THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

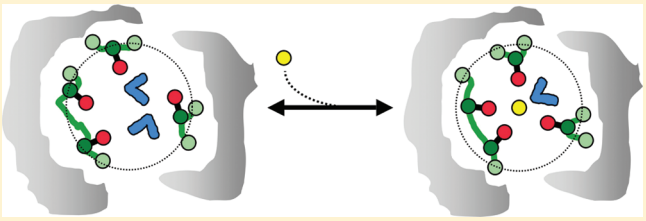
pubs.acs.org/JPCB

Ion Binding Sites and Their Representations by Reduced Models

Benoît Roux*

Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, University of Chicago, Chicago, Illinois 60637, United States

ABSTRACT: The binding of small metal ions to complex macromolecular structures is typically dominated by strong local interactions of the ion with its nearest ligands. Progress in understanding the molecular determinants of ion selectivity can often be achieved by considering simplified reduced models comprised of only the most important ion-coordinating ligands. Although the main ingredients underlying simplified reduced models are intuitively clear, a formal statistical mechanical treatment is nonetheless necessary in order to draw meaningful conclusions about complex macromolecular systems. By construction, reduced models only treat the ion and the nearest coordinating ligands explicitly. The influence of the missing atoms from the protein or the solvent is incorporated indirectly. Quasi-chemical theory offers one example of how to carry out such a separation in the case of ion solvation in bulk liquids, and in several ways, a statistical mechanical formulation of reduced binding site models for macromolecules is expected to follow a similar route. However, there are also important differences when the ion-coordinating moieties are not solvent molecules from a bulk phase but are molecular ligands covalently bonded to a macromolecular structure. Here, a statistical mechanical formulation of reduced binding site models is elaborated to address these issues. The formulation provides a useful framework to construct reduced binding site models, and define the average effect from the surroundings on the ion and the nearest coordinating ligands.



1. INTRODUCTION

Small metal ions are a fundamental component to the structure and function of biological systems. Ion binding is involved in the folding of proteins and nucleic acids, enzyme catalysis, and numerous cellular signaling processes. Although detailed computations have an central role to play in trying to understand these complex systems, simpler models can often help by providing a clearer conceptual view and highlighting the most important factors.[1−14] By virtue of their charge and their small radius, ions interact very strongly with their nearest neighbors. For this reason, complex systems can often be understood from the local physical features of the ion's environment. For example, a simple and appealing explanation of ion selectivity, frequently invoked in the field of "host-guest" chemistry[15,16] and with ion channels and transporters,[17,18] relies primarily on the ligand geometry forming a binding site. In this local view, which relies on a precise ligand geometry at the sub-Ångström level, the binding site provides a cavity of the appropriate size to bind one specific ion but is unable to adapt to an ion of a different size. Alternate mechanisms proposed to explain ion selectivity, also emphasizing local features of a binding site, have focused on the chemical properties and the number of ion-coordinating ligands (for a review, see ref 14).

Progress in understanding the molecular determinants of ion selectivity in macromolecular binding sites may be achieved from theoretical studies based on simplified reduced models comprised of only the nearest ion-coordinating ligands. The concept of a simplified reduced ion binding site model is illustrated schematically in Figure 1. By construction, reduced binding site models only treat the ion and the nearest coordinating ligands explicitly. The influence of the missing atoms from the protein or the solvent is incorporated indirectly. At first sight, the validity of a mechanism explaining ion selectivity from local interactions may seem counterintuitive. After all, a charged ion makes strong electrostatic interactions with its surroundings over very long distances. Nevertheless, electrostatic interactions are not expected to directly affect the relative free energy of ions carrying the same charge. Current views of ion selectivity focusing on the local physical features of the environment owe much to simple physics-based treatments of ion solvation,[19−25] or their extension to describe the pairing of monatomic ions[26−28] and molecular solutes.[29,30] Nevertheless, a formal statistical mechanical treatment is necessary in order to draw meaningful conclusions about complex macromolecular systems from reduced models.

A typical starting point is to physically separate the system into an "inner" and "outer" region. Quasi-chemical theory (QCT) offers one example of how to carry out this type of formal separation in the case of a solute immersed in a bulk liquid.[19,20,25] The general concept goes back to the method of "local configurations" formulated by Bethe, Guggenheim, Fowler, and Kirkwood,[31−35] which was extended by Pratt and co-workers to study solvation.[19,20] In QCT, the system is rigorously partitioned into a small spherical subvolume centered on the solute and a second region corresponding to

Special Issue: B: Harold A. Scheraga Festschrift

Received: January 22, 2012
Revised: April 3, 2012
Published: April 12, 2012

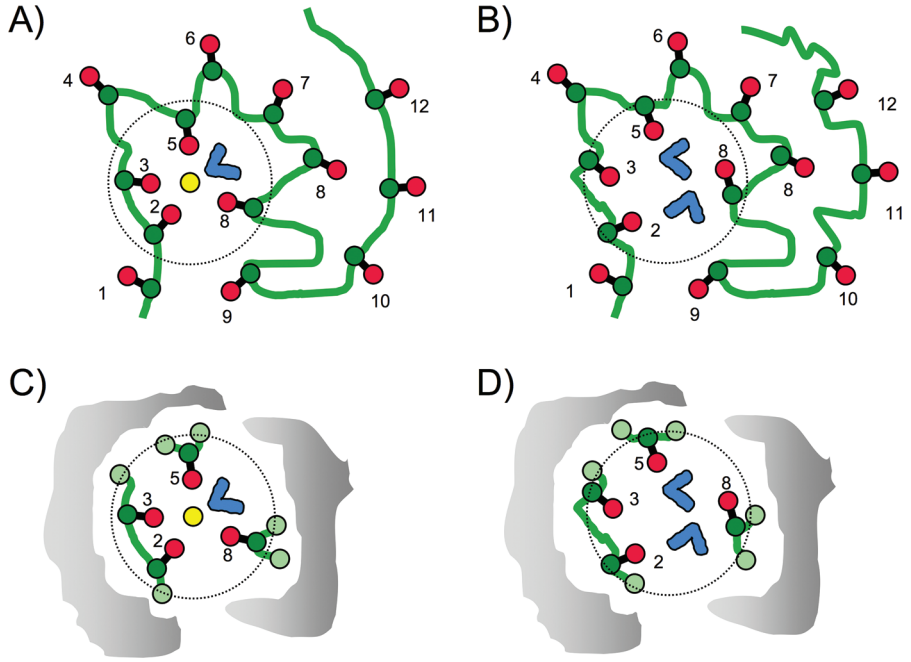


Figure 1. Schematic illustration of configurations of an ion coordinated by n ligands bonded to a macromolecule of N possible ligands with a fixed topological structure. The holo and apo states are respectively shown in A and B for the full length system and in C and D for the reduced model. The ion (yellow circle), ligands (green-red dumbells), and water molecules (blue sticks) are shown. The covalent topological structure of the macromolecule, with its predetermined connectivity, is depicted schematically as a thick green line. In the reduced model, the inner subvolume region is shown as a thin dotted line, and the influence of the surroundings is schematically depicted as a gray area (C and D). In the reduced model, the ligands that are covalently bonded to the rest of the macromolecule must be truncated into chemically reasonable moieties (illustrated by the pale green spheres shown in C and D). The set of ligands {2,3,5,8} was selected to be part of the reduced model. A step-function configuration restriction is applied to prevent any ligands from the set {1,4,6,7,9,10,11,12} from entering the inner subvolume.

the remaining bulk liquid phase.[19,20,25] Most thermodynamic properties can then be expressed as sum over simple terms. For example, the excess chemical potential of the solute can be expressed as $-k_BT\ \ln[P_0\sum_{n\geq 0}K_n\rho^n]$, where the quantities K_n correspond to equilibrium association constants for one solute and n solvent molecules to the inner region, ρ is the density of the unperturbed bulk solvent, and P_0 is the probability of finding zero solvent molecules in the subvolume in the absence of the solute. Although the notation has slightly changed over the years, the standard QCT equation to describe the solvation free energy of a solute immersed in a bulk phase liquid always has the same form; see eq 11 in ref 5, eq 12 in ref 19, eq 7.8 in ref 20, eq 1 in ref 22, eq 3.1 in ref 24, eq 11 in ref 25, eq 1 in ref 36, and eq 38 in the present article. The name quasi-chemical comes from the observation that the underlying equations for all relevant thermodynamic properties have a structure that is similar to that of simple chemical equilibrium between states,[20] and in this regard, the equilibrium association constants K_n associated with the clusters are a critical component of the formulation. This treatment constitutes what we shall refer to as the standard sum-over-clusters bulk phase form of QCT. The latter was shown to be formally equivalent to a treatment of ion solvation based on the small system grand canonical ensemble (SSGCE).[25] In practical applications, the equilibrium association constants are often evaluated by treating the solute and the n solvent molecules as an isolated cluster or fragment in vacuum, within the rigid-rotor harmonic oscillator (RRHO) approximation, using ab initio methods.[21,22,24] The influence of the remaining solvent in the outer region is treated using a computationally inexpensive method such as continuum dielectric.[21,22,24] In the case of ion hydration, direct comparison with the results from free energy MD simulations shows that the approach yields excellent results for small cations such as Li^+ and Na^+, but the accuracy becomes more limited in the case of a larger ion such as K^+.[25]

In several ways, a statistical mechanical formulation of reduced models of ion binding site in macromolecules is expected to follow a route similar to QCT. Indeed, a number of studies of selectivity in ion channels were based directly on the standard form of QCT[4,5,13] or SSGCE.[8] However, there are some critical differences with ion solvation in bulk phase that must be noted. The most important is that the molecular ligands coordinating the ion in a binding site are covalently bonded to the macromolecular structure, whereas solvent molecules from a bulk fluid are identical and free to exchange with one another. Thus, while QCT may offer a useful approach to study ion solvation, modeling binding sites of macromolecules by applying a theoretical formulation especially designed to describe solvation in a bulk fluid raises a number of fundamental issues that need to be addressed.

Here, a general statistical mechanical formulation of reduced binding site models is elaborated to address these issues. The formulation provides a useful framework to construct reduced binding site models and rigorously define the average effect from the surroundings on the ion and the nearest coordinating ligands. The article begins by briefly reviewing the statistical thermodynamics formulation of molecular association and equilibrium binding free energy. The treatment of configurational restraints is discussed. The equilibrium binding free energy of an ion is formally expressed in terms of a reduced subset of ligands. Issues of selectivity are naturally treated by comparing the absolute binding free energies of ions taken from

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

this development. The possibility of a variable number of water molecules in the ion binding site is explicitly accounted for in the formal development. The harmonic approximations based on the assumption of small oscillations and fluctuations is discussed in the context of the reduced model. Finally, we discuss the similarities and differences with QCT.

2. THEORETICAL DEVELOPMENTS

2.1. Statistical Mechanics of Equilibrium Binding. It is useful to briefly recall the statistical mechanical formulation describing the equilibrium binding of a ligand to a macromolecule. It should be noted that ligand normally refers to the small molecule that binds to the larger molecule, whereas in the context of ion binding, "ligands" is generally used to refer to the molecular groups that form a strong coordination complex with the ion. To avoid any confusion, we will refer to the associating entity as the binder (b), and refer to the ion-coordinating groups as the ligands (l) in the following. The equilibrium binding constant can be expressed as[37−41]

$$K_b = \frac{\int_{holo} d\mathbf{R}\, e^{-\beta U(\mathbf{R})}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}') e^{-\beta U(\mathbf{R})}} \quad (1)$$

where $\mathbf{R}$ represents all the coordinates of the ligand, macromolecule, and solvent, $U(\mathbf{R})$ is the total potential energy of the system, $\beta \equiv 1/k_B T$, $\mathbf{r}_b$ is the position of the center-of-mass of the binder, and $\mathbf{r}'$ is some arbitrary position far away in the bulk region.[37,42,40] Equilibrium ion binding is a special case of eq 1 for which the binder is a monatomic species with no internal or orientational degrees of freedom (b → ion). In writing eq 1, it was assumed without loss of generality that the macromolecular receptor cannot freely translate or rotate in solution; this is equivalent to the assumption that the coordinates $\mathbf{R}$ are expressed in terms of a reference frame that is defined with respect to the macromolecule. The equilibrium association constant K_b has dimension of volume (typically expressed in Å^3). The standard binding free energy is defined as $\Delta G_b^{(\circ)} \equiv -k_B T \ln[K_b C^{(\circ)}]$, where $C^{(\circ)}$ is the standard concentration (1 mol/L = 1/ 1660 Å^3). Equation 1 can be generalized to represent the equilibrium binding constant of n identical binders to the macromolecule[37]

$$K_b(n) = \frac{1}{n!} \frac{\int_{holo} d\mathbf{R}\, e^{-\beta U}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_{b_1} - \mathbf{r}'_1)...\delta(\mathbf{r}_{b_n} - \mathbf{r}'_n) e^{-\beta U}} \quad (2)$$

where $\mathbf{r}_{b_i}$ and $\mathbf{r}'_i$ are the position of the center-of-mass of the ith binder and some arbitrary position far away in the bulk region, respectively (the different $\mathbf{r}'_i$ are distant from one another and uncorrelated). This form is particularly useful to treat the association of water molecules to an ion binding site (see below). The equilibrium binding constant eq 1, or it generalized form eq 2, can be re-expressed in a convenient form by introducing an intermediate state in which the center-of-mass of the binder is fixed at some point $\mathbf{r}^g$ in the gas phase

$$K_b = \frac{\int_{holo} d\mathbf{R}\, e^{-\beta U(\mathbf{R})}}{\int d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}^g) e^{-\beta U(\mathbf{R})}} \frac{\int d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}^g) e^{-\beta U(\mathbf{R})}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}') e^{-\beta U(\mathbf{R})}} = \tilde{K}_b e^{\beta G_{bulk}} \quad (3)$$

where G_{bulk} is the free energy to transfer the binder from the gas phase to the bulk solution (i.e., the absolute solvation free energy of the binder), and $\tilde{K}_b$ is the equilibrium binding constant with the binder starting from the gas phase. This defines the binding free energy associated with the site, $\Delta G_{site}^{(\circ)} = k_B T \ln[\tilde{K}_b C^{(\circ)}]$. An equivalent expression can be obtained by choosing an intermediate state described by the potential energy function U_0, in which the binder is artificially decoupled from its surroundings

$$K_b = \frac{\int_{holo} d\mathbf{R}\, e^{-\beta U}}{\int d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}'') e^{-\beta U_0}} \frac{\int d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}'') e^{-\beta U_0}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}') e^{-\beta U}} = \tilde{K}_b e^{\beta G_{bulk}} \quad (4)$$

where $\mathbf{r}'$ and $\mathbf{r}''$ are arbitrary positions in the bulk and gas phase, respectively.

2.2. Configurational Restrictions. In the above expressions, the subscript "holo" implies that the multidimensional integral should include only to those configurations where the binder is considered to be associated to the macromolecule, and exclude all other configurations. From a formal point of view, this means that a configurational restriction is implicitly included in the argument of the multidimensional integrals

$$\int_{holo} d\mathbf{R}... \equiv \int d\mathbf{R}\, H_{holo}(\mathbf{R})... \quad (5)$$

where $H_{holo}(\mathbf{R})$ is a multidimensional step-function equal to 1 when the configuration fulfills the criteria, and 0 otherwise. Formally, the equilibrium binding constant defined by eq 1 should be insensitive to the precise definition of the step-function H_{holo}. This implies that the latter must not be overly restrictive in order to capture all the possible binding modes of the binder. On the other hand, if one is interested in focusing on a narrowly defined subset of configurations, then it may be necessary to impose additional configurational restriction; see eq 6 below. The subscript "apo" in the denominator means that the integral is restricted to exclude all contributions from configurations with the binder in complex with the macromolecule. This is partly redundant because the delta function $\delta(\mathbf{r}_b - \mathbf{r}')$ is already imposing this condition. Binding free energy computations often consider the situation at infinite dilution, in which there is only one macromolecular receptor and one binder. However, if there is a finite concentration of binders in the system then the configurational restriction H_{apo} is needed to to exclude all other binders from the binding site. This becomes important, for example, when one is considering the association of exactly n water molecules into the binding site of the macromolecule. To formally define such a state, it is necessary to prevent additional water molecules from entering the binding site.[37] In the following, the explicit notation will be used whenever it is needed to avoid confusion.

It is often useful to introduce additional configurational restrictions in eqs 1 and 2 to breakdown the equilibrium binding constants into state-specific quantities. For the sake of simplicity, this will be illustrated with the case of a single binder, although similar expressions could be developed for the case of multiple binders. Let us assume that the step-function $H^{(s)}(\mathbf{R})$ is equal to 1 when the system is in a state s, and 0 otherwise. The binding constant of a binder can then be expressed as

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

$$K_b = \frac{P^{(s)}_{apo}}{P^{(s)}_{holo}} K^{(s)}_b \tag{6}$$

where $P^{(s)}_{holo} = \langle H^{(s)} \rangle_{holo}$ and $P^{(s)}_{apo} = \langle H^{(s)} \rangle_{apo}$ represent the probability to find the system in the state s when the complex is formed (holo) or not formed (apo), respectively, and $K^{(s)}_b$ is the binding constant under the constraint that the system must be restricted to configurations corresponding to the state s

$$K^{(s)}_b = \frac{\int_{holo} d\mathbf{R}\, H^{(s)}(\mathbf{R})\, e^{-\beta U}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}') H^{(s)}(\mathbf{R})\, e^{-\beta U}} \tag{7}$$

The formulation based on eqs 6 and 7 is general, and can be used for a wide range of situations. If a complete set of states s can be defined such that $\sum_s H^{(s)} = 1$, then the binding constant can also be expressed as

$$K_b = \sum_s P^{(s)}_{apo} K^{(s)}_b \tag{8}$$

A related strategy consists in introducing some restraining potential u_c to bias the conformation of the receptor or the binders; typically the restraining potential is a function of a subset of all coordinates $\mathbf{R}$. The equilibrium binding free energy can then be rewritten as

$$K_b = e^{-\beta[G^{holo}_c - G^{apo}_c]} K^{(u_c)}_b \tag{9}$$

where the equilibrium binding constant under the presence of the restraining potential is

$$K^{(u_c)}_b = \frac{\int_{holo} d\mathbf{R}\, e^{-\beta[u_c + U]}}{\int_{apo} d\mathbf{R}\, \delta(\mathbf{r}_b - \mathbf{r}')\, e^{-\beta[u_c + U]}} \tag{10}$$

and $G^{apo}_c = -k_B T \ln[\langle e^{-\beta u_c} \rangle_{apo}]$ and $G^{holo}_c = -k_B T \ln[\langle e^{-\beta u_c} \rangle_{holo}]$ represent the free energies for introducing the configurational restriction in the apo and holo state, respectively. In previous applications, such conformational free energy contributions have been calculated by using biased umbrella sampling simulations along some preidentified order parameter. A convenient choice has often been a quadratic restraining potential based on $\chi[\mathbf{X}, \mathbf{X}_{ref}]$, the root-mean-square deviations (rmsd) of the configuration for a subset of atoms, $\mathbf{X}$, relative to a reference configuration, $\mathbf{X}_{ref}$, $k(\chi[\mathbf{X}; \mathbf{X}_{ref}])^2$. The reference configuration $\mathbf{X}_{ref}$ of the subset of atoms can be chosen in several ways, and can be binder-specific. For example, it could be the average configuration, the energy-minimized configuration, or the most probable configuration.

2.3. Ion Selectivity. Equilibrium ion selectivity arises naturally from the definition of the binding constant eq 1. Assuming that the binding constants for ion type i and j are K^i_{ion} and K^j_{ion}, respectively, the free energy difference that controls ion selectivity is

$$\Delta\Delta G^{ij} = -k_B T \ln[K^i_{ion}/K^j_{ion}] \tag{11}$$

Using the decoupling of eqs 3 or 4, this can be rewritten as

$$\begin{aligned} \Delta\Delta G^{ij} &= -k_B T \ln[\tilde{K}^i_{ion}/\tilde{K}^j_{ion}] - [G^i_{bulk} - G^j_{bulk}] \\ &= \Delta G^{ij}_{site} - \Delta G^{ij}_{bulk} \end{aligned} \tag{12}$$

where the first and second terms correspond to the free energy difference between ion i and ion j in the binding site and in the bulk solvent, respectively. The difference in solvation free energy ΔG^{ij}_{bulk} is an offset constant and all the information about the selectivity of a given binding site is included in ΔG^{ij}_{site}.

2.4. Water Molecules in the Ion Binding Site. It may be advantageous to explicitly keep track of the water molecules coordinating the ion in the binding site. Such a treatment is particularly important if the number of water molecules differs for the bound (holo) and unbound (apo) state. From a formal point of view, this is a special case of the configurational restriction that was discussed above and which led to eq 6. Let us define the function $n'(\mathbf{R})$, equal to the total number of water molecules with their center-of-mass $\mathbf{r}_{w_i}$ within a distance R away from the center of the binding site. The configurational restriction can be constructed from the discrete Kroenecker delta function $\delta_{n,n'(\mathbf{R})}$, which equal to 1 for all configurations comprising exactly n water molecules in the binding site, and is equal to 0 otherwise. The equilibrium binding constant can be rewritten as

$$K_{ion} = \frac{P_{apo}(n)}{P_{holo}(n)} K_{ion}(n) \tag{13}$$

where $P_{apo}(n) = \langle \delta_{n,n'(\mathbf{R})} \rangle_{apo}$ and $P_{holo}(n) = \langle \delta_{n,n'(\mathbf{R})} \rangle_{holo}$ are the probability that n water molecules are located in the binding site, in the apo and holo states respectively, and $K_{ion}(n)$ is the equilibrium binding constant of the ion when there are exactly n water molecules in the binding site. Because the probabilities are normalized by construction, $\sum_n \delta_{n,n'} = 1$, the sum over all possible values of n constitute a complete set of states eq 8 can be used to express the ion binding constant. The probability of finding n water molecules in the binding site can, itself, be formulated as an equilibrium binding problem, leading to the expressions[37]

$$P_{apo}(n) = \frac{(\bar{\rho}_w)^n e^{n\beta\Delta\mu_w} \tilde{K}^{apo}_w(n)}{\sum_{m=0} (\bar{\rho}_w)^m e^{m\beta\Delta\mu_w} \tilde{K}^{apo}_w(m)} \tag{14}$$

and

$$P_{holo}(n) = \frac{(\bar{\rho}_w)^n e^{n\beta\Delta\mu_w} \tilde{K}^{holo}_w(n)}{\sum_{m=0} (\bar{\rho}_w)^m e^{m\beta\Delta\mu_w} \tilde{K}^{holo}_w(m)} \tag{15}$$

where $\bar{\rho}_w$ is the bulk density of water, $\Delta\mu_w$ is the excess chemical potential of a water molecule in the bulk phase, and $\tilde{K}^{holo}_w(n)$ and $\tilde{K}^{apo}_w(n)$ are the equilibrium association constants of water molecules from the gas phase to the holo (ion-bound) and apo (ion-free) states, respectively. Following eqs 4 and 2, the association of n water to the ion binding site in the macromolecule to the holo state can be written as

$$\tilde{K}^{(holo)}_w(n) = \frac{1}{n!} \frac{\int_{holo} d\mathbf{R}\, H^{(in)}_{\{w_1,\ldots,w_n\}} H^{(out)}_{\{w_{n+1},\ldots\}}\, e^{-\beta U}}{\int_{holo} d\mathbf{R}\, \delta(\mathbf{r}_{w_1} - \mathbf{r}^g_1)\cdots\delta(\mathbf{r}_{w_n} - \mathbf{r}^g_n) H^{(out)}_{\{w_{n+1},\ldots\}}\, e^{-\beta U}} \tag{16}$$

where $H^{(in)}_{\{w_1,\ldots,w_n\}}$ and $H^{(out)}_{\{w_{n+1},\ldots\}}$ are step-functions equal to 1 only when the stated condition is realized, and 0 otherwise. A similar expression can be written for the apo state. In the numerator, the water molecules 1 to n are restricted to the binding site region and all other water molecules are restricted to be in the bulk; in the denominator, all water molecules are restricted to the bulk. For the special case of $n = 0$, the equilibrium constant of the ion to the binding site is

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

$$K_{\text{ion}} = K_{\text{ion}}(0)\frac{\sum_{m=0}(\bar{\rho}_{\text{w}})^m e^{m\beta\Delta\mu_{\text{w}}}\tilde{K}_{\text{w}}^{\text{holo}}(m)}{\sum_{m=0}(\bar{\rho}_{\text{w}})^m e^{m\beta\Delta\mu_{\text{w}}}\tilde{K}_{\text{w}}^{\text{apo}}(m)} \quad (17)$$

Using eqs 14 and 15, this expression can be rewritten as

$$K_{\text{ion}} = \tilde{K}_{\text{ion}}(0)e^{-\beta G_{\text{bulk}}}P_{\text{apo}}(0)\sum_{m=0}(\bar{\rho}_{\text{w}})^m e^{m\beta\Delta\mu_{\text{w}}}\tilde{K}_{\text{w}}^{\text{holo}}(m) \quad (18)$$

where

$$P_{\text{apo}}(0) = \frac{1}{\sum_{m=0}(\bar{\rho}_{\text{w}})^m e^{m\beta\Delta\mu_{\text{w}}}\tilde{K}_{\text{w}}^{\text{apo}}(m)} \quad (19)$$

is the probability that there are zero water molecules occupying the binding site when the ion is not present, and $\tilde{K}_{\text{ion}}(0)$ is the binding constant of the ion from vacuum to a site that contains zero water molecules. Equation 18 bears a striking similarity with the expression from quasi-chemical theory (QCT) for the solvation free energy of an ion in bulk solvent (see ref 25 for a similar notation). We return to this subject in section 2.7.

2.5. Reduced Model with Nearest Ion-Coordinating Groups. The ion-coordinating ligands in the binding site are a critical component of the inner region of any reduced model. A necessary first step for a formulation of a reduced model is a formal separation of these ligands from the rest of the macromolecule. However, as they are part of a covalently linked macromolecule, this procedure will necessarily leave dangling covalent bonds at the boundary of the inner and outer regions. In practice, this problem could be resolved by using one of the schemes used in QM/MM simulations combining quantum mechanical (QM) ab initio and a molecular mechanical (MM) force field representations for different parts of a system.[43−48] To proceed further, one must then introduce a configurational restriction associated with the set of ligands,$\{l, ..., l_n\}$, that are implicated in ion coordination. The implication is that all ligands, other than the chosen set $\{l_1, ..., l_n\}$, must remain outside some predefined inner region. This requirement is enforced by the step-function $H^{(\text{out})}_{\{l_{n+1},...\}}$, equal to 1 if the ligand $\{l_{n+1}, ...\}$ are located in the outer region, and 0 otherwise. Once a configurational restriction has been properly defined, unbiased equilibrium results can be calculated on the basis of eq 6. The probability that this condition is violated is

$$P^{\text{apo}}_{\{l_{n+1},...\}} = \langle H^{(\text{out})}_{\{l_{n+1},...\}}\rangle_{\text{apo}} \quad (20)$$

for the apo state, and

$$P^{\text{holo}}_{\{l_{n+1},...\}} = \langle H^{(\text{out})}_{\{l_{n+1},...\}}\rangle_{\text{holo}} \quad (21)$$

for the holo state. A quantitative criterion for choosing a representative set of ion-coordinating ligands in an optimal way can be formulated from eq 6. If the ratio

$$f_{\text{excl}} = \frac{P^{\text{apo}}_{\{l_{n+1},...\}}}{P^{\text{holo}}_{\{l_{n+1},...\}}} \quad (22)$$

is equal or close to 1, then the set $\{l_1, ..., l_n\}$ is highly representative of the binding site and excluding any of the other ligands from the inner region via the configurational restriction $H^{(\text{out})}$ becomes almost a formality.

Once a particular set of n ligands is chosen, we distinguish between three main components in the total potential energy of the system: the ion (i), the selected set of ion-coordinating ligands (l), and the rest of the system (r). The ion and the coordinating ligands are part of the inner region, while the rest of the system is part of the outer region. Let the coordinates of the ion and the n coordinating ligand be represented by $\mathbf{X} \equiv \{\mathbf{r}_{\text{ion}}, \mathbf{x}_1, ..., \mathbf{x}_n\}$, and those for the remaining atoms be represented by $\mathbf{Y}$. We write the total potential energy U as the sum

$$U(\mathbf{R}) = u_{\text{il}}(\mathbf{X}) + u_{\text{ll}}(\mathbf{X}) + u_{\text{ir}}(\mathbf{X}, \mathbf{Y}) + u_{\text{lr}}(\mathbf{X}, \mathbf{Y}) + u_{\text{rr}}(\mathbf{Y}) \quad (23)$$

While the functional form in eq 23 arises naturally for a pairwise decomposable molecular mechanical force field, it is always possible to reconstruct the total potential energy as a sum of separate terms, even in the context of a nonpairwise additive energy surface. It should be noted that in order to enable the conceptual separation of ligands from the macromolecule to which they are covalently bonded implicitly requires the ability to break chemical bonds. Although the formal problem can easily be overcome when the system is modeled using a molecular mechanical force field (i.e., there is a bonded list of atoms), a proper formulation of the separation process is needed to define the inner and outer regions. As in QM/MM simulation methodologies, the coupling between the ligands and the rest of the macromolecule is expected to involve MM nonbonded interactions (van der Waals and electrostatics) as well as a number of MM covalent interactions (bonds, angles, dihedrals);[43−48] all of these terms are included in $u_{\text{lr}}(\mathbf{X}, \mathbf{Y})$. Following from this separation, the equilibrium binding constant can be expressed according to eq 6 as

$$K_{\text{ion}} = f_{\text{excl}}\, e^{\beta G_{\text{bulk}}}\frac{\int'_{\text{holo}} d\mathbf{R}\, e^{-\beta[u_{\text{il}}+u_{\text{ll}}+u_{\text{ir}}+u_{\text{lr}}+u_{\text{rr}}]}}{\int' d\mathbf{R}\, \delta(\mathbf{r}_{\text{ion}} - \mathbf{r}')\, e^{-\beta[u_{\text{ll}}+u_{\text{lr}}+u_{\text{rr}}]}} \quad (24)$$

where the prime on the integral is a short-hand notation for the configurational restriction

$$\int' d\mathbf{R}... \equiv \int d\mathbf{R}\, H^{(\text{out})}_{\{l_{n+1},...\}}... \quad (25)$$

Integrating out the $\mathbf{Y}$ degrees of freedom, the binding free energy can be expressed as

$$K_{\text{ion}} = f_{\text{excl}}\, e^{\beta G_{\text{bulk}}}\frac{Z_{\{\text{ion},l_1,...,l_n\}}}{Z_{\{l_1,...,l_n\}}} \quad (26)$$

where $Z_{\{\text{ion},l_1,...l_n\}}$ and $Z_{\{l_1,...l_n\}}$ are Boltzmann configurational integrals for the complex comprising the n ligands with and without the ion, respectively,

$$Z_{\{\text{ion},l_1,...,l_n\}} = \int_{\text{holo}} d\mathbf{r}_{\text{ion}}\, d\mathbf{x}_1 ...\, d\mathbf{x}_n\, e^{-\beta W_{\{\text{ion},l_1,...,l_n\}}(\mathbf{r}_{\text{ion}},\mathbf{x}_1,...,\mathbf{x}_n)} \quad (27)$$

with $W_{\{\text{ion},l_1,...l_n\}} \equiv [u_{\text{il}} + u_{\text{ll}} + \Delta W_{\text{ir}} + \Delta W_{\text{lr}}]$, and

$$Z_{\{l_1,...,l_n\}} = \int d\mathbf{x}_1 ...\, d\mathbf{x}_n\, e^{-\beta W_{\{l_1,...,l_n\}}(\mathbf{x}_1,...,\mathbf{x}_n)} \quad (28)$$

with $W_{\{l_1,...l_n\}}Z_{\{\text{ion},l_1,...l_n\}} \equiv [u_{\text{ll}} + \Delta W_{\text{lr}}]$. The potential of mean force (PMF) ΔW_{ir} is the contribution coming from the interaction between the ion and the rest

$$e^{-\beta\Delta W_{\text{ir}}(\mathbf{r}_{\text{ion}},\mathbf{x}_1,...,\mathbf{x}_n)} = \frac{\int'_{\text{holo}} d\mathbf{Y}\, e^{-\beta[u_{\text{ir}}+u_{\text{lr}}+u_{\text{rr}}]}}{\int'_{\text{holo}} d\mathbf{Y}\, e^{-\beta[u_{\text{lr}}+u_{\text{rr}}]}} \quad (29)$$

and ΔW_{lr} is defined as the contribution coming from the interaction between the ligands and the rest

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

$$e^{-\beta[\Delta W_{lr}(\mathbf{x}_1,...,\mathbf{x}_n)-\Delta W_{lr}^{ref}]} = \frac{\int' d\mathbf{Y}\, e^{-\beta[u_{lr}(\mathbf{X},\mathbf{Y})+u_{rr}(\mathbf{Y})]}}{\int' d\mathbf{Y}\, e^{-\beta[u_{lr}(\mathbf{x}_1^{ref},...,\mathbf{x}_n^{ref},\mathbf{Y})+u_{rr}(\mathbf{Y})]}} \quad (30)$$

ΔW_{lr} represents the influence of the rest of the system onto the selected ligands coordinating the ion in the inner region. In particular, this includes the effect of the covalent bonded terms (bonds, angles and dihedrals) between the ligands and the macromolecule. The PMF ΔW_{lr} is defined relative to an arbitrary offset constant, ΔW_{lr}^{ref}. Typically, it is advantageous to set the value of the PMF to zero at its absolute minimum, $\{\mathbf{x}_1^{ref}, ..., \mathbf{x}_n^{ref}\}$, such that it will be positive for all other configurations. This also simplifies the expressions resulting from small harmonic fluctuations (see below). Nevertheless, the actual value of the offset constant is unimportant since it cancels out in the expression for the binding constant, eq 26. It is important to note that the function $\Delta W_{lr}(\mathbf{x}_1, ..., \mathbf{x}_n)$ cannot be defined on the basis of a complete decoupling of the ligands with the rest of the system

$$e^{-\beta\Delta W_{lr}(\mathbf{X})} \neq \frac{\int d\mathbf{Y}\, e^{-\beta[u_{lr}(\mathbf{X},\mathbf{Y})+u_{rr}(\mathbf{Y})]}}{\int d\mathbf{Y}\, e^{-\beta u_{rr}(\mathbf{Y})}} \quad (31)$$

as is done with respect to the ion-rest interaction $\Delta W_{ir}(\mathbf{r}_{ion}, \mathbf{x}_1, ..., \mathbf{x}_n)$ in eq 29. Though mathematically possible, a complete decoupling of the ligands from the rest of the macromolecule as proposed elsewhere[13] is essentially unphysical and leads to a poorly defined thermodynamic state.[49]

Equilibrium ion selectivity can be treated from the perspective of the reduced model. Following eq 11, the free energy difference between ion i and j in the binding site is

$$e^{-\beta\Delta G_{site}^{ij}} = \frac{f_{excl}^{i} Z_{\{ion,l_1,...,l_n\}}^{i}/Z_{\{l_1,...,l_n\}}}{f_{excl}^{j} Z_{\{ion,l_1,...,l_m\}}^{j}/Z_{\{l_1,...,l_m\}}} \quad (32)$$

where the short-hand notation indicates the functional dependence of ion i on the n ligands coordinates $\{\mathbf{r}_{ion}, \mathbf{x}_1, ..., \mathbf{x}_n\}$, or ion j on the m ligands coordinates $\{\mathbf{r}_{ion}, \mathbf{x}_1, ..., \mathbf{x}_m\}$. If the sets of ligands are chosen wisely, it is expected that the probabilities restricting all other ligands will be equal to one. Although the formalism is general and the sets of ligand do not need to match exactly, it is reasonable to expect that the set of ligands for ion i and ion j will overlap to some extent for a given binding site under consideration. If the reduced set of n coordinating ligands is identical for ion i and j then the ion-independent configurational integrals for the apo state, $Z_{\{l_1,...l_n\}}$ and $Z_{\{l_1,...l_n\}}$, cancel out, yielding considerable simplification, see eq 2 in ref 11. However, it is important to note that this is possible only if the number and selection of nearest ligands is the same for ion i and j. If the reduced models of the two ions do not include the same set of ligands, then it is critical to keep track of the configurational integral for both the holo and apo states properly, and to consider the ratio of the absolute binding constants for ion i and j (including all the steps corresponding to imposed configurational restrictions) based on eq 32. Water molecules in the ion binding site present a special case, which can be treated via eq 18 on the basis of the equilibrium association constants, $\tilde{K}_w^{apo}(m)$ and $\tilde{K}_w^{holo}(m)$, that m water molecules occupy the ion binding site. It is also possible to extend the present theoretical formulation to express these equilibrium association constants in terms of reduced models (see the Appendix).

2.6. Simplified Reduced Models. In many cases, ion binding involves very strong interactions, leading to an ion-ligand complex able to undergo relatively small thermal fluctuations $\Delta\mathbf{X}$ around a dominant energy-minimum configuration $\mathbf{X}_{min}$. An attractive treatment in this case is to approximate the configurational integrals using a local quadratic expansion of the potential energy surface around the energy minimum. Letting W represent the total potential energy relevant for the configurational integral Z for a given reduced system, the quadratic approximation generally assumes the form

$$W(\mathbf{X}) \approx W_{min} + \frac{1}{2}\Delta\mathbf{X}^t\cdot\mathbf{W}\cdot\Delta\mathbf{X} \quad (33)$$

where $\Delta\mathbf{X} = (\mathbf{X} - \mathbf{X}_{min})$ is the relative displacement, W_{min} is the minimum energy, and $\mathbf{W}$ is the second derivative matrix of the PMF evaluated at $\mathbf{X}_{min}$. The configurational integral can then be approximated as

$$\begin{aligned} Z &= \int_{holo} d\mathbf{X}\, e^{-\beta W(\mathbf{X})} \\ &\approx e^{-\beta W_{min}} \int d(\Delta\mathbf{X})\, e^{-\beta(\Delta\mathbf{X}^t\cdot\mathbf{W}\cdot\Delta\mathbf{X})/2} \\ &= e^{-\beta W_{min}} \prod_{i=1}^{N} \left(\frac{2\pi k_B T}{\lambda_i}\right)^{1/2} \end{aligned} \quad (34)$$

where λ_i are the N eigenvalues of the matrix $\mathbf{W}$. It is noteworthy that all the eigenvalues are expected to be nonzero since the reduced model is constructed with respect to a frame of reference that is fixed relative to the orientation of the macromolecule (i.e., free translation or rotation of the degrees of freedom $\mathbf{X}$ do not occur).

A common difficulty with the harmonic approximation concerns the existence of multiple energy minima.[50] Introducing a restraining potential u_c may offer an effective solution to the problems encountered if there are multiple local energy local minima. An alternative approach to address the issue of multiple local energy local minima is to introduce a step-function, $I(\mathbf{X})$, restricting the configuration to a single harmonic energy well.[13] As discussed above, the unbiased ion binding constant involving such configurational restrictions can be expressed as

$$K_{ion} = \frac{\langle I\rangle_{apo}}{\langle I\rangle_{holo}} K_{ion}^{(I)} \quad (35)$$

where $K_b^{(I)}$ is the equilibrium binding constant for a system in which all configurations have been restricted by the step-function I. If the anharmonicity is strong and there are multiple energy minima very close to one another, the function $I(\mathbf{X})$ must be narrowly defined to encompass a single energy well, which can then be treated by the local harmonic approximation. This implies that the average of the narrowly defined step-function, $\langle I\rangle_{apo}$ and $\langle I\rangle_{holo}$ will be much smaller than 1, which may pose some difficulties in trying to evaluate them accurately. An additional problem might occurs if the restriction imposed by the step-function I becomes so narrow that it affects the Gaussian integral

$$\int d(\Delta\mathbf{X}) I\, e^{-\beta(\Delta\mathbf{X}^t\cdot\mathbf{W}\cdot\Delta\mathbf{X})/2} \neq \prod_{i=1}^{N} \left(\frac{2\pi k_B T}{\lambda_i}\right)^{1/2} \quad (36)$$

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

In this case, the familiar closed-form expressions for the harmonic partition function may not be valid and a correction to account for the configurations excluded by the function I would be needed. In contrast, the configurational restraining potential u_c can be chosen to be arbitrarily stiff and yet differentiable, thus ensuring the Gauss integral is performed only on a near-harmonic energy surface. Thus, the introduction of a restraining potential u_c represents a general solution to a well-known limitation of the harmonic approximation, which is the neglect of anharmonicity and the treatment of multiple energy minima. This approach has some similarities with the usage of soft cutoffs to define the indicator function of the inner region in QCT.[51] Nevertheless, these treatments requires the correct evaluation of the constraint free energies, which may be difficult to calculate.

A different approach to circumvent the limitations of the harmonic approximation is to try to incorporate some of the dominant anharmonic features of the system into the reduced model.[11] This can be achieved by exploiting the information from detailed all-atom MD simulations. To this end, we define the confining contribution, $\Delta W_c(\mathbf{X})$, for each atom k in the reduced model such that it corresponds to the smallest possible spherical volume V_k that encompasses all the dynamic excursions of that atom in the reference frame of the protein (no global translation/rotation). Each spherical volume V_k is parametrically defined by the position of the center $\overline{\mathbf{r}}_k$ and radius R_k; those parameters can be extracted from MD simulations of the full system. Accordingly, the architectural confinement contribution is defined as

$$\Delta W_c(\mathbf{X}) \equiv -k_B T \ln[\prod_k \Theta(|\mathbf{r}_k - \overline{\mathbf{r}}_k| - R_k)] \tag{37}$$

where $\mathbf{r}_k$ is the position of the kth atom, and the Θ's are Heaviside step-functions. Mathematically, the atomic Heaviside functions $\Theta(|\mathbf{r}_k - \overline{\mathbf{r}}_k| - R_k)$ support a spherical region of radius R_k centered on $\overline{\mathbf{r}}_k$ for each atom k. For each atom k, the radius R_k can be extracted from the spontaneous excursions of this atom during all-atom MD with ion i and j. R_k is independent of ion type. In the reduced model, the movement of the atom k is "confined" to remain inside this spherical region, but is free otherwise. Typically, the radius of the volumes V_k estimated from all-atom MD is typically on the order of 1.0−1.5 Å.[11] This simple choice defines probability distributions for a minimal default model that incorporates the generic effect of architectural confinement by the surrounding protein structure in an idealized fashion (without the protein, the volume V_k would be unbounded and the ligands would be allowed to translate away in space). Remaining architectural effects, $\Delta W_{lr} - \Delta W_c$, can be introduced systematically into the reduced model by relying a cumulant expansion of the atomic fluctuations evaluated from from all-atom MD simulations of the binding site.[11]

2.7. Relation to Quasi-Chemical Theory. As noted previously, eq 18 is similar to the expression from quasi-chemical theory (QCT) for the solvation free energy of an ion in bulk solvent.[19,20,25] This approach has been used to examine the hydration of Li^+, Na^+, and K^+.[21,22,24] QCT has also been employed to discuss ion selectivity in proteins binding sites.[4,5] QCT typically consists in defining a spherical inner shell region of radius R, based on a distance criterion between the ion and the solvent molecule. In previous applications of QCT, the solvation free energy of the ion has typically been expressed as if the complex comprising one ion and n solvent molecules is first formed in vacuum (or in the gas phase), and is subsequently inserted into the bulk solvent environment. This leads to the standard QCT expression for the solvation of an ion in a bulk phase liquid[25]

$$e^{-\beta G^i_{bulk}} = P^i_0(0) e^{-\beta G^i(\text{cluster})} \sum_{n=0} e^{n\beta\Delta\mu_s} (\overline{\rho}_s)^n \tilde{K}^i_s(n) \tag{38}$$

where $P^i_0(0)$ is the probability that zero solvent molecules are found to occupy the spherical region of radius R_i in the absence of the ion i (the inner sphere), $\overline{\rho}_s$ is the density of the unperturbed bulk solvent, $\Delta\mu_s$ is the excess chemical potential of a solvent molecule in the unperturbed bulk phase, $\tilde{K}^i_s(n)$ is the vacuum equilibrium binding constant for n solvent molecules to associate as a cluster within a spherical region centered on the ion i, and $G^i(\text{cluster})$ is the solvation free energy for inserting this cluster into a spherical region excluding all other solvent molecules from the bulk. Similarly, eq 18 could also be rewritten by using a vacuum or gas-phase-like intermediate state. QCT reformulates the solvation free energy of an ion in bulk solvent into a mathematical form that is reminiscent of a chemical equilibrium leading to the formation of a molecular cluster, hence its name quasi-chemical. QCT adopts a particularly simple form if it is assumed that the clusters undergo only small harmonic fluctuations around a single energy minimum, which leads to the RRHO approximation. The central quantities of interest, which can be evaluated readily with standard ab initio programs, are the gas phase association constants for the clusters with one ion and n water molecules. This leads to the standard primitive form of QCT (pQCT), which provides the simple closed-form expressions for the solvation free energy of an ion i in a bulk liquid.[25] In many studies of ion solvation based on QCT, the sum over n in eq 38 was approximated by a single dominating coordination state.[21,22,24,36]

The standard form of QCT based on eq 38 is not directly applicable to a protein binding site because solvent molecules are part of a bulk liquid and free to translate in space, whereas the ion-coordinating ligands of a binding site are covalently tethered to the rest of the macromolecule. An appropriate QCT-like formulation, designed to describe ion binding to a macromolecule, might take the form

$$K^i_{ion} = f^i_{excl} e^{\beta G^i_{bulk}} \frac{\tilde{K}^i_{\{ion,l_1,\ldots,l_n\}}}{\tilde{K}_{\{l_1,\ldots,l_n\}}} \tag{39}$$

where $\tilde{K}^i_{\{ion,l_1,\ldots,l_n\}}$ would be the equilibrium binding constant for assembling the ion and its n ligands into the macromolecule starting from a reference state where the ion and the ligands are decoupled from the surroundings (holo state), and $\tilde{K}_{\{l_1,\ldots l_n\}}$ would be the equilibrium binding constant for assembling the n ligands into the macromolecule starting from a reference state where the ligands are decoupled from the surroundings (apo state). For the ion, a suitable reference gas phase state can be defined, where it is fully decoupled from its surroundings. However, it is not straightforward to construct a similar reference state for the n ligands, see eq 31. In reality, the physical dissociation of the ligands from the macromolecule would give rise to a highly unstable situation, probably causing its unfolding.[49] This would, in turn, lead to a poorly defined statistical mechanical reference state used to define $\tilde{K}^i_{\{ion,l_1,\ldots,l_n\}}$ and $\tilde{K}_{\{l_1,\ldots l_n\}}$. As indicated by eq 31, a reference state in which

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

the microscopic interaction between the ligands and the macromolecule $u_{lr}(\mathbf{X},\mathbf{Y})$ is switched off can be written,[13] but the mathematical expression represents a highly artificial situation.[49] Thus, while they have the form of equilibrium binding constants, the quantities $\tilde{K}^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}$ and $\tilde{K}_{\{l_1,\ldots l_n\}}$ must be constructed in terms of a virtual association process. This is in contrast with the treatment of the solvent molecules occupying the inner region in the QCT theory for ion solvation in a bulk liquid as in eq 38, or the water molecules occupying the binding site as in eq 18. Because these solvent molecules are not covalently bound to the ion in the inner region, it is possible to define a reference state in which they are in the gas phase. While the configurations of the ligands of the macromolecule may be altered by the ion binding process, they remain tethered at all time to the rest of the macromolecular structure. Thus, it is necessary to define a virtual decoupled state that serves as a reference. This can be achieved on the basis of the two PMFs ΔW_{ir} and ΔW_{lr} that were defined in eqs 29 and 30. The QCT-like pseudoassociation constants are then

$$\tilde{K}^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}} = \frac{Z^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}}{\Pi^{n}_{\alpha=1} Z_{l_\alpha}} \tag{40}$$

and

$$\tilde{K}_{\{l_1,\ldots,l_n\}} = \frac{Z_{\{l_1,\ldots,l_n\}}}{\Pi^{n}_{\alpha=1} Z_{\{l_\alpha\}}} \tag{41}$$

where Z_{l_α} is the configurational integral for the ligand α decoupled from the influence of the ion or the macromolecule and isolated in vacuum

$$Z_{\{l_\alpha\}} = \int \mathrm{d}\mathbf{x}_\alpha \delta(\mathbf{r}_{l_\alpha} - \mathbf{r}^{\mathrm{g}})\, \mathrm{e}^{-\beta u_{l_\alpha}(\mathbf{x}_\alpha)} \tag{42}$$

The $Z_{\{l_\alpha\}}$ could be calculated with the classic RRHO approximation (n.b., for a nonlinear rigid molecule, this expression is equal to $8\pi^2$). It is noteworthy that eq 42 does not need to include an excess chemical potential of the ligand, $\Delta\mu_{l_\alpha}$, or a bulk density of a phase of ligands, $\overline{\rho}_{l_\alpha}$. This is in contrast with the quantities $\Delta\mu_s$ and $\overline{\rho}_s$ appearing in the QCT theory for ion solvation in bulk liquid water based on eq 38. Once these formalities are settled, addressing issues of ion selectivity is straightforward by considering the ratio of the pseudoassociation constants for ion i and j, i.e.

$$\begin{aligned}\mathrm{e}^{-\beta\Delta G^{ij}_{\mathrm{site}}} &= \frac{f^{i}_{\mathrm{excl}} \tilde{K}^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}/\tilde{K}_{\{l_1,\ldots,l_n\}}}{f^{j}_{\mathrm{excl}} \tilde{K}^{j}_{\{\mathrm{ion},l_1,\ldots,l_m\}}/\tilde{K}_{\{l_1,\ldots,l_m\}}} \\ &= \frac{f^{i}_{\mathrm{excl}} Z^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}/Z_{\{l_1,\ldots,l_n\}}}{f^{j}_{\mathrm{excl}} Z^{j}_{\{\mathrm{ion},l_1,\ldots,l_m\}}/Z_{\{l_1,\ldots,l_m\}}}\end{aligned} \tag{43}$$

and all of the configurational integrals for the ligand α, decoupled from the influence of the ion or the macromolecule and isolated in vacuum, cancel out. From this perspective, the construction of QCT-like equilibrium binding constants via eqs 39−42 allows one to recover the set of familiar RRHO gas-phase partition functions.

Because the bound complex ought to contain all the critical information about the ion-coordination structure by the ligands, it is tempting to focus exclusively on the ion-ligand binding constants $\tilde{K}^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}$ and $\tilde{K}^{j}_{\{\mathrm{ion},l_1,\ldots,l_n\}}$, to analyze the factors controlling selectivity in ion binding sites. This route is intuitively appealing because these are the only molecular fragments that actually include the ions i and j together with their respective coordinating ligands. The situation is similar with a treatment of ion solvation based on the standard form of QCT, eq 38, which also relies on the equilibrium constants $\tilde{K}^{i}_{s}(n)$ of ion−solvent clusters in vacuum. However, it is important to recall that many of the formal simplifications leading to eq 38 in the case of ion solvation in a bulk liquid are not applicable to the case of ion binding to the ligands of a macromolecule.[49] For instance, it is noteworthy that in eq 38, all statistical mechanical information of the system with the ion decoupled from the surroundings is incorporated into the probability $P^{i}_{0}(0)$ to find no solvent molecule inside the subvolume of radius R corresponding to the inner region. Although the latter can be expressed as a sum of configurational integrals with n solvent molecules and no ion, it is typically evaluated directly via MD simulations or empirical approximations in application of QCT. As a result of this resummation, configurational integral corresponding to a cluster of n solvent molecules without the ion do not appear explicitly in eq 38. The only remaining quantity that requires an explicit evaluation are the $\tilde{K}^{i}(n)$, which corresponds to the equilibrium constants for the formation of a cluster comprising one ion and n solvent molecules in vacuum. A similar route cannot easily be taken for the ligands forming an ion binding site in a macromolecule because the configurational integrals for the n ligands in the absence of the ion cannot be resummed as a simple probability $P^{i}_{0}(0)$ as in eq 19.

The ensuing formal complication is perhaps illustrated most clearly when the relative free energy of ions i and j is expressed on the basis of reduced models comprising a different number of ligands. For the sake of simplicity, let us consider the case where the set of ligands 1 to n are the same for the ions i and j but that the reduced model for ion j has one additional ligand. The selectivity of binding sites can be expressed as

$$\mathrm{e}^{-\beta\Delta G^{ij}_{\mathrm{site}}} = A\frac{\tilde{K}^{i}_{\{\mathrm{ion},l_1,\ldots,l_n\}}}{\tilde{K}^{j}_{\{\mathrm{ion},l_1,\ldots,l_{n+1}\}}} \tag{44}$$

where the prefactor A is

$$\begin{aligned}A &= \frac{\tilde{K}_{\{l_1,\ldots,l_{n+1}\}}}{\tilde{K}_{\{l_1,\ldots,l_n\}}} \\ &= \frac{Z_{\{l_1,\ldots,l_{n+1}\}}}{Z_{\{l_1,\ldots,l_n\}} Z_{\{l_{n+1}\}}}\end{aligned} \tag{45}$$

Effectively, the prefactor A represents a virtual (quasi-chemical) equilibrium association constant for binding the $(n+1)$th additional ligand to the reduced model of the apo state with n ligands (A has dimension of volume for one additional ligand). It is possible to derive a closed-form expression from eq 45 by integrating out the $(n+1)$th ligand submitted to the spatial restriction that it must lie within the outer region

$$A = \mathrm{e}^{-\beta\Delta g} \int \mathrm{d}\mathbf{r}\, h(\mathbf{r}) \tag{46}$$

where Δg is the free energy for inserting the $(n+1)$th ligand into the reduced model with its center-of-mass at some optimal reference position $\mathbf{r}_{\mathrm{ref}}$

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

$$e^{-\beta\Delta g} = \frac{\int d\mathbf{x}_1 \ldots d\mathbf{x}_{n+1}\, \delta(\mathbf{r}_{l_{n+1}} - \mathbf{r}_{ref})\, e^{-\beta W_{\{l_1,\ldots,l_{n+1}\}}}}{\int d\mathbf{x}_1 \ldots d\mathbf{x}_{n+1}\, \delta(\mathbf{r}_{l_{n+1}} - \mathbf{r}_{ref})\, e^{-\beta[W_{\{l_1,\ldots,l_n\}} + u_{\{l_{n+1}\}}]}} \quad (47)$$

and $h(\mathbf{r})$ is the distribution function of the center-of-mass of the ligand (no global translation or rotation of the macromolecule are allowed)

$$h(\mathbf{r}) = \frac{\int d\mathbf{x}_1 \ldots d\mathbf{x}_{n+1}\, \delta(\mathbf{r}_{l_{n+1}} - \mathbf{r})\, e^{-\beta W_{\{l_1,\ldots,l_{n+1}\}}}}{\int d\mathbf{x}_1 \ldots d\mathbf{x}_{n+1}\, \delta(\mathbf{r}_{l_{n+1}} - \mathbf{r}_{ref})\, e^{-\beta W_{\{l_1,\ldots,l_{n+1}\}}}} \quad (48)$$

Both Δg and $h(\mathbf{r})$ incorporate short-range covalent interactions (bonds, angles, and dihedrals) as well as nonbonded interactions (van der Waals and electrostatics) of the $(n + 1)$th ligand with the surrounding macromolecule. By construction, $h(\mathbf{r}_{ref}) = 1$, and $h(\mathbf{r}) \leq 1$ for all $\mathbf{r}$ if $\mathbf{r}_{ref}$ is chosen to be the optimal position for inserting the ligand. It follows that $h(\mathbf{r})$ is a dimensionless function supporting a small volume $v = \int d\mathbf{r}\, h(\mathbf{r})$, corresponding to the allowed spatial fluctuations of the center-of-mass of the $(n + 1)$th ligand (in the apo state). Integrating eq 46 leads to the form of the prefactor, $A = v \exp[-\beta\Delta g]$. This procedure could be repeated any number of times if m ligands need to be integrated out. If these m ligands are chemically identical, it is tempting to write the relative free energy of ion i and j in the binding site as

$$e^{-\beta\Delta G^{ij}_{site}} \approx v^m e^{-m\beta\Delta g} \frac{\tilde{K}^i_{\{ion,l_1,\ldots,l_n\}}}{\tilde{K}^j_{\{ion,l_1,\ldots,l_{n+m}\}}} \quad (49)$$

This approximation is based on the assumption that correlations between the ligands in the apo state can be ignored in calculating the prefactor, i.e., the value of the fluctuation volume v and of the interaction free energy Δg is the same for the addition of the $(n + 1)$th ligand, $(n + 2)$th ligand, etc. To proceed any further, it would be necessary to estimate the parameters v and Δg directly from the properties of the particular macromolecule under investigation.

3. DISCUSSION

A formulation of reduced binding site models comprising a number of ion-coordinating ligands was provided. The general idea is depicted schematically in Figure 1. In the reduced model, the "near" degrees of freedom are treated explicitly, while the influence of the surrounding is implicitly incorporated via the PMFs ΔW_{ir} and ΔW_{lr}. The equilibrium binding constant of an ion for the binding site was expressed in terms of the reduced model. The rigorous definition of these two PMFs, via eqs 29 and 30, and the equilibrium binding constant of an ion expressed in terms of the reduced model via eq 26 represent the main formal results presented here. Ion selectivity can be treated by simply examining the ratio of the binding constant of different ions, each treated with their proper reduced model. Although the formulation developed here has exlusively been concerned with the binding of a single ion, it could in principle be generalized to treat the cooperative binding of two ions by considering reduced models that encompasses two coupled sites. However, such a study would have to proceed with caution because accounting for the cooperativity correctly would put an increased demand on the accuracy and realism of the reduced model. In the following, several issues that deserve a special attention will be discussed.

3.1. Formal Reduction. While the concept of a reduced model is intuitively obvious, many choices are possible, leading to important differences in formulations and implementations. For example, one could "tag" any putative ligand coordinator that comes within a certain distance R from the bound ion, and define a complete set of states comprising exactly n ligands. Such a strict restriction on the number of ligand is particularly useful to treat ion solvation. Summing over all values of n recovers all possible set of nearest ion-coordinating ligands. Furthermore, by virtue of the equivalence of all solvent molecules, the analysis leads to a formulation analogous to a small system grand canonical ensemble.[52] The ion solvation free energy is then expressed as a sum over a set of reduced models, i.e., small clusters comprising exactly one ion and n solvent molecules surrounded bulk liquid outside the sphere of radius R . This approach leads to quasi-chemical theory (QCT), see ref 25, or to a small system grand canonical ensemble (SSGCE) development.[8] However, a number of simplifications are possible in the case of ion solvation because any permutations over states with exactly n solvent molecules located inside the inner region yield equivalent configurational integrals of equal statistical mechanical Boltzmann weight. Such a procedure is not useful in the case of an ion binding site because even chemically identical ligand coordinators (e.g., backbone carbonyl groups of a protein) are no longer indistinguishable once they are covalently tethered to a macromolecule. Atoms and residues that are part of a biological macromolecule do not spontaneously swap their position on the experimental time scale of observation (except perhaps labile hydrogens), and statistical mechanical averages should be restricted to the sets of configurations where these chemical bonds are not breaking and forming. Thus, even if n ligands are chemically identical, one cannot assume that they produce $n!$ equivalent configurational integrals upon permutation. Considering the reduced model shown in Figure 1 as an example, the configurational probability of having the set of four ligands {2,3,5,8} as part of the inner region must obviously be different from the configurational probability of having ligands {1,2,3,4} within the inner region. The lack of equivalent permutation among these four ligands is also why no factor of 4! should be associated with the configurational integrals of this reduced model, see eq 26. The situation is different with respect to the water molecules, because they are genuinely equivalent. In this case, there should be a factor of $n!$ if n water molecules are in the binding site, see eq 18. Confusion over this type of issue has led to incorrect combinatorial factors in previous attempts at formulating reduced models of ion binding sites;[49] for example, see the development leading to eq A7 in ref 8. In the present formulation, the set of ion-coordinating ligands is prechosen to construct the reduced model; their number does not vary. The only exception concerns the presence of water molecules in the ion binding site. Extending the theoretical formulation via eq 18, the equilibrium association constants of m water molecules to the ion binding site (apo or holo) can also be expressed within a similar reduced model formalism (the generalization of these results is outlined briefly in the Appendix).

3.2. Selection of Relevant Ligand Coordinators. Because of the nonequivalence of the ligand coordinators from a macromolecule, it is advantageous to describe each ion-coordination state from a complete set of ligands $\{l_1, \ldots, l_n\}$ that are involved. It would be possible to introduce a family of configurational restrictions in order to define the complete ensemble of all allowed ion-coordination states, but this is not

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

necessary. Here, the reduced binding site model was formulated the basis of the configurational restriction $H^{(\mathrm{out})}_{\{l_{n+1},\ldots\}}$, introduced in eqs 20 and 21. This configurational restriction excludes any ligands, other than the chosen set $\{l_1, \ldots, l_n\}$, from the inner region. In Figure 1C and D, the set of ligands {2,3,5,8} are part of the reduced model, and a configuration restriction is applied to prevent the remaining ligands, {1,4,6,7,9,10,11,12}, from entering the inner subvolume. In principle, a reduced model could be defined without excluding these other ligands, but this means that their influence would then be implicitly incorporated in the PMFs $W_{\{\mathrm{ion},l_1,\ldots l_n\}}$ and $W_{\{l_1,\ldots l_n\}}$. Preventing any other ligands from directly participating to the binding site strengthens the significance of the reduced model because it allows one to have complete control over its content. While the present formulation imposes a configurational restriction $H^{(\mathrm{out})}_{\{l_{n+1},\ldots\}}$ excluding any ligands other than the chosen set $\{l_1, \ldots, l_n\}$, from the inner region, there is no corresponding restriction on the chosen set of ligands. For example, the set of ligands {2,3,5,8} in Figure 1C and D are not strictly confined to stay within the inner region. This is in contrast with the developments leading to eq 16, where the tagged n water molecules must strictly remain within the inner region via the function $H^{(\mathrm{in})}_{\{\mathrm{w}_1,\ldots\mathrm{w}_n\}}$. Such a spatial restriction in the case of eq 16 is useful because a sum over the configurations with n water molecules then forms a complete set. This is also critical in the derivation of QCT for ion solvation.[25] However, while a similar restriction $H^{(\mathrm{in})}_{\{l_1,\ldots l_n\}}$ could also be applied to the ligands of a macromolecule, it offers no particular advantage. In fact, a formulation without such a restriction is simpler and more convenient. It may be noted that the exclusion factor f_{excl} defined in eq 22 may be considerably affected if essential ion-coordinating ligands are excluded from the inner region, but that it remains unchanged even if superfluous ligands are included to the reduced model. Which ligands must be included in a meaningful reduced model is critical to allow the proper range of chemical composition and coordination state. As an example, excluding the strongly interacting ligand #2 in Figure 1C and D would certainly cause the exclusion factor f_{excl} to be smaller than 1, yielding a formally correct yet awkward reduction of the site. As another example, one could build a reduced model of a binding site of the KcsA channel with 8 carbonyl ligands, and study the binding of a Na^+ in this model. If the reduced model has been build correctly, then the dynamical behavior and range of flexibility should match what is seen in all-atom simulations. Namely, the Na^+ will most likely prefer to be directly coordinated by 4 carbonyls. This can still happen in a reduced model with 8 carbonyls, as long as the reduced model is build correctly with accurate representations for the two PMFs, ΔW_{ir} and ΔW_{lr}. However, if one construct a reduced model with only 4 carbonyls and then tries to study the binding of K^+, coordination by 8 carbonyl ligands is explicitly ruled out. Although the influence of the missing ligands ought to be incorporated implicitly with the two PMFs, ΔW_{ir} and ΔW_{lr}, this is not reaching the goal of treating all the critical degrees of freedom and nearest ligands explicitly. Therefore, it is clear that a reduced model can be build incorrectly and yield misleading results. This is particular true if critical ligands are left out. On this basis, it seems preferable to select the set $\{l_1, \ldots, l_n\}$ as generously as can be afforded, and treat the exclusion of any ligand with caution when constructing the reduced model. The direct participation of water molecules within the binding site is one exception to this rule. If ion coordination involves a small number of water molecules, i.e., less than 2 or 3, then it seems preferable to deliberately monitor their presence as discussed in section 2.4. On the other hand, the reduced model requires a larger number of water molecule, it may be necessary to resort to a scheme based on grand canonical Monte Carlo (GCMC) to achieve proper sampling of the relevant coordinate states. As shown elsewhere, eq 18 and a GCMC sampling scheme are equivalent.[53]

3.3. Significance of QCT Applied to Protein Sites. The present analysis can help clarify some critical aspects of previous applications of the standard form of QCT to study ion selectivity of protein binding sites.[4,5] In principle, an application of eq 38 to a protein binding site raises a number of questions because the theory relies on the fact that solvent molecules are identical and free to exchange with one another in the bulk phase. These assumptions are not satisfied in the case of a protein binding site because the ion-coordinating ligands of a binding site are covalently tethered to the rest of the macromolecule.[49] To apply the bulk phase form of QCT based on eq 38 to a protein binding site, one must "map" the features of a protein binding site onto the properties $\Delta\mu$ and $\overline{\rho}$ of some hypothetical reference fluid. This is the approach that was taken to examine the selectivity of the binding site in the KcsA K^+ channel.[4,5] The ion-coordinating backbone carbonyl ligands were represented as formamide (form) molecules and the interior of the protein was represented as a quasi-liquid formamide phase with a low dielectric constant. To evaluate the selectivity of the channel for K^+ over Na^+, the free energy of a K^+ ion coordinated by 8 formamide ligands molecules was compared to the free energy of a Na^+ ion coordinated by 6 formamide ligand molecules. The free energy of the ion was estimated using the standard form of QCT as expressed in eq 38, with the sum limited to a single cluster with n ligands. All the configurational integrals were evaluated using the RRHO approximation as ion and formamide clusters in vacuum. The particular number of coordinating ligands, 6 for Na^+ and 8 for K^+, were chosen because it made it possible to optimize the geometry of the ion-formamide clusters in the gas phase using energy minimization. A necessary requirement for the application of the RRHO approximation is the existence of a stable, energy-minimized configuration with a Hessian matrix comprising only real and positive-definite eigenvalues. In this context, the relative free energy of Na^+ and K^+ in the reduced model of the binding site was expressed as,[4]

$$\begin{aligned}\Delta G_{\mathrm{site}}(\mathrm{Na}^+, \mathrm{K}^+) &= -k_{\mathrm{B}}T \ln\left[\frac{(\overline{\rho}_{\mathrm{form}} e^{\beta\mu^{\mathrm{ex}}_{\mathrm{form}}})^6 \tilde{K}^{\mathrm{Na}^+}_{\mathrm{form}}(n=6)}{(\overline{\rho}_{\mathrm{form}} e^{\beta\mu^{\mathrm{ex}}_{\mathrm{form}}})^8 \tilde{K}^{\mathrm{K}^+}_{\mathrm{form}}(n=8)}\right] \\ &= -k_{\mathrm{B}}T \ln\left[\frac{1}{(\overline{\rho}_{\mathrm{form}} e^{\beta\mu^{\mathrm{ex}}_{\mathrm{form}}})^2} \frac{\tilde{K}^{\mathrm{Na}^+}_{\mathrm{form}}(n=6)}{\tilde{K}^{\mathrm{K}^+}_{\mathrm{form}}(n=8)}\right] \end{aligned} \quad (50)$$

where $\tilde{K}^{\mathrm{Na}^+}_{\mathrm{form}}(6)$ and $\tilde{K}^{\mathrm{K}^+}_{\mathrm{form}}(8)$ are the vacuum ion-cluster QCT configurational integrals, of Na^+ or K^+ with 8 and 6 form molecules, respectively (contribution from ΔW_{ir} and ΔW_{lr} were thus ignored in $\tilde{K}$). The parameters $\overline{\rho}_{\mathrm{form}}$ and $\Delta\mu_{\mathrm{form}}$ were associated with a bulk-like phase meant to represent the concentration of carbonyl ligands in the interior of the protein; $\overline{\rho}_{\mathrm{form}}$ was taken near the bulk density of liquid formamide (22.8 M or 0.014 $Å^{-3}$). Depending on the assumed dielectric

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

constant of the surrounding protein phase, $\Delta\mu_{form}$ varied from 0 kcal/mol (for a medium with $\epsilon = 1$) to −10.7 kcal/mol (for a medium of high dielectric coefficient). The concept of "quasi-liquid" environment for the ion binding site, responsible for driving up the coordination preference of the ion, was defined as the conditions yielding decreased electrostatic energy penalties for "ligand extraction" from the low dielectric surrounding protein phase.[4,5]

While no rational explanation was offered in ref 4 to justify an application of the standard form of QCT based on eq 38 to model the ion binding sites in the KcsA channel, the underlying physical assumptions can now be understood from eqs 49 and 50. The identity $\exp[\beta\Delta\mu_b]/\overline{\rho}_b \equiv \nu \exp[\beta\Delta g]$ shows that the formamide molecule representing the carbonyl ligands of the channel are not extracted from some effective bulk-like liquid formamide phase of density $\overline{\rho}_{form}$ with excess chemical potential $\Delta\mu_{form}$; they are incorporated into the reduced model of the apo state via a virtual binding process according to eq 45. Here, Δg is the free energy for inserting one formamide ligand into the reduced model while ν correspond to its allowed spatial fluctuations; the power-law form with $m = 2$ in eq 50 means that correlations are ignored during the virtual binding process. Clearly, the numerical value attributed to the prefactor $A = \nu \exp[\beta\Delta g]$ carries some very specific assumptions about the character of the binding site under consideration. It is thus of interest to contrast the plausible range of the two parameters, ν and Δg, with the properties of the macromolecule and binding site under investigation. Based on MD simulations of proteins, ν might vary from 0.125 Å^3 for a ligand in nearly rigid binding site, up to about 3 Å^3 for a ligand in a fairly flexible binding site. Estimating a plausible range for Δg is more difficult. Presumably, it could vary from a favorable to an unfavorable situation, with a range of energy that corresponds roughly to the hydrogen bonding energy of the ligand, e.g., from about +2 kcal/mol to −6 kcal/mol. Thus, plausible values of the prefactor $A = \nu \exp[\beta\Delta g]$ could range from about 10^{-4} Å^3 to about 10^{+5} Å^3, leading to a maximum possible impact on the order of about 9 kcal/mol on the relative free energy ΔG^{ij}_{site} in eq 44. A site for which the prefactor is $\gg 1$ means that there is a high free energy reward for including the additional $(n + 1)$th ligand in the reduced model of n ligands. In contrast, a site for which the prefactor is $\ll 1$ means that there is a high free energy cost for including this ligand in the reduced model. The numerical values of prefactor that were considered in ref 4 ranged roughly from 73 to 10^{10} Å^3; note that the bulk density of liquid formamide (0.014 Å^{-3}) corresponds to a relatively large volume $\nu = 1/\overline{\rho}_{form}$ of 72.8 Å^3. In this regime, the prefactor is always $\gg 1$. In other words, a considerable free energy cost for recruiting an additional ligand to the reduced model was implicitly assumed.[4,5] The relative free energy would have the opposite sign if the binding site was nearly rigid or architecturally restrained to retain a fixed number of ligands, a situation that was not considered. Such a binding site, with a somewhat limited conformational freedom, would be consistent with the atomistic rms fluctuations of the backbone carbonyl ligands displayed by detailed MD simulations of the KcsA K^+ channel (on the order of 0.8 Å). This analysis shows that conclusions of ref [4] based on eq 50 were strongly affected through the choice of ν and Δg. These difficulties can be avoided. As shown by eq 43, a QCT-like formulation constitutes an unnecessary formal detour toward a virtual association process in the treatment of ion binding to a macromolecule.

3.4. Simplified Reduced Models. It is of interest to see how eq 26, which provides an exact expression for the equilibrium binding constant of an ion in terms of the reduced model, can be approximated and put into a simpler form to help clarify the mechanisms underlying the ion selectivity of a binding site. For example, if the total PMF is assumed to be dominated by a single harmonic free energy well according to eq 33, the multidimensional integrals of eq 26 can be evaluated by assuming small quadratic fluctuations around a dominant configuration using eq 34. An additional approximation consists in assuming that the properties of the binding site are dominated by the ion-ligand and ligand−ligand interactions, $u_{il} + u_{ll}$, and that the contribution from the two PMFs, ΔW_{ir} and ΔW_{lr}, can be treated perturbatively. In this case, the first step is to determine the minimum energy configuration of the ion-ligand complex isolated in vacuum, $\mathbf{X}_{min}$, and then evaluate the remaining terms for the same configuration.

Clearly, neglecting the long-range interactions between the ion and the rest of the system can be a reasonable approximation when comparing the relative free energy of two ions i and j carrying the same charge (e.g., Na^+ and K^+). These long-range contributions are incorporated via the PMF ΔW_{ir} defined by eq 29. The accuracy of this approximation, which has been used in several studies based on reduced models (see ref 11), can easily be verified in practice by showing that $\Delta W^i_{ir} \approx \Delta W^j_{ir}$. On the other hand, neglecting the PMF ΔW_{lr} in eq 34 requires careful considerations. In this case, some of the eigenvalues in eq 34 become identically equal to zero and it becomes necessary to use the rigid rotor and harmonic oscillator (RRHO) partition function. While the above approximate treatment allows one to express the equilibrium association free energy of the ion for the binding site in closed form, it can lead to serious errors. In the real system the ligands are covalently attached to the macromolecule and their deviations are restricted. Without ΔW_{lr}, all architectural influence from the surroundings macromolecule is ignored and the ion and the ligands forming the binding site are treated as an isolated molecular fragment in vacuum. In particular, it is possible that that the configuration at the minimum energy for the isolated fragment will depart considerably from the structure observed in the binding site of the macromolecule. Yet, energy minimization of the ion-ligand cluster cannot be avoided in practice because the Hessian matrix of the reduced model must be evaluated at an minimum energy to apply the harmonic approximation based on eq 34. In that sense, the need to have a proper energy minimum to evaluate the RRHO partition functions can lead to large and unrealistic structural deviations when it is applied to the isolated ion-ligand complex.

The results based on an isolated fragment in vacuum are necessarily of limited significance and must be interpreted with caution. Nevertheless, ab initio studies of isolated ion-ligand fragments can sometimes yield useful information about a binding site.[1,4,13,54−56] For example, it is fascinating to observe that the optimized geometry of one K^+ ion bound to four diglycine mimicking the backbone carbonyl ligands of the binding site S2 from the selectivity filter of the KcsA channel depart by only 0.28 Å root-mean-square relative to the high-resolution crystallographic structure.[1] This supports the notion that the S2 site is well adapted to K^+ coordination and that the binding of this ion is not accompanied by any structural strain energy in the macromolecule. In contrast, large structural deviations are observed when similar models are optimized in

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

the presence of a bound Na^+ ion.[4,55] It has been suggested that such large deviations provide fundamental theoretical evidence that the K^+ channel is specifically optimized for the conduction and releasing of K^+ ions.[4,13] Althought this suggestion is intriguing, it is important to keep in mind that the properties of the isolated ion-ligand complex in vacuum, without any account of ΔW_{lr}, could have a very limited relevance to the behavior of the ligands in the real binding site.

Fundamentally, these difficulties reflect the fact that isolated ion-ligand fragments in vacuum offer a poor and unrealistic representation of the binding site in the native environment of the macromolecule. In most cases, it is likely that the influence of the PMF ΔW_{lr} is too important to be neglected. One possible avenue to prevent large deviations is to apply a differentiable restraining potential u_c, which can then be unbiased properly according to eq 9. The effect of the restraining potential would, of course, have to be taken into account in calculating the Gaussian integral based on eq 34. This strategy also offers an effective solution to the problems encountered if there are multiple local energy local minima. As long as the free energy cost of the imposing the configurational restriction is correctly accounted for, the choice of the restraining potential u_c is essentially a matter of convenience. For example, if u_c is a steep quadratic potential centered on one of the dominant configurations, it is possible to evaluate the configurational integral accurately by assuming that there are only small harmonic fluctuations around a unique dominant configuration. An alternative route is to try to incorporate some of the dominant features of the PMF ΔW_{lr} into the reduced model.[11]

3.5. Strategies for Practical Applications. While the formulation in terms of a reduced model expresses the quantities of interest in terms of configurational integrals that involve a smaller number of degrees of freedom than the large initial system, one must still design a practical computational strategy for their evaluation. One possibility is to use direct free energy perturbation molecular dynamics (FEP/MD) simulation methods. Although FEP/MD simulations imposes some computational burden to achieve an appropriate statistical convergence, they can yield essentially exact results for a given reduced model. Alternatively, the configurational integrals could be evaluated numerically in closed form using a harmonic approximation (see above).

In practical applications one may envision to evaluate the configurational free energy contributions G_c^{holo} and G_c^{apo} via MD simulations based on a classical force field, and proceed with the treatment of the reduced model using ab initio electronic structure methods. To a first approximation, the free energy contribution from $\Delta W_{lr}(\mathbf{X})$ could be ignored. However, while the configuration of the reduced system is largely dominated by u_c if the rmsd restraining potential is chosen sufficiently stiff, it is important to realize that the ratio the value of $\Delta W_{lr}(\mathbf{X})$ correspond to a free energy associated with the surrounding architecture of the macromolecule. This quantity, which is difficult to estimate quantitatively, may not be negligible in the case of a structural rigid or stiff binding site.

It is important to ascertain the validity and accuracy of the reduced models in mixed quantum mechanical−molecular mechanical treatments. A possible method would be to first compute the free energy difference between ion i and j from all-atom FEP/MD simulations with a given force field, and then recompute the free energy dierences via the RRHO reduced model formalism with the same force field. An obvious necessary condition is that the two routes must yield nearly the same result. If the two routes do not yield nearly identical result then this is a strong indication that some critical element is missing from the reduced models and that the latter are incomplete. For example, a proper treatment may require a different number of water molecules for ions i and j, as described in eq 17.

When the number of ligands is not the same for the reduced model of ions i and j, the apo states with different number of ligands contribute and must be considered. The main challenge in this case is whether the reduced models will be sufficiently accurate to treat the four situations: ions i or j in the holo state, and two different apo states. However, when the number of ligands is the same for the reduced model of of ions i and j, the two apo states cancel out and one needs only to consider the ratio of the configurational integral for the holo state, i.e., the molecular fragments comprised of the ion and its coordinating ligands. In this case, one might hope that there are cancelation of errors and that the results will be less sensitive to the details of the reduced models. Ultimately, it is critical to validate the reduced model via a direct comparison with the results of FEP/MD simulations on the complete system as described above.

4. SUMMARY

A general statistical mechanical theory of reduced models for ion binding sites in macromolecules was formulated. The equilibrium binding free energy of an ion to a macromolecule was formally expressed in terms a reduced subset of ligands. A special attention was given to the issues related to the constraints posed by the covalent structure. The possibility of a variable number of water molecules in the ion binding site as explicitly taken into account into the formal development. The formulation provides a useful framework to incorporate the average effect from the surroundings on the ion and the nearest coordinating ligands by defining the two PMFs, ΔW_{ir} and ΔW_{lr}, corresponding to the coupling between the ion (i) and the ligand (l) with the rest (r) of the system. The harmonic approximations based on the assumption of small oscillations and fluctuations was discussed. The framework provided an opportunity to clarify the similarities and differences with quasi-chemical theory designed to treat solvation in bulk liquids.

It is appropriate to end with a word of caution. Although it is certainly helpful have correct statistical mechanical expressions in trying to formulate the microscopic basis of a reduced model for an ion binding site, this does not guarantee meaningful conclusions about every possible situations. Ultimately, one must also exert good physical judgment in choosing the key microscopic ingredients at the basis of the reduced model. This involves the preselection of the important ligands and the construction of realistic approximations for the two PMFs, ΔW_{ir} and ΔW_{lr}. A reasonable guideline is that the reduced model should be able to provide a genuinely accurate mimicry of the behavior observed in all-atom MD simulations of the system of interest. Only then would a meaningful assessment of ion selectivity be accessible from an analysis of a reduced model. With these limitations in mind, it seems difficult to envision how selectivity of a binding site could be assessed a priory on the basis of reduced models without some information from MD simulations about the coordination states. Therefore, the main utility of reduced models should be primarily to serve as a tool to help conceptualize the observations made from detailed all-atom MD simulations. In this regard, it is hoped that the present framework will serve in

dx.doi.org/10.1021/jp3007365 | J. Phys. Chem. B 2012, 116, 6966−6979

future studies of ion binding to complex biological macromolecular structures.

APPENDIX

Following a similar route to reformulate eq 17 in terms of the reduced models, the binding constant of the ion could be expressed as

$$K_{\text{ion}} = \mathrm{e}^{-\beta G_{\text{bulk}}} P_{\text{apo}}(0) \sum_{m=0} (\overline{\rho}_{\text{w}})^m \mathrm{e}^{m\beta\Delta\mu_{\text{w}}} \frac{Z_{\{\text{ion},\text{w}_1,\ldots,\text{w}_m,l_1,\ldots,l_n\}}}{Z_{\{l_1,\ldots,l_n\}} (Z_{\{\text{w}\}})^m} \tag{51}$$

where $Z_{\{\text{w}\}}$ is the configurational integral of an isolated water molecule in the gas phase

$$Z_{\{\text{w}\}} = \int \mathrm{d}\mathbf{x}_{\text{w}}\, \delta(\mathbf{r}_{\text{w}} - \mathbf{r}^{\text{g}}) \mathrm{e}^{-\beta u_{\text{w}}(\mathbf{x}_{\text{w}})} \tag{52}$$

and $Z_{\{\text{ion},\text{w}_1,\ldots,\text{w}_m,l_1,\ldots,l_n\}}$ is the configurational integral

$$Z_{\{\text{ion},\text{w}_1,\ldots,\text{w}_m,l_1,\ldots,l_n\}} = \frac{1}{m!} \int_{\text{holo}} \mathrm{d}\mathbf{r}_{\text{ion}}\, \mathrm{d}\mathbf{x}_{\text{w}_1} \ldots \mathrm{d}\mathbf{x}_{\text{w}_m}\, \mathrm{d}\mathbf{x}_{l_1} \ldots \mathrm{d}\mathbf{x}_{l_n}\, \mathrm{e}^{-\beta W_{\{\text{ion},\text{w}_1,\ldots,\text{w}_m,l_1,\ldots,l_n\}}} \tag{53}$$

defined from the total PMF of the ion–water complex in the binding site

$$W_{\{\text{ion},\text{w}_1,\ldots,\text{w}_m,l_1,\ldots,l_n\}} \equiv u_{\text{iw}} + u_{\text{il}} + u_{\text{ww}} + u_{\text{wl}} + u_{\text{ll}} + \Delta W_{\text{ir}} + \Delta W_{\text{wr}} + \Delta W_{\text{lr}} \tag{54}$$

comprising the ion–water, ion–ligands, water–water, water–ligands, and ligand–ligand interactions, together with the interactions of the ion, water, and ligands with the remainder (r) of the systems.

AUTHOR INFORMATION

Corresponding Author

*Phone: (773) 493-5303. E-mail: roux@uchicago.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was funded by Grant GM-62342 from the National Institute of Health (NIH). Helpful discussions with Chris N. Rowley and Haibo Yu are gratefully acknowledged. The author is beholden to Lawrence Pratt and Susan Rempe for generously sharing their scholarly views.

REFERENCES

(1) Noskov, S.; Berneche, S.; Roux, B. *Nature* **2004**, *431*, 830–834.
(2) Asthagiri, D.; Pratt, L.; Paulaitis, M. *J. Chem. Phys.* **2006**, *125*, 24701.
(3) Noskov, S.; Roux, B. *J. Gen. Physiol.* **2007**, *129*, 135–143.
(4) Varma, S.; Rempe, S. *Biophys. J.* **2007**, *93*, 1093–1099.
(5) Varma, S.; Rempe, S. B. *J. Am. Chem. Soc.* **2008**, *130*, 15405–15419.
(6) Noskov, S. Y.; Roux, B. *J. Mol. Biol.* **2008**, *377*, 804–818.
(7) Dixit, P. D.; Merchant, S.; Asthagiri, D. *Biophys. J.* **2009**, *96*, 2138–2145.
(8) Bostick, D. L.; Brooks, C. L. *Biophys. J.* **2009**, *96*, 4470–4492.
(9) Asthagiri, D.; Dixit, P.; Merchant, S.; Paulaitis, M.; Pratt, L.; Rempe, S.; Varma, S. *Chem. Phys. Lett.* **2010**, *485*, 1–7.
(10) Roux, B. *Biophys. J.* **2010**, *98*, 2877–2885.
(11) Yu, H.; Noskov, S. Y.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 20329–20334.
(12) Yu, H.; Ratheal, I.; Artigas, P.; Roux, B. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1159–1163.
(13) Rogers, D. M.; Rempe, S. B. *J. Phys. Chem. B* **2011**, *115*, 9116–9129.
(14) Roux, B.; Berneche, S.; Egwolf, B.; Lev, B.; Noskov, S. Y.; Rowley, C. N.; Yu, H. *J. Gen. Physiol.* **2011**, *137*, 415–426.
(15) Pedersen, C.; Frensdor, H. *Angew. Chem., Intl. Ed.* **1972**, 16.
(16) Dietrich, B. *J. Chem. Educ.* **1985**, 954–964.
(17) Bezanilla, F.; Armstrong, C. M. *J. Gen. Physiol.* **1972**, *60*, 588–608.
(18) Gouaux, E.; Mackinnon, R. *Science* **2005**, *310*, 1461–1465.
(19) Pratt, L.; Laviolette, R. *Mol. Phys.* **1998**, *94*, 909–915.
(20) Pratt, L.; Beck, T.; Paulatis, M. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: Cambridge, U.K., 2006.
(21) Rempe, S.; Pratt, L.; Hummer, G.; Kress, J.; Martin, R.; Redondo, A. *J. Am. Chem. Soc.* **2000**, *122*, 966–967.
(22) Rempe, S.; Pratt, L. *Fluid Phase Equilib.* **2001**, *183*, 121–132.
(23) Griffith, J.; Scheraga, H. *J. Mol. Struct. Theochem* **2004**, *711*, 33–48.
(24) Rempe, S.; Asthagiri, D.; Pratt, L. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1966–1969.
(25) Roux, B.; Yu, H. *J. Chem. Phys.* **2010**, *132*, 234101.
(26) Paterson, Y.; Nemethy, G.; Scheraga, H. *J. Sol. Chem.* **1982**, *11*, 831–856.
(27) No, K.; Nam, K.; Scheraga, H. *J. Am. Chem. Soc.* **1997**, *119*, 12917–12922.
(28) Cho, K.; No, K.; Scheraga, H. *J. Phys. Chem. A* **2000**, *104*, 6505–6509.
(29) Maksimiak, K.; Rodziewicz-Motowidlo, S.; Czaplewski, C.; Liwo, A.; Scheraga, H. *J. Phys. Chem. B* **2003**, *107*, 13496–13504.
(30) Makowski, M.; Liwo, A.; Maksimiak, K.; Makowska, J.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 2917–2924.
(31) Bethe, H. *Proc. R. Soc.* **1935**, *150*, 552–575.
(32) Guggenheim, E. *Proc. R. Soc. London A* **1935**, *148*, 304–312.
(33) Kirkwood, J. *J. Chem. Phys.* **1940**, *8*, 623–627.
(34) Bethe, H.; Kirkwood, J. *J. Chem. Phys.* **1939**, *7*, 578–582.
(35) Fowler, R.; Guggenheim, E. *Proc. R. Soc. London A* **1940**, *174*, 189–206.
(36) Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E.; Rempe, S. B. *J. Am. Chem. Soc.* **2004**, *126*, 1285–1289.
(37) Roux, B.; Nina, M.; Pomes, R.; Smith, J. *Biophys. J.* **1996**, *71*, 670–681.
(38) Gilson, M.; Given, J.; Bush, B.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
(39) Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399–10404.
(40) Deng, Y.; Roux, B. *J. Chem. Theory Comp.* **2006**, *2*, 1255–1273.
(41) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2009**, 2234–2246.
(42) Woo, H.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.
(43) Singh, U.; Kollman, P. *J. Comput. Chem.* **1986**, *7*, 718–730.
(44) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
(45) Gao, J.; Amara, P.; Alhambra, C.; Field, M. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
(46) Amara, P.; Field, M.; Alhambra, C.; Gao, J. *Theo. Chem. Acc.* **2000**, *104*, 336–343.
(47) Lin, H.; Truhlar, D. *J. Phys. Chem. A* **2005**, *109*, 3991–4004.
(48) Ihrig, A. C.; Schiffmann, C.; Sebastiani, D. *J. Chem. Phys.* **2007**, *135*, 214107.
(49) Roux, B. arXiv:1202.4939v1 2012.
(50) Yu, H.; Roux, B. *Biophys. J.* **2009**, *97*, L15–17.
(51) Chempath, S.; Pratt, L. R.; Paulaitis, M. E. *J. Chem. Phys.* **2009**, *130*, 054113.
(52) Gerardo Soto-Campos, G.; Corti, D. S.; Reiss, H. *J. Chem. Phys.* **1998**, *108*, 2563–2570.
(53) Woo, H.; Dinner, A.; Roux, B. *J. Chem. Phys.* **2004**, *121*, 6392–6400.

(54) Ban, F.; Kusalik, P.; Weaver, D. F. *J. Am. Chem. Soc.* **2004**, *126*, 4711−4716.
(55) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2009**, *131*, 8092−8101.
(56) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2010**, *132*, 2321−2332.