybean trypsin inhibitor. The bound structure is in magenta, and the unbound structure is in blue. The bound/complex structure is 1avw,<sup>60</sup> and the unbound trypsin structure is 2a31.<sup>61</sup> Phe41 keeps its conformation while undergoing a core-to-surface transition, with a change in the RASA from 16.9% in the unbound state to 28.6% in the bound state due to the conformational change in Lys60, which has two alternative unbound conformations. Lys224 keeps its conformation while increasing its RASA from 24.1% to 28.2% due to the conformational change in Tyr217 ( $\Delta$ RASA = 14.3%). Leu99 changes the RASA from 19.9% to 34.4% due to its own conformational change and the change in Asn97 ( $\Delta$ RASA = 16.2%).

explained by the fact that the SDs of the mean rates are 1 or 2 orders of magnitude smaller than the difference between the rates of the core-to-surface and surface-to-core interface transitions.

The largest difference between the rates is observed in six nonpolar residues, Ala, Val, Pro, Ile, Leu, and Phe, and a polar Tyr. At protein interfaces, Tyr often has been identified as a hot spot.<sup>64</sup> The bias in the core-to-surface transitions toward the nonpolar residues suggests that protein-protein interactions may perturb an unbound interface to increase the nonpolar interface area, thus increasing the hydrophobic contribution to the binding free energy (a major force stabilizing protein complexes). To test this hypothesis, we calculated average changes of polar and nonpolar interface areas induced by binding. On average, the binding increases both the polar and nonpolar interface areas in the complex. However, the increase of the nonpolar area is greater for all classes of complexes:  $\Delta S_p = 17.4 \pm 7.5 \text{ \AA}^2$  and  $\Delta S_n = 38.3 \pm 17.2 \text{ \AA}^2$  for antibody/antigen;  $\Delta S_p = 25.9 \pm 7.8 \text{ \AA}^2$  and  $\Delta S_n = 32.9 \pm 12.8 \text{ \AA}^2$  for enzyme/inhibitor; and  $\Delta S_p = 12.9 \pm 6.1 \text{ \AA}^2$  and  $\Delta S_n = 24.2 \pm 9.7 \text{ \AA}^2$  for other ( $\Delta S_{n,p}$  is the change of the nonpolar or polar interface area).

Two typical scenarios of the core-to-surface transitions were observed, as illustrated in Figs. 5



**Fig. 6.** Core-to-surface interface transitions in TEM-1  $\beta$ -lactamase induced by  $\beta$ -lactamase inhibitor protein-II. The bound structure is in magenta, and the unbound one is in blue. The bound/complex structure is 1jtd,<sup>62</sup> and the unbound TEM-1 is 1m40.<sup>63</sup> Leu102 keeps its conformation while undergoing core-to-surface transition, with a change in the RASA from 18.6% in the unbound state to 30% in the bound state due to the conformational change in Gln99, which has two alternative unbound conformations. Pro107 keeps its conformation while changing the RASA from 22.3% to 34.1% due to the conformational changes in Tyr105 ( $\Delta$ RASA = 49.9%) and Lys111, which have two alternative unbound conformations. Met129 keeps its conformation, but changes the RASA from 17.4% to 44.7% due to the conformational changes in Tyr105 and Lys215, which have two alternative unbound conformations. Glu104 changes the RASA by 30.3%.

and 6. In the first scenario, a core side chain does not change conformation, but other residues in the vicinity change conformations to increase the side-chain surface (e.g., Phe41 and Lys224 in Fig. 5, and Leu102, Pro107, and Met129 in Fig. 6). In the second scenario, both the side chain and its neighbors change their conformations (e.g., Leu99 in Fig. 5). A case where a core side chain undergoes a conformational change and its structural neighbors within 5 Å stay unchanged was not observed.

### Implications for protein docking

Side-chain flexibility is an important factor in the development of protein docking approaches.<sup>45,46,49,65–67</sup> Thus, comparison of conformational changes in experimentally determined structures with those modeled by computational approaches is useful for assessing the flexible docking methodology. Protein–protein docking decoy sets contain the structures of predicted complexes, both incorrect (false-positive matches) and near-native ones. For comparison with the observed unbound-to-bound conformational changes in the X-ray structures, we calculated the average RMSDs and RSDs between the native unbound structures and the top-ranked docked structures, as well as the most accurate (closest to the co-crystal structure) near-native docked structures in the “bound perturbations” and “unbound perturbations” sets by Gray *et al.*<sup>68</sup> These sets were chosen because they were generated by a protocol that involves explicit modeling of the side chain flexibility in binding. The modeling of the side chains in the “bound perturbation” and “unbound perturbation” sets is based on the experimentally determined bound and unbound conformations, respectively, of the backbone.

We performed the test of statistical significance of the difference between experimentally and computationally determined conformational changes in 18 amino acid types (Supplementary Tables S6–S9). The results showed that the difference between the average RSDs/RMSDs in the benchmark set and in the “bound” decoy set for the most accurate near-native structures is statistically significant, at the significance level of 5%, for 15 (RSD) and 17 (RMSD) amino acids in the core, 13 (RSD) and 11 (RMSD) at the non-interface surface, and 7 (RSD) and 4 (RMSD) at the interface. The test of the average conformational changes in the “unbound” decoy set gave similar results. The difference between the average RSDs/RMSDs in the benchmark set and the unbound decoy set for the most accurate near-native structures is statistically significant for 15 (RSD) and 14 (RMSD) amino acid types in the core, 13 (RSD) and 13 (RMSD) non-interface surface amino acids, and 7 (RSD) and 5 (RMSD) interface amino acids. The lack of statistical significance may indicate that the modeled chains follow the trend of the unbound-

to-bound changes in the X-ray set. However, the likely reason for the higher percentage of statistically indistinguishable comparisons at the surface, and especially at the interface, is the smaller number of the amino acids due to the smaller size of the interface *versus* non-interface surface *versus* core.

Despite the statistical differences between the average unbound-to-bound and unbound-to-decoy conformational changes, we found high correlations between these changes in terms of RSD/RMSD, with correlation coefficients of 0.90–0.96. The correlation remained high when the top-ranked structures were substituted for the most accurate ones. The high correlation reflects the larger flexibility (conformational freedom) of the longer side chains, similar to the above interpretation of the high correlation between the RSD/RMSD values in the core and at the non-interface and interface surface in the X-ray structures. Modeling of the longer side chains results in larger structural amplitudes, and thus larger values of RSD/RMSD, which correlate well with the larger RSD/RMSD for longer side chains in crystal structures (Fig. 1). Thus, such high correlation is not necessarily an indication of accurate modeling. The fact that the average RSD/RMSD unbound-to-bound values in the X-ray set and the corresponding average unbound-to-decoy values in the modeled set in most cases are statistically significantly different indicates that there is room for improvement in modeling.

Generally, the side chain rotamer libraries, often used in flexible docking, may not be sufficient for an adequate modeling of the side-chain conformational changes upon binding. More efficient techniques for modeling of the side-chain flexibility could utilize the average (or the most probable) changes of the dihedral angles (Fig. 3) in local sampling and the probabilities of the inter-rotamer conformational transitions (Fig. 2; Kirys *et al.*, in preparation).

### Conclusions

Knowledge of the conformational changes upon protein binding is essential for understanding the molecular mechanisms of life processes and our ability to model cell phenomena. The study focuses on the side-chain conformational changes in protein heterodimerization. The results indicate that short and long side chains have propensities for different mechanisms of the conformational changes. Long side chains with three or more dihedral angles are often subject to the induced fit mechanism resulting in a conformational transition. Shorter residues with one or two dihedral angles typically undergo local conformational changes not leading to a conformational transition. A relationship between the local readjustments and the equilibrium fluctuations of a side chain around its unbound conformation is

suggested. The local readjustments were dominant in 11% of the complexes. All other complexes were subject to the interplay of the induced fit and the local readjustments. We showed that most of the side chains undergo larger changes in the dihedral angle most distant from the backbone. The amino acids with symmetric aromatic (Phe and Tyr) and charged (Asp and Glu) groups show the opposite trend where the near-backbone dihedral angles change the most. The frequencies of the core-to-surface interface transitions of six nonpolar residues and Tyr are larger than the frequencies of the inverse surface-to-core transitions. The binding increases both polar and nonpolar interface areas. However, the increase of the nonpolar area is larger for all considered classes of the protein complexes, suggesting that the protein association perturbs the unbound interfaces to increase the hydrophobic contribution to the binding free energy. To test modeling approaches to side-chain flexibility in protein docking, conformational changes in the X-ray set were compared with those in the bound and unbound docking decoy sets. The results lead to a better understanding of the conformational changes in proteins and suggest directions for more efficient conformational sampling in docking protocols.

## Methods

The results were obtained on a nonredundant benchmark set of 233 non-obligate protein–protein complexes from the DOCKGROUND resource<sup>†</sup>.<sup>40,69</sup> The set contains unbound structures of both proteins for 99 complexes and the unbound structure of one of the proteins for 134 complexes. The structures were selected from PDB, based on the following criteria: sequence identity between bound and unbound structures >97%, sequence identity between complexes <30%, and homomultimers and crystal packing complexes excluded.

Conformational changes between the unbound and bound conformations were considered for each of the protein side chains in the set. The conformational changes were expressed in terms of the RMSD of the atom coordinates and the RSD of the dihedral angles

$$\Delta\chi = \left[ \sum_{i=1}^n D_i^2(\chi_i^b, \chi_i^u) \right]^{1/2} \quad (1)$$

where  $n$  is the number of the dihedral angles in a side chain,  $i$  is the index of a dihedral angle  $\chi_i$ , and  $b$  and  $u$  indicate bound and unbound conformations, respectively. Function

$$D_i(\chi_i^b, \chi_i^u) = \begin{cases} |\chi_i^b - \chi_i^u|, & \text{if } |\chi_i^b - \chi_i^u| \leq 180^\circ \\ 360^\circ - |\chi_i^b - \chi_i^u|, & \text{if } |\chi_i^b - \chi_i^u| > 180^\circ \end{cases} \quad (2)$$

gives the shortest distance between the dihedral angles on the circle. The values of the dihedral angles were taken

from  $0^\circ$  to  $360^\circ$ , except the last angles in Phe, Tyr, Asp, and Glu, which were taken from  $0^\circ$  to  $180^\circ$  due to the symmetry of aromatic and charged groups.<sup>70</sup> The dihedral angles analyzed for Arg were  $\chi_{1-4}$  because the tip of the side chain containing  $\chi_5$  is planar. The dihedral angles were determined using the Dang program<sup>‡</sup>. The conformational changes were placed in 18 groups corresponding to standard amino acids (Gly and Ala were not considered). The average conformational changes and the SDs were computed for the interface residues (surface residues at protein interfaces), non-interface surface residues, and core residues. Surface residues were defined as those with relative solvent-accessible surface area (RSA) >25%, as determined by NACCESS.<sup>71</sup> Interface residues were defined as those losing >1 Å of their surface upon binding.<sup>53</sup>

Supplementary materials related to this article can be found online at [doi:10.1016/j.jmb.2011.02.030](http://dx.doi.org/10.1016/j.jmb.2011.02.030)

## Acknowledgements

This study was supported by grant R01 GM074255 from the NIH. We thank Dr. Simon C. Lovell for code converting the dihedral angles into Cartesian coordinates of the atoms.

## References

- Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
- Betts, M. J. & Sternberg, M. J. E. (1999). An analysis of conformational changes on protein–protein association: Implications for predictive docking. *Protein Eng.* **12**, 271–283.
- Li, X., Keskin, O., Ma, B., Nussinov, R. & Liang, J. (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.
- Davis, I. W., Arendall, W. B., Richardson, D. C. & Richardson, J. S. (2006). The backrub motion: How protein backbone shrugs when a sidechain dances. *Structure*, **14**, 265–274.
- Bhardwaj, N. & Gerstein, M. (2009). Relating protein conformational changes to packing efficiency and disorder. *Protein Sci.* **18**, 1230–1240.
- Guharoy, M., Janin, J. & Robert, C. H. (2010). Side-chain rotamer transitions at protein–protein interfaces. *Proteins*, **78**, 3219–3225.
- Kidd, B. A., Baker, D. & Thomas, W. E. (2009). Computation of conformational coupling in allosteric proteins. *PLoS Comp. Biol.* **e1000484**.

<sup>†</sup> <http://dockground.bioinformatics.ku.edu>

<sup>‡</sup> <http://kinemage.biochem.duke.edu/software/dang.php>

8. Najmanovich, R., Kuttner, J., Sobolev, V. & Edelman, M. (2000). Side-chain flexibility in proteins upon ligand binding. *Proteins*, **39**, 261–268.
9. Gunasekaran, K. & Nussinov, R. (2007). How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **365**, 257–273.
10. Gutteridge, A. & Thornton, J. M. (2005). Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **346**, 21–28.
11. Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science*, **267**, 383–386.
12. Wlodarski, T. & Zagrovic, B. (2009). Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl Acad. Sci. USA*, **106**, 19346–19351.
13. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A. *et al.* (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438**, 117–121.
14. Zhang, X. J., Wozniak, J. A. & Matthews, B. W. (1995). Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J. Mol. Biol.* **250**, 527–552.
15. Bui, J. M., Radic, Z., Taylor, P. & McCammon, J. A. (2006). Conformational transitions in protein–protein association: Binding of fasciculin-2 to acetylcholinesterase. *Biophys. J.* **90**, 3280–3287.
16. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F. A., Becker, S. *et al.* (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, **320**, 1471–1475.
17. Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dt. Chem. Ges.* **27**, 2985–2993.
18. Koshland, D. E. (1959). Enzyme flexibility and enzyme action. *J. Cellular Compar. Physiol.* **54**, 245–258.
19. Changeux, J. P. & Edelstein, S. J. (2005). Allosteric mechanisms of signal transduction. *Science*, **308**, 1424–1428.
20. Tsai, C. J., Kumar, S., Ma, B. & Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci.* **8**, 1181–1190.
21. Ma, B., Kumar, S., Tsai, C. J. & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng.* **12**, 713–721.
22. Tsai, C. J. T., Ma, B. M. & Nussinov, R. N. (1999). Folding and binding cascades: Shifts in energy landscapes. *Proc. Natl Acad. Soc. USA*, **96**, 9970–9972.
23. Csermely, P., Palotai, R. & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: An extended view of binding events. *Trends Biochem. Sci.* **35**, 539–546.
24. Bosshard, H. R. (2001). Molecular recognition by induced fit: How fit is the concept? *News Physiol. Sci.* **16**, 171–173.
25. Boehr, D. D. & Wright, P. E. (2008). How do proteins interact? *Science*, **320**, 1429–1430.
26. Boehr, D. D., Nussinov, R. & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5**, 789–796.
27. Zhou, H. X. (2010). From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* **98**, L15–L17.
28. Janin, J., Henrick, K., Moult, J., Ten Eyck, L., Sternberg, M. J. E., Vajda, S. *et al.* (2003). CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins*, **52**, 2–9.
29. Berman, H., Henrick, K. & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Struct. Biol.* **10**, 980.
30. Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry*, **17**, 4277–4285.
31. Janin, J. & Wodak, S. (1978). Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125**, 357–386.
32. Dunbrack, R. L. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681.
33. Bromberg, S. & Dill, K. A. (1994). Side-chain entropy and packing in proteins. *Protein Sci.* **3**, 997–1009.
34. Schrauber, H., Eisenhaber, F. & Argos, P. (1993). Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. *J. Mol. Biol.* **230**, 592–612.
35. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. (2000). The penultimate rotamer library. *Proteins*, **40**, 389–408.
36. Nayeem, A. & Scheraga, H. A. (1994). A statistical analysis of side-chain conformations in proteins: Comparison with ECEPP predictions. *J. Protein Chem.* **13**, 283–296.
37. Ponder, J. W. & Richards, F. M. (1987). Internal packing and protein structural classes. *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. LII, pp. 421–428. Cold Spring Harbor Laboratory.
38. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
39. Hwang, H., Pierce, B., Mintseris, J., Janin, J. & Weng, Z. (2008). Protein–protein docking benchmark version 3.0. *Proteins*, **73**, 705–709.
40. Gao, Y., Douguet, D., Tovchigrechko, A. & Vakser, I. A. (2007). DOCKGROUND system of databases for protein recognition studies: Unbound structures for docking. *Proteins*, **69**, 845–851.
41. Goh, C. S., Milburn, D. & Gerstein, M. (2004). Conformational changes associated with protein–protein interactions. *Curr. Opin. Struct. Biol.* **14**, 104–109.
42. Mezei, M. (1998). Chameleon sequences in the PDB. *Protein Eng.* **14**, 411–414.
43. Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60.
44. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S. *et al.* (2001). Intrinsically disordered proteins. *J. Mol. Graph. Mod.* **19**, 26–59.
45. Gray, J. J. (2006). High-resolution protein–protein docking. *Curr. Opin. Struct. Biol.* **16**, 183–193.
46. Andrusier, N., Mashiah, E., Nussinov, R. & Wolfson, H. J. (2008). Principles of flexible protein–protein docking. *Proteins*, **73**, 271–289.
47. Ritchie, D. W. (2008). Recent progress and future directions in protein–protein docking. *Curr. Protein Peptide Sci.* **9**, 1–15.

48. Camacho, C. J. & Vajda, S. (2002). Protein-protein association kinetics and protein docking. *Curr. Opin. Struct. Biol.* **12**, 36–40.
49. Bonvin, A. M. J. J. (2006). Flexible protein-protein docking. *Curr. Opin. Struct. Biol.* **16**, 194–200.
50. Rajamani, D., Thiel, S., Vajda, S. & Camacho, C. J. (2004). Anchor residues in protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.
51. Smith, G. R., Sternberg, M. J. E. & Bates, P. A. (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.
52. Grunberg, R., Nilges, M. & Leckner, J. (2006). Flexibility and conformational entropy in protein-protein binding. *Structure*, **14**, 683–693.
53. Yogurtcu, O. N., Erdemli, S. B., Nussinov, R., Turkay, M. & Keskin, O. (2008). Restricted mobility of conserved residues in protein-protein interfaces in molecular simulations. *Biophys. J.* **94**, 3475–3485.
54. Kimura, S. R., Brower, R. C., Vajda, S. & Camacho, C. J. (2001). Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.* **80**, 635–642.
55. Ruvinsky, A. M. & Vakser, I. A. (2010). Sequence composition and environment effects on residue fluctuations in protein structures. *J. Chem. Phys.* **133**, 155101.
56. Glaser, F., Steinberg, D., Vakser, I. A. & Ben-Tal, N. (2001). Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins*, **43**, 89–102.
57. Ma, B., Elkayam, T., Wolfson, H. & Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl Acad. Sci. USA*, **100**, 5772–5777.
58. Finkelstein, A. V. & Ptitsyn, O. B. (2002). *Protein Physics—A Course of Lectures*. Academic Press, .
59. Fong, J. H., Shoemaker, B. A., Garbuzynskiy, S. O., Lobanov, M. Y., Galzitskaya, O. V. & Panchenko, A. R. (2009). Intrinsic disorder in protein interactions: Insights from a comprehensive structural analysis. *PLoS Comp. Biol.* **5**, e1000316.
60. Song, H. K. & Suh, S. W. (1998). Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *J. Mol. Biol.* **275**, 347–363.
61. Transue, T. R., Gabel, S. A. & London, R. E. (2006). NMR and crystallographic characterization of adventitious borate binding by trypsin. *Bioconjugate Chem.* **17**, 300–308.
62. Lim, D., Park, H. U., De Castro, L., Kang, S. G., Lee, H. S., Jensen, S. *et al.* (2001). Crystal structure and kinetic analysis of  $\beta$ -lactamase inhibitor protein-II in complex with TEM-1  $\beta$ -lactamase. *Nature Struct. Biol.* **8**, 848–852.
63. Minasov, G., Wang, X. & Shoichet, B. K. (2002). An ultrahigh resolution structure of TEM-1  $\beta$ -lactamase suggests a role for Glu166 as the general base in acylation. *J. Am. Chem. Soc.* **124**, 5333–5340.
64. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
65. Zacharias, M. (2010). Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.* **20**, 180–186.
66. Lensink, M. F. & Wodak, S. J. (2010). Docking and scoring protein interactions: CAPRI 2009. *Proteins*, **78**, 3073–3084.
67. Vakser, I. A. & Kundrotas, P. (2008). Predicting 3D structures of protein-protein complexes. *Curr. Pharm. Biotech.* **9**, 57–66.
68. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A. & Baker, D. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299.
69. Douguet, D., Chen, H. C., Tovchigrechko, A. & Vakser, I. A. (2006). DOCKGROUND resource for studying protein-protein interfaces. *Bioinformatics*, **22**, 2612–2618.
70. Dunbrack, R. L. & Karplus, M. (1993). Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.* **230**, 543–574.
71. Hubbard, S. J. & Thornton, J. M. (1993). NACCESS. Computer Program, Department of Biochemistry and Molecular Biology, University College London.