

pH-Dependent Association of Proteins. The Test Case of Monoclonal Antibody HyHEL-5 and Its Antigen Hen Egg White Lysozyme

Maciej Długosz,^{*,†} Jan M. Antosiewicz,[‡] and Joanna Trylska[†]

Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Żwirki i Wigury 93, Warsaw 02-089, Poland, and Division of Biophysics, Department of Physics, University of Warsaw, Żwirki i Wigury 93, Warsaw 02-089, Poland

Received: July 19, 2009; Revised Manuscript Received: October 1, 2009

We describe a method for determining diffusion-controlled rate constants for protein–protein association that explicitly includes the solution pH. The method combines the transient-complex theory for computing electrostatically enhanced association rates with an approach based on a rigorous thermodynamic cycle and partition functions for energy levels characterizing protonation states of associating proteins and their complexes. To test our method, we determine the pH-dependent kinetics of association of the HyHEL-5 antibody with its antigen hen egg white lysozyme. It was shown experimentally that their association rate constant depends on pH, increasing linearly in the pH range 6–8 and saturating or even exhibiting a flat maximum in the pH range 8–10. The presented methodology leads to a qualitative agreement with the experimental data. Our approach allows one to study diffusion-controlled protein–protein association under different pH conditions by taking into account the ensembles of protonation states rather than just the most probable protonation state of each protein.

Introduction

All proteins contain ionizable groups that exchange protons with their environment. In solution, proteins fluctuate between available protonation states whose distribution is specific to a given pH. Protein ionizable groups are involved in intraprotein, protein–solvent, protein–protein, and protein–ligand interactions influencing protein solubility, folding, and catalytic activity.¹ A proton transfer between ionizable groups and the surrounding solvent is a process with a pH-dependent equilibrium. To quantitatively describe such protonation equilibria in proteins, pK_a 's of ionizable groups are often used. The pK_a of a specific ionizable group in a single protein or in a protein–protein complex can be significantly shifted relative to the pK_a of the corresponding isolated group in solution.^{2,3} In addition, shifts in pK_a 's between native and denatured states of proteins or between unbound and complexed forms are a source of the pH dependence of macromolecular stability or binding.⁴ Because the exchange of protons modifies protein charge density, protein electrostatic properties are strongly related to the pH. Therefore, taking into account the protonation equilibria in proteins is essential for modeling their pH-dependent properties and functioning.

Association of proteins with each other and with other biomolecules is an important step in many biological processes. pH-dependent protonation equilibria determine both local charge distributions and protein net charges influencing protein–protein interactions and the kinetics of binding. Uptake or release of protons by protein ionizable groups can accompany their association and cause the association rate constant to be dependent on pH.

Effects of pH on protein association have not been widely studied. Standard theoretical techniques used to estimate rates

of protein–protein association (such as Brownian dynamics) only partially take into account the pH effects.^{5,6} Associating proteins are assigned single protonation states which are the most probable ones at a given pH either for isolated molecules or for their complexes. The facts that each protein exists as an ensemble of protonation states (with a number of different protonation states readily accessible by thermal agitations if their energies are sufficiently close to the energy of the most probable state) and that the protonation preferences of each of the associating proteins can be changed upon binding are neglected. Additionally, the pH at which the association occurs is not explicitly included during computations.

It was previously shown that the electrostatically enhanced rate of a diffusion-limited association can be modeled with reasonable accuracy by^{7–9}

$$k_a = k_{ao} e^{-\langle E_{el} \rangle^*/k_B T} \quad (1)$$

where k_a and k_{ao} are, respectively, the association rates in the presence and absence of electrostatic interactions, $\langle E_{el} \rangle^*$ is the average electrostatic interaction energy in a transient-complex ensemble that separates the bound and unbound states of proteins and is computed with the Poisson–Boltzmann (PB) model,¹⁰ and $k_B T$ is the thermal energy. This approach was successfully applied to study the influence of ionic strength and mutations on protein–protein and protein–RNA association kinetics.^{7–9} In this study, we extend this model by explicitly introducing the protonation degrees of freedom and the solution pH. To evaluate E_{el} , we simultaneously use the partition functions for the protonation states of proteins (both isolated and complexed) and the thermodynamic cycle, providing a common reference level for the computed energies.

Even though protonation equilibria are crucial for many biological processes, the amount of available experimental data of protein–protein association kinetics under varying pH is

* Corresponding author. E-mail: mdlugosz@icm.edu.pl.

† Interdisciplinary Centre for Mathematical and Computational Modelling.

‡ Department of Physics.

limited. We chose to test our approach on a pair of proteins: the antihen egg lysozyme monoclonal antibody (HyHEL-5) and its antigen, hen egg lysozyme (HEL) (Figure 1). The large number of ionizable groups in both proteins and extensive titrations of mutually influencing residues makes the chosen system an interesting and challenging test case. The kinetics of association of these proteins under various conditions was previously studied by stopped-flow fluorescence polarization.¹² It was found that the association rate of HyHEL-5/HEL complex varies with pH in the range 6.0–10.0. We determined the association rate constant in the HyHEL-5/HEL system using both the new method and the standard one. We show that including the protonation degrees of freedom and pH while evaluating the electrostatic interaction energies leads to significantly better agreement with experiment than the standard methodology.

Methods

Transient-Complex Theory for Protein Association. Diffusion-Limited Association Rate Constant. The stereospecific binding of two proteins can be described as a consecutive two-step process. First, proteins (A and B) approach each other via translational and rotational diffusion and reach a region in configurational space where their separations and orientations are close to those observed in the final complex. Such configurations of proteins constitute the transient-complex ensemble (C^*). Second, the conformational rearrangement of proteins leads to the stereospecific final complex (C):⁷



The overall association rate is:

$$k_{\text{on}} = \frac{k_a k_b}{k_{-a} + k_b} \quad (3)$$

If the process is diffusion-controlled, $k_b \gg k_{-a}$ and $k_{\text{on}} \approx k_a$ approximation holds. In the transient-complex theory, the transient-complex ensemble represents configurations that are as close to the native complex as possible without the need to consider short-range interactions and conformational rearrangements. If the two proteins in the transient complex are severely constrained in relative translation and rotation, the diffusion-controlled rate constant to reach the transient complex, k_a is given with eq 1.

Specification of the Transient-Complex. The transient-complex ensemble is located at the outer boundary of the native-complex interaction energy well. The procedure to generate and identify the transient-complex ensemble follows that described previously.^{8,9} Proteins, in a configuration as in the native complex, are treated as rigid bodies with only six degrees of freedom: three for relative translation and three for relative rotation. The larger of the two proteins, HyHEL-5 is fixed in the laboratory frame. HEL translational and rotational coordinates are defined relative to the native-complex configuration built from the X-ray structure. Translation is represented in spherical coordinates by a relative displacement vector $\vec{\rho}(\rho, \theta, \phi)$. Its norm, ρ , measures

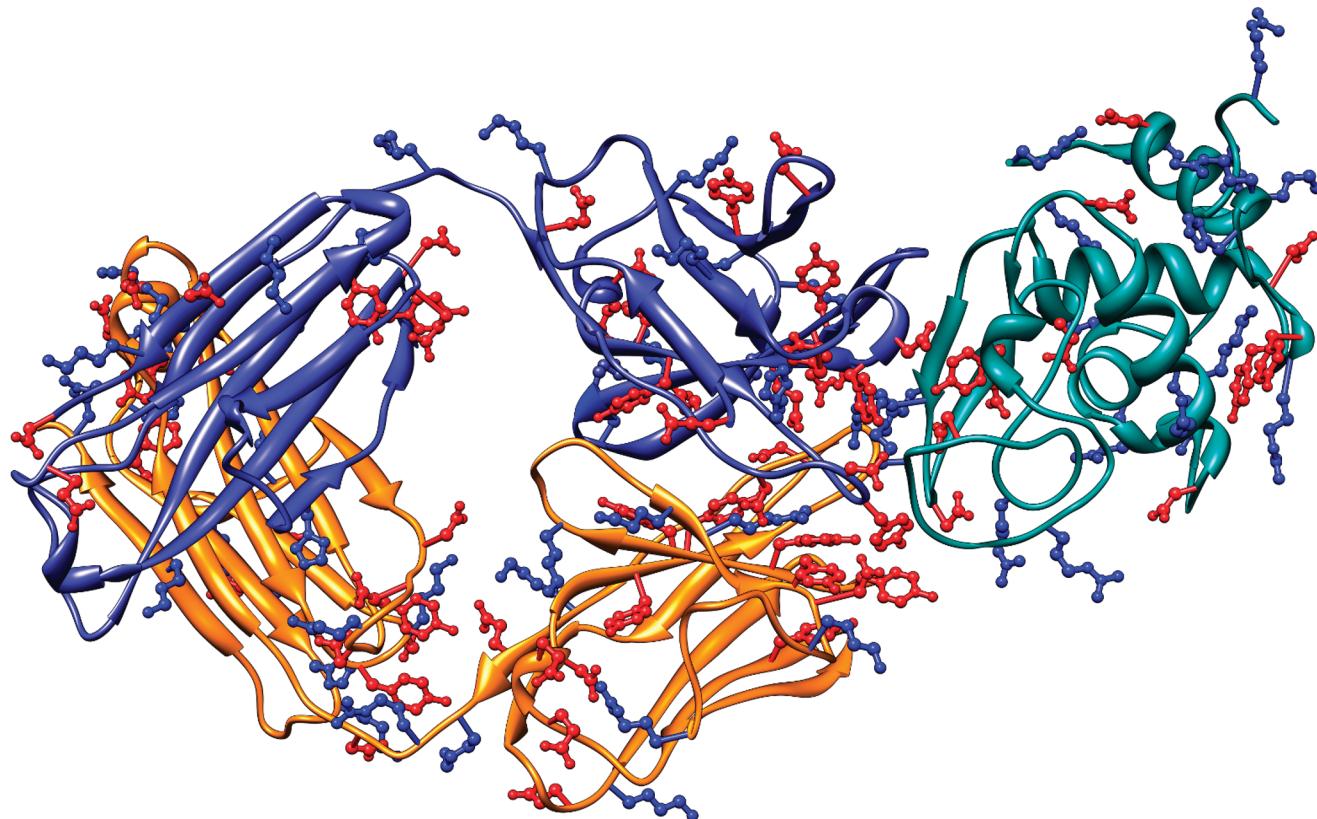


Figure 1. Cartoon model of the HyHEL-5/HEL complex. Cyan, HEL; orange, HyHEL-5 light chain; blue, HyHEL-5 heavy chain. Ionizable groups are shown in blue (basic) and red (acidic) with the ball and stick representation. Drawing was done using the UCSF CHIMERA.¹¹

the separation between the surfaces of the two proteins. Rotation is represented by three Euler angles (α, β, γ) of the HEL-fixed frame, relative to the laboratory frame. The α angle describes the rotation around a HEL-fixed unit vector; β and γ are polar and azimuthal angles of the HEL-fixed vector in the laboratory frame. Coordinates are sampled from $\rho \in [\rho_0, \rho_0 + 6]$ (where ρ_0 denotes the distance measured in the X-ray structure), $\theta \in [0, 180^\circ]$, $\phi \in [0, 360^\circ]$, $\alpha \in [-180^\circ, 180^\circ]$, $\beta \in [-180^\circ, 180^\circ]$ and $\gamma \in [0, 180^\circ]$.

We generate random configurations of the two proteins by sampling the global coordinates ($\rho, \theta, \phi, \alpha, \beta, \gamma$) from uniform distributions so that all configurations are equally probable. Only the nonoverlapping configurations are accepted and stored for further analysis. The overlaps are detected by testing all atoms of HyHEL-5 against all atoms of HEL. The distances among three classes of atoms are considered: hydrogens, polar (oxygens and nitrogens), and nonpolar (carbons and other atoms). We assume that an overlap between atoms within the same or different classes occurs when their distance in the generated configuration is smaller than the minimum distance of such a pair in the X-ray native structure.

Nonoverlapping configurations and contacts (native and non-native) are used to locate the boundary of the native-complex interaction energy well, as previously described.^{8,9} We choose a number of distinct atom pairs belonging to the interface between proteins, and these native pairs form a list used to define contacts. Each atom is assigned a contact radius computed as half of the distance from its contact partner in the native structure of the complex. Native contacts occur if for a given nonoverlapping configuration the distance between a native pair of atoms is smaller than the sum of its native distance plus 3.5 Å. Nonnative contacts are defined when any of the two atoms not forming a native pair are within the distance smaller than the sum of their contact radii plus 2.5 Å. All steps of the procedure leading to the transient-complex ensemble were performed with a homemade software.

As stated above, the transient-complex ensemble separates the bound and unbound states of proteins. Its location can be identified by measuring the translational and rotational freedom of proteins in nonoverlapping configurations. This can be done using $\sigma_\alpha(N)$, the standard deviation of the α angle among all nonoverlapping configurations for a contact level, N ,^{8,9} or Ξ , the difference between $\sigma_\alpha(N)$ and its average at all contact levels lower than N . The maximum of Ξ coincides with the transient complex. The level of contacts characteristic for the transient-complex ensemble can be therefore identified by examining Ξ as a function of N :^{8,9}

$$\Xi(N) = \langle \sigma_\alpha(N') \rangle_{N' < N} - \sigma_\alpha(N) \quad (4)$$

Calculation of the pH-Dependent Electrostatic Interaction Energy. To compute E_{el} that includes protonation degrees of freedom, we use the previously described thermodynamic cycle.¹³ The electrostatic interaction energy for each configuration of proteins is

$$E_{\text{el}} = G_{\text{el}}^{\text{o,C}} - (G_{\text{el}}^{\text{o,A}} + G_{\text{el}}^{\text{o,B}}) + \Delta G_{\text{el}}^{\text{nn}} \quad (5)$$

where $G_{\text{el}}^{\text{o,A}}$, $G_{\text{el}}^{\text{o,B}}$, and $G_{\text{el}}^{\text{o,C}}$ are electrostatic free energies of isolated proteins and of their configuration taken from the transient-complex ensemble. These terms explicitly include the solution pH as described further on. The last term, $\Delta G_{\text{el}}^{\text{nn}}$, computed for molecules with all ionizable groups of proteins

in their neutral states, is pH-independent,

$$\Delta G_{\text{el}}^{\text{nn}} = \frac{1}{2} \sum_{i=1}^{N_A+N_B} q_i \phi_i - \left(\frac{1}{2} \sum_{i=1}^{N_A} q_i \phi_i + \frac{1}{2} \sum_{i=1}^{N_B} q_i \phi_i \right) \quad (6)$$

$\Delta G_{\text{el}}^{\text{nn}}$ is evaluated using the traditional PB approach where the solute is represented as a low-dielectric region with N_A , N_B , and $N_A + N_B$ partial atomic charges located at positions of atomic nuclei. The solute is surrounded by a high-dielectric continuum solvent containing dissolved electrolyte. The distribution of mobile charges (ions) in solvent is described according to the Boltzmann law. Dielectric polarizability is included implicitly by dielectric constants of the media. In eq 6, ϕ_i is the electrostatic potential at the location of the charge q_i computed as a sum of Coulombic potentials due to all charges except q_i and the total reaction field at the location of the charge q_i resulting from the solute–solvent dielectric boundary.

The electrostatic free energy of a molecule with M ionizable groups can be computed as

$$G_{\text{el}}^{\text{o}} = -RT \sum_m^{2^M} \exp\left(-\frac{G_m}{RT}\right) \quad (7)$$

where G_m is the free energy level associated with the m th protonation vector $\{x_1^m, x_2^m, \dots, x_M^m\}_{m=1,2, \dots, 2^M}$ (where x_i^m is either 0 or 1) and R is the gas constant. At a given solution pH and temperature, the free energy of the molecular state characterized by the m th protonation vector is¹⁴

$$G_m \equiv G(x_1^m, \dots, x_M^m, \text{pH}, T) = 2.303RT \sum_{i=1}^M x_i^m \gamma_i (\text{pH} - \text{p}K_{\text{a},i}^{\text{mod}}) + \sum_{i=1}^M x_i^m \Delta \Delta G_{ii} + \sum_{i=1}^{M-1} \sum_{j=i+1}^M x_i^m x_j^m \Delta \Delta G_{ij}^{\text{int}} \quad (8)$$

where $\text{p}K_{\text{a},i}^{\text{mod}}$ is the negative decimal logarithm of the acidic dissociation constant of the isolated i th ionizable group (known from experiments or quantum mechanical computations for small model compounds), $\Delta \Delta G_{ii}$ is the free energy change for ionizing protein's i th site when all other ionizable groups are electrically neutral relative to the analogous change in the model compound, $\Delta \Delta G_{ij}^{\text{int}}$ is the interaction energy between ionizable groups i and j in their ionized (charged) state. Zero energy state is defined with all ionizable groups neutral. γ_i is equal to -1 for acidic and $+1$ for basic residues. Terms $\Delta \Delta G_{ii}$ and $\Delta \Delta G_{ij}^{\text{int}}$ can be evaluated using the PB method as described elsewhere.^{15,16}

It was shown before¹³ that it is possible to accurately compute the electrostatic free energy of a molecule with M ionizable groups without including all 2^M energy levels. Moreover, to determine the electrostatic free energy with accuracy better than 0.01 kcal/mol, only the energy levels from the lowest one (i. e., the most probable) G_1 up to $RT \ln(2) M$ (L states) above has to be known:

$$G_{\text{el}}^{\text{o}} = G_1 - RT \sum_{m=1}^L \exp\left(-\frac{G_m - G_1}{RT}\right) \quad (9)$$

Computing G_{el}^{o} at a given pH, separately for HEL, HyHEL-5 and their transient-complex configurations, together with $\Delta G_{\text{el}}^{\text{nn}}$

allows one to determine pH-dependent electrostatic interaction energy defined by eq 5.

For comparison, we also considered a “traditional” or standard approach in which one calculates E_{el} for each configuration of the transient-complex ensemble assigning protonation states consistent with the lowest energy state determined for selected pH and for the X-ray structure of the complex,

$$E_{\text{el}}^{\text{trad}} = G_{\text{el},1,\text{X}}^{\text{o,C}} - (G_{\text{el},1,\text{X}}^{\text{o,A}} + G_{\text{el},1,\text{X}}^{\text{o,B}}) + \Delta G_{\text{el}}^{\text{nn}} \quad (10)$$

where indices 1 and X denote that only the lowest protonation state of the X-ray native complex is considered.

Technical Details of Computations. HyHEL-5/HEL Complex. The coordinates of the HyHEL-5/HEL complex (Figure 1) determined at 1.7 Å resolution were taken from the Protein Data Bank (accession code 1YQV¹⁷). To obtain fully protonated molecules, hydrogen atoms were added, and their positions were optimized using the GROMACS package.¹⁸ Partial charges and radii were assigned to atoms according to the OPLS-AA force field.¹⁹

Transient-Complex Ensemble. Total number of 3.8×10^7 random configurations of HyHEL-5 and HEL was generated. All these configurations were tested for clashes, and 417 894 nonoverlapping ones were accumulated. By analyzing contacts in these nonoverlapping configurations, the transient-complex ensemble was characterized as described above.

Calculations of the Electrostatic Interaction Energy. Terms $\Delta\Delta G_{ii}$ and $\Delta\Delta G_{ij}^{\text{int}}$ (eq 8) for HEL, HyHEL-5, and each of their transient-complex configurations were evaluated using MEAD (macroscopic electrostatics with atomic detail).¹⁵ For each configuration, the lowest energy protonation states of proteins at a given pH and temperature were determined using the modified Monte Carlo software of Antosiewicz.²⁰ To evaluate G_{el} 's, up to 10^6 lowest-energy states were registered. The net charges of molecules as a function of pH were calculated with the HYBRID²¹ software. Calculations were performed for isolated HEL and HyHEL-5 and for their complex in a single configuration as in the X-ray native structure. MEAD-calculated terms $\Delta\Delta G_{ii}$ and $\Delta\Delta G_{ij}^{\text{int}}$ were used as an input for HYBRID. The net charges were computed as averages over ensembles of protonation states evaluated at given pH.

$\Delta G_{\text{el}}^{\text{nn}}$ and $E_{\text{el}}^{\text{trad}}$ for each of the transient-complex configurations were computed with APBS²² by solving the PB equation on the 225 Å × 225 Å × 289 Å grid with final spacing of 0.48 Å. All electrostatic computations were performed at 150 mM ionic strength and temperature of 298.15 K. Dielectric constants were set to 4 for proteins and 78.54 for the solvent (water). The solute–solvent dielectric boundary was defined as the van der Waals surface. It was previously shown that for oppositely charged molecules, the PB calculations with the van der Waals surface definition of a dielectric boundary give negative electrostatic interaction energies, but with the solvent accessible surface definition, the electrostatic interaction energies can be positive.²³

Results

Localization of the Transient-Complex Ensemble. The transient complex is located at the outer boundary of the bound state. As described in the Methods section, a unique definition of the transient-complex ensemble is possible using σ_{α} (the standard deviation of the rotation angle α) that measures the degree of rotational freedom in the set of generated nonoverlapping configurations. Figure 2 plots the number of contacts

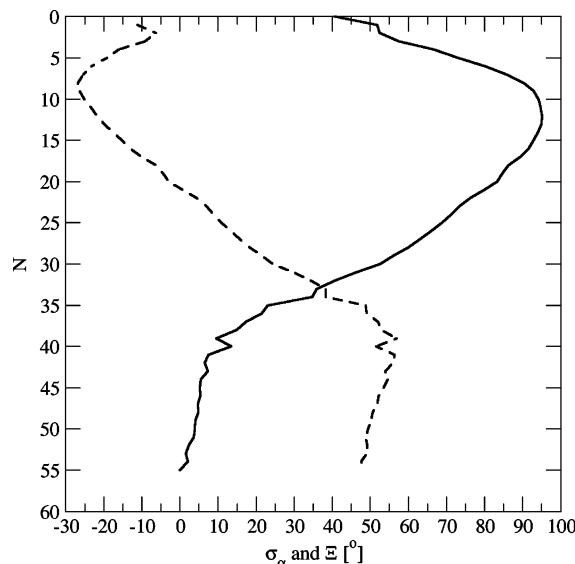


Figure 2. The standard deviation of α , σ_{α} (continuous line), as a function of the number of contacts (N) determined for nonoverlapping configurations of HyHEL-5/HEL. The bound and unbound states are characterized by a large and small number of contacts, respectively. Transition from the bound state to the unbound one corresponds to a sharp increase in σ_{α} . The transient-complex ensemble (and its contact number) is located on the basis of the position of the maximum of the Ξ function (dashed line), as described elsewhere.⁸

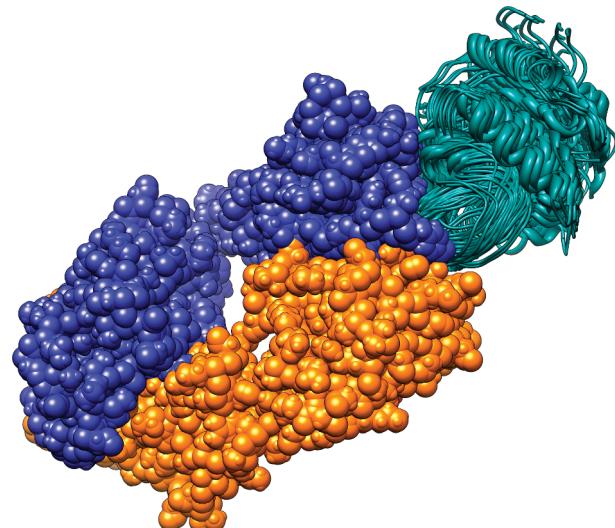


Figure 3. Representative configurations in the transient complex of HyHEL-5 (shown as van der Waals spheres) and HEL (cartoon models). Drawing was done using CHIMERA.¹¹

(N) registered in 417 894 configurations of HyHEL-5/HEL that passed the clash test versus σ_{α} computed for a given contact level. The Ξ function defined in eq 4, also presented in Figure 2, shows a maximum for 39 contacts.

The number of configurations from the nonoverlapping set with the contact level of 39 was 351; they were classified as the transient-complex ensemble. Representative examples of the transient-complex configurations are shown in Figure 3.

Scatter plots with the number of contacts as a function of relative translation ρ and rotation angle α prepared based on all random HyHEL-5/HEL configurations that passed the clash test are presented in Figure 4. For configurations characterized by 39 contacts, the relative translation ρ ranges from 3.7 to 6.6 Å with an average of 4.6 Å. The standard deviation of ρ observed in the transient-complex configurations is 0.5 Å. The

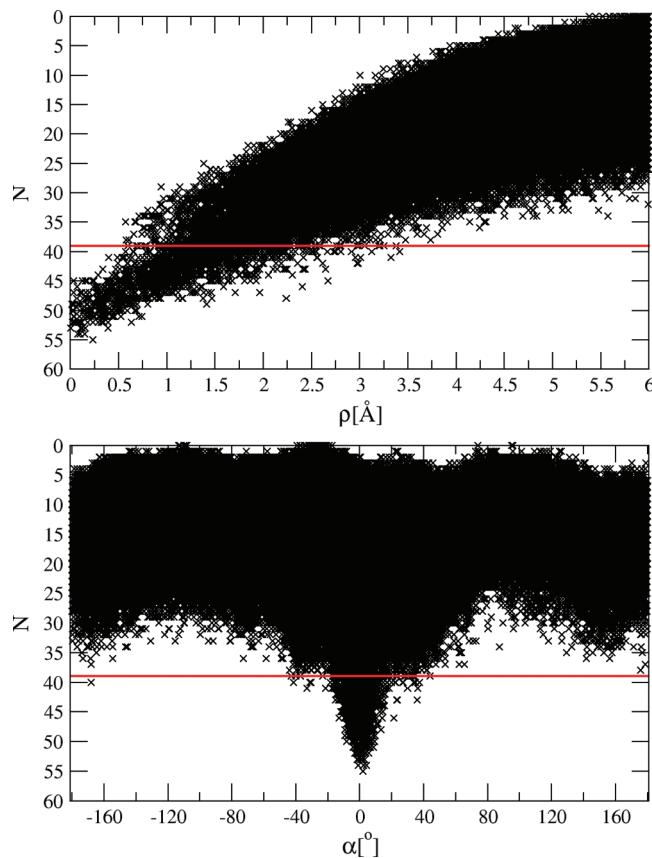


Figure 4. Scatter plot of the contact level N versus relative separation ρ (top graph, with ρ_0 , the separation measured in the X-ray complex, set to 0) and versus the rotation angle, α (bottom graph), for the HyHEL-5/HEL system. The number of points in each plot is 417 894. The transient-complex ensemble with $N = 39$ is marked with the red line.

rotation angle, α , characterizing the transient-complex ensemble falls between -45° and 45° ; however, we also observe weakly populated configurations with $\alpha \sim -180^\circ$ and $\alpha \sim 180^\circ$. The $N(\alpha)$ plot is asymmetric mostly around $\alpha \sim -90^\circ$ and $\alpha \sim 90^\circ$. Additionally, configurations with α around $\pm 90^\circ$ are

characterized by a smaller number of contacts. The average α computed over the transient-complex configurations is -1.2° (with a standard deviation of 9.6°), indicating that positive and negative α values are nearly equally sampled.

Net Charges of HEL and HyHEL-5 versus pH. The total number of ionizable groups is 32 in HEL and 101 in HyHEL-5. Both our and previous calculations of residual pK_a 's in HEL, HyHEL-5, and the HyHEL-5/HEL complex²⁴ show that pK_a 's of several residues located at the binding interface, including those involved in the Arg-Glu salt links, change significantly upon association. In agreement with ref 24, we find that most acidic residues titrate in the pH range below 5.0, whereas most basic residues titrate in the pH range above 9.0.

For pH between 5.0 and 10.0, we computed the net charges of molecules as an average over the ensemble of protonation states. The net charge of isolated HEL varies between $+10e$ and $+3e$; that of isolated HyHEL-5, between $+10e$ and $-15e$; and of HyHEL-5/HEL complex (as in the X-ray structure), between $+18e$ and $-13e$ (Figure 5). The resulting isoelectric points, pI , (pH of zero net charge) are ~ 11 for HEL (11.2 determined experimentally²⁵), ~ 7 for HyHEL-5 (consistent with a pI of mouse immunoglobulines usually between 6.0 and 7.5^{26,27}), and ~ 9 for their complex.

At pH above 7.0, HEL and HyHEL-5 are oppositely charged. Experiments show that the maximal binding rate is observed for pH ~ 8.0 .²⁴ At pH below 7, both molecules are positively charged, but binding is still observed because local interactions between positively and negatively charged regions of both molecules and screening of interactions due to the nonzero ionic strength overcome the overall repulsion resulting from their positive net charges.

Electrostatic Interaction Energy. According to eq 9, the electrostatic free energy of a protein with multiple ionizable groups can be described as the sum of its free energy in the lowest protonation state and contributions from all other protonation states accessible at a given temperature. The latter contributions decrease the electrostatic free energy below its value in the lowest state, G_1 (see eq 9). If the number of easily accessible states is significant, the drop in the free energy below the lowest protonation state becomes substantial.

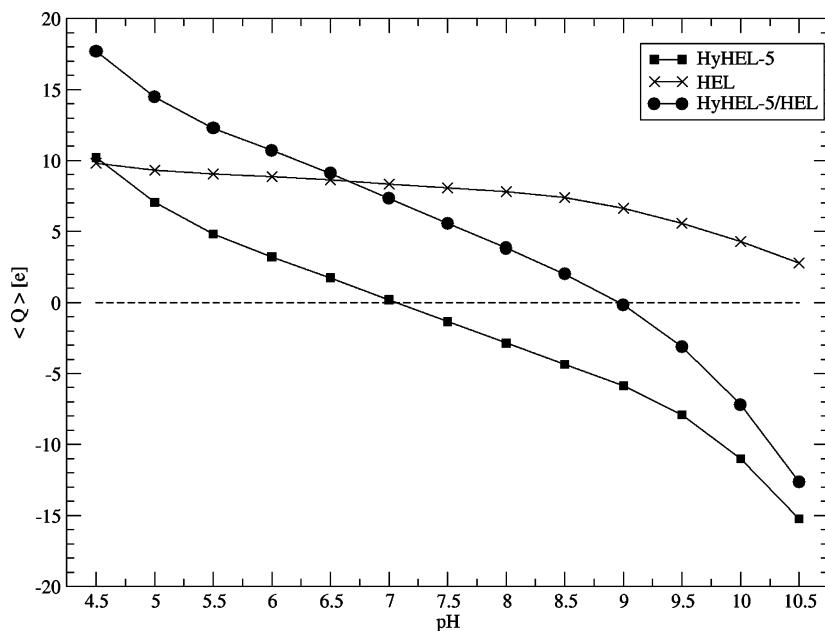


Figure 5. Average net charges of HEL, HyHEL-5, and their complex (X-ray structure) computed as a function of solution pH.

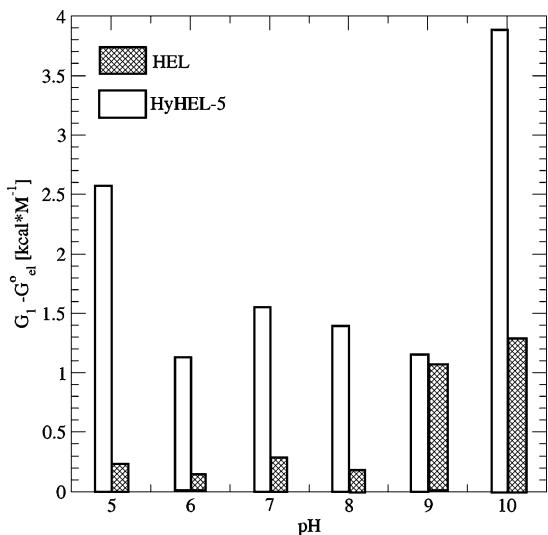


Figure 6. The difference between the energy of the lowest (i.e., the most probable) protonation state, G_1 , and the electrostatic free energy, G_{el}^o , computed for HEL and HyHEL-5 as a function of pH.

Figure 6 presents electrostatic free energy differences between the lowest protonation state G_1 and G_{el}^o of eq 9 computed for HEL and HyHEL-5 in the pH range 5–10. In general, the free energy differences observed for HEL are smaller than for HyHEL-5, which can be explained by the smaller number of ionizable groups in HEL (32) than in HyHEL-5 (101). Large electrostatic free energy differences are found at low (2.6 kcal/mol at pH 5) and high pH (3.9 kcal/mol at pH 10) for the HyHEL-5 system and at high pH for HEL (1.3 kcal/mol at pH 10). Acidic ionizable groups titrate below pH 5, and basic groups start releasing protons at pH above 9. As a consequence, the number of thermally accessible protonation states is raised at low and high pH. The number of accessible protonation states in the pH between 6 and 8 is lower because protonation/deprotonation of acidic/basic groups is less probable. This accounts for the smaller decrease in the electrostatic free energy. For example, the number of states (L in eq 9) needed to evaluate the electrostatic free energy of HEL with accuracy better than 0.01 kcal/mol is 258, 142, and 1157 for pH 5, 7, and 10, respectively. For HyHEL-5, the same accuracy is obtained with 68 846, 21 110 and 185 897 states, respectively. Upon association, the number of states accessible for both proteins and their distributions are likely to change. The number of states included in calculations of electrostatic free energies of transient-complex configurations varies (depending on a particular configuration and pH) from about 28 000 (at moderate pH) to almost 500 000 (at pH 10).

For transient-complex configurations, we present the distributions of electrostatic interaction energies, E_{el} (eq 5) and E_{el}^{trad} (eq 10) at different pH (see Figure 7). In general, the E_{el} distributions are wider and their centers are more shifted upon changing pH than the distributions of E_{el}^{trad} . The E_{el} and E_{el}^{trad} averages computed over the transient-complex ensemble and their corresponding Boltzmann factors differ while varying the pH (Figure 8). As shown in Figure 8, the E_{el}^{trad} averages are lower than E_{el} . Binding is less favorable when the protonation degrees of freedom and pH are explicitly considered for evaluating the interaction energies. Largest differences between average E_{el} and E_{el}^{trad} are obtained at extreme pH (5 and 10). At moderate pH 8, acidic and basic residues are not likely to change their protonation states. Probabilities of proton uptake or release by acids or bases are low, both in isolated proteins and in their

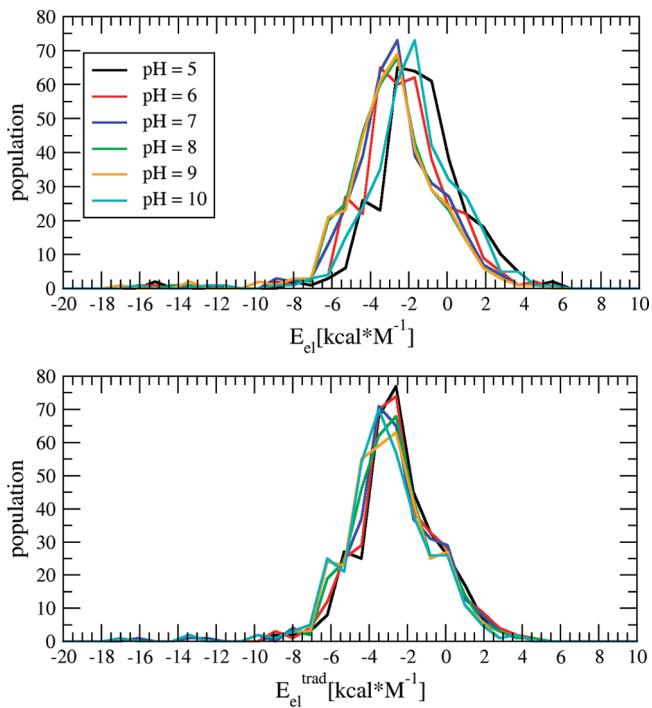


Figure 7. Electrostatic interaction energy distributions computed for transient-complex configurations at various pHs. Top plot, including the protonation degrees of freedom (E_{el}); bottom plot, after assigning a single protonation state (i.e., the most probable one at a given pH for the X-ray complex) to all transient-complex configurations (E_{el}^{trad}).

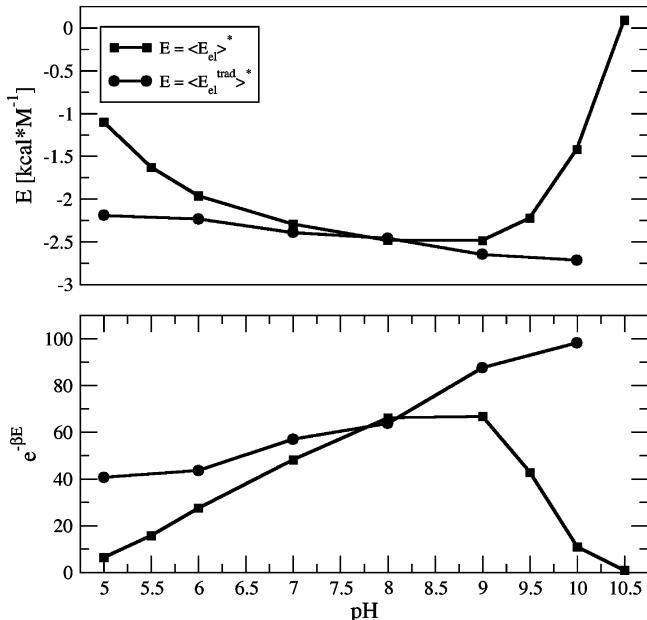


Figure 8. Averages of E_{el} and E_{el}^{trad} (top) and their Boltzmann factors (bottom) computed over the ensemble of transient-complex configurations at different pHs.

complexes; therefore, we observe small differences between average E_{el}^{trad} and E_{el} . It should be noted that in the pH range 5.0–10.0, both methods show favorable binding. However, whereas in the traditional approach, binding becomes more favorable upon increasing pH, our method gives the most negative E_{el} for pH 9.0. Additionally, when pH is raised to 10.5, E_{el} becomes positive, interactions in transient-complex ensemble become repulsive, and the association is retarded.

Rate of Association versus pH. According to experiments,¹² the association rate constant for the interaction of HyHEL-5

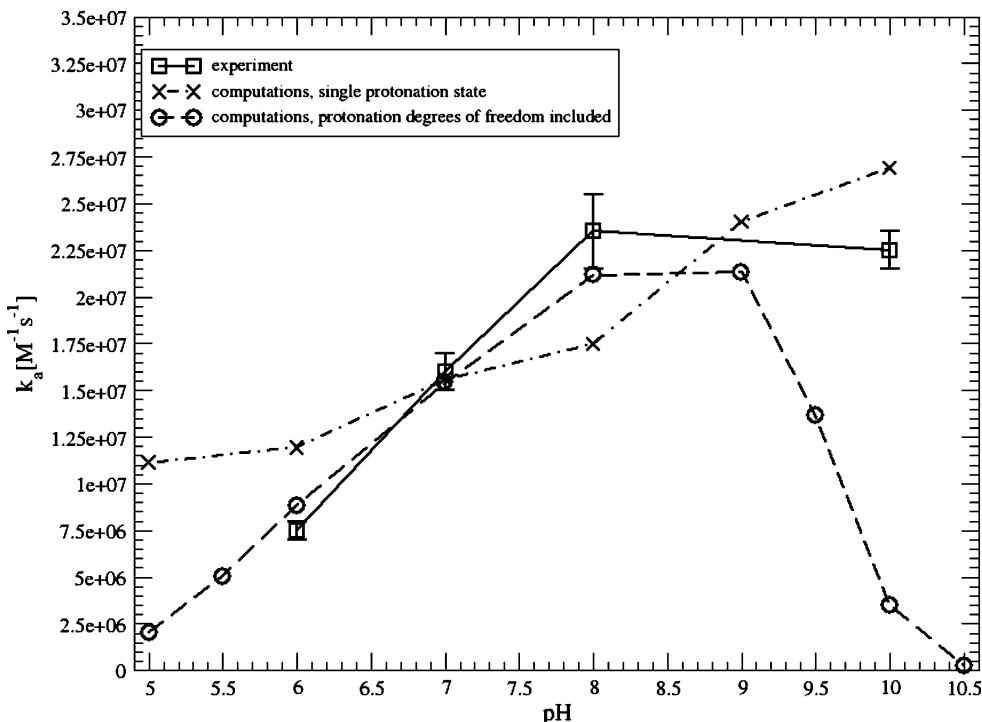


Figure 9. The association rate of HEL with HyHEL-5 computed as a function of pH. Experimental values are taken from ref 12; measurements were conducted at 150 mM ionic strength and a temperature of 298.15 K.

with HEL varies with pH in the range 6.0–10.0 (see Figure 9). The experimentally measured k_a linearly increases with pH with a 3-fold rise in magnitude in the pH range between 6.0 and 8.0. Above pH 8.0, k_a decreases, but experiments were performed only at pH 10, so it is possible that the measured association rate exhibits a maximum between pH 8 and 10.

To compare our calculations with experimental data, the basal rate for association, k_{a0} (eq 1), that is, the association rate in the absence of electrostatic interactions, needs to be determined. Although it is possible to estimate the basal rate theoretically by simulating the transitional and rotational movement of proteins in the absence of electrostatic interactions (as described elsewhere^{7–9}), we treated k_{a0} as an adjustable parameter. We determined its value by finding a multiplicative factor that gives the best fit of the computed Boltzmann factors $e^{-\beta \langle E_{el} \rangle^*}$ and $e^{-\beta \langle E_{el}^{trad} \rangle^*}$ (Figure 8) to the experimental curve in the pH range 6.0–8.0. The resulting basal rates are $2.74 \times 10^5 \pm 0.80 \times 10^5$ $1/M \cdot s$ (based on fitting of $e^{-\beta \langle E_{el} \rangle^*}$) and $3.19 \times 10^5 \pm 0.35 \times 10^5$ $1/M \cdot s$ (based on fitting of $e^{-\beta \langle E_{el} \rangle^*}$). Both values fall into the 10^5 – 10^6 $1/M \cdot s$ range that is appropriate for stereospecific protein–protein association.^{28,29}

As shown in Figure 9, the traditional approach fails to reproduce the experimentally observed trend in the association kinetics. The slope of the computed curve in the pH region 6.0–8.0 is too small as compared to the slope of the experimental curve. Moreover, the association rate constant computed with the traditional approach increases almost linearly in the whole studied pH range.

A much better result is obtained when protonation degrees of freedom and solution pH are explicitly considered. However, while in the pH range 6.0–9.0, the fit to experimental data is almost perfect, and the drop in the association rate observed above pH 9.0 is significantly larger than the experimental one. This may be due to a number of reasons. The presented approach neglects intraprotein conformational degrees of freedom. Although protonation equilibria are treated explicitly and rigor-

ously, the fact that proteins are frozen in their conformations found in the X-ray structure of the complex affects the results. Protein conformational freedom can be important, particularly at low and high pH because titration of acidic or basic residues might be accompanied by local changes in the protein structure. Moreover, even some local changes in protein conformation (e.g., the side-chain reorientation) affect the computed energies of different protonation states.

Other limitations arise from the PB method. The dielectric model and the choice of solute dielectric constant may influence the computed shifts in the pK_a 's of groups in the protein relative to their pK_a 's in the corresponding model compound.²⁴ Nevertheless, different dielectric models and dielectric constants are usually justified as a way to indirectly consider the internal flexibility of molecules.

Conclusions

We presented a computational approach to study kinetics of protein–protein association that explicitly takes into account proteins' protonation equilibria and solution pH. We combined the transient-complex theory for protein association^{8,9,29} with explicit evaluation of electrostatic free energies of molecules with multiple protonation equilibria.¹³ The method was applied to determine the association kinetics of the HyHEL-5/HEL complex as a function of pH. Our results corroborate reasonably well with experimental data.¹² The standard approach that we also tested fails to reproduce the pH-dependent trend in the association kinetics. Whereas the standard approach gives almost monotonically increasing dependence of the association rate in the studied pH range, a maximum at pH ~ 8 is observed in experiment.

The presented methodology is not free of limitations. Associating proteins are fixed in their crystal structure configurations, and even local changes in protein conformation can affect the computed interaction energies. However, other methods

widely used to study protein–protein association, such as Brownian dynamics or PARE (predicting association rate enhancement),³⁰ also suffer from this limitation. Moreover, these methods employ a less rigorous treatment of electrostatic interactions (as compared to the PB method); that is, test charges or effective charge models and the Debye–Hückel approximation^{30,31} that may lead to additional uncertainties in predictions of association rate constants.

Because proteins are represented with single structures, the coupling between protonation and conformational degrees of freedom is neglected. It is possible to include conformational flexibility during calculation of pH-dependent protein properties using either continuum electrostatic models (for example, the MCCE method of Alexov and Gunner³²) or empirical approaches (the PROPKA method developed in the Jensen group³³). However, to incorporate conformational degrees of freedom in the presented methodology one would need to redefine the formalism used to compute the pH-dependent energy of interaction. Such modifications are underway.

We use the PB model to evaluate the electrostatic energies of protonation states. This model often overestimates the shifts in the pK_a 's of ionizable groups in the protein relative to their pK_a 's of the model compound. The results of the PB model may also depend on the force field used to assign atomic partial charges. Although different methods exist to predict pK_a 's and some of them outperform the PB approach,³⁴ this is the only method that allows for rigorous evaluation of standard electrostatic free energies of molecules with multiple ionization sites.

The method used for calculation of free energy levels associated with different protonation states is based on the linear form of the PB equation. The electrostatic potential in the linear PB equation is a superposition of electrostatic potentials of each partial charge of the molecule. This property allows one to precalculate the interaction matrix,¹⁶ $\Delta\Delta G_{ij}^{\text{int}}$ (of the order equal to the number of ionizable groups in the molecule) that is later used to evaluate free energy levels according to eq 8. Whereas the fact that the linear PB equation is used might prevent applications of our method to highly charged systems, the usage of nonlinear PB equation requires performing PB computations on-the-fly with Monte Carlo generation of energy levels; such a task is computationally prohibitive or even impossible.

Notwithstanding its limitations, our approach can predict the association kinetics; its sensitivity to solution pH; and help to analyze, understand, and rationalize experimental observations. By comparing the presented method with the traditional one, we find that the latter approach is reliable in the pH region where protonation/deprotonation events have low probability (neutral pH in the case of HyHEL-5 and HEL). However, in pH regions where protonation/deprotonation is more likely to occur (acidic and basic pH ranges in the studied case), our approach provides a more physically sound description and leads to a quantitatively better agreement with the experiment. The next step would be to check predictive capabilities of our approach.

Acknowledgment. M.D. and J.T. acknowledge support from the Foundation for Polish Science, University of Warsaw (115/30/E-343/BST1345/ICM/2008 and G31-4), Polish Ministry of Science and Higher Education (N N301 245236), and Fogarty International Center (National Institutes of Health Research Grant no. R03 TW07318). J.A. is supported by the University of Warsaw (BST). M.D. thanks Huang-Xiang Zhou and Sanbo Qin for helpful discussions.

References and Notes

- (1) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. *J. Biol. Chem.* **2009**, *284* (20), 13285–13285.
- (2) Alexov, E. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 572–584.
- (3) Mason, A.; Jensen, J. *Proteins* **2008**, *71*, 81–91.
- (4) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
- (5) Ermakova, E. *Biophys. Chem.* **2007**, *130*, 26–31.
- (6) Ermakova, E. *J. Mol. Model.* **2005**, *12*, 34–41.
- (7) Vijayakumar, M.; Wong, K.-Y.; Schreiber, G.; Fersht, A.; Szabo, A.; Zhou, H.-X. *J. Mol. Biol.* **1998**, *278*, 1015–1024.
- (8) Alsallaq, R.; Zhou, H.-X. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 320–335.
- (9) Qin, S.; Zhou, H.-X. *J. Phys. Chem. B* **2008**, *112*, 5955–5960.
- (10) Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recognit.* **2002**, *15* (6), 377–392.
- (11) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (12) Xavier, K. A.; Willson, R. *Biophys. J.* **1998**, *74*, 2036–2045.
- (13) Pilat, Z.; Antosiewicz, J. *J. Phys. Chem. B* **2008**, *112*, 15074–15085.
- (14) Yang, A.-S.; Gunner, M.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1993**, *15* (3), 252–265.
- (15) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
- (16) Antosiewicz, J.; Briggs, J.; Elcock, A.; Gilson, M.; McCammon, J. *J. Comput. Chem.* **1996**, *17*, 1633–1644.
- (17) Cohen, G.; Silverton, E.; Padlan, E.; Dyda, F.; Wibbenmeyer, J.; Willson, R.; Davies, D. *Acta Crystallogr., Sect. D* **2005**, *61*, 628–633.
- (18) Hess, A.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (19) Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (20) Antosiewicz, J. *Biophys. J.* **1995**, *69* (4), 1344–1354.
- (21) Gilson, M. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 266–282.
- (22) Baker, N.; Sept, D.; Joseph, S.; Holst, M.; McCammon, J. *Proc. Natl. Sci. U.S.A.* **2001**, *98*, 10037–10041.
- (23) Dong, F.; Zhou, H.-X. *Proteins* **2006**, *65*, 87–102.
- (24) McDonald, M.; Willson, R.; McCammon, J. *Protein Eng.* **1995**, *8* (9), 915–924.
- (25) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192–2198.
- (26) Hamilton, R. G.; Roeber, M.; Reimer, C. B.; Rodkey, L. S. *Electrophoresis* **1987**, *8*, 127–134.
- (27) Raman, C. S.; Jemmerson, R.; Nall, B.; Allen, M. *Biochemistry* **1992**, *31*, 10370–10379.
- (28) Northrup, S.; Erickson, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3338–3342.
- (29) Zhou, H.-X. *Biophys. J.* **1997**, *73*, 2441–2445.
- (30) Selzer, T.; Schreiber, G. *J. Mol. Biol.* **1999**, *287*, 409–419.
- (31) Gabdoulline, R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868–3878.
- (32) Alexov, G.; Gunner, M. *Biophys. J.* **1997**, *72*, 2075–2093.
- (33) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704–721.
- (34) Davies, M.; Toseland, C.; Moss, D.; Flower, D. *BMC Biochemistry* **2006**, *7*, 18–30.

JP906829Z