9-2006

Electrostatic Properties of Protein-Protein Complexes

Petras J. Kundrotas
Clemson University

Emil Alexov
Clemson University, ealexov@clemson.edu

Electrostatic Properties of Protein-Protein Complexes

Petras J. Kundrotas and Emil Alexov
Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634

ABSTRACT Statistical electrostatic analysis of 37 protein-protein complexes extracted from the previously developed database of protein complexes (ProtCom, http://www.ces.clemson.edu/compbio/protcom) is presented. It is shown that small interfaces have a higher content of charged and polar groups compared to large interfaces. In a vast majority of the cases the average pK_a shifts for acidic residues induced by the complex formation are negative, indicating that complex formation stabilizes their ionizable states, whereas the histidines are predicted to destabilize the complex. The individual pK_a shifts show the same tendency since 80% of the interfacial acidic groups were found to lower their pK_as, whereas only 25% of histidines raise their pK_a upon the complex formation. The interfacial groups have been divided into three sets according to the mechanism of their pK_a shift, and statistical analysis of each set was performed. It was shown that the optimum pH values (pH of maximal stability) of the complex tend to be the same as the optimum pH values of the complex components. This finding can be used in the homology-based prediction of the 3D structures of protein complexes, especially when one needs to evaluate and rank putative models. It is more likely for a model to be correct if both components of the model complex and the entire complex have the same or at least similar values of the optimum pH.

INTRODUCTION

To understand the mechanism of the protein-protein binding and to alter it, if necessary, one needs to know the three-dimensional (3D) structure of the complexes. However, only a small fraction of the existing complexes have experimentally determined 3D structures. Apparently, there is a need for developing computational methods for modeling the 3D structures of protein complexes; and development of such techniques will benefit from studies revealing basic characteristics of known 3D structures of protein complexes. Currently the complex structures are predicted by docking (1) and binding simulations (2) by fitting to electron microscopy data (3) or by threading (4). Using a scoring function based on statistically delivered pairing propensities, Sali and coworkers (5) evaluated putative complexes made up of domains that were found to be homologous to the query sequences. Aloy and Russell (6,7) have used existing 3D structures of protein complexes to deliver empirical potentials and then used these potentials to score possible pairs between two protein families. Recently we have shown (P. J. Kundrotas and E. Alexov, unpublished) that even standard homology-based methodology based solely on sequence information is capable of producing models with a high rate of correctly predicted interfacial residues. However in many cases the same pair of sequences with unknown structures (query sequences) produces several models, and tools are needed to evaluate and rank these models.

The protein-protein complexes with experimentally available 3D structures and their interfaces were studied either statistically or energetically. Statistical properties of the protein interfaces have been analyzed extensively, and obtained results were used to predict binding interfaces (9–11). Protein-protein interfaces were also predicted by Rost and co-workers using a local sequence (12). Two sets of transient protein-protein dimers, 16 transient homo- and 23 heterodimers, were statistically studied in Nooren and Thornton (13) and it was found that transient heterodimers tend to have larger interface areas compared to that of homodimers, but no significant differences were revealed for other parameters, such as interface polarity, hydrophobicity, planarity, contacts net, etc. In another study, the amino acid composition and pairing frequencies were shown to be different for six different types of interfaces (12).

The role of electrostatic interactions on the protein-protein association was also broadly studied, and it was shown (see, e.g., Hu et al. (14) and references therein) that electrostatic interactions play a more important role in the protein binding than they do in folding. It was found experimentally (15,16) and computationally (17) that most of the polar and charged residues on the protein-protein interfaces are ''hot spots'', i.e., their replacement with alanine residue critically affects protein-binding affinity. It was shown that redesigning charged interfacial residues results in a complex with better affinity (18,19). Charge complementarity appears to also be an important factor affecting the binding affinity as it was demonstrated in the case of the barnase-barstar complex (20,21). Recent experimental studies (22) of the complex TEM1-b-lactamase have demonstrated that the interface between proteins is built by modules, with main interfacial interactions being within the module and with only a few interactions between different modules. The role of electrostatic interactions in formation of protein-protein interfaces was thoroughly studied in Sheinerman and Honig (23) using the set of four proteins, and it was concluded that relative the contributions of electrostatic and hydrophobic forces to the binding depends on the interfacial area. One of the largest series of works devoted to computation of electrostatic properties for different groups of complexes is that by

Submitted March 29, 2006, and accepted for publication June 2, 2006.
Address reprint requests to Emil Alexov, E-mail: ealexov@clemson.edu.

0006-3495/06/09/1724/13 $2.00
doi: 10.1529/biophysj.106.086025

McCammon and co-workers (24–29). In an earlier review (25) the role of salt bridges across the complex interfaces has been scrutinized. In the following works, the electrostatic calculations were performed for the binding of antibiotics to a small ribosome unit (26) for the enzyme active sites (27), for the kinase-ligand binding (28), and for the protein kinase systems (29). One of the findings in those works was that interfacial acidic residues in enzyme active sites are the proton donors and consequently have their pK_a values shifted toward neutral or sometimes even toward the basic pH region.

Formation of a complex may change the pK_a values of titratable residues in respect to the pK_as in the unbound monomers, especially if these residues are located within the interface of the complex. These pK_a shifts can be used as an indicator of electrostatic energy contribution of a particular residue to the monomer's binding. For instance, an acidic residue, which has its pK_a shifted upon the complex formation toward acidic pHs (negative pK_a shift), stabilizes the complex as compared to the complex with this residue replaced by a noncharged (e.g., Ala) group. A basic group has the same stabilizing effect if its pK_a value shifts upon the complex formation toward basic pHs (positive pK_a shift). However, the contribution of the pairwise electrostatic interactions to the complex formation cannot be estimated from the total pK_a shift alone. The total pK_a shift consists of two major energy components: the desolvation penalty and the pairwise interactions across the interface. Since the desolvation penalty is always unfavorable, the total pK_a shift reflects how well electrostatic interactions compensate for the burial of interfacial charges caused by the complex formation.

A primary focus of our work is on electrostatic energies and statistics of pK_a shifts in protein-protein complexes. We take advantage of a recently developed database of protein-protein complexes (ProtCom) (P. J. Kundrotas and E. Alexov, unpublished), which, after applying some selection rules, provided us with a set of 37 protein-protein complexes. Such a number of complexes is large enough for statistical studies and at the same time computationally tractable for full-scale calculations of the electrostatic energies and pK_a values of ionizable groups. We aimed at revealing the correlations in the interfacial and macroscopic properties which than can be further used for evaluation and ranking of the putative models obtained in homology-based prediction of the 3D structures of protein complexes.

METHODS

Protein-protein complexes used in the study

We extracted the set of protein-protein complexes from the previously developed ProtCom database (www.ces.clemson.edu/compbio/protcom). The detailed description of the database will be published elsewhere (P. J. Kundrotas and E. Alexov, unpublished) and here we mention only that it contains 251 two-chain complexes and 708 two-domain structures with experimentally determined 3D structures. Any sequence of the database entries has <95% identity to any other sequence therein. In this study we used only the structures of protein-protein complexes. For the sake of computational efficiency we have chosen only complexes with both monomers having between 100 and 200 amino acids. This resulted in 37 complexes with diverse functions (see Table 1), which include 9 antibody-antigen complexes, 8 enzyme-inhibitors structures, 5 transport proteins, and 12 other ensembles like binding complexes, receptors, etc.

Electrostatic calculations

The pH dependence of the stability of the complexes and their monomers was calculated using the formula (31)

$$\Delta G(\mathrm{pH}_1) = 2.3\,RT \int_{\mathrm{pH}_0}^{\mathrm{pH}_1} \Delta Q(\mathrm{pH})\, d(\mathrm{pH}), \qquad (1)$$

where $\Delta Q(\mathrm{pH}) = Q_f(\mathrm{pH}) - Q_u(\mathrm{pH})$ is the difference between molecule net charges in folded, $Q_f(\mathrm{pH})$, and unfolded, $Q_u(\mathrm{pH})$, states, R is the universal gas constant, and T is the temperature (in K). The unfolded state of the proteins was modeled in this study as a chain of noninteracting residues, and we calculated the net charge of unfolded protein molecules using the Henderson-Hasselbalch equation

$$Q_u(\mathrm{pH}) = \sum_{i=1}^{N} \frac{10^{-2.3y(i)(\mathrm{pH}-\mathrm{pK_a}(i))}}{1 + 10^{-2.3y(i)(\mathrm{pH}-\mathrm{pK_a}(i))}} \qquad (2)$$

with the set of the standard pK_a values for the titratable amino acids (32). In Eq. 2, $y(i)$ is $+1$ for basic groups and -1 for acidic groups, and the summation runs over all N titratable residues. Since in this work we will compare the electrostatic properties of the complexes and the monomers, which have the same unfolded states, the model of the unfolded state is not crucial.

The charge of the folded state, $Q_f(\mathrm{pH})$, and the pK_a values and their components (desolvation penalty and interactions with other charges) for ionizable groups were calculated by means of the multi-conformation continuum electrostatics (MCCE) method as previously described (32–34) (URL: www.sci.ccny.cuny.edu/~mcce/). Calculations were performed for all 37 selected complexes for the entire complex and for each of the monomers. The structures of the monomers were kept as they are in the complex, i.e., no conformational changes during the complex formation were considered in this study. The internal dielectric constant of protein moiety was set to 4 and the ionic strength was 0.15 M. The calculations were carried out in the pH interval from 0 to 14.

The main implication of Eq. 1 for our study is that it is possible to determine the pH value at which ΔG has a minimum (the pH of maximal stability (35) or optimum pH) solely from the $\Delta Q(\mathrm{pH})$ dependence. Indeed, $\Delta G(\mathrm{pH})$ has a minimum at the pH value where $\Delta Q(\mathrm{pH})$ crosses zero. In most cases, the $\Delta Q(\mathrm{pH})$ is a monotonic function and crosses the zero point only at a particular pH. Nevertheless, in some cases, $\Delta Q(\mathrm{pH})$ crosses the zero line more than once, which results in several local $\Delta G(\mathrm{pH})$ minima. In such cases optimum pH corresponds to the pH where $\Delta G(\mathrm{pH})$ has a global minimum.

The pK_a shifts are conveniently reported in pH units, and the energies of interactions are given in kcal/mol. The conversion coefficient between pH units and kcal/mol is 1.36 at room temperature.

Calculations of interfacial properties

In this study we define that two residues are on the complex interface (contacting residues) if they belong to different monomers and the distance (in the complex) between their closest atoms is less than the sum of corresponding van der Waals radii plus the size of one water molecule. Defining interfacial residues in such a way produces a residue list that may be larger compared to a list produced using change in solvent accessible surface area (SASA) of a residue upon complex formation. We argue, however, that

TABLE 1 Protein complexes chosen from the ProtCom database for the pH-optimum calculations

Pdb ID	Title	Function*	Number of groups†		Reference
			Larger part	Smaller part	
1c3a	Flavocetin-A from habu snake venom	O	135	125	(45)
1c40	Bar-headed goose hemoglobin	T	146	141	(46)
1doa	GTP-binding protein CDC42 in complex with RHOGDI	O	200	190	(47)
1ds6	RAC-RHOGDI complex	O	181	179	(48)
1f3v	Complex N-terminal domain of TRADD/TRAFF domain of TRAF2	O	164	158	(49)
1fgv	FV fragment of anti-CD18 antibody H52	A	120	107	(50)
1jat	MMS2/UBC13 ubiquitin conjugating complex	E	152	132	(51)
1of5	MEX67-MTR2 complex	T	154	128	(52)
1ow3	RHOA.GDP.MGF3 in complex with RHOGAP	E	196	179	(52)
1p4i	SCFV against peptide GCN4	A	112	110	(53)
1sq2	Nurse shark antigen receptor variable domain in complex with lysozyme	A	129	112	(54)
1th8	Bacillus anti-σ factor SpoIIAB in complex with anti-anti-σ SpoIIAA	E	132	115	(55)
1tnr	Tumor necrosis factor R55 complexed with tumor necrosis factor-β	O	144	139	(56)
1tx4	RHO/RHOGAP/GDP.AL F4 complex	O	196	174	(57)
1us8	RAD50 signature motif	E	140	129	(58)
1v1p	SSL from *Staphylococcus Aureus*	A	191	191	(59)
1veu	P14/MP1 complex	E	119	118	(60)
1vfa	FV fragment of mouse monoclonal antibody D1.3	A	116	108	(61)
1wmu	Hemoglobin D from the aldabra giant tortoise	T	146	141	(62)
1x2w	APO-HABU IX-BP	O	129	123	(63)
1xex	SMC protein	O	166	161	(64)
1xfp	CAB-Lys3 in complex with hen egg white lysozyme	A	131	129	(65)
1xxw	Heterodimer of phospholipase A2 from *Naja sattifera*	E	119	119	(66)
1y75	Catalytically inactive phospholipase A2	E	118	118	(67)
1ycs	P53-53BP2 complex	O	193	191	(68)
1z92	Interleukin-2 with its α receptor	O	123	121	(69)
1zbd	G-protein RAB3A complexed with the effector domain of RABPHILIN-3A	O	171	119	(70)
1zbx	ORC1P-SIR1P complex	O	197	121	(71)
1zhi	Complex of the *S. Cerevisiae* ORC1 and SIR1 interacting domains	O	195	125	(72)
2a5g	Cholera toxin A1 subunit bound to ARF6(Q67L)	E	185	162	(73)
2b59	The type II cohesion dockerin complex	O	166	156	(74)
2bfv	Monoclonal antibody fragment FV4155 from *E. Coli.*	A	119	112	(75)
2bjm	SPE7: Anthrome complex	A	120	110	(76)
2dlf	FV fragment from antidansyl switch variant antibody IGG2A(S)	O	120	113	(77)
2mhb	Hemoglobin, horse, aquo met	T	146	141	(78)
2ngr	Transition state complex for GTP hydrolysis by CDC42	O	196	191	(79)
3hhb	Hemoglobin (deoxy)	T	146	141	(80)

*Abbreviations for protein functions: A, antibody-antigen; E, enzyme-inhibitor; T, transport; O, others.
†Numbers of groups were determined as numbers of C_α atoms for the corresponding complex components in the ATOM section of pdb files.

distance criteria are better in recovering interfacial residues, since a SASA-based definition can miss many buried residues that are close to the complex interface and therefore can make contacts with a residue in another monomer.

The interface area, S_I, was calculated as the difference in the gross SASA of the monomers, $S_A + S_B$ (hereafter we denote the larger monomer as A and the smaller one as B) and the SASA of the complex, S_{AB}:

$$S_I = \frac{1}{2}(S_A + S_B - S_{AB}). \quad (3)$$

The SASAs were calculated by means of the SURFV program (36) developed in the Honig lab (URL: trantor.bioc.columbia.edu) with a water probe radius of 1.4 Å and with default atomic radii. The relative interfacial area (RS_I) is defined as the ratio S_I/S_{AB}.

For the statistical analysis of the pK_a shift, we have defined the fraction of interfacial residues of specific type Z with a particular property Y:

$$n_I^Z(Y) = \frac{N_I^Z(Y)}{N_I^Z}, \quad (4)$$

where $N_I^Z(Y)$ is the number of interfacial groups of type Z having a property Y and N_I^Z specifies the total number of interfacial residues of type Z. The property Y stands for the type of the pK_a shift (large, small, positive, or negative) and Z stands for acidic (Glu plus Asp residues) or His residues. For this quantity we considered only entire complexes, and hence both the numerator and denominator in Eq. 4 reflect the total number of interfacial acidic residues in the entire complex.

We have also used the relative density of residues with a certain property defined as

$$n_{rel}(Y) = \frac{N_I(Y)}{S_I} \quad (5)$$

with the $N_I(Y)$ being the total number of interfacial residues having the property Y, which in this case stands for the type of the residue with respect to the physical chemical classification of the amino acids (hydrophobic, Y=H; polar, Y=P; charged, Y=C). Henceforth, the ''macroscopic'' quantities $n_{rel}(H)$, $n_{rel}(P)$, and $n_{rel}(C)$ will be referred to as hydrophobicity, polarity, and relative charge density of the complex interfaces, respectively.

RESULTS AND DISCUSSIONS

Statistical analysis of the interfacial residues

Since our data set comprises a limited number of proteins, we first addressed the issue of how representative this set is compared to the universe of all proteins. For this purpose, we calculated occurrence frequencies for all 20 types of amino acids in the data set. Then we compared these frequencies to the ''background'' frequencies for all known proteins in the SWISS-PROT sequence database (as of February 2006). It was found (see supplementary materials) that the amino acid composition of the complexes studied in this work does not differ statistically from the background distribution and therefore our data set can be considered a representative one.

In addition, we studied the frequencies of interfacial residues. We found that the occurrence frequencies on the complex interfaces do not differ considerably from the compositions of both entire complexes and of all known proteins. Several hydrophobic (alanine, valine, and leucine) and polar (threonine) groups are underrepresented in our protein set, whereas other polar (tyrosine, asparagine, and tryptophan) groups and glycine are overrepresented. A similar conclusion was reached in Lu et al. (37) on a much larger set of dimers (768 entries) that includes also homodimers and interfaces created from the multi-chain entries in the Protein Data Bank (PDB). The absolute numbers and trends for interfacial residues being under- or overrepresented (except for tryptophane and glycine) are very similar to those observed in this work. The similar values for occurrence frequencies of interfacial residues were also reported in Glaser et al. (38) on the nonredundant set of 621 protein-protein interfaces. Overrepresentation of the tyrosine residue was also observed in the work (14). However a different conclusion was reached in the study by Ofran and Rost (12) where six types of protein-protein interfaces were analyzed for the set of 1812 pdb structures. For that set of proteins, the residue composition on all types of interface (including interfaces of heterocomplexes which correspond to the interface type in our protein set) differ considerably from the ''background'' frequencies for a majority of the amino acids, which most likely is related to a different way of defining interfacial residues.

Statistics over pK_a shifts

We performed the MCCE calculations for each of our 37 complexes and for each component of each complex separately. The primary goal of this study is to investigate the role of electrostatics on the complex formation. It is convenient for this purpose to study the shifts of pK_a values of residues on interfaces of complexes or, in other words, to measure how the ionization energy of the interfacial residues change upon complex formation. We left out of our considerations the lysine and arginine groups since they usually have pK_a values in the very basic pH region far away from physiological pH range where association/dissociation of the complexes takes place. Basic histidine residues can have their pK_a values in the pH region of complex formation and therefore they are included in our consideration. However, the frequency of occurrence for histidine residues is low (~3.5%) compared to the aspartic and glutamic acids (~7%) and thus one should keep in mind that tendencies observed for histidine residues, which will be discussed below, could be statistically not very reliable.

As a measure of ''macroscopic'' influence of charge-charge interactions on complex formation, it is tempting to study an average shift of pK_a values for a certain protein defined as

$$\langle \Delta pK_a^Z \rangle = \frac{1}{N_I^Z} \sum_i \Delta pK_a^Z(i), \tag{6}$$

where $\Delta pK_a^Z(i)$ is the pK_a shift for the ith interfacial residue of type Z and summation runs over all N_I^Z interfacial groups. The data plotted in Fig. 1 show a slight tendency for the increase of the relative polarity and charge density of the interfaces to induce predominantly positive pK_a shifts for both acidic and histidine residues. Another, the most pronounced, effect observed in Fig. 1 is that that the majority of the proteins have negative $\langle \Delta pK_a^Z \rangle$. For the interfacial acidic groups, negative $\langle \Delta pK_a^Z \rangle$ values were observed for 24 complexes among the total 37 studied complexes, resulting in 65% of the cases. This

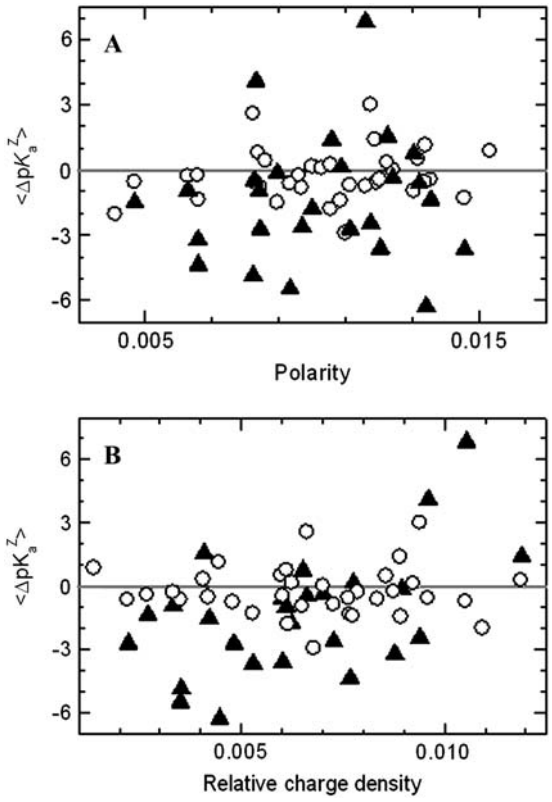

FIGURE 1 Average pK_a shifts (Eq. 6) for acidic (*open circles*) and histidine (*solid triangles*) interfacial residues of protein complexes in Table 1 plotted versus polarity (*panel A*) and relative charge density (*panel B*) of interfaces. Solid gray lines in all panels show zero level and are guides for the eye.

tendency is even more pronounced for the histidine residues. Out of 21 complexes with histidine residue at the interface, 16 have negative $\langle \Delta pK_a^Z \rangle$, which is 76% of the cases. This brings forward a conclusion that acidic groups have on average a stabilizing effect on the complex formation, whereas histidines tend to have an unfavorable contribution (the same sign of $\langle \Delta pK_a^Z \rangle$ for the acidic and for the histidine groups means an opposite effect).

When a complex is formed, the total pK_a shift, ΔpK_a, for a residue includes several components, two of them being the most important for the complex formation. First is the desolvation penalty, ΔpK_{sol}, due to the removal of favorable interactions with water molecules ($\Delta pK_{sol} > 0$). The second major component in ΔpK_a stems from the charge-charge interactions mostly across the interface and usually has an opposite sign compared to ΔpK_{sol}. Interplay of those components results in either stabilizing ($\Delta pK_a < 0$) or destabilizing ($\Delta pK_a > 0$) electrostatic contribution of an interfacial group to the complex formation. Panel *A* in Fig. 2 shows the distribution of positive and negative pK_a shifts for the acidic groups of the considered proteins. As seen, the vast majority (~80%) of the interfacial acidic groups have a total stabilizing contribution into association of the complex. Still ~20% of the groups have a destabilizing contribution. Due to the relatively small number of interfacial histidines in our data set (52 histidines, compared to the 253 interfacial acidic residues), we did not plot the same general plots as in Fig. 2 for histidines, but some general tendencies for individual histidine residues will be discussed below.

Thus, for assessing the total electrostatic contribution for a whole complex, an analysis of the pK_a shifts with respect to their magnitude is needed. In this study we defined the pK_a shift as small if $|pK_a| < 1$ and large if $|pK_a| > 1$ and analyzed the distributions for small and large pK_a shifts (Fig. 2 *B*). The distribution of the large pK_a shift peaks at 50%, whereas the distribution of small pK_a shift results in a broad maximum ranging from 20% to 50%. The finding that 50% of interfacial acidic groups have their pK_as shifted by more than one pH unit indicates that individual charged groups, in principle, can contribute noticeably into the protein binding, although their total contribution is anticipated to be rather small compared to the contribution of the hydrophobic effect, for instance.

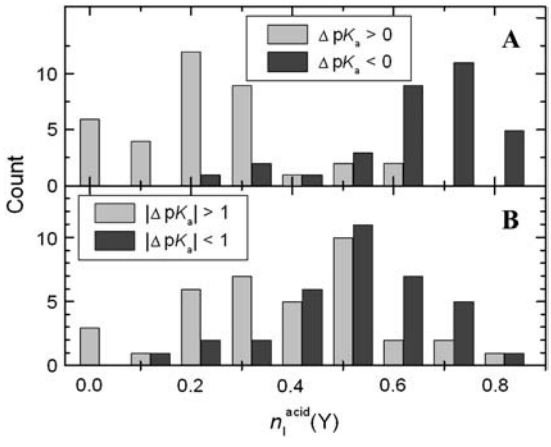

FIGURE 2 Counts of fractions of interfacial acidic groups with positive and negative pK_a shifts (*panel A*) and with small and large pK_a shifts (*panel B*) in the protein set from Table 1.

Origin of the pK_a shifts upon complex formation

For detailed analysis we compare the energy components of the pK_a shifts for interfacial acidic and histidine residues. We noticed that large pK_a shifts stem from three distinct situations. A residue can form favorable interaction(s) with opposite charged residues belonging to the other complex component. The energy of such interactions may be much stronger than the energy loss due to the desolvation penalty upon complex formation. A spectacular example of such a situation is demonstrated in Fig. 3, where a fragment of interface between two chains of protein flavocetin-A from habu snake venom (protein pdb ID 1c3a) is shown. Asp-087 residing on the loop in chain A enters upon the complex formation into a pocket formed by three positively charged histidines. The distances between the nitrogen atoms of these histidines and the oxygens of the aspartic acid is quite large for strong interactions, but this pocket is capable of ''guiding'' the aspartic acid toward the Lys-322 of chain B, with which the salt bridge is formed. In a hypothetical monomeric state, the pK_a value of the Asp-087 residue is calculated to be only slightly perturbed from the standard pK_a value. This residue is preburied ($\Delta pK_{sol} = 1.27$), but this shift is almost compen-

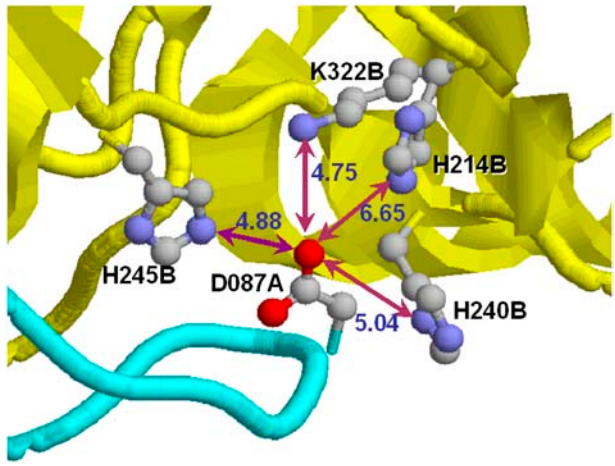

FIGURE 3 Fragment of interfacial structure for the flavocetin-A from habu snake venom (protein pdb ID 1c3a). The backbone of chain A is shown in cyan, and the backbone of chain B is drawn in yellow. Side chains of the amino acids contributing to favorable electrostatic interactions are presented as their ball-and-stick models with residues' names and numbers shown next to them. Arrows mark closest oxygen-nitrogen pairs of these residues with numbers displaying distances in angstroms.

sated for by the formation of the hydrogen bond with the neighboring Ser-089 residue ($\Delta pK_{pol} = -1.64$). In the complex the degree of burial for this residue remains almost the same as in the hypothetical ''free'' state, resulting almost to the same desolvation energy $\Delta pK_{sol} = 1.09$. However in the complex, the Asp-089 residue has a pK_a value of -1.13, i.e., at all physiologically relevant pHs this residue remains charged. This large shift originates from the favorable interactions formed with the His-214 (energy of interaction is -1.2 kcal/mol), His-240 (-3.0 kcal/mol), His-245 (-2.4 kcal/mol), and Lys-322 (-3.0 kcal/mol) residues. The ''mirror'' situation (when chains A and B change places) is almost the same except that two out of three ''pocket'' histidine residues are replaced by tyrosine and glutamic acid and the interactions with these residues are of hydrogen-bond origin rather than pure charge-charge interactions. Hereafter we will refer to a residue which participates in favorable interactions across the complex interface and does not pay a large desolvation penalty upon complex formation as the I-type residue. Such a case was previously analyzed (20,21,23), and it was shown that ''preburied'' residues may favor the binding.

Another situation is shown in Fig. 4 for the complex between the origin recognition subunit and regulatory protein SIR1 (PDB ID 1zbx). The residue Asp-503B exhibits a large pK_a shift (from 5.46 in the ''free'' state down to -0.27 in the bound state) upon complex formation too, but the origin of this shift is qualitatively different. The Asp-503B residue even in the ''free'' state is highly buried ($\Delta pK_{sol} =$ 5.46) by the hydrophobic surrounding (Leu-520B, Phe-494B, Ile-510B) and has the potential to form a strong salt bridge with the Arg-493B residue. However the Arg-493B residue is fully exposed to the solvent and has large conformational flexibility. MCCE has detected five possible conformations for this residue and all of them have a considerable population at physiological pH. Therefore the average energy of the Asp-503B-Arg-493B interaction is only ~ -3.5 kcal/mol. Such a relatively low value and few favorable interactions with other permanent charges ($\Delta pK_{pol} = -1.64$) give rise to the almost unperturbed pK_a value for the Asp-503B residue in the ''free'' state. However, the presence of the other chain in the complex reduces greatly the conformational flexibility of the Arg-493B residue, ''pushing'' it close to Asp-503B. Now only two Arg-493B conformations are populated at the physiological pH and the average energy of interactions is now ~ -13 kcal/mol. The desolvation penalty and the interaction with the permanent charges do not differ considerably for the bounded state ($\Delta pK_{sol} = 5.16$ and $\Delta pK_{pol} = -2.62$) thereby resulting in the large pK_a shift for the Asp-503B residue upon the complex formation. Note that the increase of the interactions with the neighboring partner results also from the change of the dielectric boundary since the interface is buried in the complex. Hereafter we will refer to a residue which strengthens its existing interactions upon complex formation as the S-type residue. The role of intramolecular interactions to binding affinity has been previously discussed in Sims et al. (29) for the protein kinase systems, and it was concluded that such interactions can in some cases play a crucial role in the protein-ligand recognition process.

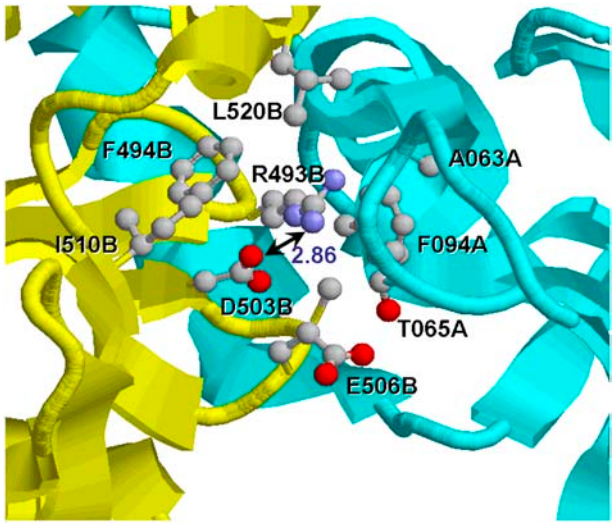

FIGURE 4 Fragment of interfacial structure for complex between origin recognition subunit 1 and regulatory protein SIR1 (protein pdb ID 1zbx). The backbone of chain A is shown in cyan, and the backbone of chain B is drawn in yellow. Side chains of the amino acids directly or indirectly contributing into electrostatic interactions are presented as their ball-and-stick models with residues names and numbers shown next to them. Arrow marks closest oxygen-nitrogen pair with number displaying distance in angstroms.

Finally, there are many residues on the interfaces which do not participate in the strong interactions with other charged residues either within the same chain or across the interface. Their pK_a shifts upon complex formation originate from small changes in desolvation penalties along with changes in many weak (<1 kcal/mol) interactions with the charged and polar groups. We called such residues the 0-residues.

Note that the two examples discussed above are for the acidic residues, but the same pairwise interaction mechanisms are, in general, valid for the histidines as well. Thus, we applied the above classification scheme (I-, S-, and 0-type residues) also for the histidines.

Distribution of different types of interfacial residues

Fig. 5 shows distributions of the pK_a shifts for all interfacial acidic residues and for each residue type described above. Without the residue categorization the distribution has an almost symmetrical bell-like shape with the center around zero shifts (Fig. 5 *A*). Categorized distributions reveal that I-type residues tend to have negative pK_a shifts (Fig. 5 *B*). The same tendency, though not so clearly pronounced, is observed for the S-type residues (see Fig. 5 *C*). The 0-residues,

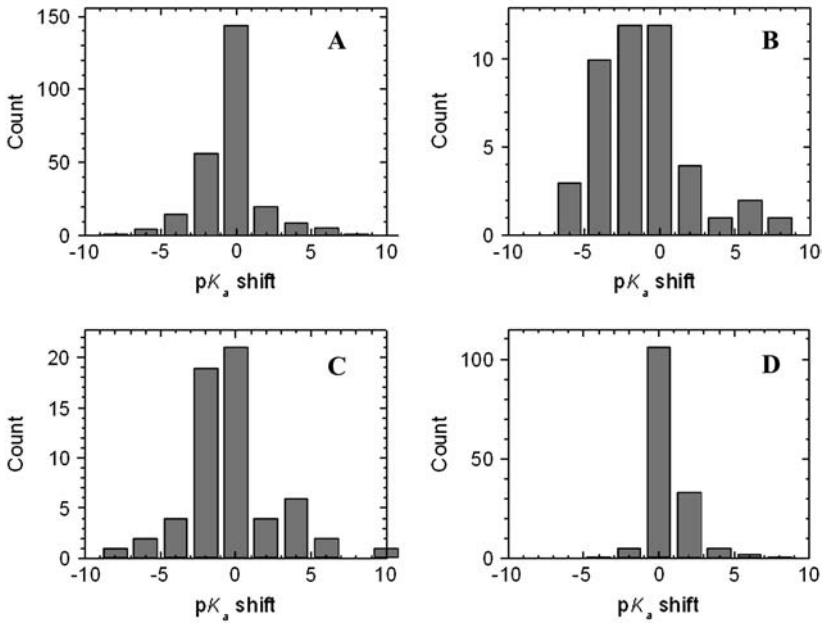

FIGURE 5 Distributions of pK_a shifts for acidic groups in protein complexes from Table 1. Panel *A* displays the distribution for all interfacial acidic groups, and panels *B*, *C*, and *D* show the distributions for I-, S-, and 0-type residues (for the definitions, see the text).

however, tend to have positive pK_a shifts as demonstrated in Fig. 5 *D*. Note that for some 0-residues the resulting pK_a shift can be large, up to 6 pH units, but this is due to the large number of small contributions rather than to strong interactions. Thus, if a charged acidic interfacial residue participates in strong charge-charge interaction (I- or S-type), a favorable increase of these interactions would outweigh an unfavorable increase in the desolvation penalty or, in other words, there will be a total stabilizing effect of electrostatic interactions on the complex formation. However, the residue should be ''preburied'', as indicated by other studies in the field (39,40). If there are no strong charge-charge interactions (0-type acidic residues), then an unfavorable increase in the desolvation penalty in most of the cases would prevail, and the gross contribution of electrostatic interactions into complex binding would be destabilizing.

Due to the small number of interfacial histidine we did not present the results graphically, but some general tendencies have been observed for histidines as well. There are nearly equal amounts of all three types of histidine residues (17 of the I-type, 15 of the S-type, and 20 of the 0-type). Also, almost equal portions of the histidine residues have either small total electrostatic effect on the complex formations (24 histidines with $|\Delta pK_a| < 1$) or a large destabilizing contribution (25 residues with $\Delta pK_a < -1$). Only three histidines have large stabilizing contributions ($\Delta pK_a > 1$) into the complex binding (all three are of the I-type). The same asymmetry holds for the I-, S-, and 0-type residues separately (e.g., nine of the S-type histidines have $|\Delta pK_a| < 1$, six residues display $\Delta pK_a < -1$, and no residues possess $\Delta pK_a > 1$). Note that most of the pairwise interactions involving interfacial histidine residues contribute favorable to complex formation, as discussed above for the acidic residues. However, in the majority of the cases the prevailing effect for the histidines is increase of residue burial upon complex formation. This effect outweighs the increase in favorable pairwise interactions, and the resulting electrostatic effect is unfavorable, as demonstrated in Fig. 1.

In accordance with the findings of Nielsen and McCammon (41), we have also observed on the interfaces of enzyme-inhibitor complexes aspartic and glutamic acids that exhibit unusually large positive pK_a shifts upon complex formation. For example, Glu-016A of the 1th8 complex has a pK_a of 15.0 (6.5); Asp-822A of the 1us8 complex shows a pK_a of 7.2 (4.5); Asp-075A of the 1veu complex has a pK_a of 10.3 (3.3); and Asp-024A and Asp-024B of the 1xxw complex possess a pK_a of 7.9 (3.9) and 11.09 (4.1), respectively (pK_a values are given here for both the bounded and the free (in parentheses) states). It is interesting to note that we observed this kind of behavior for other types of complexes in our data set as well. For instance, Glu-385A and Asp-463A of the 1of5 complex also exhibit a large positive pK_a shift from 3.9 and 2.2 in the ''free'' state up to 8.5 and 7.2 in the bounded state, respectively.

The analysis of the complexes in our data set revealed three types of interfaces: a), interfaces with 0- and S-type groups, b), interfaces with 0- and I-type groups, and c), interfaces with all three types of residues presented on the interfaces. In this context it is interesting to correlate the presence of the S- and I-type groups with the functions of the proteins in the data set. Table 2 displays such a correlation with respect to the four groups of functions in our data set. Note that the antibody-antigen complexes do not have any preference in residue type whereas enzyme-inhibitor complexes tend to have at least one I-type residue and transport proteins are inclined to have at least one S-type residue. Other (including signaling,

TABLE 2 Correlation between protein function and type of acidic residues on the complex interface

Function of protein	Number of proteins in the set			
	Total with given function	With at least one I-type group	With only S-type groups	With only 0-type groups
Antibody-antigen	9	2	3	4
Enzyme-inhibitor	8	7	0	1
Transport	5	0	5	0
Others	15	10	5	0

DNA-binding, etc.) complexes tend to have both I- and S-type interfacial residues but ''dislike'' interfaces with the 0-residues only.

pH dependence of the net charge for complexes and their components

Most proteins have maximal stability at a particular pH (optimum pH). It was demonstrated earlier (35) that optimum pH can be determined by finding the pH value at which a protein has a minimum in the free energy of folding or where the difference, ΔQ, in the net charges of the folded and unfolded state is zero (see Eq. 1). We have obtained the ΔQ(pH) dependency for each of the complexes in our data set and for each monomer separately. Two typical sets of calculation results are shown in Fig. 6 for two protein complexes, one of them being the 1c3a dimer with the I-type interfacial group demonstrated in Fig. 3, whereas another example is for the 1zbx complex with the S-type group displayed in Fig. 4.

As seen, the difference in electrostatic interactions of the interfacial acidic groups for these two proteins is not reflected in the ΔQ(pH) dependence. In both cases one can clearly see three distinguishable pH regions. In the first one, in the acidic pH range, $\Delta Q < 0$ (proton release) and has a minimum corresponding to the titration of acidic groups with pK_a values lower than standard pK_a values. The third region, at very basic pHs where $\Delta Q > 0$ (proton uptake) and has maximum is due to the titration of arginine and lysine groups with ''normally'' perturbed pK_a values (pK_a in the folded state is larger than pK_a in the unfolded state).

The most interesting region (the second region) is at the intermediate pH range, where titration of histidine groups along with acidic and basic groups with unusual pK_a shifts takes place. This region is the most interesting since many proteins function here. Optimum pH is usually at the border between the first and second regions (see Fig. 6, for example) but sometimes might be shifted toward basic pH values when, for instance, there are several groups with ''abnormal'' pK_a shifts (toward acidic pH for basic groups and toward basic pHs for the acidic groups). It should be noted, however, that ΔQ values in this pH region are often determined by titration of residues with their pK_a values spread sometimes uniformly over the whole region (contrary to the acidic and basic regions where pK_a values usually tend to be grouped). Therefore this region is sensitive to numerical instabilities, and statistical noise of the algorithm and analysis of the data in this region may not be trivial.

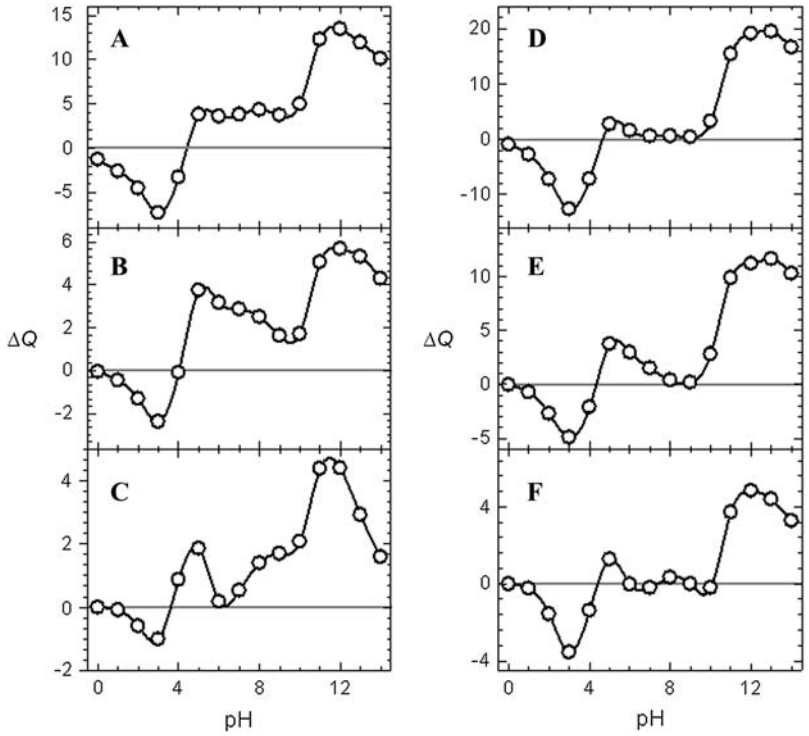

FIGURE 6 pH-dependence of difference between charges in folded and unfolded states, ΔQ, for flavocetin-A from habu snake venom (protein pdb ID 1c3a, *panels A–C*) and for the complex between origin recognition subunit 1 and regulatory protein SIR1 (protein pdb ID 1zbx, *panels D–F*). Panels *A* and *D* show the ΔQ (pH) function for the whole complex, and panels *B* and *E* (*C* and *F*) display this dependence for the larger (smaller) part of the complex. Solid black lines are guides for the eye and solid gray lines represent the $\Delta Q = 0$ level.

Correlation between optimum pH values of the complexes and their components

To achieve the main goal of this study, namely to investigate possibilities of using electrostatic properties for evaluation of structural models, we have analyzed how the optimum pH of the complex is correlated with the optimum pH of the components and how the pH of the components are correlated between themselves. The optimum pH was calculated as discussed above. The results of MCCE calculations are shown in Fig. 7 for 31 proteins. The remaining six proteins do not exhibit clearly pronounced ΔQ(pH) dependence (which indicates that titratable residues are not involved in strong interactions) and therefore were left out of the figure. As seen, most of the points are located close to the $y = x$ straight line, which indicates that optimum pH values of the complex tend to be close to optimum pH values of the monomers. There are several exceptions with the two most prominent (1tx4 and 1ycs) labeled in Fig. 7 and discussed in detail below.

Our calculations indicate that the components of the 1tx4 complex (a complex between GTPASE-activating protein RHOGAP and transforming protein RHOA, which functions as a regulator of phosphorylation pathways) have very distinct optimum pH values. For the larger monomer (RHOGAP), this value is in the basic region (pH optimum = 10.0), whereas the smaller component (RHOA) has optimum pH in the acidic region (pH optimum = 4.5). For the entire complex it was difficult to determine optimum pH unambiguously since ΔQ was close to zero in the wide pH region (5–10), and in Fig. 7 we report the lowest pH where ΔQ crosses the zero line. For another ''offbeat'' complex, the DNA-binding complex P53-53BP2 (pdb ID 1ycs), the situation is the opposite. The larger component has optimum pH in the acidic pH range (optimum pH 3.8), whereas the smaller component shows optimum pH in the basic region (optimum pH 10.0). Similarly it was difficult to determine optimum pH for the entire complex due to the ΔQ proximity to zero in the wide pH region (in this case even with multiple zero crossings; and the first crossing is reported in Fig. 7). In both complexes the ''basic'' value of the optimum pH was caused by several histidines with pK_a values shifted to the acidic region (His-115A, pK_a 3.6; His-140A, pK_a 4.04 for 1tx4 and His-365B, pK_a 2.96; His-398B, pK_a 0.23 for 1ycs).

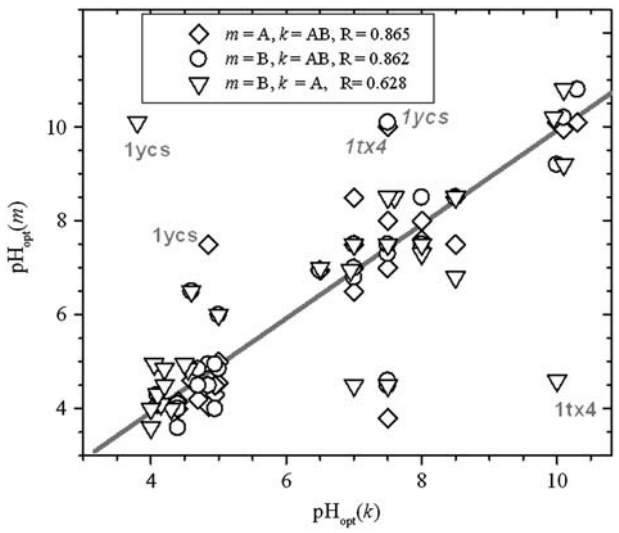

FIGURE 7 pH optimum of the components k of complexes in Table 1 plotted versus pH optimum of the components m of the complexes. Different symbols correspond to the different k and m combinations. Thick gray line is the function $f(x) = x$ and the correlation coefficients, R, with respect to that function for different k and m combinations are also displayed in the legend. Gray labels at points denote pdb IDs for two ''offset'' proteins discussed in the text. In the case of coinciding points, the 1ycs label corresponds to the circle, and the 1tx4 label belongs to the diamond.

Interfacial acidic residues in both complexes participate in strong S- or I-type interactions, but pK_a shifts for these residues either tend to compensate for each other (1tx4) or are not large (1ycs). Therefore, the ΔQ(pH) for the entire complex only slightly shifts toward zero compared to the ΔQ(pH) for the ''basic'' component. However, such a shift is large enough to introduce a statistical uncertainty in determination of optimum pH for the whole complex. Remarkably, those two complexes have large asymmetry in the composition of interfacial titratable residues. The larger part of the 1tx4 complex has only one acidic and six basic residues on the interface, whereas for the smaller component the situation is the opposite (eight acidic and three basic residues). For the 1ycs complex the interface contains only nine acidic and no basic residues for the larger part and two acidic and six basic groups for the smaller part. For other protein complexes in the set there was no such clear asymmetry observed. However, this asymmetry is not so clearly pronounced when the residual composition is considered for the whole sequence of monomers. The A components contain in total 19 and 35 acidic along with 14 and 11 basic groups for the 1tx4 and 1ycs complexes, correspondingly, whereas the B components possess 30 and 19 acidic and 24 and 22 basic groups. At this stage we could not draw definite conclusions on whether the asymmetry in the interface composition is related to the large discrepancy in the optimum pH values for the complex and monomers or could be simply an artifact of our methodology.

Acidic, basic, and neutral optimum pH values

The values for optimum pH are clearly clustered in the three pH regions, acidic, basic, and neutral (see Fig. 7). Table 3 shows the number of proteins in the data set within a given functional group having an optimum pH within a certain pH region. As seen, there is a weak correlation between function and the pH region of the optimum pH. Antibody-antigen complexes are almost uniformly spread over the whole pH interval; the enzyme-inhibitor complexes have a weak tendency to have an optimum pH in the neutral range; transport proteins are leaning to the basic pH region, and other complexes have a propensity for the neutral pH region. A more

TABLE 3 Correlation between protein function and pH-optimum of the complex and its components

	Number of proteins in the set with pH-optimum in			
Function of protein	Acidic region	Basic region	Neutral region	No pH dependence
Antibody-antigen	3	1	2	3
Enzyme-inhibitor	2	0	4	2
Transport	0	1	4	0
Others	7	2	5	1

TABLE 4 Correlation between pH-optimum and type of acidic residues on the complex interface

	Number of proteins in the set			
pH-optimum region	Total in the given pH region	With at least one I-type group	With only S-type groups	With only 0-type groups
Acidic	12	7	4	1
Basic	4	3	1	0
Neutral	15	7	8	0
No pH dependence	6	2	1	3

pronounced correlation is observed between the pH region of the optimum pH and type of interfacial residues (Table 4). Complexes having acidic and basic optimum pH tend to have at least one I-type group, whereas complexes having neutral optimum pH tend to form interfaces with either S-type groups or with I-type groups. As also seen from Table 4, there are only a few proteins in our data set with only 0-type interfacial groups and most of them belong to the complexes which do not exhibit pH dependence of ΔQ. Finally, Table 5 displays relationships between the region of the optimum pH and the ratio of the acidic and basic groups in the components A and B, defined in the following manner:

$$R_{\mathrm{A(B)}} = \frac{N_{\mathrm{A(B)}}^{\mathrm{acidic}}}{N_{\mathrm{A(B)}}^{\mathrm{basic}}}, \tag{7}$$

where $N_{\mathrm{A(B)}}^{\mathrm{acidic(basic)}}$ is the number of the acidic (basic) groups in the component A(B) of the complex. We grouped the $R_{\mathrm{A(B)}}$ values into the three regions: >1, <1, and $\sim = 1$ (tolerance level here was ± 1 residue). Not surprisingly, most proteins with the acidic optimum pH have at least one component with prevailing acidic groups. Interestingly most proteins with neutral optimum pH also have a tendency to have one component being acidic.

CONCLUDING REMARKS

We have studied statistical and electrostatic properties of the interfaces for a set of 37 complexes extracted from our previously developed database of protein-protein complexes. It was found that residue composition on the interfaces statistically differs from the composition on the entire surface of the components. Although the charged and polar residues tend to be in abundance and the hydrophobic residues are underrepresented on the whole surface, the frequencies of occurrence for the interfacial residues do not differ much from the ''background'' frequencies for the whole body of the protein universe. Thus, finding surface patches with residual composition not different from the ''background'' frequencies of occurrence may hint where the putative interface of a model complex is likely to be located. Effectively this means finding interfacial patches with higher hydrophobicity than normal, an approach that is originally taken by Abagyan and co-workers (42,43).

The interplay between electrostatic and hydrophobic interaction results in a tendency, observed previously by Honig and co-workers, for larger interfacial areas to be more hydrophobic whereas small interfaces are rich in charged and polar groups. Thus, the role of electrostatics is more pronounced within complexes with small interfaces (23).

The contribution of the individual amino acids on the binding affinity also shows a clear trend. It was found that, in a vast majority of the cases, the average pK_a shifts for acidic residues are negative, indicating that complex formation stabilizes the ionizable states. The individual pK_a shifts show the same tendency since 80% of the interfacial acidic groups were found to lower their pK_as upon complex formation. It should be clarified that this does not necessary indicate that these groups favor the complex formation, but their eventual mutation to Ala may decrease the binding affinity (see Lee and Tidor (20) and Sheinerman and Honig (23) for detailed discussions).

The tendency observed for the interfacial histidine groups is the opposite, i.e., the majority of the groups have a

TABLE 5 Correlation between pH-optimum and ratio of acidic and basic groups

	Number of proteins in the set with given ratio of acidic and basic groups*					
pH-optimum region	$R_A > 1$ $R_B > 1$	$R_A > 1$ $R_B \sim 1$	$R_A \sim 1$ $R_B \sim 1$	$R_A > 1^{\dagger}$ $R_B < 1$	$R_A \sim 1$ $R_B < 1$	$R_A < 1$ $R_B < 1$
Acidic	4 (2)	5 (4)	0 (4)	1 (1)	1 (1)	1 (0)
Basic	0 (0)	0 (3)	1 (1)	1 (0)	1 (0)	1 (0)
Neutral	2 (0)	6 (1)	1 (3)	4 (6)	2 (5)	0 (0)
No pH dependence	1 (0)	2 (2)	1 (3)	1 (1)	1 (0)	0 (0)

*Numbers in parentheses are obtained with ratio of acidic and basic groups calculated only for interfacial residues.
†Numbers in this column include figures for $R_A < 1$ and $R_B > 1$.

destabilizing electrostatic effect on the complex formation. Moreover, many histidine groups were calculated to have such low pK_a values in the complex that they are predicted to remain uncharged in any pH range where association/dissociation of the complexes take place.

Two types of effects causing a large pK_a shift at the complex formation were identified. The first type involves a preburied residue that establishes new favorable interactions in the complex. This type of residue was pointed out by other authors as a residue that may contribute to the binding affinity (40). The second major effect is the change of the strength of existing interactions in a monomeric state caused by both the change of the dielectric boundary and side-chain rearrangement upon complex formation. These two types of effects were correlated with the function of the complexes in the data set.

One of the most important findings is the correlation of the optimum pH of the complex and the monomers. A formation of a complex requires that the monomers meet in the same compartment of the cell and thus the monomers should be able to tolerate the conditions there, in our case the specific pH (44) of the compartment. The resulting complex should be able to tolerate these conditions too. We showed that the optimum pH values of the complex tend to be the same as the optimum pH values of the complex components. This finding can be used in the homology-based prediction of the 3D structures of protein complexes, especially when one needs to evaluate and rank putative models. It is more likely for a model to be correct if both components of the model complex and the entire complex have the same or at least similar values of the optimum pH.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors thank Barry Honig for the continuous support. We thank Amber Allardice for the technical assistance with the project and Lucas Sawle for proofreading the manuscript.

E.A. acknowledges the startup funds from Clemson University.

REFERENCES

1. Smith, G. R., and M. J. E. Sternberg. 2002. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* 12:28–35.
2. McCammon, J. A. 1998. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* 8:245–249.
3. Topf, M., and A. Sali. 2005. Combining electron microscopy and comparative modeling. *Curr. Opin. Chem. Biol.* 15:578–585.
4. Lu, L., H. Lu, and J. Skolnick. 2005. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins.* 15:350–364.
5. Peieper, U., N. Eswar, H. Braberg, M. Madhusudhan, F. Davis, A. Stuart, N. Mirkovich, A. Rossi, M. Marti-renom, A. Fiser, B. Webb, D. Greenblatt, C. Huang, T. Ferrin, and A. Sali. 2004. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 32:217–222.
6. Aloy, P., B. Bottcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A. Gavin, P. Bork, G. Superti-Furga, L. Serrano, and R. Russell. 2004. Structure-based assembly of protein complexes in yeast. *Science.* 303:2026–2029.
7. Aloy, P., and R. B. Russell. 2002. Interrogating protein interaction networks through structural biology. *Proc. Natl. Acad. Sci. USA.* 99:5896–5901.
8. Reference deleted in proof.
9. Bordner, A. J., and R. Abagyan. 2005. Statistical analysis and prediction of protein-protein interfaces. *Proteins.* 60:353–366.
10. Jones, S., and J. Thornton. 1997. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.* 272:113–143.
11. Jones, S., and J. Thornton. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA.* 93:13–20.
12. Ofran, Y., and B. Rost. 2003. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett.* 544:236–239.
13. Nooren, I. M. A., and J. M. Thornton. 2003. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* 325:991–1018.
14. Hu, Z. J., B. Y. Ma, H. Wolfson, and R. Nussinov. 2000. Conservation of polar residues as hot spots at protein interfaces. *Proteins.* 39:331–342.
15. Clackson, T., and J. A. Wells. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science.* 267:383–386.
16. Lowman, H. B., B. C. Cunningham, and J. A. Wells. 1991. Mutational analysis and protein engineering of receptor-binding determinants in human placental lactogen. *J. Biol. Chem.* 266:10982–10988.
17. Massova, I., and P. A. Kollman. 1999. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121:8133–8143.
18. Green, D. F., and B. Tidor. 2005. Design of improved protein inhibitors of HIV-1 cell entry: optimization of electrostatic interactions at the binding interface. *Proteins.* 60:644–657.
19. Kangas, E., and B. Tidor. 1999. Charge optimization leads to favorable electrostatic binding free energy. *Phys. Rev. E.* 59:5958–5961.
20. Lee, L. P., and B. Tidor. 2001. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* 10:362–377.
21. Lee, L. P., and B. Tidor. 2001. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.* 8:73–76.
22. 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.
23. Sheinerman, F. B., and B. Honig. 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.* 318: 161–177.
24. Cerutti, D. S., L. F. Ten Eyck, and J. A. McCammon. 2005. Rapid estimation of solvation energy for simulations of protein-protein association. *J. Chem. Theory Comput.* 1:143–152.
25. Elcock, A. H., D. Sept, and J. A. McCammon. 2001. Computer simulation of protein-protein interactions. *J. Phys. Chem. B.* 105:1504–1518.
26. Ma, C. S., N. A. Baker, S. Joseph, and J. A. McCammon. 2002. Binding of aminoglycoside antibiotics to the small ribosomal subunit: a continuum electrostatics investigation. *J. Am. Chem. Soc.* 124:1438–1442.
27. Nielsen, J. E., and J. A. McCammon. 2003. Calculating pK_a values in enzyme active sites. *Protein Sci.* 12:1894–1901.
28. Sims, P. A., C. F. Wong, and J. A. McCammon. 2004. Charge optimization of the interface between protein kinases and their ligands. *J. Comput. Chem.* 25:1416–1429.
29. Sims, P. A., C. F. Wong, D. Vuga, J. A. McCammon, and B. M. Sefton. 2005. Relative contributions of desolvation, inter- and intramolecular interactions to binding affinity in protein kinase systems. *J. Comput. Chem.* 26:668–681.

30. Reference deleted in proof.

31. Tanford, C. 1968. Protein denaturation, Part A and Part B. *Adv. Protein Chem.* 23:121–282.

32. Alexov, E., and M. Gunner. 1997. Incorporating protein conformation flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* 74:2075–2093.

33. Georgescu, R., E. Alexov, and M. Gunner. 2002. Combining conformational flexibility and continuum electrostatics for calculating residue pK_a's in proteins. *Biophys. J.* 83:1731–1748.

34. Alexov, E. 2003. The role of the protein side chain fluctuations on the strength of pair wise electrostatic interactions. Comparing experimental with computed pK_a's. *Proteins.* 50:94–103.

35. Alexov, E. 2004. Numerical calculations of the pH of maximal protein stability—the effect of the sequence composition and three-dimensional structure. *Eur. J. Biochem.* 271:173–185.

36. Sridharan, S., A. Nicholls, and B. Honig. 1992. A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* 61:A174. (Abstr.)

37. Lu, H., L. Lu, and J. Skolnick. 2003. Development of unified statistical potentials describing protein-protein interactions. *Biophys. J.* 84:1895–1901.

38. Glaser, F., D. Steinberg, I. Vakser, and N. Ben-Tal. 2001. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins.* 43:89–102.

39. Lee, L., and B. Tidor. 2001. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* 10:362–377.

40. Norel, R., F. Sheinerman, D. Petrey, and B. Honig. 2001. Electrostatic contribution to protein-protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* 10:2147–2161.

41. Nielsen, J., and A. McCammon. 2003. On the evaluation and optimization of protein x-ray structures for pK_a calculations. *Protein Sci.* 12: 313–326.

42. Fernandez-Recio, J., M. Totrov, C. Skorodumov, and R. Abagyan. 2005. Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins.* 58:134–143.

43. Fernandez-Recio, J., R. Abagyan, and M. Totrov. 2005. Improving CAPRI predictions: optimized desolvation for rigid-body docking. *Proteins.* 60:308–313.

44. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. Watson. 1994. Molecular Biology of the Cell. Garland Publishing, New York.

45. Fukuda, K., H. Mizuno, H. Atoda, and T. Morita. 2000. Crystal structure of flavocetin-A, a platelet glycoprotein Ib-binding protein, reveals a novel cyclic tetramer of C-type lectin-like heterodimers. *Biochemistry.* 39:1915–1923.

46. Liu, W., P. Flynn, E. Fuentes, J. Kranz, M. McCormick, and J. Wand. 2001. Main chain and side chain dynamics of oxidized flavodoxin from cyanobacterium anabaena. *Biochemistry.* 40:14744–14753.

47. Albrecht, M., D. Hoffmann, B. O. Evert, I. Schmitt, U. Wullner, and T. Lengauer. 2003. Structural modeling of ataxin-3 reveals distant homology to adaptins. *Proteins.* 50:355–370.

48. Scheffzek, K., I. Stephan, O. N. Jensen, D. Illenberger, and P. Gierschik. 2000. The Rac-RhoGDI complex and the structural basis for the regulation of Rho proteins by RhoGDI. *Nat. Struct. Biol.* 7:122–126.

49. Park, Y. C., H. Ye, C. Hsia, D. Segal, R. L. Rich, H. C. Liou, D. G. Myszka, and H. Wu. 2000. A novel mechanism of TRAF signaling revealed by structural and functional analyses of the TRADD-TRAF2 interaction. *Cell.* 101:777–787.

50. Eigenbrot, C., T. Gonzalez, J. Mayeda, P. Carter, W. Werther, T. Hotaling, J. Fox, and J. Kessler. 1994. X-ray structures of fragments from binding and nonbinding versions of a humanized anti-cd18 antibody—structural indications of the key role of V(H) residues 59 to 65. *Proteins.* 18:49–62.

51. VanDemark, A. P., R. M. Hofmann, C. Tsui, C. M. Pickart, and C. Wolberger. 2001. Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell.* 105:711–720.

52. Graham, D. L., P. N. Lowe, G. W. Grime, M. Marsh, K. Rittinger, S. J. Smerdon, S. J. Gamblin, and J. F. Eccleston. 2002. MgF3—as a transition state analog of phosphoryl transfer. *Chem. Biol.* 9:375–381.

53. Zahnd, C., S. Spinelli, B. Luginbuhl, P. Amstutz, C. Cambillau, and A. Pluckthun. 2004. Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J. Biol. Chem.* 279:18870–18877.

54. Stanfield, R. L., H. Dooley, M. F. Flajnik, and I. A. Wilson. 2004. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science.* 305:1770–1773.

55. Masuda, S., K. S. Murakami, S. Olson, J. Donigian, F. Leon, S. A. Darst, and E. A. Campbell. 2004. Crystal structures of the ADP and ATP bound forms of the Bacillus anti-sigma factor SpoIIAB in complex with the anti-anti-sigma SpoIIAA. *J. Mol. Biol.* 340:941–956.

56. Banner, D. W., A. Darcy, W. Janes, R. Gentz, H. J. Schoenfeld, C. Broger, H. Loetscher, and W. Lesslauer. 1993. Crystal-structure of the soluble human 55 kd TNF receptor-human TNF-beta complex—implications for TNF receptor activation. *Cell.* 73:431–445.

57. Rittinger, K., P. A. Walker, J. F. Eccleston, S. J. Smerdon, and S. J. Gamblin. 1997. Structure at 1.65 angstrom of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature.* 389:758–762.

58. Moncalian, G., B. Lengsfeld, V. Bhaskara, K. P. Hopfner, A. Karcher, E. Alden, J. A. Tainer, and T. T. Paull. 2004. The Rad50 signature motif: essential to ATP binding and biological function. *J. Mol. Biol.* 335:937–951.

59. Al-Shangiti, A. M., C. E. Naylor, S. P. Nair, D. C. Briggs, B. Henderson, and B. M. Chain. 2004. Structural relationships and cellular tropism of staphylococcal superantigen-like proteins. *Infect. Immun.* 72:4261–4270.

60. Kurzbauer, R., D. Teis, M. E. G. de Araujo, S. Maurer-Stroh, F. Eisenhaber, G. P. Bourenkov, H. D. Barkunik, M. Hekman, U. R. Rapp, L. A. Huber, and T. Clausen. 2004. Crystal structure of the p14/MP1 scaffolding complex: how a twin couple attaches mitogen-activated protein kinase signaling to late endosomes. *Proc. Natl. Acad. Sci. USA.* 101:10984–10989.

61. Bhat, T. N., G. A. Bentley, T. O. Fischmann, G. Boulot, and R. J. Poljak. 1990. Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen-binding. *Nature.* 347:483–485.

62. Kuwada, T., T. Hasegawa, I. Satoh, K. Ishikawa, and F. Shishikura. 2003. Crystallization and preliminary x-ray diffraction study of hemoglobin D from the Aldabra giant tortoise, Geochelone gigantea. *Protein Pept. Lett.* 10:422–425.

63. Suzuki, N., Z. Fujimoto, T. Morita, A. Fukamizu, and H. Mizuno. 2005. pH-dependent structural changes at Ca2+-binding sites of coagulation factor IX-binding protein. *J. Mol. Biol.* 353:80–87.

64. Lammens, A., A. Schele, and K. P. Hopfner. 2004. Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. *Curr. Biol.* 14:1778–1782.

65. De Genst, E., F. Handelberg, A. Van Meirhaeghe, S. Vynck, R. Loris, L. Wyns, and S. Muyldermans. 2004. Chemical basis for the affinity maturation of a camel single domain antibody. *J. Biol. Chem.* 279:53593–53601.

66. Jabeen, T., S. Sharma, N. Singh, A. Bhushan, and T. P. Singh. 2005. Structure of the zinc-saturated C-terminal lobe of bovine lactoferrin at 2.0 angstrom resolution. *Acta Crystallogr. D.* 61:1107–1115.

67. Jabeen, T., S. Sharma, N. Singh, R. K. Singh, A. K. Verma, M. Paramasivam, A. Srinivasan, and T. P. Singh. 2005. Structure of the zinc-induced heterodimer of two calcium-free isoforms of phospholipase A(2) from Naja naja sagittifera at 2.7 angstrom resolution. *Acta Crystallogr. D.* 61:302–308.

68. Gorina, S., and N. P. Pavletich. 1996. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science.* 274:1001–1005.

69. Rickert, M., X. Q. Wang, M. J. Boulanger, N. Goriatcheva, and K. C. Garcia. 2005. The structure of interleukin-2 complexed with its alpha receptor. *Science.* 308:1477–1480.

70. Ostermeier, C., and A. T. Brunger. 1999. Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of Rabphilin-3A. *Cell.* 96:363–374.

71. Hsu, H. C., B. Stillman, and R. M. Xu. 2005. Structural basis for origin recognition complex 1 protein-silence information regulator 1 protein interaction in epigenetic silencing. *Proc. Natl. Acad. Sci. USA.* 102:8519–8524.

72. Rual, J., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. Berriz, F. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. Goldberg, L. Zhang, S. Wong, G. Franklin, S. Li, J. Alabala, J. L. Lim, C. Fraughton, E. Llamosas, S. Cevik, C. Bex, P. Lamesch, R. Sikorski, J. Vandenhaute, H. Zoghbi, A. Smoyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. Cusick, D. Hill, F. Roth, and M. Vidal. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* 437:1173–1178.

73. O'Neal, C. J., M. G. Jobling, R. K. Holmes, and W. G. J. Hol. 2005. Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science.* 309:1093–1096.

74. Adams, J. J., G. Pal, Z. C. Jia, and S. P. Smith. 2006. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl. Acad. Sci. USA.* 103:305–310.

75. Trinh, C. H., S. D. Hemmington, M. E. Verhoeyen, and S. E. V. Phillips. 1997. Antibody fragment Fv4155 bound to two closely related steroid hormones: the structural basis of fine specificity. *Structure.* 5:937–948.

76. James, L. C., and D. S. Tawfik. 2005. Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proc. Natl. Acad. Sci. USA.* 102:12730–12735.

77. Nakasako, M., H. Takahashi, N. Shimba, I. Shimada, and Y. Arata. 1999. The pH-dependent structural variation of complementarity-determining region H3 in the crystal structures of the Fv fragment from an anti-dansyl monoclonal antibody. *J. Mol. Biol.* 291:117–134.

78. Ladner, R. C., E. J. Heidner, and M. F. Perutz. 1977. The structure of horse methaemoglobin at 2.0 Å resolution. *J. Mol. Biol.* 114:385–413.

79. Nassar, N., G. R. Hoffman, D. Manor, J. C. Clardy, and R. A. Cerione. 1998. Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nat. Struct. Biol.* 5:1047–1052.

80. Fermi, G., M. F. Perutz, B. Shaanan, and R. Fourme. 1984. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* 175:159–174.