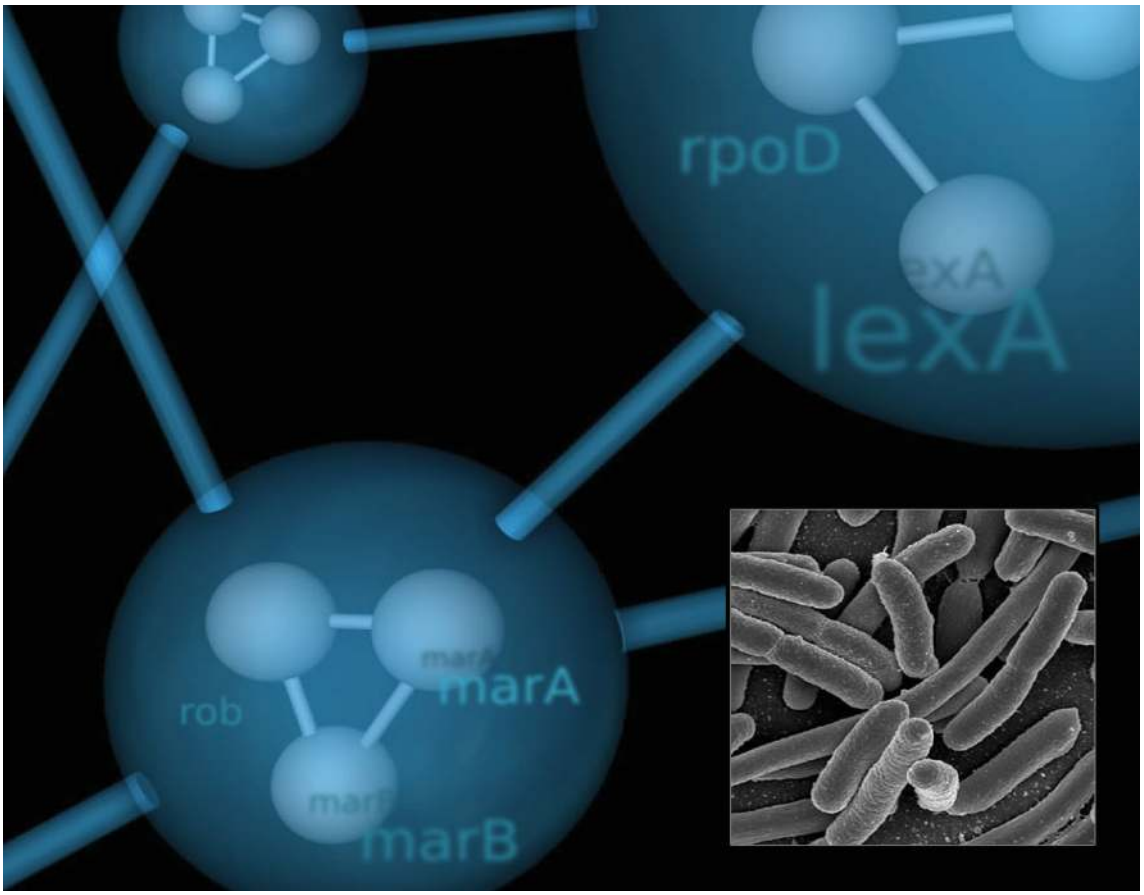
Molecular BioSystems

This article was published as part of the

Computational and Systems Biology themed issue

Please take a look at the full table of contents to access the other papers in this issue.



Energy based approach for understanding the recognition mechanism in protein–protein complexes†

M. Michael Gromiha,* Kiyonobu Yokota and Kazuhiko Fukui

Received 27th February 2009, Accepted 28th May 2009
First published as an Advance Article on the web 7th July 2009
DOI: 10.1039/b904161n

Protein–protein interactions play an essential role in the regulation of various cellular processes. Understanding the recognition mechanism of protein–protein complexes is a challenging task in molecular and computational biology. In this work, we have developed an energy based approach for identifying the binding sites and important residues for binding in protein–protein complexes. The new approach is different from the traditional distance based contacts in which the repulsive interactions are treated as binding sites as well as the contacts within a specific cutoff have been treated in the same way. We found that the residues and residue-pairs with charged and aromatic side chains are important for binding. These residues influence to form cation–π, electrostatic and aromatic interactions. Our observation has been verified with the experimental binding specificity of protein–protein complexes and found good agreement with experiments. Based on these results we have proposed a novel mechanism for the recognition of protein–protein complexes: the charged and aromatic residues in receptor and ligand initiate recognition by making suitable interactions between them; the neighboring hydrophobic residues assist the stability of complex along with other hydrogen bonding partners by the polar residues. Further, the propensity of residues in the binding sites of receptors and ligands, atomic contributions and the influence on secondary structure will be discussed.

Introduction

Protein–protein interactions are important in most of the cellular processes in life. Hence understanding the mechanism of protein–protein recognition at the molecular level is of practical interest and has direct applications to functional genomics. The two major approaches to this problem are large-scale studies on protein–protein interaction networks and investigations on the general principles of recognition and prediction of their binding sites. Unraveling the mechanism of protein–protein recognition is a fundamental problem and it would aid in function prediction and drug design.

The availability of numerous numbers of protein–protein complexes enables researchers to analyze the binding sites in terms of amino acid composition, preference of residues, secondary structures, solvent accessibility, electrostatic patches, hydrophobic contacts, hydrogen bonding networks and so on. Ofran and Rost[1] showed that the interface residues are dominated with specific amino acid residues and amino acid composition is a good feature for predicting the type of interaction interface. The recognition sites comprise of at least one patch that contains a core made of buried interface atoms, surrounded by a rim of atoms that remain accessible to solvent in the complex and have a distinctive amino acid composition.[2]

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan.
E-mail: michael-gromiha@aist.go.jp; Fax: +81-3-3599-8081; Tel: +81-3-3599-8046

† This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

The importance of amino acid residues has also been analyzed with the conservation of residues across protein–protein interfaces.[3,4] In addition, the roles of electrostatic interactions and binding energy hotspots have been reported for the binding of protein–protein complexes.[5–7] Recently, protein–protein interactions have been mapped on protein sequences, which suggest that the hotspots can be predicted from amino acid sequences.[8] Furthermore, protein–protein interactions have been studied in terms of efficient clustering,[9] stability calculations,[10] conformational changes and docking simulations.[11] Shoemaker and Panchenko[12,13] reviewed several concepts of protein–protein interactions in terms of experimental techniques, databases, prediction of protein–protein and domain interactions.

On the other hand, several methods have been proposed for identifying the binding sites in protein–protein complexes. Jones and Thornton[14] used surface patches for predicting protein–protein interaction sites. Fernandez-Recio *et al.*[15] developed a method based on docking energy landscapes for predicting protein–protein interactions. Potapov *et al.*[16] utilized molecular architecture and naturally occurring template fragments in the Protein Data Bank for identifying protein–protein interface. The combination of sequence and structural features as well as the information on nine consecutive residues, secondary structure of the central residue and average properties based on solvent accessibility, protrusion and depth has also been employed for detecting the binding sites from amino acid sequence.[17] Albou *et al.*[18] used the alpha shape of a molecule for computing surface residues. Grosdidier and Fernandez-Reico[19] identified the hotspot residues in protein–protein complexes using a computational docking method. Gong and Blundell[20] derived "environment

specific substitution tables'' by discarding the functional residues and utilized them to predict the protein–protein interaction sites. Dukka and Livesay[21] used phylogenetic motifs for predicting position-specific functional residues. Najmanovich *et al.*[22] implemented a graph-matching-based method for detecting 3D atomic similarities, which have been used to analyze protein binding sites. Shulman-Pelag *et al.*[23,24] constructed a method based on multiple alignment for detecting binding sites in protein–protein complexes. It recognizes the spatially conserved physico-chemical interactions, which often involve energetically important hot-spot residues that are crucial for protein–protein associations. Ertekin *et al.*[25] proposed a method based on the fluctuation behavior of residues to predict the putative protein binding sites. Halperin *et al.*[26,27] used phage display libraries and structurally conserved residues for identifying the interface residues. Li *et al.*[28] showed that the hotspots are often located in the complemented pockets in protein–protein complexes. Hannenhalli and Wang[29] derived a positional weight matrix from a set of experimentally derived binding sites and utilize the information for binding site prediction. Further, machine learning techniques have been widely used to identify the binding sites in protein–protein complexes.[30–33] Keskin *et al.*[34,35] extensively reviewed the principles of protein–protein interactions as well as the organization, cooperativity and mapping in a bottom-up systems biology approach.

In most of these studies, binding sites have been defined with a criteria based on the contacts between amino acid residues in two partners of protein–protein complexes. The atomic contacts between C_α atoms, C_β atoms, any atoms in a residue as well as the distances of 5–7 Å have been used to assign the contacts. These criteria include the repulsive interactions in which two residues are close to each other. In addition, the residue pairs with different distances have been treated in a same manner. Other methods employed shape complimentarity, homologs, conservation and amino acid properties. In this work, we have developed a new approach based on interaction energy for defining the binding sites. We observed that the binding sites are dominated with aromatic and charged residues indicating the importance of electrostatic, aromatic and cation–π interactions. Further, we have analyzed the preference of interacting partners and the preference of residues in binding segments of different lengths. Based on the results obtained in the present study we have proposed a novel mechanism for understanding the recognition of protein–protein complexes.

The identified binding sites using energy based approach may cluster together within densely packed hot regions and they form networks of interactions.[34] Mapping of protein–protein interactions assists in predicting protein function and in the construction of interaction maps. Through the network of protein–protein interactions, one can map cellular pathways and their intricate cross-connectivity.[36]

Materials and methods

Dataset

In this study, we developed a dataset of heterodimer protein–protein complexes from the crystal structures in 25-09-2007 release of Protein Data Bank.[37] In order to avoid artificial bias and to get clean data, we excluded the complexes that contain 'Mutation' in COMPND fields and non-standard amino acids without MSE in SEQRES field. The heterodimers were selected using the following procedure: (i) BIOMOLECULE information in REMARK 350 fields contain ''dimer'' and (ii) the two chains are not identical (sequence identity $<100\%$). The redundancy among heterodimer sequences was removed by using ALIGN program[38] so that no two sequences in the resulting dataset have more than 25% sequence identity. To get a more precise heterodimer dataset, we extracted the complexes with the following criteria: (i) >50 residues in both chains, (ii) $<10\%$ missing regions and (iii) <3.0 Å resolution. As a result, we got a dataset of 153 pairs of heterodimer protein–protein complexes. In each protein–protein complex, proteins with high and low molecular weights are termed as receptors and ligands, respectively. It is noteworthy that the current dataset is completely non-redundant whereas other datasets[39] have the complexes with same (redundant) receptors and different ligands. We have also carried out the analysis with a benchmark dataset of 124 protein–protein complexes[40] and we observed similar results. Hence, we report the results obtained with a dataset of 153 protein–protein complexes developed in this work.

Computation of interaction energy

We have calculated the interaction free-energy between atoms in protein–protein complexes using AMBER potential,[41] which is widely used to analyze and understand the recognition mechanism in protein complexes.[42] It is given by:

$$E_{\text{inter}} = \Sigma\,[(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6}) + q_i q_j/\varepsilon r_{ij}] \qquad (1)$$

where $A_{ij} = \varepsilon_{ij}^*(R_{ij}^*)^{12}$ and $B_{ij} = 2\,\varepsilon_{ij}^*(R_{ij}^*)^6$; $R_{ij}^* = (R_i^* + R_j^*)$ and $\varepsilon_{ij}^* = (\varepsilon_i^*\,\varepsilon_j^*)^{1/2}$; R^* and ε^* are, respectively, the van der Waals radius and well depth and these parameters are obtained from Cornell *et al.*;[41] q_i and q_j are, respectively, the charges for the atoms i and j, and r_{ij} is the distance between them. We have used the distant dependent dielectric constant ($\varepsilon = r_{ij}$) to take account of the dielectric damping effect of the Coulomb interactions, as used in other studies on protein complexes.[42]

Binding propensity

The binding propensity for the 20 amino acid residues in both receptors and ligands in protein–protein complexes has been developed as follows: we have computed the frequency of occurrence of amino acid residues in binding sites (f_b) and in the receptor (ligand) as a whole (f_t). The binding propensity (P_{bind}) is calculated using the equation:

$$P_{\text{bind}}(i) = f_b(i)/f_t(i) \qquad (2)$$

where, i represents each of the 20 amino acid residues.

Binding segments

The residues identified as binding sites have been analyzed in terms of binding segments. It is based on the number of consecutive binding residues in amino acid sequences.

For example, a 4-residue binding segment has a stretch of four consecutive binding residues. We have analyzed the behavior of the 20 amino residues in binding segments with one, two, three, four and more than four residues.

Residue-pair preference

We have computed the preference of residues to be involved in the interactions between receptors and ligands using the equation:[43]

$$\text{Pair}(i,j) = \Sigma N_{ij} / (\Sigma N_i + \Sigma N_j) \quad (3)$$

where i and j stand for the interacting residues in receptors and ligands, respectively. $N_{i,j}$ is the number of interacting residues of type i in receptors and j in ligands. ΣN_i and ΣN_j are the total number of residues of type i and j, respectively in receptors and ligands.

Results and discussion

Occurrence of amino acid residues at various ranges of interaction free energies

In a protein–protein complex, we have computed the interaction energy (eqn (1)) of each residue in a receptor with all residues in ligand. We have repeated the calculations for all the complexes and analyzed the interaction energies of all the residues in intervals of 0.1 from −15 to 5 kcal mol^{-1}. The frequency of occurrence of residues in receptors at different intervals of interaction free energies (from −2 to 1 kcal mol^{-1}) are displayed in Fig. 1. In this figure, we present the results for both the fraction of residues and total percentage of residues at each interval. We observed that 7.7% of the residues have strong interactions with ligands and the interaction free energy is less than −2 kcal mol^{-1}. On the other hand, 6.2% of residues have repulsive energies and 77% of the residues have the interaction energy in the range of −0.3 to 0 kcal mol^{-1}, which might be due to the presence of residues that are far away in 3D structures. Among 48 657 residues 5255 of them have the interaction free-energy less than −1 kcal mol^{-1}. Interestingly, we observed similar number of residues (4957) in ligands that are interacting with receptors. As the total number of residues in ligands are almost half of that in receptors the percentage of interacting residues are twice to that in receptors. Similar characteristics are observed for the binding site residues obtained with the contacts between residues in receptors and ligands in protein–protein complexes.

We have compared the results obtained with the energy criteria used in this work and the criteria with different cutoff distances for defining binding site residues[44–46] in Table 1. We noticed that the number and percentage of binding site residues obtained with energy based approach is similar to the one defined with the distance of 5 Å between any heavy atoms in receptors and ligands. However, the analysis of binding site residues obtained in these approaches showed significant differences between them. Only 1459 residues are common to each other and this result indicates the importance of considering the energy between different atoms to define the binding residues. In addition, 4% of the residues have strong repulsive energies and all these residues have been identified as

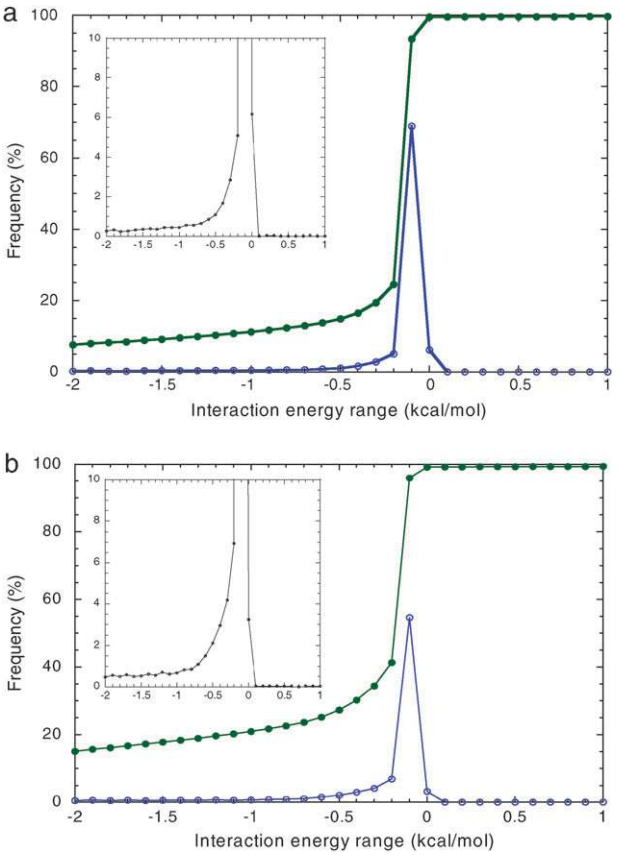


Fig. 1 Occurrence of amino acid residues in different ranges of interaction energies: (a) receptors and (b) ligands. The open and closed circles show the fraction and total percentage of residues. The expanded data for the percentage fraction of residues in different ranges of interaction energies is shown in the inset.

Table 1 Number and percentage of binding site residues using different methods

Criterion	Cutoff	N_{bind}	$\%_{bind}$	Reference
Energy	<1 kcal/mol	5255	10.8	Present work
Energy	<0.5 kcal/mol	6718	13.8	Present work
C_α distance	6 Å	1972	4.0	Keskin *et al.*[45]
C_β distance	6 Å	3449	7.1	Glaser *et al.*[46]
Any heavy atoms	5 Å	6644	13.6	Li *et al.*[44]

binding residues in distance based criteria, which are not probable binding residues in protein–protein complexes.

Behavior of different types of amino acid residues

We have classified the residues into aromatic, sulfur containing, charged, hydrophobic and polar and the percentage of residues at different energy intervals are shown in Fig. 2. We observed that a certain percentage of polar residues contribute to interact between receptors and ligands irrespective of the percentage of residues present in protein–protein complexes. Although the occurrence of Ser is 6.1% and that of Asn is 4.4% both of them contribute similar level (0.5–.6%) to binding sites as seen in Fig. 2a. The binding site aromatic residues mainly depend on the percentage of residues in the complexes as seen in Fig. 2b. Among charged residues the

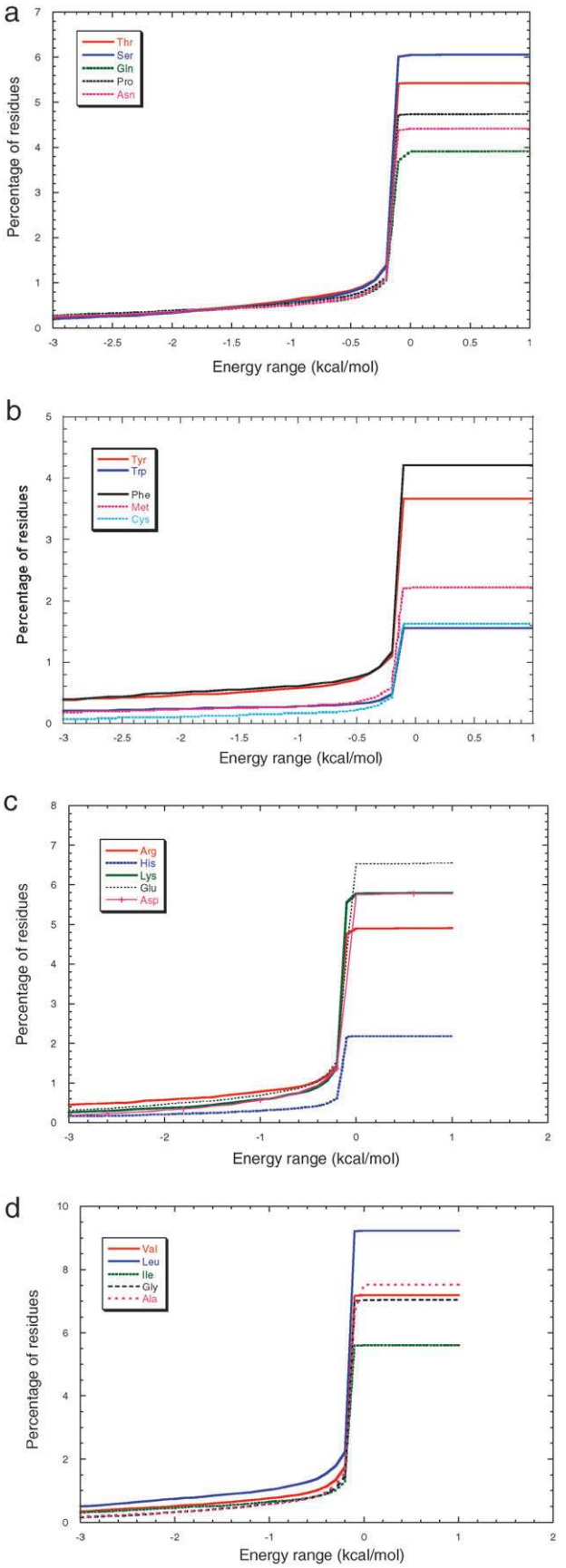


Fig. 2 Contribution of the 20 amino acid residues at different free-energy ranges: (a) polar, (b) aromatic and sulfur containing, (c) charged and (d) hydrophobic.

Table 2 Binding propensity of amino acid residues (a) in receptors, (b) in ligands

Residue	N_{bind}	N_{tot}	Propensity
a			
ALA	264	3673	7.19
ASP	261	2831	9.22
CYS	82	800	10.25
GLU	323	3233	9.99
PHE	292	2057	14.20
GLY	270	3441	7.85
HIS	140	1066	13.13
ILE	304	2735	11.12
LYS	271	2840	9.54
LEU	482	4510	10.69
MET	132	1084	12.18
ASN	240	2161	11.11
PRO	262	2317	11.31
GLN	244	1917	12.73
ARG	366	2414	15.16
SER	272	2968	9.16
THR	287	2653	10.82
VAL	354	3505	10.10
TRP	132	764	17.28
TYR	277	1807	15.33
b			
ALA	272	1856	14.66
ASP	239	1395	17.13
CYS	66	443	14.90
GLU	288	1712	16.82
PHE	262	1003	26.12
GLY	275	1703	16.15
HIS	148	539	27.46
ILE	325	1399	23.23
LYS	266	1508	17.64
LEU	477	2225	21.44
MET	128	518	24.71
ASN	219	999	21.92
PRO	231	1154	20.02
GLN	233	1064	21.90
ARG	334	1172	28.50
SER	258	1519	16.98
THR	261	1362	19.16
VAL	330	1779	18.55
TRP	110	293	37.54
TYR	235	806	29.16

contribution of Arg is the highest followed by Glu. Lys and Asp contribute similar to each other (Fig. 2c). The contribution of Leu is higher than other hydrophobic residues (Fig. 2d) as it has the highest occurrence in protein–protein complexes.

Binding residue propensity in protein–protein complexes

We have computed the binding propensity in both receptors and ligands and the results are presented in Table 2. We observed that the aromatic as well as positively charged residues contribute significantly to the interaction between receptors and ligands. Interestingly, the behavior is similar in both receptors and ligands. This result indicates the importance of cation–π, aromatic and electrostatic interactions for the recognition of protein–protein complexes. The highly favorable pairs of interacting residues also reveal the presence of several pairs formed by aromatic and charged residues. Fig. 3 shows the interaction between the pair of residues R1112 and W144 in NarGH complex (1R27) in which the interaction energy is 6.2 kcal mol^{-1}.

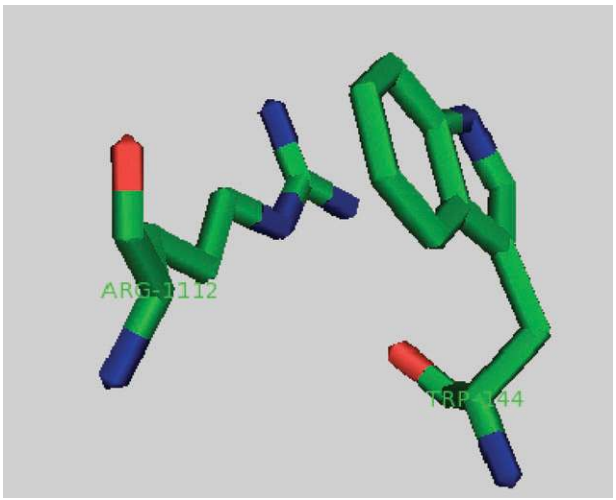


Fig. 3 Interaction between R1112 and W144 in 1R27 complex.

Binding segments in protein–protein complexes

We have analyzed the binding residues in terms of a "continuous stretch" in the amino acid sequence. The lengths of continuous binding residues are termed as binding segments. The binding segments have been analyzed in receptors and ligands in protein–protein complexes and the results are presented in Fig. 4. We observed that about half of the binding segments are accommodated with single amino acid residues so that the neighboring residues are not favorable to binding. Similar trend is also observed with other criteria used to identify the binding sites. The two and three-residue segments have 25% and 8% of the binding segments, respectively. This result differs from the binding segments of protein–DNA complexes in which most of the binding segments have 3–4 amino acid residues.[47]

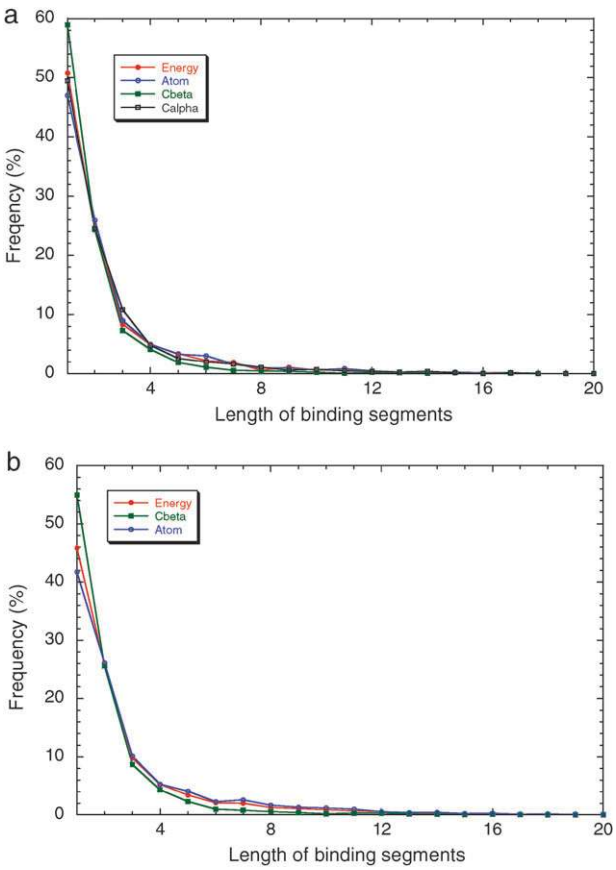


Fig. 4 Frequency of occurrence of residues at various lengths of binding segments. Energy, atom, Cbeta and Calpha denote the binding residues obtained with energetic contribution (present study), distance criteria using all atoms, C_β and C_α atoms, respectively.

Propensity of residues in different binding segments

We have further analyzed the influence of amino acid residues involved in binding in terms of binding segments. The results obtained with receptors and ligands are displayed in Fig. 5. We noticed that the single residue binding segments are mainly dominated with aromatic and positively charged residues in both receptors (Fig. 5a) and ligands (Fig. 5b). Hydrophobic residues, Ala, Ile, Leu and Val involve in the 2-residue binding segments. Further analysis indicates the presence of hydrophobic and charged residues in two-residue segments. Polar residues present in all the binding segments indicating their importance to support the interaction between receptors and ligands at all regions of binding. Due to the flexibility of Gly it occurs mainly in the binding segments with more than 2 residues.

Preference of secondary structure in binding sites

The analysis on the preference of residues occurring in different secondary structural elements indicates that most of the regular secondary structures have a similar level of

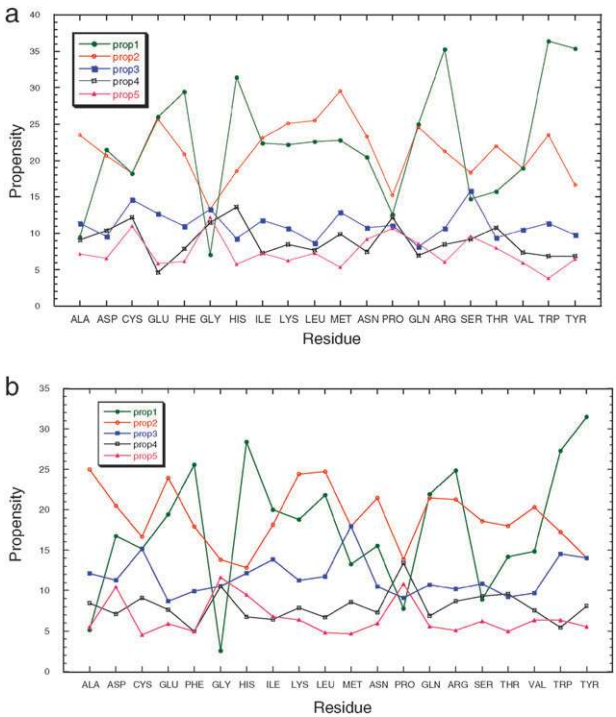


Fig. 5 Propensity of binding residues in different binding segments. Prop1, prop2, prop3, prop4 and prop5 represent the propensity of residues in binding segments of length, 1, 2, 3, 4 and >4 residues, respectively.

Table 3 Average energetic contribution due to different atoms

Atom	Receptor	Ligand
Main chain		
C	−0.29	−0.31
N	−0.26	−0.29
O	−0.31	−0.34
Side chain		
C	−0.46	−0.51
N	−0.51	−0.56
O	−0.44	−0.48
S	−0.78	−0.81

preference (about 10% in receptors and 20% in ligands). However, we noticed that both the residues in 2-residue binding segments prefer to have similar secondary structure (helix or strand or coil).

Contribution of different types of atoms in protein–protein interactions

In order to understand the importance of main chain and side chain atoms for binding we have analyzed the contribution of interaction free energy due to different atoms in protein–protein complexes. We have classified the atoms into seven groups, three for the main chain (C, N and O) and four for the side chain (C, N, O and S). The average interaction energy for each atom is presented in Table 3. We observed that the average contribution is similar in both receptors and ligands. The comparative analysis on main chain and side chain atoms showed that the contributions due to side chain atoms are significantly stronger than that due to main chain atoms. This analysis indicates the importance of specific amino acid residues for recognition in protein–protein complexes.

Preference of interacting residues in protein–protein complexes

We have also analyzed the preference for interacting residues in receptors and ligands by calculating the residue pair preference at the binding sites. The topmost residue pairs are Asp–Arg, Arg–Asp, Cys–Cys, Glu–Arg, Arg–Glu, Phe–Phe, Tyr–Trp, Trp–Tyr, His–Trp, Arg–Tyr, Tyr–Arg, Arg–Trp *etc*. This analysis shows that the pairwise interactions are mainly formed with electrostatic, cation-π and aromatic interactions. In addition these residues are formed between the interacting partners in protein–protein complexes irrespective of receptors or ligands.

Comparison with experiments

We have compared the results obtained in this work with experimental binding energies of protein–protein complexes. This has been done with the data on changes in binding free energy change upon amino acid substitutions. The search on protein–protein interactions thermodynamic database[48] showed the presence of 217 interactions, which have the difference in binding free energy of >2 kcal mol^{-1}. Generally residues that can cause the binding free energy of >2 kcal mol^{-1} are identified as hotspots. Further analysis on 217 interactions revealed that 68 of them are unique. We have analyzed all the unique interactions and we observed that 38 residues are charged, 32 of them are positive charged and aromatic. On the other hand only seven residues are hydrophobic. This result demonstrates the importance of electrostatic, cation-π and aromatic interactions for the recognition of protein–protein complexes. Our computational analysis revealed the importance of these interactions, showing the good agreement with experiments.

We have also verified the importance of such interactions with different specific complexes, such as the E6AP–UbcH7, interleukin4–receptor and Ras–Rap complexes.

(i) E6AP–UbcH7 complex. Eletr and Kuhlman[49] measured the binding free energies of 49 mutants in E6AP–UbcH7 complex (1C4Z) and 15 of them are identified as hotspots. We have analyzed all the hotspot residues and observed that 10 residues are positively charged/aromatic, 9 are positively/negatively charged and 3 are hydrophobic. Further, the replacement of F63A altered the binding free energy of 3 kcal mol^{-1}. We have analyzed the energetic contribution of F63 in 1C4Z and found that F63 in UbcH7 makes a strong aromatic interaction with Y694 in E6AP and the interaction free energy is -1.2 kcal mol^{-1}. Fig. 6a shows the aromatic interactions between the residues F63 and Y694 in E6AP–UbcH7 complex.

(ii) Interleukin4 receptor binding protein complex. Zhang *et al.*[50] carried out binding experiments on interleukin4 receptor binding protein complex (1IAR) and reported binding free energies for 29 mutants. The analysis on ΔΔG values shows the presence of 11 hotspots. Interestingly, six of them are cation-π interaction forming residues, 3 are charged and 2 are hydrophobic residues. We have analyzed the contribution due to different residues and the interaction between Y127 in interleukin4 and R85 in binding protein is shown in Fig. 6b. We observed the presence of a cation-π interaction and the substitution of Y127A changed the binding free energy to 2.2 kcal mol^{-1}.

(iii) Ras–Rap complex. Kiel *et al.*[51] studied the thermodynamic behavior of binding in the Ras–Rap complex and measured the binding free energy of 27 mutants in both Ras and Rap proteins. We noticed the presence of six hotspots and four of them involved charged residues, Lys, Arg and Asp. Five residues have the capability of cation-π interactions and there is no residue with hydrophobic behavior. In this experiment Kiel *et al.*[51] reported that the replacement of K32A in Rap with wild type Ras altered the free energy of 2.5 kcal mol^{-1} whereas the substitution of D238A in Ras with wild type Rap contributed to the free energy of 3.9 kcal mol^{-1}. We have analyzed the interaction between K32 and D238 and the contribution towards electrostatic interaction is shown in Fig. 6c. This analysis verifies the importance of electrostatic interactions obtained in this work with experimental observations.

Mechanism for protein–protein recognition

The structural analysis on protein–protein complexes and the importance of specific amino acid residues for binding suggested the following mechanism for protein–protein recognition. The aromatic residues and Arg have high binding propensity in receptors and ligands (Table 2a and 2b) and these residues are dominant in single residue binding segments. The negatively charged residues also have high proportions in

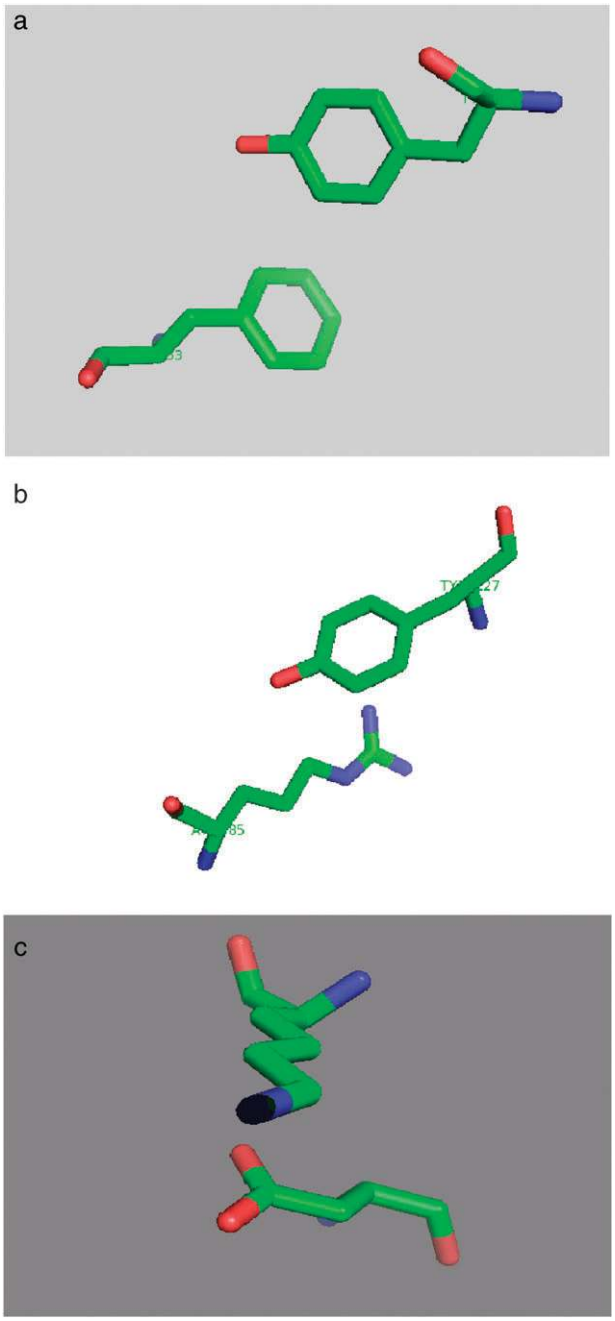


Fig. 6 Interactions between receptors and ligands in different complexes: (a) E6AP–UbcH7 complex, (b) Interleukin4 receptor binding protein complex and (c) Ras–Rap complex.

single residue binding segments. The contributions due to the atoms in side chains are higher than main chains. These results prompted the importance of electrostatic, aromatic and cation-π interactions, which may be initiating factors to protein–protein recognition.

We have compared the importance of these interactions with experiments and other computational studies reported in the literature. Kiel *et al.*[51] demonstrated the importance of long-range electrostatic interaction for the formation of the Ras–Rap complex. Zhang *et al.*[50] reported the importance of ion pairs and hydrogen bonds in the shield of hydrophobic

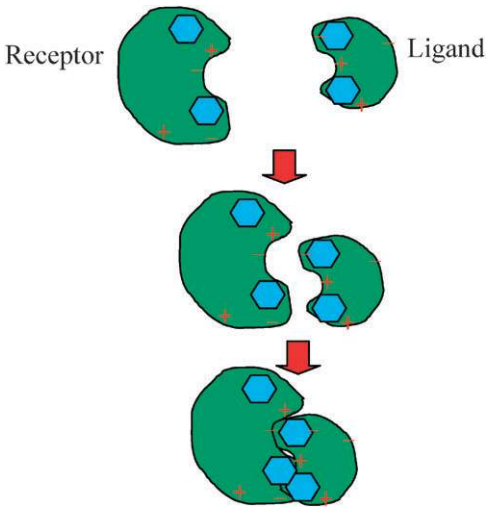


Fig. 7 Recognition mechanism in protein–protein complexes: +,– and hexagon indicates the positively charged, negatively charged and aromatic residues, respectively.

groups in the recognition of interleukin4 receptor binding complex. The importance of salt bridges, hydrogen bonds and hydrophobic interactions are revealed in the E6AP–UbcH7 complex.[49]

del Sol and O'Meara[52] analyzed 48 protein–protein complexes through small world network and reported that Trp, Arg, Gln and Asp serve as key residues in protein–protein interactions. These residues belong to form different interactions observed in the present work. Crowley and Golovin[53] reported that cation-π interactions are dominant in protein–protein interface followed by electrostatic interactions. Ansari and Helms[54] showed the pairing between charged residues. Considering the experimental and theoretical reports and the results obtained in the present study we propose that the charged and aromatic residues initiate the interactions in the form of electrostatic, aromatic and cation-π interactions (**Step 1;** Fig. 7). The two-residue binding segments are dominated with hydrophobic residues, which may contribute to hydrophobic interactions (**Step 2;** Fig. 7). The polar residues are widely present in different binding segments and have moderate binding propensities. These residues make necessary hydrogen bonds and keep the interacting partners to form the complex and maintain the stability of the complex (**Step 3;** Fig. 7). The three-step mechanism helps to understand the binding of protein–protein complexes apart from the major contribution due to hydrophobic interactions.

Protein–protein interactions in the context of systems biology

The principles of protein–protein interactions may lead toward two practical goals:[34] (i) once the major components of the stabilizing interactions are understood, it may facilitate designing drugs to block the critical interactions in cases where binding leads to disease[55] and (ii) it should facilitate prediction of interactions. Mapping of protein–protein interactions assists in predicting protein function and in the construction of interaction maps. Hence, it may be possible to map the functional networks of a proteome and all macromolecules in the cell. This will help to understand the controlling

of gene expression. Such a procedure may be viewed as a bottom-up-strategy (starting from specific contacts between molecules and building a system to create a map) in systems biology.

Conclusions

We have proposed a novel approach for identifying the binding residues in protein–protein complexes using energy based approach. Our results showed the importance of residues that are involved in electrostatic, cation-π and aromatic interactions for the formation of protein–protein complexes. Further analysis revealed that most of the binding segments are accommodated with single residues followed by two-residue segments with hydrophobic and polar residues. The importance of specific amino acid resides have also been verified with the contributions of atoms in main and side chains of residues in receptors and ligands. We have compared the results obtained in this work with the experimental binding energies of protein–protein complexes and we observed a good agreement between them. Based on the results we have proposed a mechanism for the recognition of protein–protein complexes. The approach may be an initial step for the bottom-up strategy in systems biology.

Acknowledgements

MMG thanks Dr Madan Babu for the invitation to contribute this work for the special issue of "Computational and Systems Biology". We acknowledge Prof. B. Jayaram and Dr S. Selvaraj for useful discussions. This research was supported by Strategic International Cooperative Program, Japan Science and Technology Agency (JST).

References

1 Y. Ofran and B. Rost, *J. Mol. Biol.*, 2003, **325**, 377–387.
2 P. Chakrabarti and J. Janin, *Proteins: Struct., Funct., Bioinf.*, 2002, **47**, 334–343.
3 M. Guharoy and P. Chakrabarti, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15447–15452.
4 B. Ma, T. Elkayam, H. Wolfson and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5772–5777.
5 F. B. Sheinerman and B. Honig, *J. Mol. Biol.*, 2002, **318**, 161–177.
6 T. Kortemme and D. Baker, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14116–14121.
7 P. J. Kundrotas and E. Alexov, *Biophys. J.*, 2006, **91**, 1724–1736.
8 Y. Ofran and B. Rost, *PLoS Comput. Biol.*, 2007, **3**, e119.
9 Z. Aung, S. H. Tan, S. K. Ng and K. L. Tan, *J. Bioinform. Comput. Biol.*, 2008, **6**, 415–33.
10 Y. Urakubo, T. Ikura and N. Ito, *Protein Sci.*, 2008, **17**, 1055–65.
11 M. F. Lensink and R. Méndez, *Curr. Pharm. Biotechnol.*, 2008, **9**, 77–86.
12 B. A. Shoemaker and A. R. Panchenko, *PLoS Comput. Biol.*, 2007, **3**, e42.
13 B. A. Shoemaker and A. R. Panchenko, *PLoS Comput. Biol.*, 2007, **3**, e43.
14 S. Jones and J. M. Thornton, *J. Mol. Biol.*, 1997, **272**, 133–143.
15 J. Fernandez-Recio, M. Totrov and R. Abagyan, *J. Mol. Biol.*, 2004, **335**, 843–865.
16 V. Potapov, D. Reichmann, R. Abramovich, D. Filchtinski, N. Zohar, D. Ben Halevy, M. Edelman, V. Sobolev and G. Schreiber, *J. Mol. Biol.*, 2008, **384**, 109–119.
17 M. Siki, S. Tomi and K. Vlahovicek, *PLoS Comput. Biol.*, 2009, **5**, e1000278.
18 L. P. Albou, B. Schwarz, O. Poch, J. M. Wurtz and D. Moras, *Proteins: Struct., Funct., Bioinf.*, 2009, **76**, 1–12.
19 S. Grosdidier and J. Fernández-Recio, *BMC Bioinformatics*, 2008, **9**, 447.
20 S. Gong and T. L. Blundell, *PLoS Comput. Biol.*, 2008, **4**, e1000179.
21 B. K. Dukka and D. R. Livesay, *Bioinformatics*, 2008, **24**, 2308–2316.
22 R. Najmanovich, N. Kurbatova and J. Thornton, *Bioinformatics*, 2008, **24**, i105–111.
23 A. Shulman-Peleg, R. Nussinov and H. J. Wolfson, *J. Mol. Biol.*, 2004, **339**, 607–633.
24 A. Shulman-Peleg, M. Shatsky, R. Nussinov and H. J. Wolfson, *Nucleic Acids Res.*, 2008, **36**, W260–264.
25 A. Ertekin, R. Nussinov and T. Haliloglu, *Protein Sci.*, 2006, **15**, 2265–2277.
26 I. Halperin, H. Wolfson and R. Nussinov, *Protein Sci.*, 2003, **12**, 1344–1359.
27 I. Halperin, H. Wolfson and R. Nussinov, *Structure*, 2004, **12**, 1027–1038.
28 X. Li, O. Keskin, B. Ma, R. Nussinov and J. Liang, *J. Mol. Biol.*, 2004, **344**, 781–795.
29 S. Hannenhalli and L. S. Wang, *Bioinformatics*, 2005, **21**, i204–212.
30 P. Fariselli, F. Pazos, A. Valencia and R. Casadio, *Eur. J. Biochem.*, 2002, **269**, 1356–1361.
31 A. Koike and T. Takagi, *Protein Eng., Des. Sel.*, 2004, **17**, 165–173.
32 I. Res, I. Mihalek and O. Lichtarge, *Bioinformatics*, 2005, **21**, 2496–501.
33 Y. Ofran and B. Rost, *Bioinformatics*, 2007, **23**, e13–e16.
34 O. Keskin, B. Ma, K. Rogale, K. Gunasekaran and R. Nussinov, *Phys. Biol.*, 2005, **2**, S24–35.
35 O. Keskin, A. Gursoy, B. Ma and R. Nussinov, *Chem. Rev.*, 2008, **108**, 1225–1244.
36 P. M. Kim, L. J. Lu, Y. Xia and M. B. Gerstein, *Science*, 2006, **314**, 1938–1941.
37 H. Berman, K. Henrick, H. Nakamura and J. L. Markley, *Nucleic Acids Res.*, 2007, **35**, D301–303.
38 S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.*, 1970, **48**, 443–453.
39 M. Guharoy and P. Chakrabarti, *Bioinformatics*, 2007, **23**, 1909–1918.
40 H. Hwang, B. Pierce, J. Mintseris, J. Janin and Z. Weng, *Proteins: Struct., Funct., Bioinf.*, 2008, **73**, 705–709.
41 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
42 F. Pichierri, M. Aida, M. M. Gromiha and A. Sarai, *J. Am. Chem. Soc.*, 1999, **121**, 6152–6157.
43 M. M. Gromiha and S. Selvaraj, *Prog. Biophys. Mol. Biol.*, 2004, **86**, 235–77.
44 W. Li, A. H. Keeble, C. Giffard, R. James, G. R. Moore and C. Kleanthous, *J. Mol. Biol.*, 2004, **337**, 743–59.
45 O. Keskin, C. J. Tsai, H. Wolfson and R. Nussinov, *Protein Sci.*, 2004, **13**, 1043–55.
46 F. Glaser, D. M. Steinberg, I. A. Vakser and N. Ben-Tal, *Proteins: Struct., Funct., Bioinf.*, 2001, **43**, 89–102.
47 S. Ahmad, M. M. Gromiha and A. Sarai, *Bioinformatics*, 2004, **20**, 477–86.
48 M. D. Kumar and M. M. Gromiha, *Nucleic Acids Res.*, 2006, **34**, D195–8.
49 Z. M. Eletr and B. Kuhlman, *J. Mol. Biol.*, 2007, **369**, 419–28.
50 J. L. Zhang, I. Simeonowa, Y. Wang and W. Sebald, *J. Mol. Biol.*, 2002, **315**, 399–407.
51 C. Kiel, L. Serrano and C. Herrmann, *J. Mol. Biol.*, 2004, **340**, 1039–58.
52 A. del Sol and P. O'Meara, *Proteins.*, 2005, **58**, 672–82.
53 P. B. Crowley and A. Golovin, *Proteins.*, 2005, **59**, 231–9.
54 S. Ansari and V. Helms, *Proteins.*, 2005, **61**, 344–55.
55 M. R. Arkin and J. A. Wells, *Nat. Rev. Drug Discov.*, 2004, **3**, 301–317.

