COMMUNICATIONS IN COMPUTATIONAL PHYSICS
Vol. **3**, No. 5, pp. 1071-1086

Commun. Comput. Phys.
May 2008

Modeling Salt Dependence of Protein-Protein Association: Linear vs Non-Linear Poisson-Boltzmann Equation

Kemper Talley, Petras Kundrotas and Emil Alexov*

Computational Biophysics and Bioinformatics, Department of Physics, Clemson University, Clemson, SC 29634, USA.

Received 30 September 2007; Accepted (in revised version) 5 January 2008

Available online 25 January 2008

Abstract. Proteins perform various biological functions in the cell by interacting and binding to other proteins, DNA, or other small molecules. These interactions occur in cellular compartments with different salt concentrations, which may also vary under different physiological conditions. The goal of this study is to investigate the effect of salt concentration on the electrostatic component of the binding free energy (hereafter, salt effect) based on a large set of 1482 protein-protein complexes, a task that has never been done before. Since the proteins are irregularly shaped objects, the calculations have been carried out by a means of finite-difference algorithm that solves Poisson-Boltzmann equation (PB) numerically. We performed simulations using both linear and non-linear PB equations and found that non-linearity, in general, does not have significant contribution into salt effect when the net charges of the protein monomers are of different polarity and are less than five electron units. However, for complexes made of monomers carrying large net charges non-linearity is an important factor, especially for homo-complexes which are made of identical units carrying the same net charge. A parameter reflecting the net charge of the monomers is proposed and used as a flag distinguishing between cases which should be treated with non-linear Poisson-Boltzmann equation and cases where linear PB produces sound results. It was also shown that the magnitude of the salt effect is not correlated with macroscopic parameters (such as net charge of the monomers, corresponding complexes, surface and number of interfacial residues) but rather is a complex phenomenon that depends on the shape and charge distribution of the molecules.

PACS: 82.20.wt, 87.15.km

Key words: Poisson-Boltzmann equation, electrostatics, protein-protein complexes, salt-dependence of binding energy, finite-difference Poisson-Boltzmann method.

*Corresponding author. *Email addresses:* ktalley@clemson.edu (K. Talley), pkundro@clemson.edu (P. Kundrotas), ealexov@clemson.edu (E. Alexov)

1 Introduction

The electrostatic potential in media containing mobile charged particles obeys the Poisson-Boltzmann equation [1–3]. Examples of such media include colloidal systems [4, 5], macromolecules [1, 2, 6, 7], and membranes [8–10] in a physiological liquid with a non-zero concentration of mobile ions (non-zero ionic strength, I). While in the case of simple geometry, it is possible to obtain an analytical solution of the Poisson-Boltzmann equation, most of the practical cases do not fall into this category. This is especially true in the case of biological macromolecules in solution with a non-zero ionic strength that represents a system with multiple dielectric regions separated by irregularly shaped boundaries. Solving the Poisson-Boltzmann equation in this case requires numerical techniques such as finite-difference methods [11–13].

Modelling the interactions of biological macromolecules in a solution containing mobile ions is an important task that allows for an understanding of the mechanisms of protein-protein binding [14–16]. Among the forces driving the association of macromolecules, the electrostatic force is primarily affected by the ion "atmosphere" and thus at the first order of approximation, the salt dependence of the binding free energy could be considered purely electrostatic in origin (see Ref. [17] for detailed discussion). Therefore, calculating the change of the electrostatic binding free energy component as a function of the ionic strength provides a good estimate of the sensitivity of protein-protein interactions to the concentration of mobile ions.

It was shown in the past that the binding free energies associated with the formation of macromolecular complexes are generally extremely sensitive to ionic strength. For example, the binding of proteins to nucleic acids and to the surface of membranes containing anionic phospholipids exhibits a strong salt dependence that has been extensively studied both experimentally and theoretically [9, 18, 19]. The underlying principles are well-understood, and the calculated salt dependence of binding free energies based on the non-linear Poisson-Boltzmann equation (NLPB) are generally in remarkable agreement with experimental measurements. The salt-dependence of protein-protein interactions has also been studied experimentally [20–23], and it has generally been found that increases in ionic strength weaken binding affinities. The linear Poisson-Boltzmann equation (LPB) has been applied with considerable success on calculations of protein-protein binding free energies [14, 24] and on studying the salt dependence of the association rate constant [15, 25]. However, systematic studies of the salt effect on the binding free energy of protein-protein complexes using LPB and NLPB on a large scale of examples are absent from the scientific literature.

From the electrostatic standpoint, biological systems can be grouped into two categories: (a) systems with entities carrying large net charge such as protein-DNA/RNA and protein-membrane complexes and (b) systems that generally do not carry a large net charge such as protein-protein complexes. Nucleic acid or phospholipids have a large and fairly uniform negative charge density which results in a large accumulation of positively charged counter-ions in the vicinity of DNA/RNA [18,26] and phospholipid mem-

branes [27]. In contrast, proteins are usually not highly charged and are close to neutral, and their charge distribution is often not uniform. Thus, from a theoretical prospective, complexes containing nucleic acids [28] or charged lipids [29] are to be treated with the NLPB whereas the general expectation is that the LPB is adequate for most protein-protein complexes [14]. Therefore one might expect that the interactions between protein molecules are more easily treatable computationally as compared to the interactions between protein and nucleic acid and/or membrane. However, the non-uniformity of the charge distribution in proteins introduces a serious complexity not encountered with nucleic acids or phospholipid membranes, and this complexity may introduce effects that the numerical algorithm should be able to capture.

In this paper we apply the finite-difference method as implemented in the Delphi program [11, 12] with linear and non-linear protocols in order to calculate distributions of the electrostatic potential in binary macromolecular systems at different salt concentrations thereby obtaining the salt dependence of the binding free energy. The dataset used in this study consists of experimentally determined X-ray structures of 209 hetero and 1273 homo protein-protein complexes at 40% sequence identity level. Such a large and diverse set of protein-protein complexes has never been studied before and assures statistical significance of the obtained results. In addition, we performed calculations for the same set of proteins, but with their structures energetically minimized by the TINKER package [30]. By these means we tested the sensitivity of our calculations with respect to small atomic deviations from the equilibrium that are often observed in X-ray protein structures. The salt dependence of the binding free energy was calculated with NLBP and LPB, and the obtained results were analyzed with respect to net charge of the complexes, net charge of the monomers, interfacial area, and functional types of the complexes.

2 Methods

2.1 Protein-protein complexes used in the study and their parameters

Protein-protein hetero-complexes subjected to the study were extracted from the ProtCom [31] database, version June 2006 which can be accessed at `http://www.ces.clemson.edu/compbio/protcom`. That version contains 1771 entries at 95% sequence identity level. To avoid the bias toward overrepresented complexes, the entries were purged with *CD-hit* [32] at 40% sequence identity level for all components of the hetero complexes, including monomers that belong to the same protein-protein complex. This resulted in 299 structures out of which 39 structures were excluded from further considerations due to large defects in the PDB files like large missing segments of polypeptide chains. Note that in this study we did not use the latest release of our PROTCOM database, version 2006-11-02, since it contains "2-chain" complexes extracted from multichain structures, and the inclusion of such structures into electrostatic calculations for binary complexes may be incorrect since electrostatic interactions across interface of any given two monomers are sensitive to the presence of the third, fourth, etc. component of the complex.

In addition we extracted all of the 2-chain complexes with sequences of monomers that were almost identical (>95%) from the PDB databank and processed them in a manner similar to that used to create the PROTCOM database. The main difference was that the purging of these homo complexes was done only for monomers belonging to different complexes. Some structures were excluded as having large structural defects and thus this resulted to 2617 structures at 40% sequence identity level.

This resulted in an initial set of 260 hetero and 2617 homo protein-protein complexes, which after removing structures for which some of the computational procedures could not reach desired accuracy was reduced to 209 hetero and 1273 homo protein-protein complexes. The interfacial area was calculated with *surfv* program [12] by subtracting the accessible surface area of the complex from the accessible surface area of the free monomers. The net charge of the complex and the free monomers were calculated assuming that all titratable amino acids are in their charged state.

2.2 Hydrogen placement and energy minimization

Some of the structures in our initial data set had structural defects, and thus all structures were subjected to the *profix* program from the Jackal package developed in Honig's lab (`http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Jackal`) in order to add missing atoms and/or sequence fragments. All the complexes that had a missing segment chain of longer than fifteen amino acids were removed from the set because building such long sequence stretches could introduce significant structural errors. The rest of the structures were subjected to the TINKER [30] software to add missing hydrogens, using the *pdbxyz.x* and the *xyzpdb.x* modules of the package with the Charmm27 [33] force field parameters. We further will refer to that set of structures as the non-minimized set.

A common approach in computing biophysical quantities using the 3D structures of biological macromolecules is to refine these structures by performing energy minimization with a particular force field. Following this strategy, we created another set of protonated structures (hereafter referred to as the minimized set) for all the PDB files described above by running the *minimize.x* module of the TINKER between the runs of the *pdbxyz.x* and the *xyzpdb.x* modules. The *minimize.x* module performs energy minimization using the Limited Memory BFGS Quasi-Newton Optimization algorithm [30]. The implicit solvent was modeled using the Still Generalized Born model [34], and the internal dielectric constant was set to 1 to be consistent with the CHARMM27 force field parameters [35] used in the calculations. A weak convergence criteria was applied (RMS gradient per atom $= 0.1$) to make computation tractable. Even in such a case, minimizing 2617 protein-protein complexes, some of which were larger than 50,000 atoms, is quite a challenge from a computational point of view. Therefore for these calculations we utilized a High Throughput distributed computing resource, *CONDOR*, originally developed at the University of Wisconsin-Madison (`www.cs.wisc.edu/condor`) which is available at Clemson University with more than 2,000 single CPUs of computational power.

Then, the minimization of all structures was completed in less than two months.

2.3 Calculation of the electrostatic component of the binding energy

The salt dependent term of the electrostatic component of the binding energy ($\Delta\Delta G_{el}$) is calculated as the difference of the electrostatic free energies of the complex and of the free molecules [14]:

$$\begin{aligned}&\Delta\Delta G_{el}(I)\\&=\Delta\Delta G_{el}^{AB}(I)-\Delta\Delta G_{el}^{A}(I)-\Delta\Delta G_{el}^{B}(I)\\&=\{\Delta G_{el}^{AB}(I)-\Delta G_{el}^{AB}(I=0)\}-\{\Delta G_{el}^{A}(I)-\Delta G_{el}^{A}(I=0)\}-\{\Delta G_{el}^{B}(I)-\Delta G_{el}^{B}(I=0)\},\end{aligned}\quad(2.1)$$

where $\Delta\Delta G_{el}^{X}(I)$ is the salt dependent component of the electrostatic energy of the complex, monomer A and monomer B, respectively ($X=AB$, A or B) at salt concentration I. These terms were calculated with the Delphi [11, 12] program as the difference between the grid energy at $I\neq 0$ and the "reference" grid energy usually calculated at smallest I, which in our case was 0.1 M [36]. The grid energy is the sum over products of charge and potential at each grid point in the finite difference lattice ($129\times129\times129$ grids). The charge at each grid point is obtained from partial atomic charges based on an extrapolation procedure [37]. All runs were analyzed with respect to the resulting "scale" and it was found that for 94% of hetero- and 75% of homo-complexes the resulting scale was larger than 1 grid/Å. The vast majority of the rest of the cases, 6% of hetero- and 25% of homo-complexes resulted in scale very close to 1grid/Å. Resolution of 1grid/Å and better was shown to produce accurate electrostatic calculations with Delphi [12, 38]. However, there were four cases, corresponding to very large homo-complexes resulting in scale smaller than 0.5grid/Å. At such a resolution, the calculated electrostatic energies may not be accurate, but these cases are only 0.4% of the total pool of homo-complexes investigated in this work.

The calculations were performed assuming that all Arg, Asp, Glu and Lys residues are ionized in both free and bound states. His residues were considered to be neutral at pH= 7 because their standard pKa is about 6.0. Keeping all ionizable residues in their default charge state is obvious simplification which could affect the results, especially in the cases of interfacial residues with unusual charge states. However, computing pKa's of ionizable groups on such a large set of protein complexes is not a trivial problem and will be postponed for further investigation. In addition, in order to reduce the complexity of the problem, the possibility of ionization changes upon complex formation, suggested either by the experimental data [39, 40], and by the theoretical simulations [41–44]) as well as the pKa shifts induced by changes in the salt concentration [41, 45], were not considered. The results were obtained with an internal dielectric constant of 2 and external dielectric constant of 80. The molecular surface was generated using a water probe with a radius of 1.4Å [46]. The force field parameters for radii and partial charges were taken from the CHARMM 27 [47] force field and the grid filling was set to 70%. Convergence criteria was rmsc $=0.0001kT/e$, as described in Delphi manual which can be obtained from

`http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:Delphi`. The results were obtained using the LPB, but were repeated with the NLPB as well. In the case of the NLPB, the free energy was calculated as described by Sharp and Honig [48] and includes electrostatic stress and osmotic pressure terms as described in Ref. [48].

Using the minimized and non-minimized sets of hetero-complexes and homo-complexes, we calculated all above energy terms at salt concentrations 0.1, 0.2, 0.5 and 1.0M. The results obtained as $I=0.1M$ were taken as a "reference" state. Since we are interested only in the electrostatic component of the binding free energy, and in particular in its salt-dependent component, the total binding free energy which includes many other terms, need not be calculated. It is then most convenient to report all values for a given protein with respect to a reference state, which in this case is $I=0.1M$. Thus, for each of the complexes we got four $\Delta\Delta G_{el}(I)$ calculated at $I=0.1$, 0.2, 0.5 and 1.0M, respectively. The salt dependence is conveniently analyzed in term of $\ln(I)$. Taking the results at $\ln(I=0.1M)$ as reference, a in-house linear regression program calculated the slopes ($\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$) of all of the complexes mentioned above and these quantities will be further analyzed throughout the paper. The electrostatic component of the energy was calculated in kcal/mol and the ionic strength was taken in mol.

3 Results

3.1 Comparison with available experimental data

The primary goal of this study is to reveal the importance of the non-linear effect on the calculation of the salt dependence of the binding energy of a large set of protein-protein complexes. We have shown [14] in the past that PB equation as implemented in Delphi [11] can correctly capture most of the non specific salt effects. However, in this study we do not perform focusing, and we are using a different grid size as compared to the previous work [14]. Thus, it is important to comprehend if and how these differences could affect the results of the calculations. Among the complexes in our dataset, which was purged to 40% sequence identity, there are two complexes (PDB ID 2thz and 1b29) which are close relatives to E9Dnase-Im9 and Barnase-Barstar complexes which were reported in our previous study [14] and have available experimental data. The slopes $\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$, calculated in this work, are 1.36 (LPB) and 1.49 (NLPB) for E9Dnase-Im9 which are very close to the results reported in Ref. [14] (1.29 (LPB) and 1.31 (NLPB)). They all are in reasonable agreement with the experimental data of 2.17 [22]. The results for the second complex calculated using PDB ID 1b29 structure are 0.35 (LPB) and 0.42 (NLPB) compared to 0.67 (LPB) and 0.74 (NLPB) in Ref. [14]. The experimental data is 0.96 [21]. This comparison indicates that large-scale calculations reported in the present paper produce results that are in reasonable agreement with both, much more time consuming calculations and available experimental data.

3.2 Role of electrostatics on salt dependence of binding energy

Proteins are complex biological macromolecules that have an irregular shape and non-homogeneous charge distribution. Because of this, the resulting distribution of ions is also non homogeneous. The effect of ions on the binding energy is the difference between the energy of interaction between ions and permanent charges in the bound and unbound states (eq. 2.1). Since in both cases, bound and unbound, the contribution of the ions to the energy is favourable, the outcome depends on the difference of two energy terms. If the electrostatics does favour protein-protein association, one can expect that the salt effects would be symmetrical, i.e. the increase of the salt concentration in some complexes would weaken the binding, whereas in the others, this would make binding stronger. In terms of the quantity we use in this work (the slope of the fitting line, $\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$, these two cases correspond to a positive and negative slope, respectively. To test such a possibility, we plot the distribution of the average calculated slopes for all hetero- and homo-complexes (Fig. 1). To simplify the presentation we average the results obtained by solving NLPB and LPB and using minimized and non minimized structures. It can be seen that the slopes are positive for a majority (65%) of the hetero-complexes (Fig. 1A) and for a significant fraction (37%) of the homo-complexes, indicating that the increase of the salt concentration makes binding weaker. From a macroscopic point of view, this suggests that most of the complexes are formed between oppositely charged monomers, so the increase in the amount of salt ions screens the favourable electrostatic interactions between monomers thereby making binding weaker. However, analysis of the net charge of the monomers shows that only $\sim$50% of hetero-complexes are made of oppositely charged monomers, thus there is no bias toward such complexes. Even more interesting is the case of homo-complexes (Fig. 1B) since both monomers carry the same net charge so from macroscopic point of view they should repel each other electrostatically and thus the increase of the salt concentration would screen unfavourable interactions, making binding stronger. However, for a significant fraction of homo-complexes ($\sim$37% of the cases) the slope is positive. To investigate the role of the net charge of the monomers we define a parameter γ,

$$\gamma = \text{sign}(q(A)q(B))\sqrt{|q(A)q(B)|}, \tag{3.1}$$

where $q(A)$ and $q(B)$ are the net charges of monomer A and B, respectively. Thus, a complex made of two oppositely charged monomers will have negative γ, and the magnitude will be proportional to the square root of the product of charges. In the case of a complex made of the same polarity monomers, γ will be a positive number.

Fig. 2 shows the average slope as a function of parameter γ. As it was discussed above, from a macroscopic point of view, complexes having negative γ should result to positive ($\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$), i.e. the increase of the salt concentration should make binding weaker. In the opposite case, the slope for complexes with positive should be negative. However, in case of hetero-complexes (Fig. 2A), the fitting line is almost horizontal and the correlation coefficient is 0.3, indicating no correlation at all. An even worse

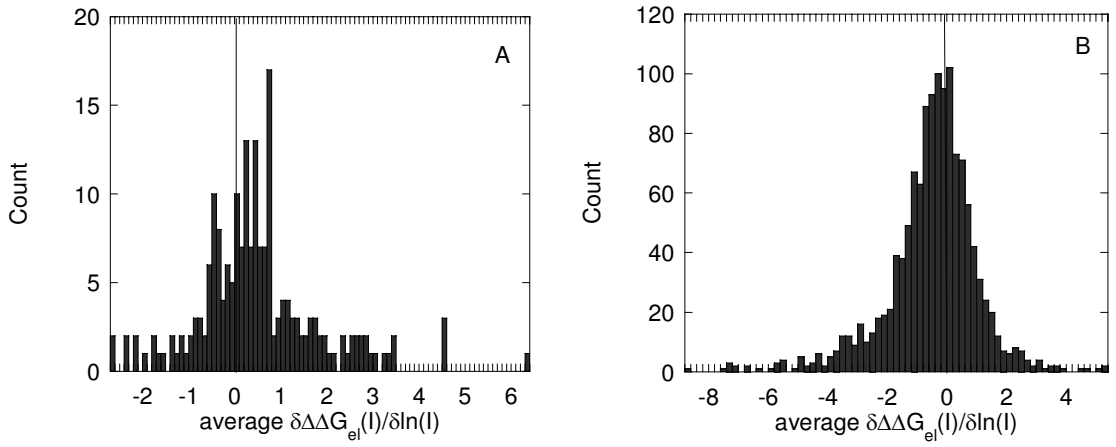

Figure 1: Distribution of the average slope ($\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$) of the fitting lines. The vertical line marks the zero line. The averaging is done using the results for NLPB and LPB, minimized and non minimized structures. Hetero-complexes are shown in panel A and homo-complexes in panel B.

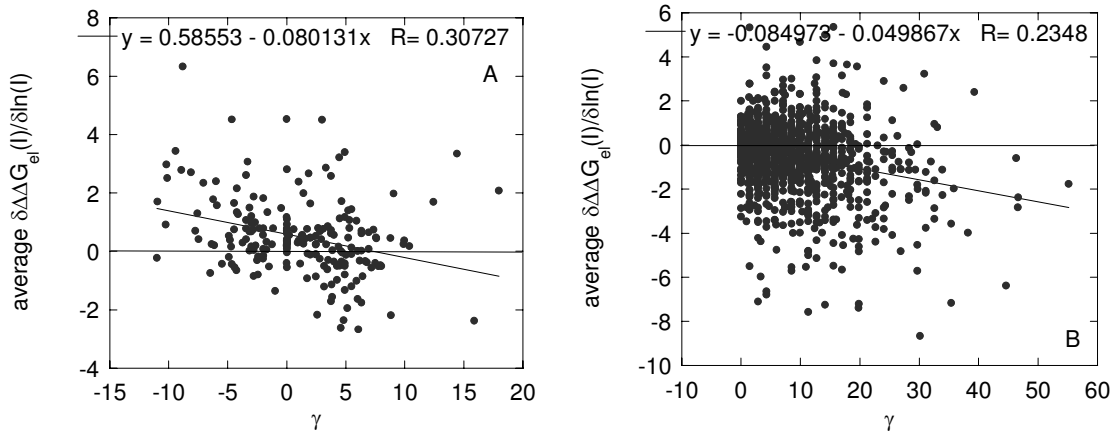

Figure 2: Average slope ($\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$) as a function of the parameter γ. The horizontal line marks the zero line. Hetero-complexes are shown in panel A and homo-complexes in panel B.

correlation is obtained in the case of homo-complexes (Fig. 2B), where the correlation coefficient is 0.23. Thus, the macroscopic parameters do not seem to be correlated with the calculated salt dependence of the binding energy. We also tried to correlate the calculated slopes with the sum net charge of the complexes, monomers and their combinations, but again no correlation was found. Similarly, the correlation coefficient between the calculated slopes and the interfacial areas is 0.13 and between the slopes and the number of interfacial residues is 0.15, which indicates no correlation at all. Thus, the salt effect is dominated by the shape and specific microscopic charge distribution in monomers and the corresponding complexes rather than by macroscopic parameters.

3.3 Effect of minimization

A crucial question in computational biophysics is how sensitive are the results with respect to feasible structural imperfections? To address this question, we performed minimization of all protein complexes used in this work. Then the resulting salt dependence of the binding energy was recalculated. The slopes of the fitting lines obtained with non-minimized and minimized structures are plotted against each other in Fig. 3, in the cases of solving NLPB and LPB, respectively for both hetero- and homo-complexes. It can be seen that the minimization has a little effect on NLPB results for hetero-complexes (Fig. 3A), while it tends to decrease the slopes calculated with LPB by 14% (Fig. 3B). However, the minimization has significant effect for homo-complexes (Figs. 3C and 3D). In both cases, NLPB and LPB, the minimization makes the calculated effect weaker by approximately 20%. The correlation coefficients in all cases are a good indicator that minimization has a similar effect on all structures and the calculated salt dependences of the binding energy. However, the largest effect of minimization is observed for the cases calculated to have large slopes. Most of these cases correspond to complexes made of highly charged monomers carrying the same polarity net charge. In such cases the electrostatic potential at the interfaces is very large due to the contributions from both monomers. The minimization adjusted the contacts across the interface of the complexes resulting in a different network of electrostatic interactions as compared to non-minimized structures. These relatively small changes in the atomic coordinates, however, resulted in a significant change of the slope $\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$ because of the large magnitude of the electrostatic potential.

3.4 Role of non-linearity

Solving the NLPB is more computationally expensive than solving the LPB and thus it is desirable to have some insights in which cases applying NLPB is necessary and in which cases it can be avoided. In addition, many applications require that the total electrostatic energy of the system can be written as a sum of the individual components, a requirement that can be satisfied if the system obeys LPB equation. Our previous work [14] on seven protein complexes demonstrated that non-linearity is not an important factor in the calculations of the salt dependence of the binding energy. However, this was shown on very limited set of data comprised mostly of hetero-complexes. Here we compare the salt dependence of the binding energy calculated on much larger set of protein complexes. The resulting slopes of the fitting lines are compared in Figs. 4A, 4B for hetero- and homo-complexes, respectively. It can be seen that many points are within the main diagonal, i.e. the slope of the fitting lines calculated with LPB and NLPB are almost the same. For these cases the non-linearity does not play a role in the calculations of the salt dependence of the binding energy. However, comparing Figs. 4A and 4B, it can be seen that the cases that fall into this category are a much smaller fraction for homo-complexes as compared to hetero-complexes. Many more points in Fig. 4B are located away from

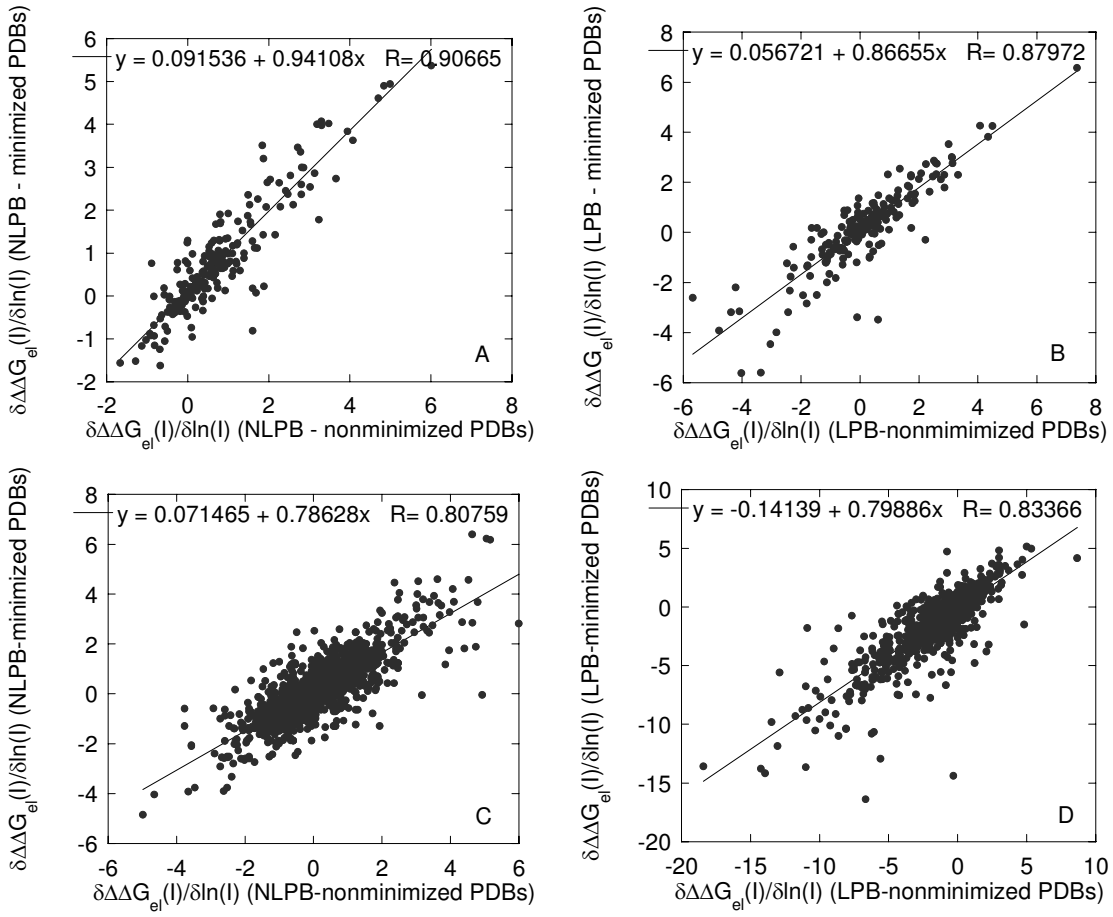

Figure 3: Fitting slopes ($\delta\Delta\Delta G_{el}(I)/\delta\ln[I]$) calculated with (A,C) NLPB and (B,D) LPB compared for non minimized and minimized structures for hetero- and homo-complexes, respectively.

the diagonal, and even a significant fraction of points for hetero-complexes (Fig. 4A) are located away from the main diagonal. The points being off the main diagonal in the upper left corner, indicate that non-linearity makes the salt dependence of the binding less favourable, i.e. makes the negative slopes less negative and positive slopes more positive (Fig.4). These results are consistent with our previous observation [14] reported on the limited set of protein complexes. This is better visualized in Fig. 5, where the distribution of the difference in the slopes (NLPB minus LPB) are plotted. It can be seen that these differences are predominantly positive (most of the bars are at positive numbers) for both hetero- (62% of cases) and homo-complexes (89% of cases). It should be pointed out, however, that most of the cases for hetero-complexes are clustered at or around zero, which confirms our observation made above that in many cases non-linearity does not play a role on the salt dependence of the binding energy in case of hetero-complexes. In contrast, the majority of results for homo-complexes are away from the zero point,

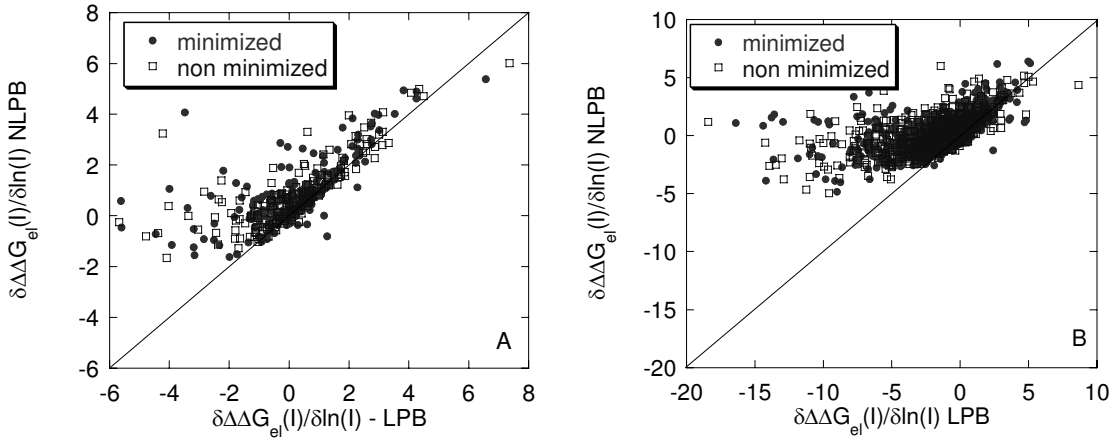

Figure 4: Comparison of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ calculated with LPB and NLPB. The straight line is the main diagonal. Points on the diagonal correspond to case where LPB and NLPB results are the same. Panel A - hetero-complexes and panel B - homo-complexes.

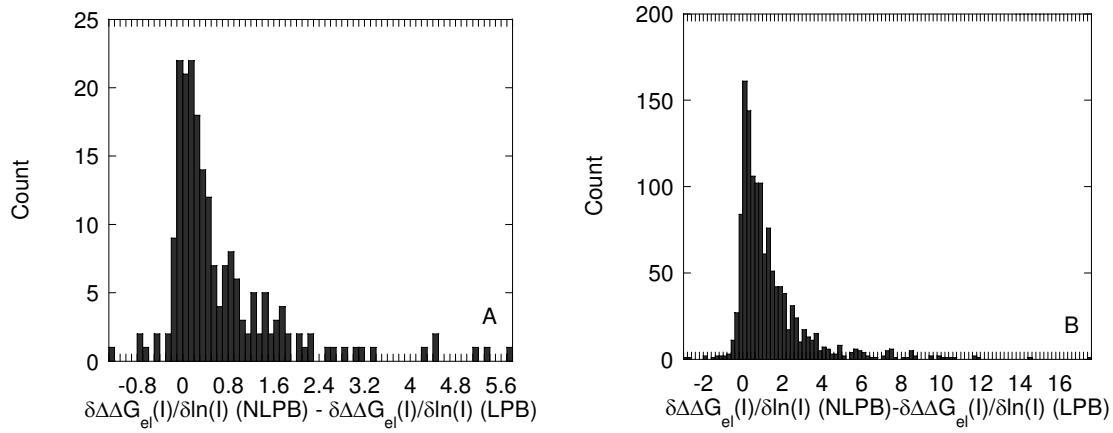

Figure 5: Distribution of difference of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ calculated with NLPB and LPB for hetero- (A) and homo-complexes (B).

confirming the importance of the non-linearity in calculating the salt dependence of the binding energy. It should be pointed out that the scales of the horizontal axes of Figs. 5A and 5B are different, because the non-linear effect is much stronger for homo-complexes as compared to hetero-complexes.

How should we distinguish between cases for which non-linearity is important and cases for which it is not, using some easily assessable macroscopic parameters that do not require extensive computing? Our previous work [14] suggested that non-linearity could be important in cases where complexes made of monomers carry a large net charge. To test such a possibility on a much larger set of data, we plotted the difference between $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ calculated with NLPB and LPB and as a function of parameter γ

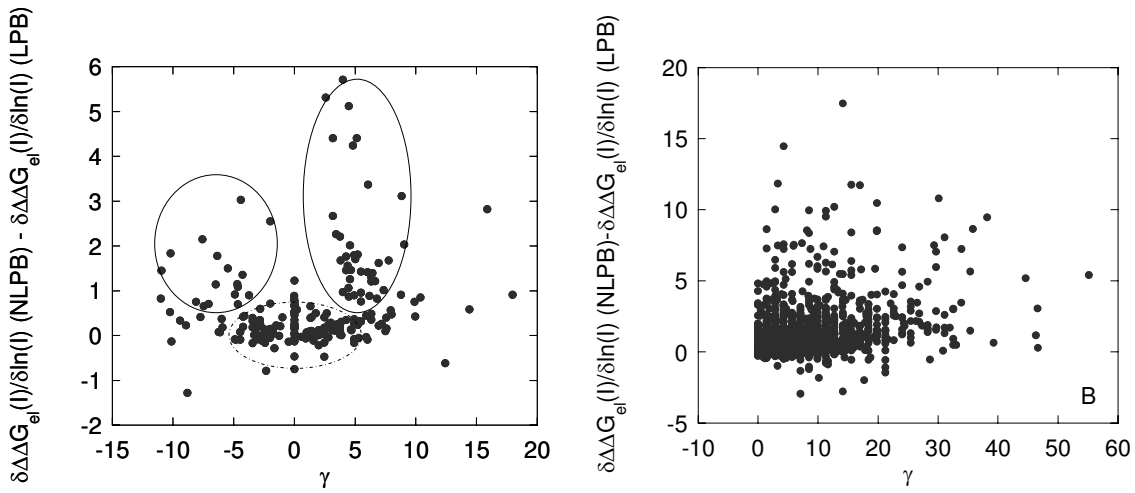

Figure 6: The difference of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ calculated with NLPB and LPB as a function of the parameter γ. Horizontal oval shows the cases for which the non-linearity is not important, while vertical ovals show the cases for which the non-linearity is important. Panel (A) shows results for hetero- and panel (B) for homo-complexes.

(Fig. 6). The hetero- and homo-complexes behave differently and thus will be discussed separately. In the case of hetero-complexes (Fig. 6A), it can be seen that cases corresponding to small γ (small net charges of the monomers), correspond to cases where the slopes of fitting lines are almost the same calculated with NLPB and LPB (the area marked with oval drawn with broken line). At γ larger than 5 or smaller than -5, the effect of non-linearity grows (the areas marked with ovals drawn with solid lines). Of course, there are several cases for which the non-linearity does not change the results although the corresponding γ is either a large positive or negative number. However, these exceptional cases are very few in number and do not affect the general conclusion that the parameter γ can be used to predict what cases should be treated with NPLB equation. In the case of homo-complexes (Fig. 6B), the non-linearity always plays an important role in the calculations of the salt dependence of the binding. Even at small γ, the difference between results calculated with NLPB and LPB is significant for most of cases. Note that the scales for vertical axes are different in Figs. 6A and 6B, because the effect is much stronger for homo-complexes as compared to hetero-complexes. Thus, the salt effects should be calculated with NLPB in the case of homo-complexes.

4 Discussion

In this paper, we studied the dependence of the electrostatic free energy of the binding of 1482 protein-protein complexes as a function of ionic strength. For the first time, a very large set of protein complexes was subjected to full scale energy minimization and electrostatic calculations. The complexes included in the calculations were very diverse

including hetero and homo complexes. The interfacial area varies from 275 Å^2 to about 7065 Å^2. The monomers of the complexes had different net charges from almost neutral to several tenth electron units. The number of interfacial residues varies from 15 to 347. Such a large and diverse set of data assures that the obtained results are statistically representative. Also in several cases for which experimental data was available, our calculations produce a slope that is similar to the experimental value.

We show in this work that the origin of the salt dependence of the binding free energy is not due to net charge effects, so two interacting proteins cannot simply be viewed as two interacting macro-ions. Rather, the effect results from the detailed charge distribution on each protein. The screening of Coulomb interactions is certainly the most obvious explanation (for detailed discussion of all energy terms see Ref. [14]). However, it should be pointed out that the binding energy is a result from the difference in the energies of the free and bound proteins. Thus, Coulomb interactions between ionizable groups buried in the interface are expected to be less sensitive to salt concentration by virtue of being removed from the solvent; however these groups were interacting with the ions prior the complex's formation. Additional complexity comes from the fact that strong electrostatic interactions are usually between groups forming ion pairs so that they will in general be closer to each other than the Debye length, and thus such interactions will not be very sensitive to ion concentration. Therefore, the electrostatic interactions that are expected to be the most sensitive to salt concentration would be between the groups that are relatively far from one another and exposed on the protein surface. However in absolute energy terms, these interactions should be relatively weak since they are already effectively screened by the high dielectric solvent. The interplay between these effects at the end determines the salt dependence of the binding energy. In addition, our attempt to correlate the slope of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ with the interfacial area or number of interfacial residues did not result in any correlation (results not shown). Thus, calculating the salt effect requires a detailed atomic model of the protein complexes and numerical solution of PB equation.

Our work demonstrated that minimization of the 3D structures does not significantly affect the average results in case of hetero-complexes. The average effect was found to be almost zero for the results obtained with NLPB and about 14% for the corresponding calculations with LPB. However, in some particular cases, the minimization was found to dramatically alter the predicted slope of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$. Although such cases are a minor fraction of the complexes considered here, this indicates that some structures undergo large structural changes under minimization, and these cases should be treated on individual basis. In contrast, the minimization was found to reduce the salt effect on the binding energy of homo-complexes by approximately 20% for both LPB and NLPB. The electrostatic potential between entities carrying the same polarity charge is much stronger than the same for oppositely charged molecules. This is the reason why the minimization has a stronger effect on the calculations involving homo-complexes. Small structural rearrangement caused by the minimization generates a significant effect because of the large amplitude of the electrostatic potential at the interface of such complexes.

One of the major findings of our investigation is that the non-linear effects are not important for the calculations of salt dependence of the binding energy of hetero-complexes. In a vast majority of the cases considered here the difference between the slope of $(\delta\Delta\Delta G_{el}(I)/\delta\ln[I])$ calculated with NLPB and LPB is within 0.5. This finding was related to a parameter γ, which reflects the net charge of the monomers. It was demonstrated that if γ is bounded between -5 and 5 for a particular protein-protein complex, then non-linearity is not expected to affect the results. However, in a significant number of cases, at which γ is outside the above mentioned limits, it is expected that non-linearity will be important factor, so to obtain correct results one should solve NLPB equation. Thus, the parameter γ serves as a flag that allows for fast determination of the necessity of applying NLPB to calculate electrostatic effect associated with protein-protein complexes. It is important to know if the electrostatic potential of a given system can be correctly calculated with LPB, since this will allow for additivity of the individual components, a feature important for many applications requiring extensive sampling. In contrast, in the case of homo-complexes no correlation was obtained between γ and the magnitude of the non-linear effect. Perhaps this is again a result of the strong electrostatic potential at the interfaces of these complexes. Such a strong potential increases the magnitude of the non-linear terms on the PB equation and they cannot be neglected. This resulted in different slopes calculated with NLPB and LPB. Therefore, the salt dependent effects for homo-complexes should be calculated by solving NLPB.

Why electrostatic energy of homo-complexes should be calculated with NLPB, while it can be correctly calculated with LPB in case of hetero-complexes. To address this question we investigated the charge complementarity at the corresponding interfaces. It was found that in 59.2% of the homo-complexes the interfaces of the monomers carry the same polarity charge. In contrast, such cases are only 19.3% of hetero-complexes. Only 8.4% of the homo-complexes calculated in this work are made of monomers carrying complementary interfacial charges, while in case of hetero-complexes this fraction is significant, 57.9%. This difference indicates that in majority of the cases, the electrostatic potential at and nearby interfaces of homo-complexes is very strong and therefore non-linearity can not be neglected. In contrast, the potential at or nearby the interfaces of the hetero-complexes is close to zero, since most of the interfaces carry opposite charges and therefore the non-linearity is not a crucial factor.

Acknowledgments

The minimization of the structures used in this work was made possible by a Condor pool deployed and maintained by Clemson Computing and Information Technology. The authors would like to acknowledge the support of the staff from the Cyber infrastructure Technology Integration group. This research was supported by an award to Clemson University from the Howard Hughes Medical Institute Undergraduate Science Education Program.

References

[1] A. Warshel, S. Russell, Quart. Rev. Biophys. 17 (1984) 283-422.
[2] K. Sharp, B. Honig, Annu. Rev. Biophys. Biophys. Chem. 19 (1990) 301-332.
[3] M. K. Gilson, M. E. Davis, B. A. Luty, J. A. McCammon, J. Phys. Chem. 97 (1993) 3591-3600.
[4] H. Ohshima, E. Mishonova, E. Alexov, Biophys. Chem. 57 (1996) 189-203.
[5] M. N. Tamashiro, V. B. Henriques, M. T. Lamy, Langmuir 21 (2005) 11005-16.
[6] A. Warshel, A. Papazyan, Curr Opin Struct Biol 8 (1998) 211-7.
[7] V. K. Misra, K. A. Sharp, R. A. Friedman, B. Honig, J. Mol. Biol. 238 (1994) 245-63.
[8] M. E. Loosley-Millman, R. P. Rand, V. A. Parsegian, Biophys. J. 40 (1982) 221-32.
[9] D. Murray, S. McLaughlin, B. Honig, J. Biol. Chem. 276 (2001) 45153-45159.
[10] D. Murray, N. Ben-Tal, B. Honig, S. McLaughlin, Structure 5 (1997) 985-989.
[11] W. Rocchia, E. Alexov, B. Honig, J. Phys. Chem. 105 (2001) 6507-6514.
[12] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, B. Honig, J. Comp. Chem. 23 (2002) 128-137.
[13] B. A. Luty, M. E. David, J. McCammon, J. Comp. Chem. 13 (1992) 1114-1118.
[14] C. Bertonati, B. Honig, E. Alexov, Biophys J 92 (2007) 1891-9.
[15] H.-X. Zhou, Biopolymers 59 (2001) 427-433.
[16] A. S. Thomas, A. H. Elcock, J. Am. Chem. Soc. 128 (2006) 7796-806.
[17] H. X. Zhou, Proteins 61 (2005) 69-78.
[18] V. K. Misra, J. L. Hecht, K. A. Sharp, R. A. Friedman, B. Honig, J. Mol. Biol. 238 (1994) 264-80.
[19] N. Ben-Tal, B. Honig, R. M. Peitzsch, G. Denisov, S. McLaughlin, Biophys. J. 71 (1996) 561-575.
[20] Z. Radic, P. D. Kirchhoff, D. M. Quinn, J. A. McCammon, P. Taylor, J. Biol. Chem. 272 (1997) 23265-77.
[21] G. Schreiber, A. R. Fersht, J. Mol. Biol. 248 (1995) 478-486.
[22] R. Wallis, G. R. Moore, R. James, C. Kleanthous, Biochemistry 34 (1995) 13743-50.
[23] P. K. Nielsen, B. C. Bonsager, C. R. Berland, B. W. Sigurskjold, B. Svensson, Biochemistry 42 (2003) 1478-87.
[24] F. Dong, M. Vijayakumar, H. X. Zhou, Biophys. J. 85 (2003) 49-60.
[25] M. Vijakumar, K.-Y. Wong, G. Schreiber, A. R. Fersht, A. Szabo, H.-X. Zhou, J. Mol. Biol. 278 (1998) 1015-1024.
[26] C. L. Tang, E. Alexov, A. M. Pyle, B. Honig, J Mol. Biol. 366 (2007) 1475-96.
[27] A. A. Gurtovenko, M. Miettinen, M. Karttunen, I. Vattulainen, J. Phys. Chem. B 109 (2005) 21126-34.
[28] K. Sharp, B. Honig, Curr. Opin. Struct. Biol. 5 (1995) 323-328.
[29] N. Ben-Tal, B. Honig, C. Miller, S. McLaughlin, Biophys. J. 73 (1997) 1717-1727.
[30] J. W. Ponder, TINKER-software tools for molecular design, Washington University: St. Luis, 1999.
[31] P. J. Kundrotas, E. Alexov, Nucleic Acids Research 35 (2007) D575-D579.
[32] W. Li, L. Jaroszewski, A. Godzik, Bioinformatics 17 (2001) 282-3.
[33] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187-217.
[34] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127-6129.
[35] A. D. MacKerell Jr., D. Bashford, M. Bellot, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau,

C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, J. Phys. Chem. 102 (1998) 3586-3616.
[36] M. K. Gilson, B. H. Honig, Proteins 3 (1988) 32-52.
[37] M. K. Gilson, K. A. Sharp, B. H. Honig, J. Comput. Chem. 9 (1987) 327-335.
[38] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A.Case, C. L. Brooks, J. Comput. Chem. 25 (2004) 265-84.
[39] J. E. Nielsen, J. A. McCammon, Protein Science 12 (2003) 1894-1901.
[40] J. Gomez, E. Freire, J. Mol. Biol. 252 (1995) 337-350.
[41] A. Elcock, A. McCammon, J. Mol. Biol. 280 (1998) 731-748.
[42] K. A. Sharp, Biophys. Chem. 61 (1996) 37-49.
[43] J. Trylska, J. Antosiewich, M. Geller, C. Hodge, R. Klabe, M. Head, M. Gilson, Protein Science 8 (1999) 180-195.
[44] E. Alexov, Proteins 56 (2004) 572-584.
[45] M. J. Cocco, Y.-H. Kao, A. T. Phillips, J. T. Lecomte, Biochemistry 31 (1992) 6481-6491.
[46] E. Alexov, Proteins 50 (2003) 94-103.
[47] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, J. Comp. Chem. 4 (1983) 187-217.
[48] K. A. Sharp, B. Honig, J. Phys. Chem 94 (1990) 7684-92.