12-2010

In Silico Modeling of pH-optimum of Protein-Protein Binding

Rooplekha C. Mitra
Clemson University

Zhe Zhang
Clemson University

Emil Alexov
Clemson University, ealexov@clemson.edu

Published in final edited form as:
Proteins. 2011 March ; 79(3): 925–936. doi:10.1002/prot.22931.

In silico modeling of pH-optimum of protein-protein binding

Rooplekha C Mitra, **Zhe Zhang**, and **Emil Alexov***
Computational Biophysics and Bioinformatics, Physics Department Clemson University, Clemson, SC 29634

Abstract

Protein-protein association is a pH-dependent process and thus the binding affinity depends on the local pH. *In vivo* the association occurs in a particular cellular compartment, where the individual monomers are supposed to meet and form a complex. Since the monomers and the complex exist in the same micro environment, it is plausible that they coevolved toward its properties, in particular, toward the characteristic subcellular pH. Here we show that the pH at which the monomers are most stable (pH-optimum) or the pH at which stability is almost pH-independent (pH-flat) of monomers are correlated with the pH-optimum of maximal affinity (pH-optimum of binding) or pH interval at which affinity is almost pH-independent (pH-flat of binding) of the complexes made of the corresponding monomers. The analysis of interfacial properties of protein complexes demonstrates that pH-dependent properties can be roughly estimated using the interface charge alone. In addition, we introduce a parameter beta, proportional to the square root of the absolute product of the net charges of monomers, and show that protein complexes characterized with small or very large beta tend to have neutral pH-optimum. Further more, protein complexes made of monomers carrying the same polarity net charge at neutral pH have either very low or very high pH-optimum of binding. These findings are used to propose empirical rule for predicting pH-optimum of binding provided that the amino acid compositions of the corresponding monomers are available.

Keywords

pH-optimum; protein binding; pKa's; electrostatics; pH-dependent effects

INTRODUCTION

Proteins never act in isolation; rather they function in a crowded cellular medium with other molecules and proteins [1]. Protein interactions are central to structural and functional organization of the cell *in vivo* and underlie many biological processes, such as metabolic control [2,3], protease inhibition [4], DNA replication and transcription [5-7], cell adhesion [8-10], hormone receptor binding [11], the action of antibody against antigen [12], regulation of gene expressions in cells [13,14] and many other important cellular processes [15-17] .Thereby, understanding many biological processes relies on the ability to understand protein-protein interactions and their mechanisms. These aspects have initialized efforts for *in silico* modeling of protein-protein association which were frequently reported [18-24], including modeling of ionic strength-(or salt-) and pH-dependent effects on the protein-protein binding free energy [25-31].

Protein–protein binding free energy has many components reflecting the complex nature of biological interactions, among them, the electrostatic energy is the only interaction that

*corresponding author, ealexov@clemson.edu, tel: (864) 656-5307, fax: (864) 656-0803.

directly depends on the pH of the solution [32-36]. Modeling electrostatic energy is complicated not only because of the presence of water molecules and plausible conformational changes occurring upon the binding, but also because the ionization states of titratable groups are not *a-priori* known or even could change upon binding [37,38]. Such ionization changes induced by the binding are referred to as proton uptake/release and perhaps are common phenomena in protein-protein association [31,33,35,39]. This notion is supported by numerous experiments showing the proton uptake/release could be as large as several proton units [33,40-42]. From a computational stand point, it is also supported by our recent study on 2887 protein complexes demonstrating that even at pH=7 more than 70% of cases were predicted to involve proton uptake/release of more than 0.5 proton units [37]. A similar observation was made by Jensen and coworkers [31] showing that protein-protein binding is frequently accompanied by a significant change in the protonation states of titratable groups in the complex relative to the separated monomers.

The change of the pKa's of ionizable groups upon the binding results in proton uptake/release and in turn in the pH dependence of the binding energy [36]. Similarly, the change of pKa's upon protein folding results in pH dependence of the folding energy [43]. The magnitude of the energy change, either binding or folding energy, upon pH variations could be significant and in some cases could alter the minimal value by more than 50% , even at physiological pH [44-48]. Obviously at pH different from physiological pH, acidic/basic denaturation may occur and protein-protein binding may be abolished as well [49-51]. However, the physiological pH varies within cellular compartments (an excellent collection of characteristic subcellular pHs can be found in the works of Warwicker and coworkers [52,53]). For example, pH is nearly neutral in the cytoplasm, in the endoplasmic reticulum, and in mitochondria; it is more acidic in vacuoles, lysosomes (as low as pH 5), and the Golgi; and it is more basic in the nucleus and peroxisomes (as high as pH 8). Proteins and protein-protein complexes can harness these differences in cellular and sub-cellular pHs for physiological purposes as outlined in a recent review by Garcia-Moreno [54]. If the charges of proteins have indeed been optimized for specific functional purposes, they may also display adaptation to specific sub-cellular conditions or pH.

The adaptation of monomeric proteins to cellular and sub-cellular characteristic pHs was extensively studied by Warwicker and coworkers [52,53] and found that the pH of maximal stability of monomeric proteins correlates with subcellular characteristic pH. It was shown that the properties of histidine residues underlie this correlation. Our previous work [39] demonstrated that structure-based approaches, as Multi Conformation Continuum Electrostatics (MCCE) [55,56], are capable of capturing pH-dependent effects, and specifically, the pH of optimal stability (pH-optimum) [57]. Combined with the observation that the pH-optimum of activity and stability are correlated, [58] this paves the way for more *in silico* studies to reveal the adaptation of monomeric proteins to cellular and sub-cellular environments, even at genome scales [54]. At the same time, the plausible adaptation of protein-protein interactions to the sub-cellular environment somehow escaped the attention of researchers, perhaps, because of the limited amount of experimental data. However, this can be avoided, since if two (or more) monomeric proteins form a complex (or assemblage) they must meet in the same sub-cellular compartment and thus the monomers and the complex should be subjected to the same local environmental conditions, including characteristic pH. Therefore, one could investigate if the pH-dependent properties of monomers and corresponding complexes are correlated, without knowing the characteristic pH of the microenvironment where they form a complex, their pH-dependent properties need simply to be the same or very similar.

In the present work, we compute the pH dependence of the proton uptake/release of binding and folding and correspondingly, the pH-dependence of the free energy of binding and

folding. Specifically, we pay attention to the pH at which these energies are maximal, the "pH-optimum" or are almost "pH-independent", the "pH-flat". We have demonstrated in the past that the pH-optimum of binding and the pH-optimum of folding of bound monomers are correlated [59], however, it was done without accounting for plausible conformational changes induced by the binding. Here we extend our investigation by applying a more realistic protocol to include the conformational changes caused by the complex formation. It is done on a set of 31 protein-protein complexes for which bound and unbound structures are experimentally available. We aim at revealing the plausible correlations between the pH-optimum/pH-flat of binding free energy and pH-optimum/pH-flat of folding of individual (unbound) monomers forming the complex and plausible relations of pH-optimum/pH-flat with interfacial and global characteristics of the corresponding proteins.

METHODS

Protein-protein complexes used in the study

Following Jensen and co-workers [31,60], we extracted a set of protein-protein complexes from the Protein-Protein Docking Benchmark (PPDB) [61] database (the Boston University benchmark set (http://zlab.bu.edu/zdock/benchmark.html)). The PPDB is a database of PDB files of protein-protein complexes and complemented with the corresponding unbound proteins. For the purpose of our study, we selected only binary complexes. The protein-protein complexes are either taken from PPDB 2.0 or from PPDB version 1.0. Frequently, the amino acid sequences of bound and unbound proteins were not identical due to the details of the expression and crystallization procedures. Such differences, which from the point of view of this study are artifacts, would introduce unwanted errors and must be avoided (if possible). This motivated us to perform sequence alignments of bound and unbound proteins using GRASP2 [62] with the goal to purge cases for which the sequences of bound and unbound proteins are quite different or having more than two polypeptide chains. In addition, entries for which the individual monomers were found to exist in different subcellular compartments (according to the assignment provided in the header of the corresponding PDB file, if any) were also deleted from the list, since they do not match our requirements. Non-annotated entries, however, were kept in the list. This resulted in 31 cases (see Table 1S in the supplementary material). The sequence alignment was also used to fix small sequence differences by the following procedure: all titratable groups (if any) which were not present in both the sequences of bound and unbound proteins were mutated *in silico* to the corresponding non-titratable group. This was done to reassure that the number and type of titratable groups prior to and after the formation of the protein-protein complex remain the same. No action was taken for non-titratable groups, i.e. they were kept as shown in the original PDB files.

pKa calculations

The net charge of either unbound proteins or the protein-protein complexes were calculated with the Multi-Conformation-Continuum-Electrostatics (MCCE) program [63]. This program is available at (URL:www.sci.ccny.cuny.edu/mcce/). The basic principles of the method have been described elsewhere [63]. The MCCE method calculates the equilibrated conformation and ionization states of protein side chains, buried waters, ions, and ligands and thus the net charge of the corresponding structure as a function of pH. The structures of the complex and of unbound monomers were subjected to MCCE calculations and the corresponding net charges as a function of pH were obtained. Thus the proton uptake/release induced by the binding is calculated as the difference between the net charge of the complex ($Q_{complex}$ (pH)) and unbound monomers, marked in the eq. (1) as Q_A (pH) and Q_B (pH) as:

$$Q_{binding}(pH) = Q_{complex}(pH) - Q_A(pH) - Q_B(pH) \quad (1)$$

where pH varies from pH=0.0 to pH=14.0.

Similarly, the proton uptake/release upon folding was calculated as the difference between the net charge of folded and unfolded protein as a function of pH. The unfolded state was modeled as an extended polypeptide chain, in accordance with our previous work [57] where we showed that such an approach reproduces the experimental data of pH-dependence of the folding of several proteins and resulted in a very good fit against experimental pH-optimum of proteins[64]. The net charge of unfolded protein molecules is calculated using the Henderson- Hasselbalch equation [65]:

$$Q_u(pH) = \sum_{i=1}^{N} \frac{10^{-23\gamma(i)(pH-pK_a(i))}}{1+10^{-23\gamma(i)(pH-pK_a(i))}} \quad (2)$$

and the summation runs over all *N* titratable residues. γ is 1 for bases and -1 for acids, pK_a is the standard pK_a value in solution of group 'i', and *pH* is the pH of the solution (see supplementary materials). Thus, proton uptake/release upon folding ($\Delta Q_{folding}(pH)$), for each unbound monomer, is:

$$\Delta Q_{folding}(pH) = Q_{folded}(pH) - Q_u(pH) \quad (3)$$

where $\Delta Q_{folded}(pH)$ is the net charge pH-dependence of unbound folded monomer.

Special attention was paid to Cys residues in the unfolded state. If Cys residues were found to form a disulfide bridge in the folded state, then they were not titrated in the unfolded state as well. The rationale for this procedure is that the disulfide bridges may still be present in the unfolded state, at least before the complete unfolding. Applying such a procedure was found to eliminate unwanted fluctuations in the proton uptake/release at neutral pH.

pH dependence of binding affinity and protein stability—The pH dependence of the stability of the unbound monomers and their binding affinity was calculated using the formula [36,43]:

$$\Delta G(pH) = 2.3RT \int_{pH_0}^{pH_1} \Delta Q(pH)\, d(pH) \quad (4)$$

where $\Delta Q(pH)$ is either $\Delta Q_{folding}$ or $\Delta Q_{binding}$, *R* is the universal gas constant, *T* is the temperature (in K), and $\Delta G(pH)$ is the pH-dependent component of the free energy of stability or binding, respectively.

The main implication of equation (4) is that the optimum pH of binding/stability may be estimated as the pH at which the corresponding ΔQ(pH) is zero. However, frequently, ΔQ(pH) is zero at several pHs, and determining the global energy minimum requires integration.

Calculations of global and interfacial properties

A variety of global and interfacial properties were calculated in attempt to deliver plausible correlations with the corresponding pH-optima or pH-bests. They are grouped into global

properties, which reflect the properties of either the protein-protein complex or unbound monomers. They are termed global because their calculations do not require knowledge of the 3D structure of the complex, and thus could be used to predict pH-dependent properties even of protein-protein complexes for which 3D structures are still not available. The second class, the interfacial properties, requires that the 3D structure of the complex is available in order to determine interfacial residues (it should be noted that principle interfaces can be predicted with quite reasonable accuracy and thus eliminating the need for a 3D structure of the complex [66-71]).

(a) *Global properties*—Using the structures of monomers, we find that the total number of amino acids in each protein and the complex is simply the sum of the amino acids of monomers. Similarly, the net negative charge of each monomer was obtained, assigning -1 to each Glu or Asp residues, assuming that they are fully ionized. The same approach was used to obtain the total positive charge accounting for Lys, Arg, and His residues as carrying +1 charge. While one may argue that if all His are ionized at physiological pH, then they represent a minor fraction in our calculations. As above, the total negative/positive charge of the complex was the sum of the charges of individual monomers. The total net charge is the sum of negative and positive charges with the appropriate sign.

(b) *Interfacial properties*—Consider a complex made of two monomers, marked as monomer "A" and monomer "B". The interface area, S_1, was calculated as the difference in the gross solvent accessible surface area (SASA) of the monomers, S_A and S_B and the SASA of the complex, S_{AB}:

$$S_I = \frac{1}{2}(S_A + S_B - S_{AB}) \tag{5}$$

The SASAs were calculated by means of the SURFV program developed in the Honig lab [72] (URL: trantor.bioc.columbia.edu) with a water probe radius of 1.4 A and with default atomic radii. A residue was considered to be interfacial if its accessibility changes by more than $5A^2$ from the unbound to bound state. Knowing the interfacial residues, we obtain the above mentioned characteristics for the interface only.

RESULTS AND DISCUSSIONS

Test case: barnase, barstar, and barnase-barstar complex

One of the complexes in our dataset is the barnase-barstar complex, one of the most extensively and experimentally studied protein-protein complex. We will use such an opportunity to illustrate several aspects of the current study.

(a) *Accuracy of the pKa calculations*—The MCCE program was extensively benchmarked in the past against large sets of experimental pKa's and it was demonstrated that it achieves very good results [56,63]. However, achieving good RMSD over a large number of pKa's does not necessarily mean that the pH-dependence of the folding/binding energy can also be successfully modeled, due to the pH-dependence typically involving only a small fraction of the titratable groups which pKa's frequently are not easy to predict. In addition, the predictions depend on the model, including the model of the unfolded state in the case of computing the pH-dependence of the folding energy. Although it was shown many times that the difference between pKa's calculated with a model representing an extended chain of amino acids (a model used in this study) and more sophisticated models [73-75] of unfolded states are only a fraction of a pH unit, still even such small effects could still contribute to the success of modeling the pH-dependence of the folding energy.

However, we have demonstrated in the past that MCCE with an extended amino acid chain model of unfolded states can successfully reproduce experimental data of pH-optimum on a set of 28 proteins [57]. The MCCE was also used to model the proton uptake/release and the pH-dependence of the binding energy for three protease-pepstatin complexes and a very good agreement with the experimental data was obtained [76]. At the same time, it should be mentioned that benchmarking the results of computer modeling against experimental data of pH-dependence of folding or binding on a large set of cases is typically obscured by not knowing the exact experimental conditions (as salt concentration, presence of other compounds, temperature of the experiment, etc) and the method of measurement (thermal denaturation, urea unfolding, etc). Binding affinity can also be measured by a variety of methods and the results may not necessary be identical [77]. Bearing in mind the complexity of the modeled phenomena, we argue that error of a pH unit or even more should be tolerable.

(b) Specific features associated with barnase-barstar complex—The experimental data of the pH-dependence of the folding free energy of barnase and barstar and the binding free energy of the barnase-barstar complex was taken from Refs. [50,51,78-81]. The corresponding pH-dependence profiles are shown in Fig. 1 together with the calculated pH-dependences. The figures demonstrate that the numerical protocol successfully predicts that the pH-optimum of the folding of (i) barnase (pH-optimum(experiment)=pH-optimum(calculated)=5.0), (ii) was almost 3 pH units offset in the case of pH-optimum of folding of barstar (pH-optimum(experiment)=7.0, pH-optimum(calculated)=9.5) and was relatively successful in modeling the pH-dependence of (iii) binding (pH-optimum(experiment)=8.0, pH-optimum(calculated)=6.0). However, in each of these cases, either experimentally determined or calculated pH-dependence shows an interesting trend: a pH interval in which the corresponding quantity (folding or binding free energy) is almost constant, i.e. pH-independent. Such a pH-region will be discussed in later sections below and will be termed “pH-flat”. If we apply some tolerance when comparing experiment and calculations by matching pH-flat regions instead of absolute pH-optimum, then each of the cases is modeled very satisfactorily.

The experimental data for the barnase-barstar complex and separate monomers allows for direct comparison of the corresponding absolute pH-optima as well. The pH-optima are 5.0, 7.0, and 8.0 for the folding of barnase and barstar and binding barnase-barstar, respectively. Obviously, in this particular case the pH-optima of binding and folding do not match, however, the corresponding pH-flat regions do overlap (pH-flat = 7.0, 7.5 and 8.0, respectively), a fact that we will explore in our analysis below.

Proton uptake/release upon binding

Our calculations confirm the observations made in previous works by Jensen and coworkers [31,36,60] and by our previous studies [37,39,59] that protein-protein complex formation is frequently accompanied with proton uptake/release. However, it should be clarified that proton uptake/release is not an absolute number, but depends on the local pH where the complex is formed. Figure 2 shows the calculated uptake/release at three different pHs. It is evident that the vast majority of the cases in our dataset do involve proton uptake or release. In some cases, the net charge change is calculated to be very large, significantly larger than in our previous calculations utilizing a bound-to-bound protocol. Analysis of the distribution of the titratable groups undergoing ionization changes upon complex formation shows that non interfacial groups contribute significantly to the proton uptake/release because of conformational changes induced by the complex formation. Typical case is transducin complex (PDB ID 1GOT), which proton uptake/release is equally contributed by interfacial and non interfacial titratable groups. Similar observation was made in our

previous work [39] which demonstrated that even ionizable groups located more than 15A away from interface could contribute to the proton uptake/release.

Comparison between the distributions calculated at different pH, reveals that at low pH (pH = 5.0), the vast majority of the complex formation is associated with proton release, while at high pH (pH = 9.0) the distribution is shifted toward positive values (proton uptake). This is to be expected from the point of view of protein electrostatics, since at low pH the protein-protein complexes are predominantly positively charged (all basic groups fully protonated and some of the acidic groups being protonated as well) and thus providing a favorable environment for the ionization of the acidic groups being fully or partially protonated in unbound monomers. At high pH (pH = 9.0) the situation is reversed, since most of the complexes are negatively charged (all acids deprotonated while some basic group may be deprotonated as well), and thus favoring the charged state of basic groups in the complex that were either fully or partially deprotonated in unbound monomers. Occasionally, some acidic groups may happen to be in very unfavorable environment in the complex, and thus also to contribute to the proton uptake.

Distribution of pH-optimum

Figure 3 shows the distribution of the pH-optimum of binding and the pH-optimum of folding the monomers. The distribution of the pH-optimum of stability of monomers is a unimodal function with a peak at approximately pH = 7, while the distribution of the pH-optimum of binding is a bimodal function. Interestingly, a similar difference was found when comparing the experimental data of pH-optimum of stability and activity of monomeric proteins [58]. Perhaps this indicates that the function, in this case, the protein-protein interactions, are more sensitive to pH changes than the structural integrity of monomeric proteins [58].

Proton uptake/release and pH-dependence of the binding/folding energy—A typical case is illustrated in the complex made of trypsin and amyloid beta-protein precursor inhibitor domain (Table 1S, PDB ID 1BRC). Figure 4a shows the proton uptake/release upon the binding ($\Delta Q_{binding}$) as a function of pH (pH = 0 – 14). It can be seen that the $\Delta Q_{binding}$ is zero at pH 7.1 (excluding the point at pH=1.8, which is an obvious artifact of the calculations) and will be termed pH-optimum of binding. In addition, it can be seen that another feature, namely that $\Delta Q_{binding}$ is very small ($\Delta Q_{binding} < 0.74$ electron units, corresponding to $\Delta G_{binding} < 1$kcal/mol) in a pH region from pH=3.0 to pH=8.0. In such a pH-region the corresponding binding free energy will be almost pH-independent and may not be very different from the binding free energy obtained at pH-optimum = 7.1. Despite of the significant progress made to improve the quality of *in silico* modeling, an error of 1kcal/mol is still tolerable. Thus to account for possibility, the mid point of this pH region, which will be termed pH-flat, will be also used in our study in parallel with pH-optimum. It will be termed pH-flat and in this case, pH-flat = 5.5. The corresponding pH-dependence of the binding free energy is also calculated with eq. (4) and is shown in Fig. 4b. It is seen that the plot is a bell shaped graph with a minimum binding energy at pH = 7.1. In the pH-flat region, pH=3.0 to pH=8.0, the $\Delta G_{binding}$ changes by less than 1kcal/mol, resulting in a shallow plateau. Similar effects can be found in the case of pH-dependence of folding free energy $\Delta G_{folding}(pH)$ of unbound monomers. As indicated in Table 1S, the 1BRC complex is constituted of unbound monomers which the corresponding PDB IDs are 1BRA and 1AAP, respectively. The proton uptake/release $\Delta Q_{folding}(pH)$ upon the folding of 1BRA and 1AAP is shown in Fig. 4c and Fig. 4e, respectively. The corresponding $\Delta G_{folding}(pH)$ is shown in Fig. 4d and Fig. 4f, for 1BRA and 1APP, respectively. It can be seen that the pH-optimum of folding for 1BRA is 5.9 (pH-optimum = 5.9), while pH-flat is 9.0. For the 1AAP, the pH-optimum is 4.8, while pH-flat is at the midpoint of the pH interval from pH =

4.0 to pH = 9.0, resulting in pH-flat = 6.5. We find that the existence of such a flat pH interval is a common phenomena for both pH-dependence of folding and binding (see Figs. 1 as well), with a tendency to occur more frequently in folding than in binding.

The goal of this work is to study plausible correlations between pH-dependence of the binding free energy of the complexes and the pH-dependence of the folding free energy of unbound monomers. We speculate that such a correlation can be revealed by comparing either pH-optimum or pH-flat. In accordance with our previous work [58], we speculate that most of the proteins and the corresponding complexes have evolved to function at a pH at which they can tolerate small pH fluctuations and maintain their function (if they are not associated with pH regulation). As it was demonstrated in Figs. 4a-f, such pH regions are either pH-optimum (the bottom of the corresponding energy profile is more or less flat) or pH-flat. Since it is not obvious which of them will be utilized is each case, for benchmarking purposes, we introduce a new quantity termed pH-best, defined below:

$$\mathrm{pH-best(X)} = \begin{cases} \mathrm{pH-optimum(Y)}, \text{if pH-optimum(X) is closer to pH-optimum(Y) than to pH-fla} \\ \mathrm{pH-flat(Y)}, \text{if pH-optimum(X) is closer to pH-flat(Y) that to pH-optimum(} \end{cases} \quad (6)$$

where “X” and “Y” stand to any combination of protein-protein complex and unbound monomers. For example, the calculations done for 1BRC complex indicate that the complex pH-optimum is 7.1 and pH-flat is 5.5. Comparing to the unbound monomers, PDB files 1BRA and 1AAP, we see that the best fit will be: pH-best(complex) = 5.5 against pH-optimum of 1BRA = 5.9 and against pH-optimum of 1AAP = 4.8.

pH-best of binding and folding energies—Figure 5a shows the pH-best of the binding free energy versus the pH-best of the folding free energy of unbound monomers and very good correlations were obtained as indicated by the correlation coefficients of 0.85 and 0.73, respectively. The pH-best for monomers are plotted against each other in Fig. 5b and the correlation coefficient is 0.68. Both plots indicate that the pH-dependence properties (as measured through pH-best) of binding and stability of individual monomers are quite similar, ultimately supporting our initial hypothesis.

The analysis of the proton uptake/release as a function of pH showed that for some cases in the dataset the proton uptake is significant, but for others is small, frequently only a fraction of an electron charge within several pH units. The error in determining the pH-optimum for the previous cases is much larger. In addition, it was noticed that in most of the cases the pH-dependence of the binding free energy was a bell-shaped curve with a single minimum and no pH-flat region. In contrast, the pH-dependence of the folding energy of unbound monomers frequently had a pH-flat region. Such a finding may reflect the fact that protein-protein interactions represent the functionality of proteins that may have evolved under strong pH constaints. In contrast, the stability (the folding energy) of protein is not directly associated with the function and need not to be so strongly tightened to a particular pH.

pH-best and global/interfacial properties

The primary focus of our investigation is protein-protein interactions and the pH-optimum/ pH-flat of binding. We calculated many global and interfacial properties, as described in the method section, and attempted to correlate them with pH-best of binding. As in our previous works [57], we found no correlation of the pH-dependence and isoelectric point (pI), neither with the mass, size, number of amino acids, and global charge of the complex or individual

monomers. However, using a parameter (parameter γ) that was modified from our previous work [82], we find that the pH-best of binding shows a significant tendency related to the macroscopic charges of individual monomers. We introduce the parameter β as:

$$\beta = \sqrt{abs\,[\,Q(A)^{*}Q\,(B)]} \tag{7}$$

and figure 6 shows the pH-best plotted against β. It can be seen that protein-protein complexes made of lightly charged monomers tend to have a pH-optimum in neutral pH regions, while with an increase of the net charge of the monomers there is a tendency that some of the complexes may have pH-optimum quite away from neutral pHs. At the end of the spectra, the highly charged monomers, the pH-optimum again tends to be in the neutral range.

Figure 6 shows another interesting trend as illustrated by the net charges of the corresponding monomers. The net charges are provided on the figure and were calculated as described in the method section, i.e. using amino acid sequence alone without structural information. It can be seen that complexes made of monomers carrying the same polarity net charge as for example Porcine pepsin/Ascaris inhibitor 3 (PDB ID 1F34) (lowest pH-optimum) and CDK2 cyclin-dependant kinase 2/Cyclin (PDB ID 1FIN) (highest pH-optimum) tend to have pH-optimum shifted away from neutral pH. As a matter of fact, none of the complexes predicted to have pH-optimum within neutral pHs (broadly defined as $6 < pH < 9$) is comprised of monomers carrying the same polarity charges, except for FAB/Taq polymerase (PDB ID 1BGX) and complexes with a monomer carrying zero net charge. This gives us the opportunity to propose an empirical rule for predicting protein complexes with pH-optimum of binding shifted away from neutral pH: they should be comprised of monomers carrying the same polarity charge and their beta parameter should not be small (beta > 3).

Figure 7 shows a plot of the pH-best of binding plotted against the interface charges. It can be seen that the correlation coefficient is significant, but there are many data points in the middle of the graph that reduce the significance of the fitting. Obviously for complexes made of lightly charged monomers, no reliable prediction can be made. The fitting polynomial, which can be used for predicting pH-best of binding, has the following functional form:

$$pH - best\,(binding) = -0.08x^{2} + 0.5x + 8.2 \tag{8}$$

where x stands for interfacial charge in electronic units.

Figure 7 inspires similar conclusions as made for the effect of the net charge of the pH-optimum of binding. It can be seen that at the wings of the distribution, i.e. very negative or very positive interfacial charges, the pH-optimum is shifted away from neutral pH. Such interfaces are composed of excess either negatively or positively charged amino acids, and their ionization states are typically depressed upon formation of the complexes, resulting in large pKa shifts, which in turn causes pH-optimum of binding to be either very low or very high.

CONCLUSIONS

The results indicate that the pH-dependent properties of protein complexes and the corresponding monomeric proteins are quite similar to each other as measured through their

pH-optima or pH-flat. Such a finding supports the idea that protein complexes and monomers have evolved to meet the constraints of the subcellular compartment where they function, and in particular, the characteristic pH. This observation can be used to annotate protein and protein complexes which subcellular location is presently unknown by transferring such an annotation from one partner to another.

Our study suggests that in most of the cases, if not always, the pH-dependence of the binding energy has a single minimum and no pH-flat region, while the pH-dependence of the folding energy typically has a pH-flat region. It is beyond the scope of this study, but it is a reasonable speculation that the monomeric proteins may be involved in more than one protein-protein interaction, and thus may have to meet another partner with slightly different pH properties.

The finding that the pH-dependence of the binding free energy can be related to the net charges of the monomers provides an opportunity of providing a clue of the pH-dependence of binding when the charges of the monomers which form the complex are known. It is important to emphasize that 3D structures of the monomers are not required to make prediction of the pH-optimum. Rather just the amino acid sequences of the monomers known to form a complex are sufficient. If the interfacial charge is known, either from existing 3D structure or other experiments, then a simple polynomial formula can be used to predict the pH-best of binding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank CCIT at Clemson University and specifically Corey Ferrier for the help with the Axiom Linux cluster. We thank Shawn Witham for reading the manuscript prior to publication and for useful suggestions. RCM would like to acknowledge fruitful discussions with Kemper Talley regarding optimum pH.

REFERENCES

1. Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. Molecular Biology of the Cell. Garland Publishing; 1994.
2. de Andrade PB, Rubi B, Frigerio F, van den Ouweland JM, Maassen JA, Maechler P. Diabetes-associated mitochondrial DNA mutation A3243G impairs cellular metabolic pathways necessary for beta cell function. Diabetologia. 2006; 49(8):1816–1826. [PubMed: 16736129]
3. Frigerio F, Casimir M, Carobbio S, Maechler P. Tissue specificity of mitochondrial glutamate pathways and the control of metabolic homeostasis. Biochim Biophys Acta. 2008; 1777(7-8):965–972. [PubMed: 18486589]
4. Bode W, Schwager P, Huber R. Structural Studies on the Pancreatic Trypsin Inhibitor-Trypsin Complex and its Free Components. Structure and Function Relationships in Serine Protease Inhibition and Catalysis. 1976:43ff.
5. Larabee JL, Hocker JR, Hanas JS. Mechanisms of inhibition of zinc-finger transcription factors by selenium compounds ebselen and selenite. J Inorg Biochem. 2009; 103(3):419–426. [PubMed: 19167089]
6. Shin CJ, Wong S, Davis MJ, Ragan MA. Protein-protein interaction as a predictor of subcellular location. BMC Syst Biol. 2009; 3:28. [PubMed: 19243629]
7. Schaal TD, Holmes MC, Rebar EJ, Case CC. Novel approaches to controlling transcription. Genet Eng (N Y). 2002; 24:137–178. [PubMed: 12416304]
8. Srivastava J, Barreiro G, Groscurth S, Gingras AR, Goult BT, Critchley DR, Kelly MJ, Jacobson MP, Barber DL. Structural model and functional significance of pH-dependent talin-actin binding

for focal adhesion remodeling. Proc Natl Acad Sci U S A. 2008; 105(38):14436–14441. [PubMed: 18780792]

9. Yan J, Yang Y, Zhang H, King C, Kan HM, Cai Y, Yuan CX, Bloom GS, Hua X. Menin interacts with IQGAP1 to enhance intercellular adhesion of beta-cells. Oncogene. 2009; 28(7):973–982. [PubMed: 19079338]
10. Fabrichny IP, Leone P, Sulzenbacher G, Comoletti D, Miller MT, Taylor P, Bourne Y, Marchot P. Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion. Neuron. 2007; 56(6):979–991. [PubMed: 18093521]
11. Chen WY, Hankinson SE, Schnitt SJ, Rosner BA, Holmes MD, Colditz GA. Association of hormone replacement therapy to estrogen and progesterone receptor status in invasive breast carcinoma. Cancer. 2004; 101(7):1490–1500. [PubMed: 15378477]
12. Sinha N, Mohan S, Lipschultz CA, Smith-Gill SJ. Differences in electrostatic properties at antibody-antigen binding sites: Implications for specificity and cross-reactivity. Biophys J. 2002; 83(6):2946–2968. [PubMed: 12496069]
13. Holmes ML, Pridans C, Nutt SL. The regulation of the B-cell gene expression programme by Pax5. Immunol Cell Biol. 2008; 86(1):47–53. [PubMed: 17998914]
14. Jamieson AC, Guan B, Cradick TJ, Xiao H, Holmes MC, Gregory PD, Carroll PM. Controlling gene expression in Drosophila using engineered zinc finger protein transcription factors. Biochem Biophys Res Commun. 2006; 348(3):873–879. [PubMed: 16899226]
15. Mota G, Manciulea M, Cosma E, Popescu I, Hirt M, Jensen-Jarolim E, Calugaru A, Galatiuc C, Regalia T, Tamandl D, Spittler A, Boltz-Nitulescu G. Human NK cells express Fc receptors for IgA which mediate signal transduction and target cell killing. Eur J Immunol. 2003; 33(8):2197–2205. [PubMed: 12884294]
16. Uetz P. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature. 2000; 403:623. [PubMed: 10688190]
17. Keskin O, Ma BY, Rogale K, Gunasekaran K, Nussinov R. Protein-protein interactions: organization, cooperativity and mapping in a bottom-up Systems Biology approach. Phys Biol. 2005; 2(2):S24–S35. [PubMed: 16204846]
18. Jones S, Thornton J. Principles of protein-protein interactions. PNAS (USA). 1996; 93:13–20. [PubMed: 8552589]
19. Nooren IMA, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. Journal of Molecular Biology. 2003; 325(5):991–1018. [PubMed: 12527304]
20. Halperin I, Wolfson H, Nussinov R. Protein-protein interactions: Coupling of structurally conserved residues and of hot spots across interfaces. implications for docking. Structure. 2004; 12(6):1027–1038. [PubMed: 15274922]
21. Keskin O, Ma BY, Nussinov R. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. Journal of Molecular Biology. 2005; 345(5):1281–1294. [PubMed: 15644221]
22. Elcock AH, Sept D, McCammon JA. Computer simulation of protein-protein interactions. J Phys Chem B. 2001; 105(8):1504–1518.
23. Massova I, Kollman PA. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. J Am Chem Soc. 1999; 121(36):8133–8143.
24. Szilagyi A, Grimm V, Arakaki AK, Skolnick J. Prediction of physical protein-protein interactions. Phys Biol. 2005; 2(1-2):S1–S16. [PubMed: 16204844]
25. Sinha N, Smith-Gill SJ. Electrostatics in protein binding and function. Curr Protein Pept Sci. 2002; 3(6):601–614. [PubMed: 12470214]
26. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D. Protein-protein docking predictions for the CAPRI experiment. Proteins. 2003; 52(1):118–122. [PubMed: 12784377]
27. Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. Biochim Biophys Acta. 2006; 1764(11):1647–1676. [PubMed: 17049320]
28. Alsallaq R, Zhou HX. Electrostatic rate enhancement and transient complex of protein-protein association. Proteins. 2008; 71(1):320–335. [PubMed: 17932929]

29. Garcia-Mayoral MF, del Pozo AM, Campos-Olivas R, Gavilanes JG, Santoro J, Rico M, Laurents DV, Bruix M. pH-Dependent conformational stability of the ribotoxin alpha-sarcin and four active site charge substitution variants. Biochemistry. 2006; 45(46):13705–13718. [PubMed: 17105190]
30. Baran KL, Chimenti MS, Schlessman JL, Fitch CA, Herbst KJ, Garcia-Moreno BE. Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. J Mol Biol. 2008; 379(5):1045–1062. [PubMed: 18499123]
31. Mason AC, Jensen JH. Protein-protein binding is often associated with changes in protonation state. Proteins. 2008; 71(1):81–91. [PubMed: 17932920]
32. Schaefer M, Sommer M, Karplus M. pH-Dependence of Protein Stability: Absolute Electrostatic Free Energy Difference between Conformations. J Phys Chem. 1997; 101:1663–1683.
33. Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson J. Dissection of the pH Dependence of Inhibitor Binding Energetics for an Aspatric Protease: Direct Measurement of the Protonation States of the Catalytic Aspartic Acid Residues. Biochemistry. 1997; 36:16166–16172. [PubMed: 9405050]
34. MacKerell AD Jr. Sommer MS, Karplus M. pH dependence of binding reactions from free energy simulations and macroscopic continuum electrostatic calculations: application to 2'GMP/3'GMP binding to ribonuclease T1 and implications for catalysis. J Mol Biol. 1995; 247(4):774–807. [PubMed: 7723031]
35. Kongsted J, Ryde U, Wydra J, Jensen JH. Prediction and rationalization of the pH dependence of the activity and stability of family 11 xylanases. Biochemistry. 2007; 46(47):13581–13592. [PubMed: 17960918]
36. Jensen JH. Calculating pH and salt dependence of protein-protein binding. Curr Pharm Biotechnol. 2008; 9(2):96–102. [PubMed: 18393866]
37. Mitra R, Shyam R, Mitra I, Miteva MA, Alexov E. Calculation of the protonation states of proteins and small molecules: Implications to ligand-receptor interactions. Current computer-aided drug design. 2008; 4:169–179.
38. Sheinerman F, Norel R, Honig B. Electrostatics Aspects of Protein-Protein Interactions. Current Opinion in Structural Biology. 2000; 10:153–159. [PubMed: 10753808]
39. Alexov E. Calculating Proton Uptake/Release and the Binding Free Energy Taking into Account Ionization and Conformation Changes Induced by Protein-Inhibitor Association. Application to Plasmepsin, Cathepsin D and Endothiapepsin-Pepstatin Complexes. Proteins. 2004; 56:572–584. [PubMed: 15229889]
40. Schreiber G, Fersht AR. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. J Mol Biol. 1995; 248:478–486. [PubMed: 7739054]
41. MacKnight ML, Gillard JM, Tollin G. Flavine-protein interactions in flavoenzymes. pH dependence of the binding of flavine mononucleotide and riboflavine to Azotobacter flavodoxin. Biochemistry. 1973; 12(21):4200–4206. [PubMed: 4745668]
42. Gramberg T, Soilleux E, Fisch T, Lalor PF, Hofmann H, Wheeldon S, Cotterill A, Wegele A, Winkler T, Adams DH, Pohlmann S. Interactions of LSECtin and DC-SIGN/DC-SIGNR with viral ligands: Differential pH dependence, internalization and virion binding. Virology. 2007
43. Yang A-S, Honig B. On the pH dependence of protein stability. J Mol Biol. 1993; 231:459–474. [PubMed: 8510157]
44. Whitten S, Garcia-Moreno B. pH Dependence of Stability of Staphyloccocal Nuclease: Evidence of Substantial Electrostatic Interactions in the Denaturated State. Biochemistry. 2000; 39:14292–14304. [PubMed: 11087378]
45. Tollinger M, Crowhurst K, Kay L, Forman-Kay J. Site-specific contributions to the pH dependence of protein stability. Proc Natl Acad Sci USA. 2003; 100(8):4545–4550. [PubMed: 12671071]
46. Bauman AT, Jaron S, Yukl ET, Burchfiel JR, Blackburn NJ. pH Dependence of peptidylglycine monooxygenase. Mechanistic implications of Cu-methionine binding dynamics. Biochemistry. 2006; 45(37):11140–11150. [PubMed: 16964975]
47. Bidwai AK, Ok EY, Erman JE. pH dependence of cyanide binding to the ferric heme domain of the direct oxygen sensor from Escherichia coli and the effect of alkaline denaturation. Biochemistry. 2008; 47(39):10458–10470. [PubMed: 18771281]

48. Matthew JB, Gurd FRN, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-Dependent Processes in Proteins. CRC Criti Rev Biochem. 1985; 18(2):91–197.
49. Anderson DE, Becktel WJ, Dahlquist FW. pH-Induced Denaturation of Proteins: A single Salt Bridges Contributes 3-5 kcal/mol to the Free Energy of Folding of T4-Lysozyme. Biochem. 1990; 29:2403–2408. [PubMed: 2337607]
50. Pace CN, Laurents DV, Erickson RE. Urea Denaturation of Barnase: pH Dependence and Characterzation of the Unfolded State. Biochemistry. 1992; 31:2728–2734. [PubMed: 1547213]
51. Khurana R, Hate A, Nath U, Udgaonkar B. pH dependence of the stability of barstar to chemical and thermal denaturation. Prot Sci. 1995; 4:1133–1144.
52. Chan P, Lovric J, Warwicker J. Subcellular pH and predicted pH-dependent features of proteins. Proteomics. 2006; 6(12):3494–3501. [PubMed: 16705750]
53. Chan P, Warwicker J. Evidence for the adaptation of protein pH-dependence to subcellular pH. BMC Biol. 2009; 7:69. [PubMed: 19849832]
54. Garcia-Moreno B. Adaptations of proteins to cellular and subcellular pH. J Biol. 2009; 8(11):98. [PubMed: 20017887]
55. Alexov E, Gunner M. Incorporating Protein Conformation Flexibility into the Calculation of pH-dependent Protein Properties. Biophys J. 1997; 74:2075–2093. [PubMed: 9129810]
56. Georgescu R, Alexov E, Gunner M. Combining Conformational Flexibility and Continuum Electrostatics for Calculating Residue pKa's in Proteins. Biophys J. 2002; 83:1731–1748. [PubMed: 12324397]
57. Alexov E. Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and 3D structure. Eur J Biochem. 2004; 271:173–185. [PubMed: 14686930]
58. Talley K, Alexov E. On the pH-optimum of activity and stability of proteins. Proteins. 2010; 78(12):2699–2706. [PubMed: 20589630]
59. Kundrotas PJ, Alexov E. Electrostatic properties of protein-protein complexes. Biophys J. 2006; 91(5):1724–1736. [PubMed: 16782791]
60. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins. 2008; 73(3):765–783. [PubMed: 18498103]
61. Chen R, Mintseris J, Janin J, Weng Z. A Protein-Protein Docking Benchmark. Proteins. 2003; 52:88–91. [PubMed: 12784372]
62. Petrey D, Honig B. “GRASP2: Visualization, surface properties and electrostatic of macromolecular structures. Methods Enzymology. 2003; 374:492–509.
63. Song, Y.; Mao, J.; Gunner, MR. MCCE2: Improving Protein pKa Calculations with Extensive Side Chain Rotamer Sampling. 2009. in press
64. Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA: a resource for enzyme data and metabolic information. TiBS. 2002; 27:54–56. [PubMed: 11796225]
65. Hasselbalch KA. Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl. Biochemische Zeitschrift. 1917; 78:112–144.
66. Davis FP, Sali A. PIBASE: a comprehensive database of structurally defined protein interfaces. Bioinformatics. 2005; 21(9):1901–1907. [PubMed: 15657096]
67. Bell RE, Ben-Tal N. In silico identification of functional protein interfaces. Compar Funct Genom. 2003; 4(4):420–423.
68. Bordner AJ, Abagyan R. Statistical analysis and prediction of protein-protein interfaces. Proteins. 2005; 60(3):353–366. [PubMed: 15906321]
69. Bradford JR, Needham CJ, Bulpitt AJ, Westhead DR. Insights into protein-protein interfaces using a Bayesian network prediction method. J Mol Biol. 2006; 362(2):365–386. [PubMed: 16919296]
70. Res I, Mihalek I, Lichtarge O. An evolution based classifier for prediction of protein interfaces without using protein structures. Bioinformatics. 2005; 21(10):2496–2501. [PubMed: 15728113]
71. Aytuna AS, Gursoy A, Keskin O. Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics. 2005; 21(12):2850–2855. [PubMed: 15855251]

72. Sridharan S, Nicholls A, Honig B. A new vertex algorithm to calculate solvent accessible surface areas. Biophys J. 1992; 61:A174.

73. Zhou H. A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins. Proc Natl Acad Sci USA. 2002; 99(6):3569–3574. [PubMed: 11891295]

74. Kundrotas PJ, Karshikoff A. Charge sequence coding in statistical modeling of unfolded proteins. Biochimica Et Biophysica Acta-Proteins and Proteomics. 2004; 1702(1):1–8.

75. Kundrotas PJ, Karshikoff A. Effects of charge-charge interactions on dimensions of unfolded proteins: A Monte Carlo study. Journal of Chemical Physics. 2003; 119(6):3574–3581.

76. Alexov E. Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein-inhibitor association: application to plasmepsin, cathepsin D and endothiapepsin-pepstatin complexes. Proteins. 2004; 56(3):572–584. [PubMed: 15229889]

77. Albeck S, Schreiber G. Biophysical characterization of the interaction of the beta-lactamase TEM-1 with its protein inhibitor BLIP. Biochemistry. 1999; 38(1):11–21. [PubMed: 9890878]

78. Oliverberg M, Arcus V, Fersht A. pKa Values of Carboxyl Groups in the Native and Denaturated States of Barnase: The pKa of the Denaturated State Are on Average 0.4 Units Lower Than Those of Model Compounds. Biochemistry. 1995; 34:9424–9433. [PubMed: 7626612]

79. Serrano L, Kellis JT, Cann P, Matouschek A, Fersht AR. The folding of an enzyme: II structure of Barnase and the contribution of different interactions to protein stability. J Mol Biol. 1992; 224:783–804. [PubMed: 1569557]

80. Martinez JC, Filimonov VV, Mateo PL, Schreiber G, Fersht AR. A calorimetric study of the thermal-stability of barstar and its interaction with barnase. Biochemistry. 1995; 34:5224–5233. [PubMed: 7711042]

81. Schreiber G, Fersht AR. Interaction of barnase with its popypeptide inhibitor barstar studied by protein engineering. Biochemistry. 1993; 32:5145–5150. [PubMed: 8494892]

82. Talley K, Alexov E. Modelling Salt Dependence of Protein-Protein Association:Linear vs Non-Linear Poisson-Bolzmann Equation. Comm in Computational Physics. 2008; 3:1071–1086.

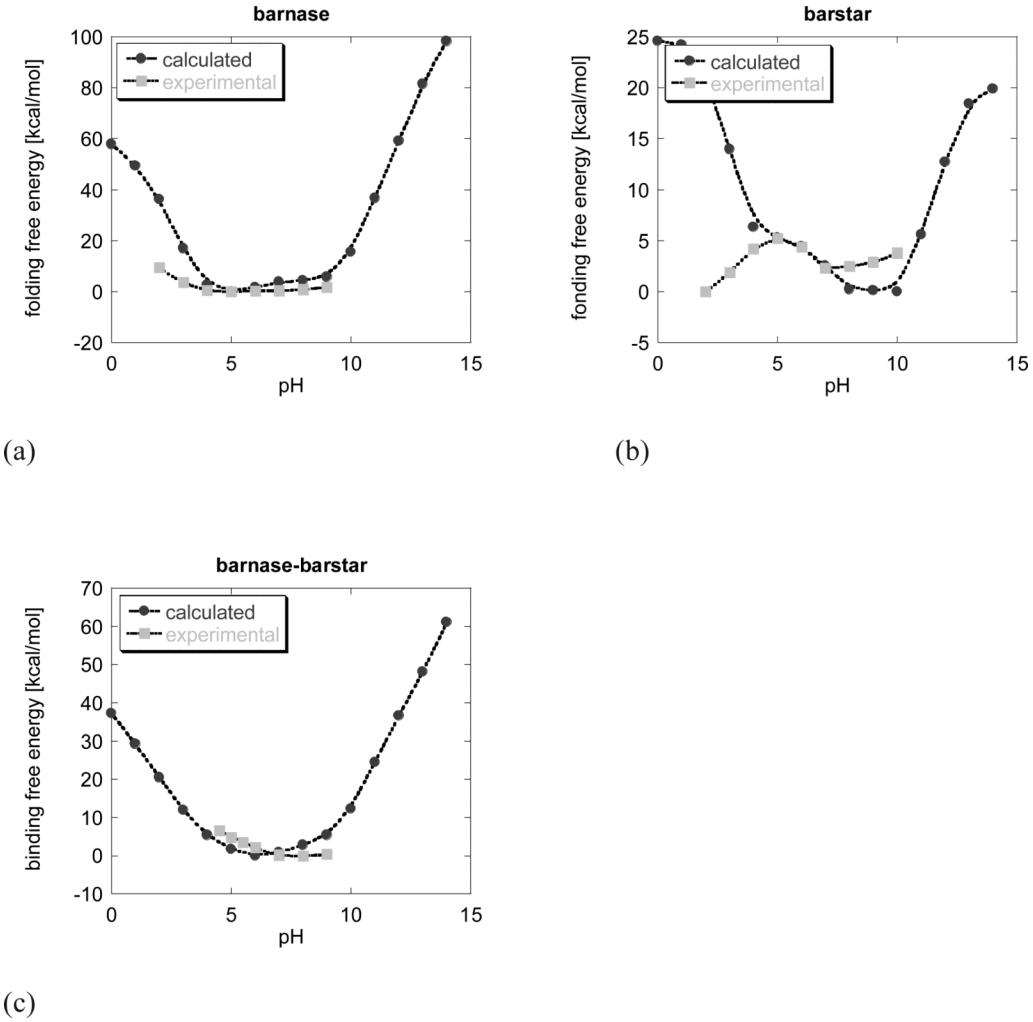

Figure 1.

The folding and binding free energies (calculated and experimental) adjusted to zero at minimum. Arrows indicate the corresponding pH-optimum: solid arrow is calculated, while empty arrow is experimental.

(a) folding free energy of barnase

(b) folding free energy of barstar. The experimental data points at pH < 5.0 indicate a stable oligomeric state of barstar, which is not relevant to our study.

(c) Binding free energy of barnase-barstar.

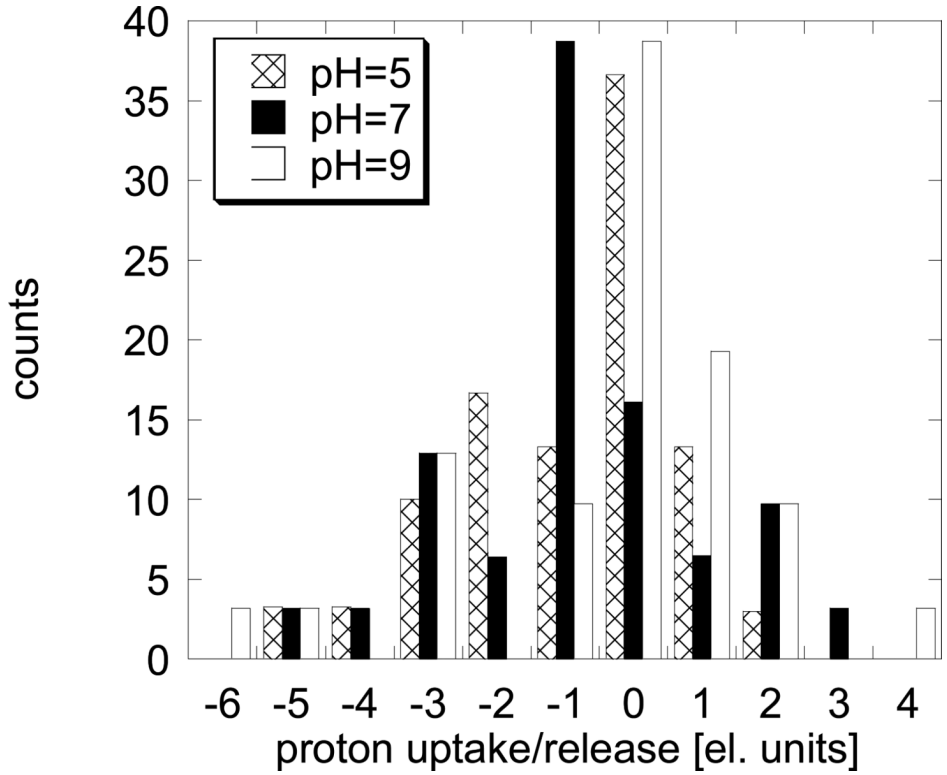

Figure 2.
Distribution plot of proton uptake/release of binding at different pH.

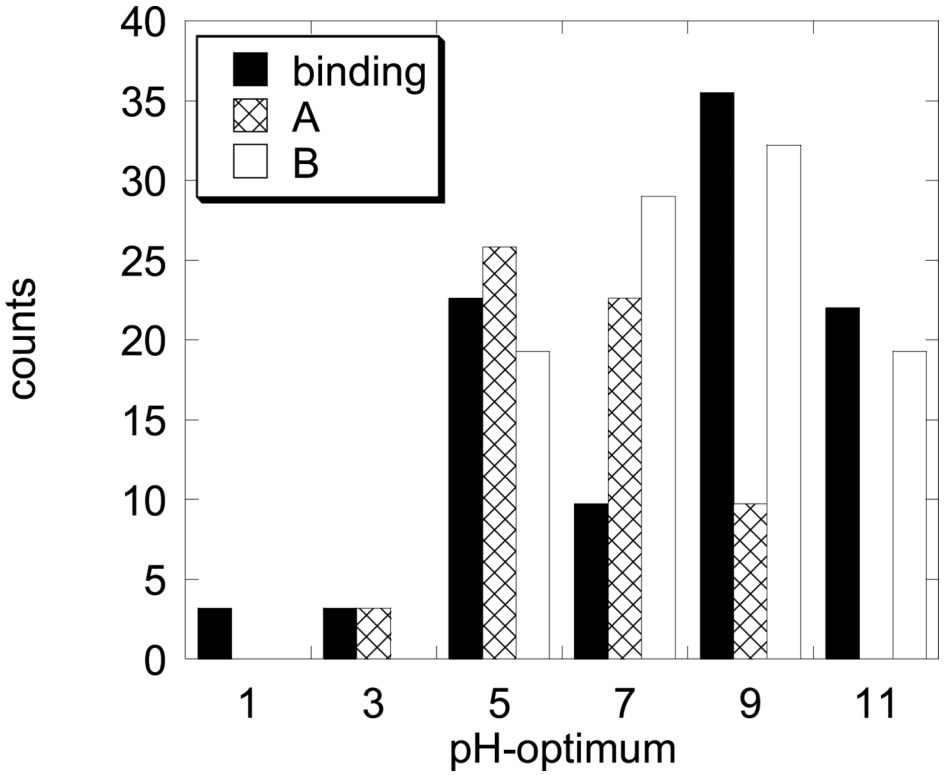

Figure 3.
Distribution plot for the pH-optimum of binding and pH-optimum of folding of the monomers, respectively.

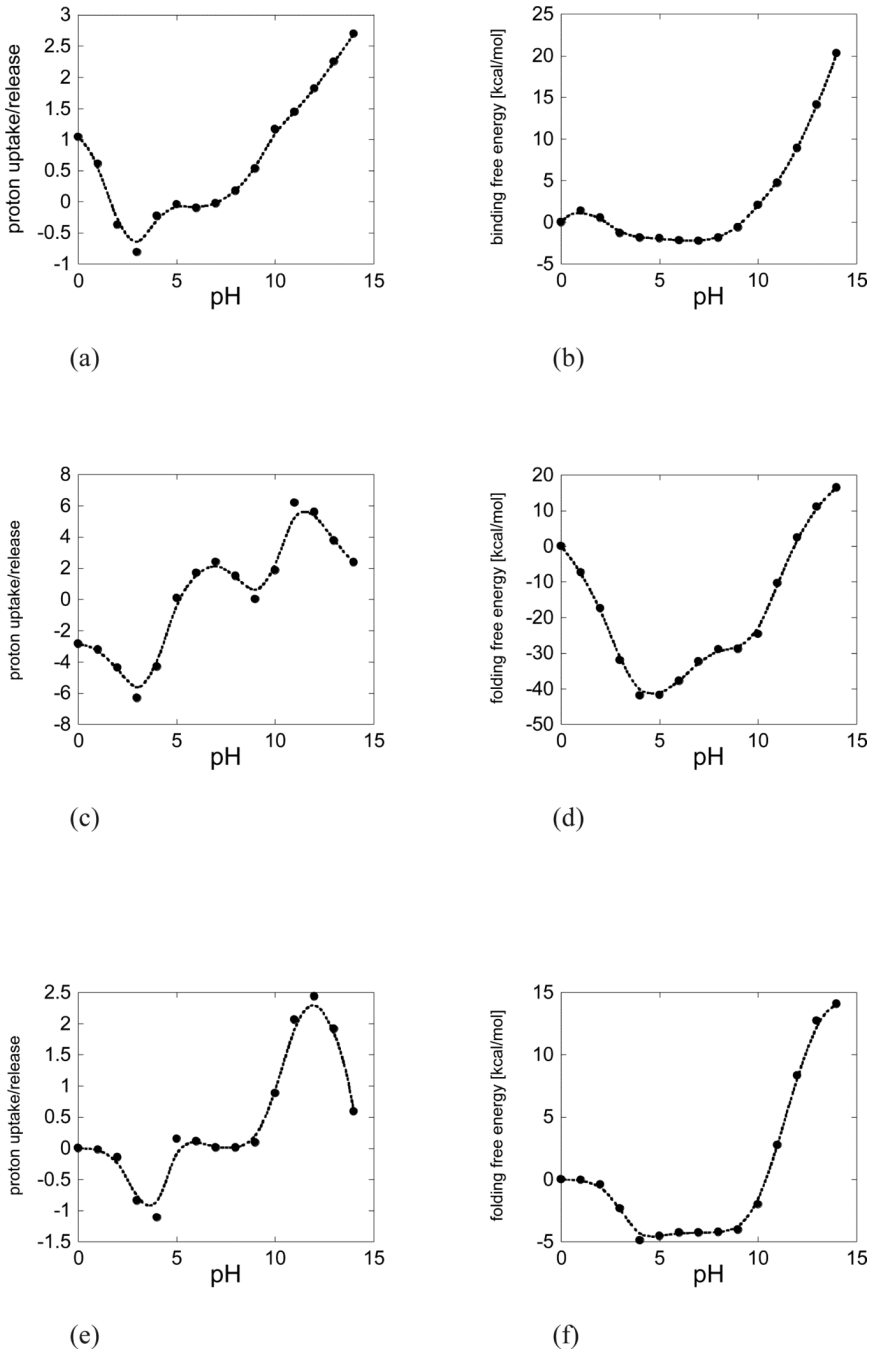

Figure 4.

Proton uptake/release and pH-dependence of the binding free energy for trypsin -and amyloid beta-protein precursor inhibitor domain complex and its monomers.

(a) Proton uptake/release upon formation of the 1BRC complex as a function of pH.

(b) pH-dependence of the binding free energy of 1BRC complex calculated with eq. (4).

(c) Proton uptake/release upon 1BRA folding as a function of pH.

(d) pH-dependence of the folding free energy of 1BRA as a function of pH.

(e) Proton uptake/release upon 1AAP folding as a function of pH.

(f) pH-dependence of the folding free energy of 1AAP as a function of pH

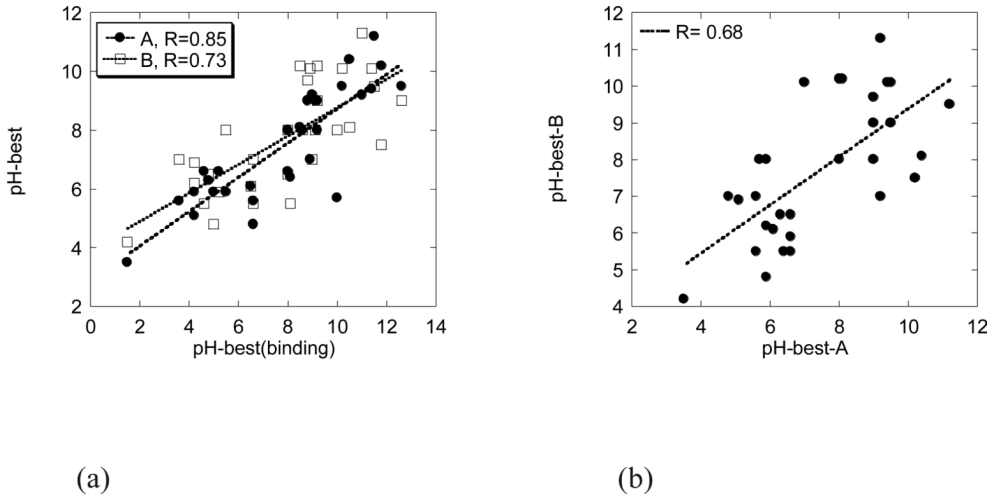

Figure 5.
Correlation between pH-best:
(a) pH-optimum of binding plotted against pH-best of folding of either A or B monomer.
(b) pH-optimum of folding of monomer A plotted against pH-best of folding of monomer B.

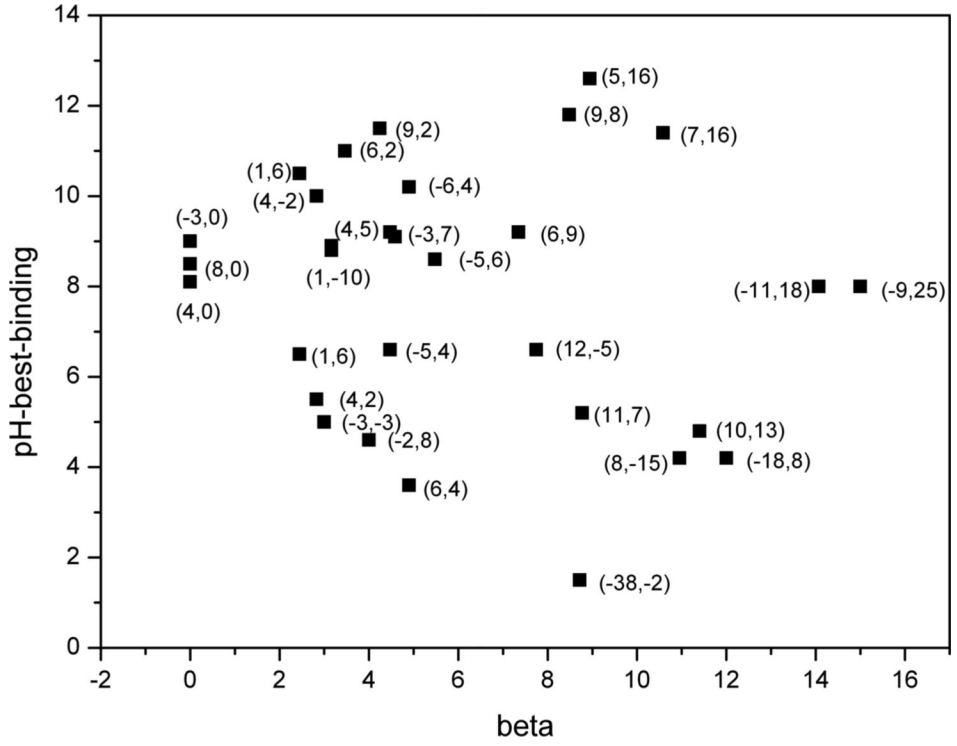

Figure 6.
The pH-best of binding plotted against the parameter beta. The net charge of the corresponding monomers estimated from their amino acid composition is shown in curved brackets for each data point.

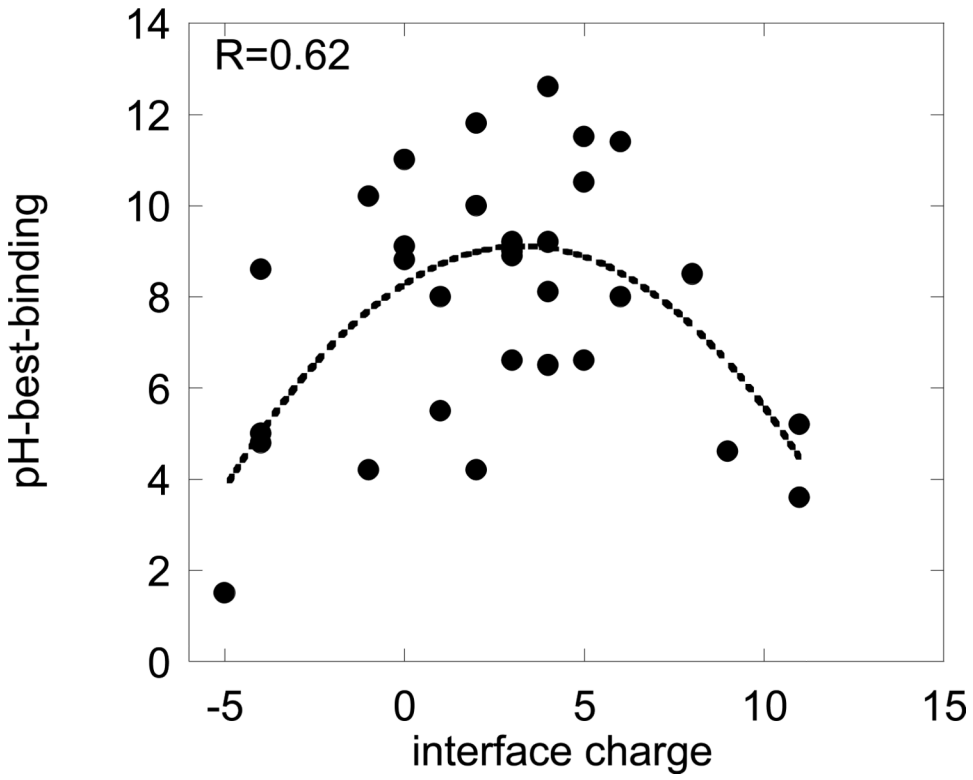

Figure 7.
The pH-best of binding plotted against interface charge.