Published in final edited form as:
J Phys Chem B. 2011 July 28; 115(29): 9116–9129. doi:10.1021/jp2012864.

Probing the Thermodynamics of Competitive Ion Binding Using Minimum Energy Structures

David M. Rogers and **Susan B. Rempe**
Center for Biological and Materials Sciences, MS 0895, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

Abstract

Ion binding is known to affect the properties of biomolecules and is directly involved in many biochemical pathways. Because of the highly polar environments where such ions are found, a quantum-mechanical treatment is preferable for understanding the energetics of competitive ion binding. Due to computational cost, a quantum mechanical treatment may involve several approximations, however, whose validity can be difficult to determine. Using thermodynamic cycles, we show how intuitive models for complicated ion binding reactions can be built up from simplified, isolated ion-ligand binding site geometries suitable for quantum mechanical treatment. First the ion binding free energies of individual, minimum energy structures determine their intrinsic ion selectivities. Next, the relative propensity for each minimum energy structure is determined locally from the balance of ion-ligand and ligand-ligand interaction energies. Finally, the environment external to the binding site exerts its influence both through long-ranged dispersive and electrostatic interactions with the binding site as well as indirectly through shifting the binding site compositional and structural preferences. The resulting picture unifies *field-strength, topological control*, and *phase activation* viewpoints into a single theory that explicitly indicates the important role of solute coordination state on overall reaction energetics. As an example, we show that the $Na^+ \rightarrow K^+$ selectivities can be recovered by correctly considering the conformational contribution to the selectivity. This can be done even when constraining configuration space to the neighborhood around a single, arbitrarily chosen, minimum energy structure. Structural regions around minima for K^+- and Na^+-water clusters are exhibited that display both *rigid/mechanical* and *disordered/entropic* selectivity mechanisms for both Na^+ and K^+. Thermodynamic consequences of the theory are discussed with an emphasis on the role of coordination structure in determining experimental properties of ions in complex biological environments.

1 Introduction

Practical models of solute behavior commonly involve the concept of a division between local and non-local response of the solution medium. The ubiquitous application of these ideas in such diverse fields as inorganic chemistry of coordination complexes, ligand binding in biological systems, solvent cage-mediated chemical reactions, proton conduction, spectroscopic response of dissolved solutes, and other condensed phase problems is well known. The common element shared by these models is the (often hidden) assumption that the exact configuration of the local solvation structure can be regarded as fixed. Such an assumption appears directly at odds with the statistical mechanical picture asserting that “Anything that can happen will (with a probability related to the free energy required).”

Correspondence to: slrempe@sandia.gov.

Despite this apparent contradiction, the local configuration point of view has played an historically important role. Two foundational questions for such a theory are the calculations of reaction free energies for inserting or transforming ionic solute molecules in fixed chemical environments and its twin, transforming chemical environments with fixed solute. The former appear as solute selectivities for particular environments, while the latter appear as weighting factors for prediction of solute properties in heterogeneous solutions. Methods for calculating ionic stability constants will increase our capacity for design and mechanistic analysis of new materials.

The evolution of local configuration methods addressing these questions can be traced back to a pairing of the rigid rotor, harmonic oscillator approximation of quantum mechanics (QM) with continuum solvation models.[1] Next, methods for including continuum solvation directly into the quantum calculations[2,3] were developed. In order to progress further and correctly capture stabilization due to charge transfer or ligand field splitting effects, it becomes necessary to include explicitly the molecules comprising the first solvation shell. However the entropic consequences of choosing explicit ligand positions in this approach have remained poorly understood, as discussed recently in Ref. 4. Even more recently, it has become possible to incorporate a local treatment of quantum effects via a hybrid quantum mechanical/molecular mechanical potential energy function,[5,6] or to correct the results of a classical molecular dynamics simulation using high-level QM calculations on sampled structures.[7–10] Clearly, a careful consideration of the thermodynamic consequences of altering structural constraints is required in order to transition between traditional views on coordination complexes and recent selectivity arguments for labile coordination structures.

The apparent contradiction between thermodynamic analyses based on local configurations compared to mechanisms avoiding such structural references has also sparked debate. Several general mechanisms have been proposed by which an isolated binding site embedded in a dielectric environment (including vacuum) may be designed to achieve selectivity for a larger potassium over a smaller sodium ion. A structureless, statistical *field-strength* viewpoint can be constructed by analyzing changes in ion selectivity induced by changes in the balance of ion-protein, ion-water, water-water, water-protein, and protein-protein interactions.[11] The resulting mechanism emphasizes shifting chemical interaction strengths to control selective ion binding. However, the balance between these effects is not intuitive and must be arrived at from separate analysis of each binding site composition. This analysis leads to strongly selective sites in only a handful of cases.1 A structureless statistical viewpoint can lead to emphasis on a particular energetic contribution (for example the dipole moment contribution from carbonyl chemistry[13]) or the importance of binding energy fluctuations.[14] These may also be relevant only in specific circumstances.

Another set of general mechanistic arguments for designing selective ion binding sites have been constructed based on a local configuration viewpoint. While this view shows the importance of *local chemical structure*, it encounters the opposite difficulty of rigorously and explicitly including application-specific long-range environmental effects. For example, a local configuration approach has shown that both water and carbonyl chemistries can be potassium selective if equivalent coordination constraints are imposed and that the absence of such constraints annihilates selectivity.[15] But applying this result in practice requires computing the external field determining the binding site composition and geometry[16,17] and presents a significant challenge. Qualitative models of the external field used recently in combination with minimum energy binding site models have shown the important role of a

[1]Of the 1077 binding sites considered in Ref. 12, the average selectivity is near zero with a standard deviation of about 2 kcal/mol. While strongly K^+ selective sites were correctly identified by weak ion-ligand attraction and strong ligand-ligand repulsion, a similar comparison for Ba^{2+} selectivity was notably missing.

local pocket characterized by decreased external dielectric in muting the free energy cost for concentrating ligands into a binding site. Such local environments provide a source of constraint on binding site composition and structure that shifts the binding site equilibrium toward more selective states, characterized by over-coordination or specific cavity size. Conversely, an environment characterized by higher dielectric imposes a free energy cost for concentrating ligands into a binding site. It can thus produce ion binding architectures like those preferred by each ion in bulk water, reducing selectivity.[18–20] While qualitative models of the environment have lead to general hypotheses for tuning selectivity in binding sites with externally specified ligand chemical potential, detailed analysis of the environments around specific binding sites still proves difficult.

Despite the challenging requirement for computing the external influence on binding site composition and geometry, the local configuration method has been applied in several quantitative studies. Consideration of minimum energy ion-water clusters in gas phase has led to highly accurate free energies of aqueous hydration for several ions.[21–23] Statistical configurational sampling in all-atom *ab initio* molecular dynamics simulations has demonstrated the overall thermodynamic equivalence between local structural and non-structural approaches to ion solvation in homogeneous water environments.[24,25] The work described above[18] represents initial efforts in method development and QM-based application. In a joint article,[26] several groups have outlined a rigorous and computationally feasible method for extending these results to inhomogeneous phases such as protein binding sites, surfaces, and solution interfaces.

To permit investigations of the behavior of coordination complexes, we aim to resolve the above issues in applying local configuration methods. The simplest path forward is to separate out the contributions of field strength, local chemical structure, and environmental influences in a rigorous way. In this report we provide such a decomposition, in the process achieving a bird's eye view of coordination complex formation and solvation through the use of an appropriate thermodynamic cycle. This allows us to identify the influence of each factor on the interplay of structure and energetics that determine chemical design principles and solution behavior of metal coordination complexes in inhomogeneous environments.

1.1 Force Field Considerations

As a preliminary step for the computational study of ion coordination structures, it is necessary to verify the molecular energy function used for ion-ligand clusters. Traditional fixed charge force field models fail in this regard,[27] consistently over-estimating the number of ligands able to occupy cation-water first solvation shells.[28] In this report, we employ the AMOEBA polarizable force field.[29] This force field has been shown to capture the structural and energetic features of principal ion-water coordination structures.[30]

Over-crowded inner shell structures typically observed using traditional force field models can lead to qualitatively different conclusions about structure *vs.* function relationships. For an example, consider the set of gas phase minima containing waters inside the first solvation shell of Na^+ and K^+ ions. For a compromise between the first minimum of the respective ion-oxygen radial distribution functions, we count only configurations with water oxygens closer than 3.1Å to the ion. One of the most widely used non-polarizable force field models for ion-protein interactions, CHARMM,[31] finds [1,1,1,1,1,9,4, and 3] different minima for Na^+-water clusters with 1–8 waters (respectively), whereas AMOEBA finds [1,1,1,1,4,4,0, and 0]. The corresponding numbers for K^+-water clusters with oxygens inside a 3.1Å cutoff are [1,1,1,1,3,9,6, and 7] (CHARMM) *vs.* [1,1,1,2,5,8,4, and 4] (AMOEBA). In addition, there is accumulating evidence for the argument that ligand polarization and possibly even charge transfer are required for a correct description of ion-ligand interaction potential energy surfaces. These effects are important for competitive ion binding thermodynamics,

especially when asymmetric cluster structures are involved.[10,32–34] Such effects are best described using a QM approach that explicitly treats valence electrons responsible for polarization and charge transfer. The quasichemical theoretical approach described below makes this possible.

1.2 Division into Local Configurations Using Quasichemical Theory

A thermodynamic approach for incorporating QM appears in quasichemical theory (QCT). QCT shows the connection between ion-coordinated and unrestrained local configuration points of view by exactly defining the requirement for formation of a solvation structure and its thermodynamic consequences. This connection is formally incorporated into Gibbsian statistical mechanics by introducing constraints on the allowed system phase space:

> C_n: The subset of allowed system configurations is limited to those in which a molecular complex $A{\cdot}X_n$ is formed.

The ambiguity of "the complex is formed" must be resolved in practice with an indicator function, $I(x;C_n)$, which is one whenever a particular point in phase space satisfies specific local structural requirements, and zero otherwise.

This structural division bears a close resemblance to the Stillinger-Weber inherent structure model for liquids.[35,36] In that model, a similar constraint was formulated for each global minimum of the liquid. Thermodynamic properties of the entire system could then be written as sums over these structural states and approximations used to estimate partition functions near energy minima. Here instead we define a constraint for each coordination complex of interest. The present definition does not require the presence of a potential energy minimum. Applying this division to flat potential energy surfaces gives a conceptually simple method for calculating the entropic consequences of choosing explicit ligand positions. This generalization is also appropriate for discussing chemical reactions occurring in arbitrary, inhomogeneous environments due to the natural appearance of the environmental potential of mean force when calculating partition functions with coordination constraints.

The local, structural division of QCT permits a discussion of solute properties in terms of alternative coordination states. Each coordination state has an intrinsic set of ion selectivities and spectroscopic properties given by a constrained average. External constraints (for example protein conformational preferences, bulk system composition, applied membrane electric field or surface tension, pressure, *etc.*) determine the overall equilibrium between alternate states in the form of a probability for formation of each state. For an example, we may choose a definition of C_4, call it $C_4(2.3)$, selecting only states where exactly two chlorides and two water oxygens are within 2.3Å of a pre-defined center. This condition is appropriate for a distinguishing states of a platinum coordination complex in solution.[37] Alternately, we may choose a definition appropriate for a sodium ion coordination complex, say C_4(tet), which is satisfied when four water oxygens are arranged tetrahedrally (with some RMSD tolerance). Or we may choose a constraint appropriate for the S2 site of the celebrated potassium-selective ion channel protein, KcsA,[38] where 8 carbonyl oxygens are located within 3.1Å of a K^+ ion. In any case, the observed system properties are a function of the averages conditional on the formation of each C_n and the set of probabilities $P(C_n)$ in an obvious way.

At this point, the non-structural thermodynamic and local configuration viewpoints may diverge for the following reason. Some ionic binding sites display a strong tendency toward local order (for example a solid crystal), whereas others are not subject to strong environmental constraints and become disordered. Associating the former with a single minimum energy structure seems natural but need not be the only recognized type of

structural order, while the latter disordered state may be, but is not necessarily, characterized by a conformational space containing many minima. Therefore a given definition of a coordination state may contain multiple local minima and be realized with high probability without contradiction. Practical investigations usually involve calculating the relative energetics and properties of several coordination states. These need not contain only one minimum energy structure, but instead represent a partitioning of phase space into alternative regions of ion-complexed and uncomplexed configurations as appropriate to the experiment at hand.

1.3 Approximations to Quasichemical Theory

Arbitrary divisions of configuration space are allowed through the definition of C_n. A consequence is that some region of configuration space can always be found that dominates the thermodynamic average by containing essentially the whole region of structures sampled during a given process (that is, any set of connected states in a thermodynamic cycle). Although not necessary in the present framework, the rigid rotor, harmonic oscillator (RRHO) approximation is often employed in QM thermodynamic calculations. However, using this approximation requires making a rigorous connection between coordination *states* and coordination *structures* by verifying that the specified constraint contains a single minimum energy structure and is well described by a harmonic expansion of the potential energy surface. A primitive approximation to quasichemical theory (QCA)[21–23,39] assumes that a stiff coordination structure is formed with high probability, and accordingly makes both the RRHO and thermodynamically dominant, or maximum term, approximations.

Despite the many advantages of the structural view outlined above, there exist serious criticisms of the idea that constrained averages could yield a correct free energy.[40] The argument is that any given coordination state, C_n, may have multiple minimum-energy structures for its solvation shell. The existence of multiple minima invalidates free energy calculations based on a single structure (that is, the RRHO approximation), and therefore minimum energy local solvation structures cannot play any role in modern statistical mechanical theories. As an example, Ref. 40 considered the definition $C_n(3.5)$, where the ion-water complex is taken as formed and the condition satisfied whenever n water centers are within 3.5Å of an ion fixed at the origin. This condition omits consideration of ligand structure inside this constraint (that is, whether or not the waters directly coordinate the ion in a chemical sense). They proceeded to test the validity of a simplistic approximation to solvation free energy that neglects thermal and entropic contributions, $\Delta E_{\min} \approx \Delta G$, and found a deviation of 1–2.5 kcal/mol. Related problems were noted when considering minimum energy structures in complicated biological environments, the worst of which is the apparent unavailability of a simple answer as to which bound ion should be used during the minimization process for studying competitive Na^+ / K^+ binding in the KcsA selectivity filter (where most authors have chosen the endogenous ligand).

In view of these recent studies, confusion clearly exists about the thermodynamic consequences of placing constraints on configuration space. The results of this report will show conclusively that the complex formation free energy can be calculated based on *any* desired constrained state when the free energy for binding site rearrangement is included. Until this point the free energy cost for placing “microdroplet” half-harmonic constraints (chosen, for example, to match the upper bound of binding site fluctuations) has been ignored by some authors. The reason this omission usually leads to correct qualitative results is that there is a negligible cost for constraining a system to the most probable biological state. However, because the cost for constraining the binding site conformation is incurred by the environment in order to produce a selective chelation complex, it is a critical element in understanding solutions nature has developed to solve the design problem.

By recognizing this constraint energy, QCT additionally produces meaningful conclusions about individual coordination (that is, minimum energy) structures. The free energy requirement for binding site rearrangement also answers the question of energy minimization protocol for simple Na^+/K^+ selectivity questions – the coordination geometry arrived at must closely resemble the native environment produced by the channel selectivity filter. This requirement is, in fact, achieved by four glycine dipeptide strands when energy minimized in the presence of K^+, but not Na^+.[18]

This report begins by introducing a statistical mechanical notation for thermodynamic cycles that facilitate discussion of ion binding thermodynamics. We then give two thermodynamic cycles for calculating solvation free energies based on QCT ideas that permit use of gas-phase QM calculations for cluster formation energies. The first is applicable to homogeneous solutions, and a small adaptation gives the second, which is applicable to non-homogeneous environments such as enzyme binding sites and spatially constrained potential of mean force calculations. We will use chemical equilibria to show the physical meaning of each term that appears. The resulting theory is applied to the problem of calculating binding site formation free energies. From our decomposition of the free energy of ionic selectivity, we find that 1) Individual coordination structures (that is, containing only one energy minimum) are natural chemical objects for carrying out inner-shell reactions; 2) *Each* structure reproduces the full cluster formation free energy when the probability of coordination structure formation is taken into account; 3) The approximation error of the maximum term method used in QCA is rigorously bounded by the disorder of the ion +binding site complex; and 4) Conditions can be found under which several useful approximations will hold, making the theory amenable to computation in ion-biomolecule binding sites.

The discussion will combine these ideas of chemical reaction constants into an insightful picture of the competing roles of intrinsic binding site structure and solute binding energetics in the determination of solvation free energies. An important conclusion is that local solvation structures are both intuitively and computationally simple physical objects of modern thermodynamic theory, appropriate for analysis of competitive ion binding in enzyme active sites.

2 Theory

Useful, intuitively simple, derivations of QCT can be expressed in terms of thermodynamic cycles.[26,34,41] To form a compact notation for such expressions, we need only a language for specifying the form of the partition function, Z, for each state in the cycle. Free energy changes between states A and B are then given by the conventional expression

$\beta\Delta F = -\ln\frac{Z_B}{Z_A}$. A natural alphabet for describing a particular partition function is the set of its constrained averages, such as energy ($\langle H\rangle$), volume ($\langle V\rangle$), or particle number ($\langle N\rangle$), as well as any constraints on the state space, for example the formation of a coordination state, C_n. For ease of exposition, the location and conformation of the solute, X, is taken as fixed throughout this report[2].

As an example, a coordination complex in a solution at constant temperature and pressure (state E of Fig. 1 and 2) has a partition function given by constrained average energy, $\langle H_N + H_X + \Delta H\rangle$, volume, $\langle V\rangle$, and coordination, C_n, yielding

[2]This condition should accordingly be included in all the partition function specifications, but is implicitly understood here.

$$Z(\langle H_N+H_X+\Delta H\rangle,\langle V\rangle,C_n)=\left(\prod_\alpha N_\alpha!\right)^{-1}\int_0^\infty dV\frac{\text{const.}}{V}\int dx^{3N}\int dp^{3N}\exp\{-\beta(PV+H_N+H_X+\Delta H)\}I(x;C_n). \quad (1)$$

Here, H_N and H_X are the Hamiltonians of the isolated N-solvent molecule system and the isolated solute, X, at a given point in phase space while ΔH represents the interaction energy, $\Delta H \equiv H_{N+X} - H_N - H_X$. The prefactor removes degenerate configurations to account for the indistinguishability of solvent molecules other than the ligand molecules specifically labeled by C_n. The last term in the integral is the indicator function taking on the value one or zero, depending on whether or not the coordination state is occupied.

For solute transfer reactions between environments, it becomes useful to invoke several of these systems containing differing chemical species and constraints. In this case, a slight abuse of the notation in Eq. 1 to center it around the identity and location of individual molecules re-casts it in familiar chemical terms. This is done by writing each molecule present in the calculation along with its physical location (that is, the "box" in a Grand Canonical Monte Carlo simulation), implying an average energy constraint in each box. Other constraints such as an average ($\bar{V}$, constant pressure) or fixed volume (V), constraints on local solvation structures, C_n, dipole moments, chemical potential, $\bar{N}$, *etc.* should be explicitly specified for each box. Thus the partition function of Eq. 1 could also be specified by

$$X(\bar{V},C_n)+N\,H_2O(\bar{V},C_n),$$

or, equivalently,

$$X\cdot(H_2O)_n(\bar{V},C_n)+(N-n)H_2O(\bar{V},C_0).$$

This notation provides a consistent and useful way to state free energy cycles involving coordination constraints as long as boxes and molecules are not created or destroyed along the way. These are understood as present in all steps of the process, though they may not contain any molecules or interact.

Using our notation, the reaction defining solvation free energy is,[26]

$$X(V_\infty^a)+N\,H_2O(\bar{V})\xrightarrow{\beta\mu_X^{ex}}X(\bar{V})+N\,H_2O(\bar{V}).$$

Here the starting position of the solute is formally a vacuum[3], and has been given a superscript to distinguish it as contained in a separate volume from (and thus non-interacting with) the constant pressure solution, $\bar{V}$.

2.1 Homogeneous QCT Process

We can proceed to write down a thermodynamic cycle mirroring the original formulation of QCT for solvation in homogeneous solutions.[42] The result is shown in Fig. 1. A direct, brute force, thermodynamic integration (TI) or free-energy perturbation (FEP) can usually be carried out to get the solvation free energy. It is usually not recognized that the results of

[3]Because the solute location is fixed, the partition function is just $\exp[-\beta E_X]$.

FEP calculations can be separated into local contributions, as shown in Fig. 1, to make conclusions about specific coordination states. An added benefit of the formulation is isolation of a chemical association step in gas phase, permitting efficient application and comparison with quantum methods that account for non-pairwise-additive effects such as polarization and charge transfer.[32] The intermediate stages are shown schematically by enclosing non-interacting molecules with dotted lines and indicating coordination and excluded volume conditions by shading. Each step has a useful chemical interpretation. Approximations used in practical calculations may also be expected to cancel partially between left (downward) and right (upward) arrows, which correspond to similar processes.

2.1.1 Ligand Extraction, (A→B)—Starting from the non-interacting system in the upper left, ligand molecules (L_n) forming the inner coordination shell are first removed from solution into separate, labeled, non-interacting volumes, V^{b_n}. The free energy cost is determined by their solvation free energies (from a standard state at density $V^{b^{-1}}$),

$$\beta\Delta\mu_{H_2O}=\ln\frac{N/\bar{V}}{1/V^b}+\beta\mu^{ex}_{H_2O}. \quad (2)$$

These quantities are usually available experimentally and have been tabulated for many important solvent force field models. Note that if the ligands were not individually distinguished in B, this step could be done in $\binom{N}{n}$ possible ways. In practice, it is more useful to label each ligand/solvent molecule in both state B and wherever the coordination condition C_n is applied, leading to a degeneracy factor change of $-\ln\binom{N}{n}-\ln n!$ in reaction A→B and corresponding addition of ln*n*! in step E→F (Eqns. 2 and 3). Taking the thermodynamic limit in the above, $\frac{N!}{(N-n)!\bar{V}^n}\rightarrow\rho^n_{H_2O}$ generates the solvent density, ρH_2O.

2.1.2 Binding Site Formation, (B→C)—The next contribution (B→C) comes from the likelihood that the given coordination geometry is formed in the absence of interactions with the solute ion (denoted by the zero subscript). Arranging the (now non-interacting) ligands into their inner-shell (IS) binding geometry, C_n, gives rise to an entropic contribution, $-n\ln V_{IS}/V^b$. This is because of the change of volume to $V^n_{IS}=\int dx^{3n}I(C_n;x^{3n})$ for purely translational constraints.

In addition, we can choose the inner-shell condition to restrict the surrounding solvent to lie outside the specified coordination complex. Logically, the indicator function $I(C_n)$ factors into a constraint on the ligand geometry times a "non-occupancy" constraint C_0 on the surrounding solvent. The corresponding free energy comes from

$$\wp_0(C_0)=\frac{Z((N-n)H_2O(\bar{V},C_0))}{Z((N-n)H_2O(\bar{V}))},$$

the probability that the rest of the solution obeys the coordination condition. This term is the origin of packing contributions to the hard-sphere solvation free energy,[41] and depends on the exact constraint geometry.

Note that the above choice to include the solvent cavity is made throughout this report in order to make a consistent comparison to the equations presented in Ref. 26 (corresponding

to the thermodynamic cycle in section2.2). However, if the constraint C_n ignores the configurations of the surrounding fluid, then the $-\ln\mathcal{P}_0(C_0)$ packing term can be moved to step D→E to make a more naturally defined $\beta\mu^{ex}_{X\cdot(H_2O)_n}$. In this way, figures 1 and 2 make use of a definition for the solvation free energy of the cluster that excludes the packing term,

$$\beta\mu^{ex}_{X\cdot(H_2O)_n} \equiv \beta\mu^{ex,*}_{X\cdot(H_2O)_n} - \ln\wp_0(C_0).$$

2.1.3 Chemical Association / Complex Formation (C→D)—Because the ligands are not interacting with the rest of solution at this stage, the chemical binding step (C→D) may be carried out in vacuum. The chemical interpretation of the free energy for this step remains well-defined because of the constraint on the locations of the ligands. In addition, this step usually produces the dominant contribution to the total thermodynamic cycle and thus carries the most relevance to selectivity of like-charged ions in the given coordination state.[18] Importantly, the rigorous separation of the solvation process has made this term amenable to high-level ab-initio calculations.[43]

Since state B contains n non-interacting ligand molecules, each in a standard state volume V^b, we may alternately envision forming a cluster from these molecules by immediately combining their volumes into a new container, C_n. The path directly from B to D thus represents a gas-phase equilibrium constant,

$$K_{X\cdot(H_2O)_n}(C_n)=\frac{Z(X(V^a_\infty)+nH_2O(V^a_\infty,C_n))}{Z(X(V^a_\infty))Z(H_2O,V^b)^n}.$$

It should be noted that this equation explicitly shows how the magnitude of the inner-shell volume V_{IS} vanishes by construction in the complete cycle and is not essential to solvation free energies.

2.1.4 Solvation of the Complex (D→E) and Spontaneous Complex Formation (F→E)—After the complex is formed in D, it is re-introduced into the rest of solution, paralleling the first step in the cycle. Because the solute location and coordination structure are fixed during D→E, only small changes in the ligand distribution within the coordination state C_n are expected. Finally, the constraint on the solute binding geometry is removed in the presence of solution in step E→F, with free energy corresponding to

$$\begin{aligned}\frac{Z(X(\bar{V})+n\,H_2O(\bar{V},C_n)+(N-n)\,H_2O(\bar{V},C_0))}{Z(X(\bar{V}),N\,H_2O(\bar{V}))}&\\ =\frac{(N-n)!^{-1}\int\int I(C_n)e^{-\beta(H_N+H_X+\Delta H)}dx^N dp^N}{N!^{-1}\int\int e^{-\beta(H_N+H_X+\Delta H)}dx^N dp^N}&\\ =n!\begin{pmatrix}N\\n\end{pmatrix}\int I(C_n)\wp_X(x^N)dx^N&\\ =n!\wp_X(C_n)\equiv\wp_X(C_n').&\end{aligned} \quad (3)$$

Here we have defined the coordination condition C_n' corresponding to C_n, but satisfied for any of the $n!$ ligand orderings, and noted that its probability is $n!$ times larger.

These probabilities are observables from an equilibrium system at state F. They indicate the relative propensity for chemical association and formation of each type of complex in that

solution. For example, suppose a metal ion is introduced between states A and F, and the solvent for both states is composed of an aqueous mixture of two different chelation compounds. The metal excess chemical potential, A→F, will be composed of contributions from both possible chelation reactions. We label each coordination complex by separate a geometric constraint, giving rise to two cycles (because of the two different bound states E). Since the individual paths A→B→C→D→E→F must give the same free energy as the overall reaction A→F, the proportion of metal bound to each complex (E→F) is completely explained by variations in the binding energetics for each coordination complex, A→E.

For comparison with prior literature, the traditional gas-phase partition coefficient appearing in step B→D is written in terms of indistinguishable ligands (C'_n),[43–46] so that it includes a degeneracy term for the coupled state.

$$K_{X\cdot(H_2O)_n}(C'_n)=\frac{Z(X(V^a_\infty)+nH_2O(V^a_\infty,C'_n))}{Z(X(V^a_\infty))Z(H_2O,V^b)^n}=K_{X\cdot(H_2O)_n}(C_n)/n!$$

2.2 Non-homogeneous QCT

Figure 2 illustrates a recently developed variation of QCT applicable to non-homogeneous solutions.[26] Similar in spirit to Fig. 1, Fig. 2 contains at its base a cluster formation free energy. Formation of the cluster can be carried out in the presence of the surrounding medium, as described by previous authors[26] ($\mu^{ex}_X(n)$), or in an isolated environment $\mu^{ex}_X(n, clust)$ – analogous to $\Delta G(C_n, \text{QCT})$ of Fig. 1.

The principle difference from previous versions of QCT is the change in the ordering of ligand removal and pre-arrangement of the coordination structure, C_n, so that only steps B and C are altered. In A→D of Fig. 1, the ligands are individually removed from a homogeneous fluid, grouped together in the (non-interacting) gas phase, and then all intermolecular interactions are turned on simultaneously. Steps A→D of Fig. 2 first form the coordination structure, C_n, then remove its interaction with the surrounding medium to bring it into the gas phase before turning on solute-to-ligand interactions. The present cycle therefore separates out thermodynamic contributions from ligand-ligand interactions implicitly in step A→B, while Fig. 1 counts both ion and ligand interactions in C→D.

An important thermodynamic component is included in the first step of both cycles; namely, a concentration contribution arising from the probability of finding the ligands in a given neighborhood around the solute. Thus the solvent density dependence of cycle 2 contributes to this probability, $\wp_0(C_n)$, which will be identical to the concentration term of cycle 1 $(-\ln\wp_0(C_0) - n\ln\frac{V^{IS}}{\bar{V}} - \ln\frac{N!}{(N-n)!})$ in the limit of negligible ligand-ligand strain force.

3 Methods

In order to achieve converged $Na^+ \rightarrow K^+$ ion mutation free energy estimates, free energy perturbation (FEP) calculations were carried out between states A→F (Fig. 3) using the AMOEBA force field by linearly interpolating the ionic Lennard-Jones radius, well depth, and polarizability – as was done in Ref. 30. We used 4 steps in-between Na^+ and K^+ parameters (for a total of 6 states). Hamiltonian replica exchange was utilized to improve sampling for these calculations. Each 6-state simulation was run for 1.5 hours using a version of Tinker[47] modified for parallel replica exchange and was able to collect at least 2000 samples spaced 1 ps in time.

Samples from all FEP intermediates were tested for satisfying each coordination constraint. This provided a direct estimate of the coordination probabilities (D→F), as well as a set of structures used to calculate conditional free energy differences ($D_{Na^+} \rightarrow D_{K^+}$ of Fig. 3) between (and expectation values at) constrained states. Data sampled from all FEP intermediates were efficiently combined using weights provided by the multi-state Bennett acceptance ratio (MBAR) technique.[48] In addition, one set of duplicate simulations were carried out in order to verify the convergence of free energy estimates to within the stated error calculated by MBAR. In all cases, the difference in free energy estimates came in very close to or below the stated uncertainties, and the average between both final estimates is reported in Fig. 6. The error-bars shown are one standard deviation, as estimated from the first simulation.

3.1 Structural Classification

All ion-water cluster structures were classified according to a set of minimum energy structures for each n in the following way. The ion position was fixed at the origin, and hydrogen positions for each water were averaged to yield $2n$ comparison points. Root mean square deviations (RMSD-s) were computed between the test and reference structures for all possible water permutations of the test structure. The total number of points used for normalizing the RMSD was $2n + 1$. Test structures were clustered to the reference structure with the lowest RMSD, and labeled by that RMSD value. The coordination condition thus arrived at is,

> $C_n(x;c_i,\mathrm{RMSD};\{c_j\})$: The structure x has lowest RMSD to c_i among all c_j and that RMSD is less than the cutoff value.

This coordination condition depends on the set of reference structures used for comparison. We arrived at a set of minima for each ion by energy minimizing and superimposing each frame from our initial 1 ns dynamics calculations. Minima with an RMSD of greater than 0.3Å were added as new reference structures, while only the lowest energy structure was retained if two fell within 0.3Å RMSD. The introduction details the number of minima found within 3.1Å. The population of structures closest to each minimum was then determined from the samples taken at thermal equilibrium and the minima with thermally averaged oxygen-ion distance larger than 3.1Å removed from further analysis. The set of sodium minima were mapped to the set of potassium minima to find corresponding structures with reasonable RMSD and which had occupancy above 5% for at least one ion. The set of potassium minima emerging from this process were used for the final characterization presented in Fig. 4 and (a), left, and (a,b), right, of Fig. 5. There was a single minimum for n=1–4, while n=5–8 had two (with three Na^+ minima mapped non-uniquely to two K^+ structures for n=6).

4 Results and Discussion

Although they participate in much more complicated thermodynamic cycles, the primitive objects of our theory are simple minimum energy structures, along with the region of conformational space in their immediate vicinity. If the subset of allowed system configurations is limited to those in the immediate vicinity of a minimum energy structure (containing n inner-shell ligands with defined geometry), the resulting free energy change for replacing a Na^+ ion with K^+ can be interpreted in a physically meaningful way. In this section, we use C_n to denote this neighborhood and $\Delta F(C_n)$ to represent the constrained $Na^+ \rightarrow K^+$ free energy change. Comparisons between $\Delta F(C_n)$ for different choices of C_n allow mechanistic arguments based on a spatial partitioning of the free energy.

To illustrate these ideas, we have chosen to break out the minimum energy structure contributions from the formation free energy of an ion-water cluster used in typical

selectivity arguments.[12,17,18] We will show step by step how a total selectivity (A→F) arises from consideration of the occupancy (D→F) and conditional free energy (C→D) of *any* coordination constraint through the appropriate thermodynamic cycle. Naturally, coordination structures that occur with higher probability prove to be the most chemically meaningful and computationally accessible, while separating out the conditional free energy under each constraint can show the structural design principles responsible for selectivity. The diversity of selectivities obtained in this way extends our insight well beyond a thermodynamic integration approach that ignores structural information.

4.1 Selectivity of an Isolated Binding Site

To show the essential features of the proposed calculation, we have computed the $Na^+ \rightarrow K^+$ mutation free energy for isolated ion-water clusters with $n = 1–8$ waters. The binding site consists of n waters whose oxygen atoms (at distances r_O from the origin) are restrained to within a radius of $R_{max} = 3.1$Å from the ion. This radius is an appropriate upper-bound for the inner-shell of both Na^+ and K^+. The ion is fixed at the origin with a flat-bottomed, harmonic potential

$$U_R(r_O) = \frac{k}{2}(r_O - R_{max})^2 I(r_O \geq R_{max}), \quad (4)$$

with force constant k=200 kcal/mol·Å^2. The indicator function, $I(\cdot)$ is zero or one as defined in the introduction. This system has the physical interpretation of ion selectivity in a water cage embedded in a nonpolar environment.

The thermodynamic cycle analogous to Fig. 1 for this system is shown in Fig. 3. Note that this cycle shares the gas phase coordination complex step, $X \cdot (H_2O)_n(V_\infty^a, C_n)[D]$, in common with the cycles illustrated in Figs. 1 and 2. It is universally arrived at via turning on or mutating ion-ligand interactions in the presence of a coordination constraint and is the hallmark of quasichemical theory. The results from this system are therefore useful as an indication of the local contributions to selectivity in diverse ion-binding environments – with the external potential and ligand density deciding the weighting between minima for each system.

Fig. 3 also shows the intrinsic connection between QCT and classical equilibrium theory in that step [B] forms the basis for arbitrary gas-phase cluster formation reactions. The constants $K_{X\cdot(H_2O)_n}$ and K_{nH_2O} are the familiar gas-phase equilibrium constants for the formation of ion-water and pure water clusters from an ideal gas standard state. The new constant, $K_{X\cdot(H_2O)_n}(C_n)$ is identical to the classical cluster formation constant except for the explicit specification of the integration region via a constraint that the ligands must satisfy. This step is essential for creating a *local* theory.

Since the choice of the coordination constraint, C_n, is arbitrary in the thermodynamic cycles above, we should expect any partitioning of configuration space to give the same total answer (A→F). Variations in the probabilities ($P_X(C_n)$) for forming each structure C_n indicate the intrinsic conformational preferences of the ion-water cluster (X=Na^+ or K^+), while variations in the conditional free energy, $\Delta F(C_n)$, give the ion selectivity belonging to a specific coordination sub-state. For our particular criteria of RMSD from a set of reference structures, labeled by c, $P_X(C_n(c,RMSD))$ is the occupancy of a region of conformation space around each reference structure. These occupancies indicate their relative importance in determining the overall free energy of cluster formation.

Fig. 4 shows the components of the free energy differences corresponding to D→F and between D(Na^+) and D(K^+) of Fig. 3. The constraint, fully explained in the Methods section,

requires that the structure be closest in RMSD to a reference structure, and have an RMSD within the cutoff. Successively larger cutoff values scale the allowed region about the minimum from a single, minimum energy structure on the left (determined in the K^+ Hamiltonian), to the whole set of structures closest to the reference on the right. For n = 1–4, only one minimum energy structure was used, so that the right side of the plot is only constrained by ligand number according to Eq. 4. For any value of the cutoff, it can clearly be seen that when the probability for coordination state occupancy is properly taken into account, the overall ionic selectivity is recovered exactly. This provides a validation of the thermodynamic cycle (Fig. 3). Some regions of the plot are not shown, however, due to insufficient sampling. For example, no structures closer than 0.2 RMSD to the minimum were visited during 2 ns of sampling for the n = 6 structure. Because the free energy for placing a constraint is the negative log of its probability of occurrence, this is expected for constraints costing over 5 kcal/mol.

The AMOEBA force field model calculates the free energy cost for replacing an aqueous Na^+ ion by K^+ at 17.3 kcal/mol,[30] in agreement with experiment.[49] As observed previously,[15] the lower cost of replacing Na^+ by K^+ in water clusters of size n = 1–8 implies all of these water clusters are barely selective for K^+ (*vs.* bulk water). Because of this weak selectivity for n=6–8, relatively small conformational constraints imposed by the environment on the metal coordination structure have the potential to shift the equilibrium in favor of either ion. This system can thus succinctly illustrate the major features of the selectivity debate.

4.1.1 Selectivity of a Minimum Energy Structure—First, for a single, fixed, ligand configuration isolated from the surrounding medium, it can easily be seen that ion selectivity (C→D) is determined solely by the simple, chemically intuitive ion-ligand interaction strength. This energy change on replacing Na^+ by K^+ appears on the far left of each plot as $\Delta F(C_n)$ at zero RMSD. In a biological environment, we must also add the response of the protein/membrane medium surrounding the binding site ($\mu^{ex}_{X\cdot(H_2O)_n}$). In most cases this contribution can be easily approximated by a mean-field, Poisson-Boltzmann type model and fits into the picture of ion-external environment interaction strength.[18] However, if a large-scale conformational change in the external medium occurs on binding, then the environmental contribution to the free energy change between Na^+ and K^+ occupied states would include this transition and become an interesting object for computation in its own right.

4.1.2 Structure, Energy, and Entropy—Second, allowing the binding site to explore a small region around the reference structure moves us to the right on the $\Delta F(C_n)$ line. Here, we begin to see entropic contributions arising from the shape of the potential energy surface near the reference structure. This entropic contribution effectively penalizes ions requiring more rigid coordination states. For n=1, it is apparent that the free energy cost, $\Delta F(C_n)$, is increasing with the cutoff because the K^+-water energy surface is more rigid around its own minimum than the Na^+-water surface. This same feature ($\Delta F(C_n) < \Delta F$) requires the free energy cost for imposing the constraint $C_1(RMSD)$ on K^+ to be less than that for imposing it on Na^+. The situation persists until the constraint is large enough to envelop the Na^+-water minimum (around 0.3Å RMSD) – at which point the constraint energy becomes effectively zero. We should therefore expect this trend for all constraints based on K^+-water minima. Unexpectedly, this trend appears to be violated for n=4. Here, at R=0.7Å, the constraint envelope is actually more easily satisfied by Na^+'s 4 waters than K^+, resulting in a higher free energy when the ligand configuration is constrained to this region. The explanation is that the K^+-4 water structure at this RMSD is more diffuse than for Na^+.

4.1.3 Absence of Structural Information—Finally, at the far right of the plot, minimal constraints are present, allowing the binding site to respond naturally to a change in its bound ligand. At this point minimal structural information is available (that is, only the "closest" matching reference structure), and it is no longer informative to explain the selectivity differences in terms of structural preference. Instead, as it has been argued, the selectivity (not considering variations in R_{max} or n[16,19]) results from all contributions, and has been described as a combination of ion-ligand attraction and ligand-ligand repulsion.[12,13] However, we should not ignore the cogent information that these interactions give rise to preferences among compositional and structural states, C_n. Because each state participates in the full equilibrium at [F] (where no constraints are present), any decomposition is equally valid. However, each state we may define will have a specific propensity for appearing and a specific ionic selectivity that determines its relevance.

If we choose to define our states using the number of waters placed in contact with the ion (and further separating these into configurations near a set of minima), then we can immediately gain structural insight into the selectivity question. Above $n = 7$, the AMOEBA force field does not find any minimum energy structures for Na^+ retaining all waters within the cutoff. Furthermore, the plot for $n = 8$ shows the stabilizing effect of a minimum for K^+, which is itself K^+ selective (vs. bulk liquid water). The present data lends support to the current consensus view that for clusters constrained to have $n = 8$ inner-shell ligands with small dipole moments, the $Na^+ \rightarrow K^+$ mutation free energy is less than bulk and the binding site selective for K^+ because Na^+ prefers a more ordered binding site with closer ligand oxygens not afforded by the high ligand density at large n. This conclusion is difficult to arrive at without structural reasoning since *ligand-ligand repulsion must act via altering the binding site conformational preferences*. Indeed, interpretation of selectivity differences through the mechanism of ligand field-strength, commonly tested by scaling/removing some set of ligand-ligand interactions, has ultimately relied on this mechanism.[50]

Further evidence is provided by examining the distribution of the water oxygens within the 3.1Å constraint radius. At $n = 8$, the most probable configuration for Na^+ is 5 waters inside 2.7Å and 3 outside. Within the set of such 5+3 configurations, the $Na^+ \rightarrow K^+$ mutation free energy is 20.9±0.1 kcal/mol – selective for Na^+ over bulk water by 3.3 kcal/mol. For K^+, the most probable configuration is 1 inside and 7 outside 2.7Å, where (for 1+7 configurations) the mutation free energy is only 11.3±0.1 kcal/mol – selective for K^+ over bulk by 6 kcal/mol. Thus, because the solvation shell of K^+ is pushed out relative to Na^+, the cost for mutating $Na^+ \rightarrow K^+$ decreases when internal structure is not imposed in the n=8 case. By reasoning in the absence of structural information, it would have not be possible to quantify these consequences of binding site architecture or understand them in the context of design.

4.2 Environmental Constraints Tuning Ion Selectivities

Because the set of coordination states were chosen based on distance from a K^+-water minimum in Fig. 4, it is not surprising to find a relatively lower cost for inserting K^+ near these special ligand configurations. This merely illustrates the structural, "snug-fit"[51] mechanism for ligand selectivity (combined with over-coordination at n=8). Accordingly, we should also expect that the ionic selectivity could be tuned in the other direction by an environment imposing the appropriate Na^+-centered constraint. Figure 5 shows the energy decomposition for $n = 4$ clusters (left) and $n = 6$ clusters (right). The upper panels give results for the K^+-centered constraints as described previously, while the lower panel for each (b,left and c,right) are re-calculated based only on distance from the most populated Na^+-water minimum. As expected, selectivity can also be achieved by structurally constraining the binding site near the appropriate Na^+-favoring geometry, even increasing the $Na^+ \rightarrow K^+$ cost above that of bulk (thus switching the binding site to Na^+-selective).

Direct structural control of the binding site conformation has been invoked to explain the K^+ selectivity of valinomycin[20,50] and the Na^+ selectivity of the LeuT NA2 site.[52]

Although it is possible to design a metal chelator by enforcing rigidity, this is not always the route taken by nature. As discussed in Ref. 15, constraining the environment around a binding site to alter the probability of forming a coordination structure $\mathcal{P}_0(C_n)$ requires the protein to incur a free energy cost during synthesis or folding because it is limiting its conformational freedom. Taking C_{site} to be the region of ligand conformational space allowed by the natural protein binding site, the free energy cost is dictated by the probability of forming C_{site} in a higher entropy mutant, $-\ln \mathcal{P}_{unconstr.}(C_{site})/\mathcal{P}_{constr.}(C_{site})$. Now assume the protein conformational space is further reduced to $C_n \subseteq C_{site}$ when bound to an ion. To the extent that C_{site} excludes ligand conformational space required by an alternate ion, the A→B cost may be directly translated into a selectivity gain by constraining the native state. The appearance of $-\ln \mathcal{P}_0(C_n)$ in A→B of Fig. 2 thus implies that the synthesis cost must be greater than or equal to the gain in selectivity. As a consequence of this energy balance, a thermodynamic cost is paid at the cellular level for maintaining ion binding sites with altered selectivities relative to the cheapest protein the cell could produce.

The discussion gives a quantitative basis to the general topological control viewpoint of a system designed to employ specific environmental constraints to achieve selectivity. Combined with the results above, this perspective shows that enforcing an 8-coordinate structure is highly desirable for excluding Na^+ from cation binding sites. Since Na^+ does not have stable 7 or 8-coordinate structures, the binding site may retain favorable flexibility while biasing the conformation well away from Na^+-selective regions. However, at lower coordination, the binding site must use a rigid/mechanical mechanism to constrain its size or shape in order to get away from structures favored by another ion.

The intrinsic conformational preferences of an isolated reference binding site including only an ion-ligand distance restraint determine the “base case,” and can be rationalized as arising from a combination of ion-ligand and ligand-ligand interactions. However, for definiteness, this idea must also be accompanied by a calculation of those conformational preferences. Ideally this would result in listing out individual reference structures, their relative populations, and how they may be influenced by the environment. Without such information it becomes difficult to apply structural reasoning consistently, as can be gleaned from the discussion in Ref. 50. By failing to address the thermodynamic consequences of altering structural constraints, recent discussions[53] have not improved this situation. If, for example, a hypothetical constraint is introduced on the positional fluctuations of the ligand atoms, its effect is to shift the binding site equilibrium toward more rigid structures *at a free energy cost to the environment*. The thermodynamic cycles introduced here place environmental constraint and selectivity contributions on equal footing, and unambiguously show the dilemma faced by nature of using all available constraints on binding site composition and structure for optimizing selectivity.

Ion binding site conformations may be restrained by a variety of interactions with the environment. As discussed at length in Refs. 20 and 17, these include conformational restraints on ligand number, position, and water accessibility. The net function of these restraints is to shift the ion-unoccupied and occupied coordination state probabilities $\mathcal{P}_0(C_n)$ and $\mathcal{P}_X(C_n)$ from their reference state, taken in this report as an isolated gas-phase cluster. If ligand exchange is possible, the reference states A and F should include variability among ligand numbers, n, at equilibrium with bulk solution.[16,20] In this case, the equilibrium between binding site compositions present at state F can be analyzed by considering steps A→D for each composition. These must combine (in the same way shown in the last section) to give the equilibrium propensities at state F for each coordination, $\mathcal{P}_X(C_n)$.

Longer-range interactions also play an important role in determining the distribution of ion-binding conformations by altering the penalty for ligand extraction and complex solvation.[18,19] For example, around 4 inner-shell waters (near the distance defined by the first maximum of the radial distribution functions) are preferred by both Na^+ and K^+ in higher, water-like, dielectric environments. However, at lower dielectrics, Na^+ prefers an inner-shell coordination of 6 waters, whereas the larger K^+ is most stable at $n = 8$. Other features of protein environments will also play a role in shifting the stability of each coordination structure. As an example, nearby charges that produce a net electric field will shift the energy levels (and Boltzmann weighting) of polarized cluster structures.

These important effects are not usually probed because of the computational difficulty of employing thermodynamic integration or FEP for changing binding site composition. However, these effects can become amenable to study with properly validated approximations.

4.3 Maximum Term Approximation

In certain circumstances, the strong ionic charge can give rise to highly rigid local coordination geometries. Because our thermodynamic cycles are invariant to the choice of C_n, it may make sense to limit investigation to the region around the most probable structure in this case.

As has been noted previously[42,43] it is possible to obtain an absolute solvation free energy using a single coordination state because only the coupled probability ($\wp_X(C_n)$) is required in the thermodynamic cycle (Fig. 1). Therefore, if the contributions A→E are obtained for each of a set of mutually exclusive and exhaustive states C_n, then the missing contribution from each is just the (necessarily negative) quantity $\beta\Delta F_{E\to F} = \ln\wp_X(C'_n)$. The most favorable $\Delta F_{A\to E}$ therefore comes from the C'_n with highest probability, $\wp_X(C'_n)$. As long as this highest coordination probability is close to unity, it is appropriate to utilize the maximum term approximation.

$$\beta\mu_X^{ex} \approx \min_{C_n}[\beta\Delta F_{A\to E}] = -\max_{C_n}\left[\ln\left(V^{b^n}K_{X\cdot(H_2O)_n}(C_n)\right) + n(\beta\mu_{H_2O}^{ex} + \ln\frac{N}{V}) - \beta\mu_{X\cdot(H_2O)_n}^{ex}\right] - \ln\wp_0(C_n) \quad (5)$$

Comparing the maximum-term approximation of Eq. 5 with Fig. 1, we see that the approximation error is given by the missing term, $\ln\wp_X(C'_n)$. Since the minimum over all states is taken in Eq.5, the approximation error is determined by the deviation of the most probable C_n from $\wp(C'_n) = 1$.

In previous studies using the maximum term approximation for bulk solvation,[21–23] the maximum term was taken as the minimum energy structure with lowest RRHO free energy. Relating this to our present results, it is appropriate to examine how the constrained free energy calculated via FEP converges to the complete free energy with increasing RMSD. From Fig. 4, it is evident that as the RMSD criterion is increased to the infinite value appropriate to the RRHO, the free energy becomes essentially exact. This occurs even for $n = 6$ and 8, where two low-lying, alternate minima are present. In the worst case for two minima, both could have equal probability, making the maximum error $k_BT\ln 2 = 0.4$ kcal/mol. For rigid metal coordination complexes or environmentally constrained protein binding sites, this difference should decrease according to the ordering of the ion+binding site complex.

Finally, we show in Fig. 6 the consequence of applying the maximum term approximation to neglect the $\wp_X(C_n)$ terms and the RRHO approximation for the free energy of complex

formation $\Delta F(C_n)$. Our results show that these approximations give reasonable selectivities (within 0.8 kcal/mol, achieved at $n = 2$) from a computationally trivial calculation up to $n = 5$, where several minima are present. Interestingly, the failure of the model at high n does not arise from errors in the maximum term approximation (since the error there is limited to 0.4 kcal/mol), but instead from roughness of the potential energy surface not captured by the harmonic approximation. Even very near a minimum, the approximate harmonic energy surface (minimum energy plus quadratic term using second derivatives computed with AMOEBA) shows noticeable deviation from the AMOEBA force field for clusters involving two or more waters (data not shown). This is due to unphysical stretching of bond and angle degrees of freedom by straightforward linear displacements along normal modes. Practitioners who use harmonic approximations of the potential energy surface anticipate this error.[54] Our previous studies have thus considered small ion-ligand distances for QCT analysis.[18,19,21–23] If this is not done, the problem of multiple, anharmonic minima arises, and less satisfactory results are obtained.[39] Anharmonic corrections,[55] internal rotors and other approximations involving redundant internal coordinates[56] may be used in the future to extend the harmonic approximation to larger clusters. However, the general success of the RRHO approximation seen in Fig. 6 may arise despite this issue if we assume a non-linear mapping exists between harmonic displacements and coordinates. In this case, the coordinate system will be slightly altered, destroying the linearity of harmonic displacements while the harmonic integral still converges to a similar answer.

5 Conclusions

We have presented thermodynamic cycles (Figs. 1 and 2) for computational analysis of competitive ion binding reactions from a chemical perspective that are appropriate for evaluation of free energies in complex heterogeneous and biological environments. These cycles give a clear interpretation to the physical forces contributing to ion selectivity as well as a well-defined scheme for carrying out computations. They have allowed us to prove rigorously that the ion mutation and conformational constraint free energies add to a constant value, independent of the choice of coordination geometry or constraint distance. In practice, several coordination constraints, corresponding to different pathways, $B \rightarrow E$, can be investigated simultaneously from unconstrained simulations at states A and F.

The physical picture gained is one of isolating individual coordination states and computing equilibrium contributions from each. Figure 1 also contains the unification of the ligand field strength, topological control, and phase activation viewpoints on general mechanisms of selectivity. Ligand-ligand interactions contribute uniquely to the gas-phase coordination entropy (B→C), while the external environment is primarily responsible for changes in solvent extraction (A→B) and hydration of the complex (D→E). Chemically specific ion-ligand interactions occur during complex formation (C→D), along with ligand-ligand interactions that effectively shift this energy for each coordination state. Remembering A→F is independent of the choice of coordination, we can mentally join this leg between all such choices. The balance in occupancy between coordination states (E→F) is then determined by the above considerations. Alternatively, ionic selectivities for a given coordination state can be simplified further by considering only D→F, as was done in Sec. 4.1. The topological control perspective then says that the environmental factors controlling complex solvation (D→E) and the topological constraints determining coordination state occupancy (E→F) completely determine the selectivity. Recall that selectivity is defined as the difference from mutation in solution ($F_A \rightarrow F_B$) to mutation in the constrained binding site ($D_A \rightarrow D_B$). This picture leads naturally to the more symmetric Fig. 2.

This picture unifies the most prominent views on the selectivity question in a way that would not be possible without consideration of individual structural states. Quasichemical

theory does this by showing the trade-off between alternative combinations of ligand chemistry, coordination number, cavity shape, rigidity constraints, and phase structure of the external environment, which are individually capable of altering ion selectivity.

One of the primary benefits of re-stating previous quasichemical theoretical developments[26] in thermodynamic cycle notation (Fig. 2) is that it immediately shows the physical manipulations and logical progression inherent in formulas designed to break out structural contributions to full solvation free energies. At the same time, it re-interprets the problem of mathematically stating a local spatial partitioning of the free energy by requiring instead an explicit statement for the intermediate states (see Sec. 2), eliminating the need for ancillary quantities such as $\mu^{\mathrm{ex}}_{\mathrm{XL}_0}$, the excess chemical potential of the ion under the constraint that no inner-shell binding is possible. Because of this, the right-hand side of Fig. 1 (equivalent to that of Fig. 2), succinctly states Eqns. 1–2 and 5–9 of Ref. 26. In addition, it becomes simpler to conceptualize minor changes to the ligand density term of the theory required for application to inhomogeneous environments. At the same time, it retains its original ability to incorporate rigorous QM free energy calculations on gas-phase coordination structures.

Thermodynamic cycles based on QCT have made it possible to decompose the energetic and structural contributions to ionic solvation free energies. Arguments based on this formalism have already been applied to deconstruct the physical forces contributing to the propensities for alternate coordination states in potassium channel selectivity filters.[18] The additive gas phase plus quasi-liquid plus concentration free energies of that work are here given by

$$\Delta G_{\mathrm{X}\cdot(\mathrm{H_2O})_n}(\mathrm{C}_n,\mathrm{QCT})+[\mu^{\mathrm{ex}}_{\mathrm{X}\cdot(\mathrm{lig})_n}-n\mu^{\mathrm{ex}}_{\mathrm{lig}}]+[-n\beta^{-1}\ln V_{\mathrm{IS}}\rho\mathrm{H_2O}]. \qquad (6)$$

As discussed in Sec. 4.3 the constant cavity formation contribution from $\mathcal{P}_0$ (C_0) does not change the relative free energies of alternate coordinations, C_n, while the variations of A→E (Eq. 6) among alternate C_n imply their relative populations, $\mathcal{P}_X(\mathrm{C}_n)$. Therefore such studies have not only predicted the relative propensities for specific binding site compositions and structures, but have also shown the structural design mechanisms utilized by nature to accomplish this tuning with minimal cost.

From Figs. 4 and 5 it is apparent that *any* choice of coordination state will lead to a correct estimate of the ion binding free energy when the full cycle is taken into account. It therefore stands to reason that if the energy landscape in the constrained region is well described by an RRHO-like approximation, then the constrained free energy can be calculated using an analytical integral. Since it is known that the RRHO approximation encounters increasing difficulty at larger cluster sizes, alternate empirical procedures may be constructed to estimate the amount of error in the harmonic approximation[57] as a function of ligand cluster size. In either case, performing the stratification of configurations into their closest minimum energy basin shows a clear path forward for applying further approximations for calculation of the chemical binding free energies $\Delta G(\mathrm{C}_n,\mathrm{QCT})$ or $\mu^{\mathrm{ex}}_{\mathrm{X}}(n,clust)$ or for applying information models[58] to packing terms A→B and E→F of Fig. 2.

The major differences between the original formulation of QCT for homogeneous solutions and the newer variant appropriate for non-homogeneous solution has been discussed in Sec. 2. Both share the ligand-constrained state [D], which is a hallmark of QCT theories. Even for n=0, this state has been usefully employed for calculations on intrinsically disordered binding sites.[34,41,57,59,60] Expanding the present formalism to include multiple ion occupancy and even protein conformational changes involves a trivial modification of the coordination constraint, C_n, to place conditions on the protein and all ions present. The same conclusions apply to solutes other than ions – for example, small molecule solvation and

biomolecular interaction. However, as more conformational transitions are constrained, the number of states also increases.

The free energy components in cycle 2, when based on small regions around single energetic minima, are key to resolving present debates on the mechanisms of ion binding and stabilization by protein environments. The ability to interpolate between single, minimum energy structures, and native-like unconstrained states for the binding site shows that the utility of $\Delta E_{\min} = \Delta F_{Na^+\rightarrow K^+}(C_n, RMSD = 0)$ lies in quantifying the ion-ligand interaction strength. This is a characteristic of individual coordination structures, and the primitive object upon which overall selectivities are built. This is not a separation into confinement and geometric contributions,[53] but rather a synthesis of the "caress of the surroundings" with the crowding of the ligands.[61] In contrast to previous speculations that including both field strength and topological control (among coordination states) contributions to ion selectivity would make the argument cumbersome,[62] this report has shown that both are indispensable for simultaneously explaining rigid/mechanical and disordered/entropic selectivity mechanisms. Similar hybrid viewpoints have also been reached by Refs. 50 and 16.

Acknowledgments

This work was supported, in part, by Sandia's LDRD program, and, in part, by the National Institutes of Health through the NIH Road Map for Medical Research. Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

1. Cramer, Christopher J.; Truhlar, Donald G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. Chem. Rev. 1999; 99(8):2161–2200. [PubMed: 11849023]
2. Miertus S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum a direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem. Phys. 1981; 55(1):117–129.
3. Cossi, Maurizio; Scalmani, Giovanni; Rega, Nadia; Barone, Vincenzo. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. J. Chem. Phys. 2002; 117(1):43–54.
4. Bryantsev, Vyacheslav S.; Diallo, Mamadou S.; Goddard, William A, III. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. J. Phys. Chem. B. 2008; 112(32):9709–9719. [PubMed: 18646800]
5. Vreven, Thom; Morokuma, Keiji. Hybrid methods: Oniom (qm:mm) and qm/mm. In: Spellmeyer, David, editor. Ann. Rep. Comp. Chem. Vol. volume 2. Elsevier; 2006 Oct. p. 35-51.
6. Lin, Hai; Truhlar, Donald G. Qm/mm: what have we learned, where are we, and where do we go from here? Theor. Chimica Acta. 2007 Feb; 117(2):185–199.
7. Wood, Robert H.; Yezdimer, Eric M.; Sakane, Shinichi; Barriocanal, Jose A.; Doren, Douglas J. Free energies of solvation with quantum mechanical interaction energies from classical mechanical simulations. J. Chem. Phys. 1999; 110(3):1329–1337.
8. Sakane, Shinichi; Yezdimer, Eric M.; Liu, Wenbin; Barriocanal, Jose A.; Doren, Douglas J.; Wood, Robert H. Exploring the ab initio/classical free energy perturbation method: The hydration free energy of water. J. Chem. Phys. 2000; 113(7):2583–2593.
9. Liu, Wenbin; Wood, Robert H.; Doren, Douglas J. Hydration free energy and potential of mean force for a model of the sodium chloride ion pair in supercritical water with ab initio solute–solvent interactions. J. Chem. Phys. 2003; 118(6):2837–2844.
10. Zhao, Zhen; Rogers, David M.; Beck, Thomas L. Polarization and charge transfer in the hydration of chloride ions. J. Chem. Phys. 2010; 132(1) 014502.
11. Bernèche, Simon; Roux, Benoît. Molecular dynamics of the KcsA K^+ channel in a bilayer membrane. Biophys. J. 2000; 78(6):2900–2917. [PubMed: 10827971]

12. Roux, Benoît. Exploring the ion selectivity properties of a large number of simplified binding site models. Biophys. J. 2010; 98(12):2877–2885. [PubMed: 20550900]
13. Noskov, Sergei Yu; Bernèche, Simon; Roux, Benoît. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. Nature. 2004; 431:830–834. [PubMed: 15483608]
14. Asthagiri D, Pratt Lawrence R, Paulaitis Michael E. Role of fluctuations in a snug-fit mechanism of KcsA channel selectivity. J. Chem. Phys. 2006; 125(2) 024701.
15. Bostick, David L.; Brooks, Charles L, III. Selectivity in K^+ channels is due to topological control of the permeant ion's coordinated state. Proc. Nat. Acad. Sci. USA. 2007; 104(22):9260–9265. [PubMed: 17519335]
16. Bostick, David L.; Arora, Karunesh; Brooks, Charles L, III. K^+/Na^+ selectivity in toy cation binding site models is determined by the 'host'. Biophys. J. 2009; 96(10):3887–3896. [PubMed: 19450462]
17. Bostick, David L.; Brooks, Charles L, III. Statistical determinants of selective ionic complexation: Ions in solvent, transport proteins, and other "hosts". Biophys. J. 2009; 96(11):4470–4492. [PubMed: 19486671]
18. Varma S, Rempe SB. Tuning ion coordination architectures to enable selective partitioning. Biophys. J. 2007; 93:1093–1099. [PubMed: 17513348]
19. Varma, Sameer; Rempe, Susan B. Structural transitions in ion coordination driven by changes in competition for ligand binding. J. Am. Chem. Soc. 2008; 130(46):15405–15419. [PubMed: 18954053]
20. Varma, Sameer; Sabo, Dubravko; Rempe, Susan B. K^+/Na^+ selectivity in K channels and valinomycin: Over-coordination *versus* cavity-size constraints. J. Mol. Biol. 2008; 376:13–22. [PubMed: 18155244]
21. Rempe, Susan B.; Pratt, Lawrence R. The hydration number of Na^+ in liquid water. Fluid Phase Equil. 2001; 183–184:121–132.
22. Rempe SB, Asthagiri D, Pratt LR. Inner shell definition and absolute hydration free energy of K(aq)+ on the basis of quasi-chemical theory and ab initio molecular dynamics. Phys. Chem. Chem. Phys. 2004; 6(8):1966–1969.
23. Asthagiri D, Pratt Lawrence R, Paulaitis Michael E, Rempe Susan B. Hydration structure and free energy of biomolecularly specific aqueous dications, including Zn^{2+} and first transition row metals. J. Amer. Chem. Soc. 2004; 126(4):1285–1289. [PubMed: 14746502]
24. Leung, Kevin; Rempe, Susan B.; von Lilienfeld, O. Anatole Ab initio molecular dynamics calculations of ion hydration free energies. J. Chem. Phys. 2009; 130:204507–204517. [PubMed: 19485457]
25. Jiao, Dian; Leung, Kevin; Rempe, Susan B.; Nenoff, Tina M. First principles calculations of atomic nickel redox potentials and dimerization free energies: A study of metal nanoparticle growth. J. Chem. Theor. Comput. 2010 in press.
26. Asthagiri D, Dixit PD, Merchant S, Paulaitis ME, Pratt LR, Rempe SB, Varma S. Ion selectivity from local configurations of ligands in solutions and ion channels. Chem. Phys. Lett. 2010; 485(1–3):1–7.
27. Varma, Sameer; Rempe, Susan B. Coordination numbers of alkali metal ions in aqueous solutions. Biophys. Chem. 2006; 124:192–199. [PubMed: 16875774]
28. Whitfield, Troy W.; Varma, Sameer; Harder, Edward; Lamoureux, Guillaume; Rempe, Susan B.; Roux, Benoît. Theoretical study of aqueous solvation of K^+ comparing ab initio, polarizable, and fixed-charge models. J. Chem. Theory Comput. 2007; 3(6):2068–2082. [PubMed: 21785577]
29. Ren, Pengyu; Ponder, Jay W. Polarizable atomic multipole water model for molecular mechanics simulation. J. Phys. Chem. B. 2003 Jun; 107(24):5933–5947.
30. Grossfield, Alan; Ren, Pengyu; Ponder, Jay W. Ion solvation thermodynamics from simulation with a polarizable force field. J. Am. Chem. Soc. 2003 Dec; 125(50):15671–15682. [PubMed: 14664617]
31. Beglov, Dmitrii; Roux, Benoît. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J. Chem. Phys. 1994; 100(12):9050–9063.

32. Varma, Sameer; Rempe, Susan B. Importance of multi-body effects in ion binding and selectivity. Biophys. J. 2010 in press.
33. Wick, Collin D.; Xantheas, Sotiris S. Computational investigation of the first solvation shell structure of interfacial and bulk aqueous chloride and iodide ions. J. Phys. Chem. B. 2009; 113(13):4141–4146. [PubMed: 19014185]
34. Rogers, David M.; Beck, Thomas L. Quasichemical and structural analysis of polarizable anion hydration. J. Chem. Phys. 2010; 132(1) 014505.
35. Stillinger, Frank H.; Weber, Thomas A. Hidden structure in liquids. Phys. Rev. A. 1982 Feb; 25(2):978–989.
36. Stillinger, Frank H.; Weber, Thomas A. Dynamics of structural transitions in liquids. Phys. Rev. A. 1983 Oct; 28(4):2408–2416.
37. Hush, Noel S.; Schamberger, Jens; Bacskay, George B. A quantum chemical computational study of the relative stabilities of cis- and trans-platinum dichloride in aqueous solution. Coord. Chem. Rev. 2005; 249(3–4):299–311.
38. MacKinnon R. Potassium channels and the atomic basis of selective ion conduction (nobel lecture). Angew. Chem. Int. Ed. Engl. 2004; 43:4265–4277. [PubMed: 15368373]
39. Roux, Benoît; Yu, Haibo. Assessing the accuracy of approximate treatments of ion hydration based on primitive quasichemical theory. J. Chem. Phys. 2010; 132(23) 234101.
40. Yu, Haibo; Roux, Benoît. On the utilization of energy minimization to the study of ion selectivity. Biophys. J. 2009; 97(8):L15–L17. [PubMed: 19843443]
41. Rogers DM, Beck TL. Modeling molecular and ionic absolute solvation free energies with quasichemical theory bounds. J. Chem. Phys. 2008; 129 134505.
42. Pratt, LR.; Rempe, SB. Quasi-chemical theory and implicit solvent models for simulations. In: Hummer, G.; Pratt, LR., editors. Simulation and Theory of Electrostatic Interactions in Solution. New York: AIP Press; 1999. p. 172-201.
43. Pratt, Lawrence R.; Rempe, Susan B. Quasi-chemical theory and implicit solvent models for simulations. arXiv:physics/9909004v1. 1999:1–30.
44. Džzidić I, Kebarle Paul. Hydration of the alkali ions in the gas phase. enthalpies and entropies of reactions $M^+(H_2O)_{n-1} + H_2O = M^+(H_2O)_n$. J Phys. Chem. 1970; 74(7):1466–1474.
45. Swope, William C.; Andersen, Hans C.; Berens, Peter H.; Wilson, Kent R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J. Chem. Phys. 1982; 76(1):637–649.
46. Welford Castleman A Jr, Guo BC, Conklin BJ. Thermochemical properties of ion complexes $Na^+(M)_n$ in the gas phase. J Amer. Chem. Soc. 1989; 111(17):6506–6510.
47. Ponder, JW. TINKER: Software tools for molecular design. 4.2 edition. St. Louis, MO: Washington University School of Medicine; 2004.
48. Shirts, Michael R.; Chodera, John D. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 2008; 129(12) 124105.
49. Tissandier, Michael D.; Cowen, Kenneth A.; Feng, Wan Yong; Gundlach, Ellen; Cohen, Michael H.; Earhart, Alan D.; Coe, James V.; Tuttle, Thomas R, Jr. The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. J. Phys. Chem. A. 1998; 102(40):7787–7794.
50. Noskov, Sergei Yu; Roux, Benoît. Ion selectivity in potassium channels. Biophys. Chem. 2006; 124(3):279–291. Ion Hydration Special Issue. [PubMed: 16843584]
51. Hille, B. Ionic Channels of Excitable Membranes. 3 edition. Sunderland, MA: Sinauer Associates; 2001.
52. Noskov, Sergei Y.; Roux, Benoît. Control of ion selectivity in LeuT: Two Na^+ binding sites with two different mechanisms. J. Mol. Biol. 2008; 377(3):804–818. [PubMed: 18280500]
53. Yu, Haibo; Noskov, Sergei Yu; Roux, Benoît. Two mechanisms of ion selectivity in protein binding sites. Proc. Nat. Acad. Sci. USA.
54. Rempe, Susan B.; Jónsson, Hannes. A computational exercise illustrating molecular vibrations and normal modes. Chem. Educator. 1998; 3(4):169–170.

55. Barone, Vincenzo. Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation. J. Chem. Phys. 2004; 120(7):3059–3065. [PubMed: 15268458]

56. Ayala, Philippe Y.; Schlegel, H. Bernhard Identification and treatment of internal rotation in normal mode vibrational analysis. J. Chem. Phys. 1998; 108:2314.

57. Sabo, Dubravko; Varma, Sameer; Martin, Marcus G.; Rempe, Susan B. Studies of the thermodynamic properties of hydrogen gas in bulk water. J. Phys. Chem. B. 2008; 112(3):867–876. [PubMed: 18154326]

58. Hummer G, Garde S, García AE, Pohorille A, Pratt LR. An information theory model of hydrophobic interactions. PNAS. 1996; 93(17):8951–8955. [PubMed: 11607700]

59. Shah JK, Asthagiri D, Pratt LR, Paulaitis ME. Balancing local order and long-ranged interactions in the molecular theory of liquid water. J. Chem. Phys. 2007; 127 144508.

60. Paliwal A, Asthagiri D, Pratt LR, Ashbaugh HS, Paulaitis ME. An analysis of molecular packing and chemical association in liquid water using quasichemical theory. J. Chem. Phys. 2006; 124 224502.

61. Jordan, Peter C. New and notable: Tuning a potassium channel–the caress of the surroundings. Biophys. J. 2007; 93(4):1091–1092. [PubMed: 17513347]

62. Yu, Haibo; Noskov, Sergei Yu; Roux, Benoît. Hydration number, topological control, and ion selectivity. J. Phys. Chem. B. 2009; 113(25):8725–8730. [PubMed: 19489546]

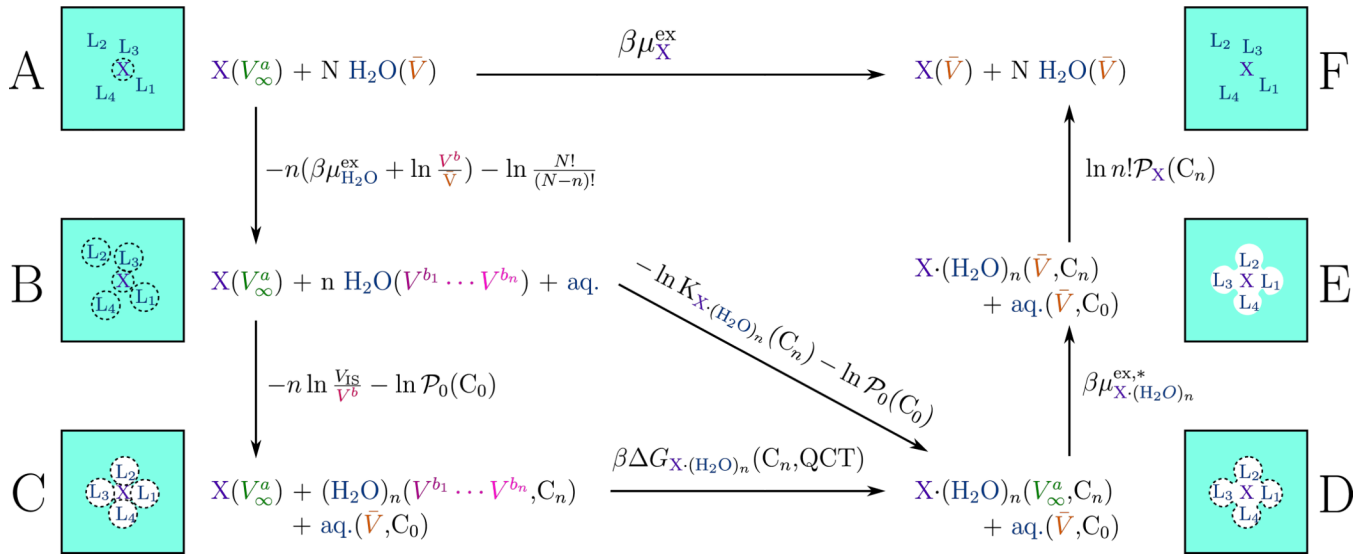

Figure 1.

Free energies of the homogeneous QCT process. The local excess chemical potential determining solute activity at a point in solution can be calculated following any path between states leading from top left to top right. The particular choice of water for the coordinating ligands, L_n, is arbitrary, as similar pathways can be defined using any combination of solution components.

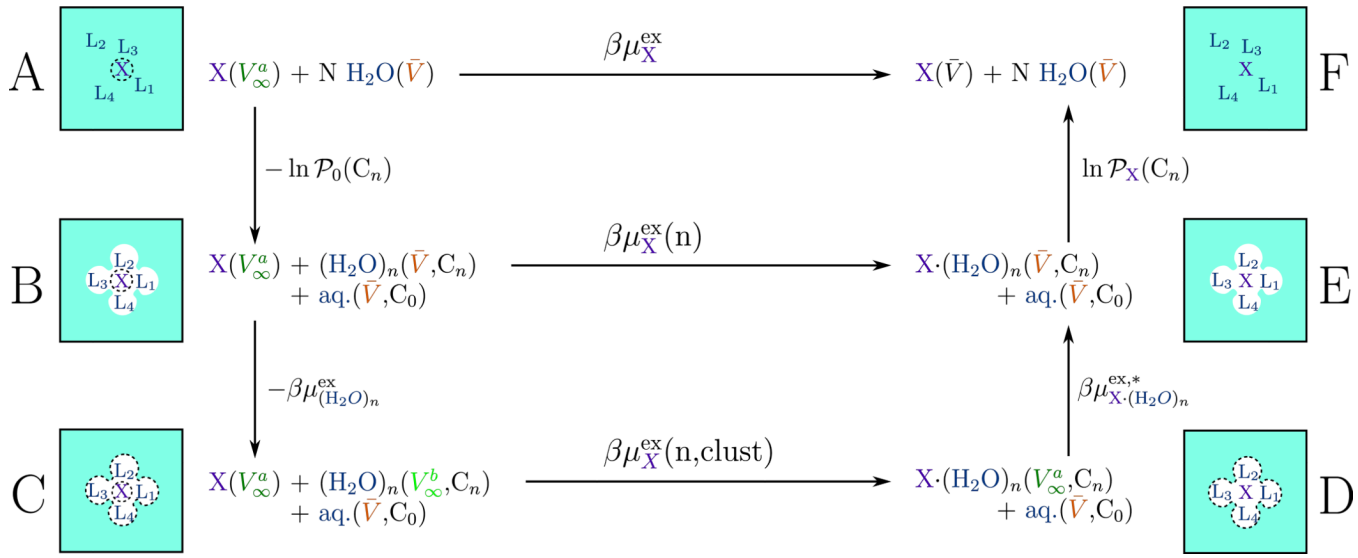

Figure 2.
Free energies of the non-homogeneous QCT process, applicable to enzyme active sites.

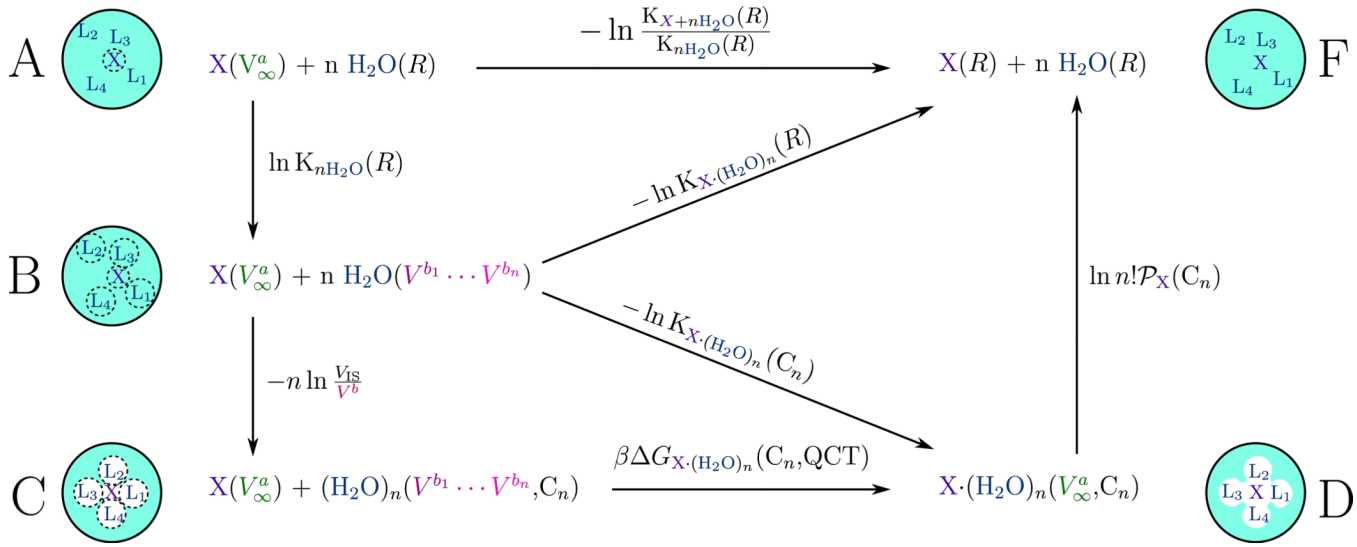

Figure 3.

QCT process for an isolated binding site. Due to the absence of a surrounding environment, ligand extraction (A→B) completely removes all interactions between waters, and insertion of the $X\cdot(H_2O)_n$ complex is trivial, merging states D and E from Fig. 1.

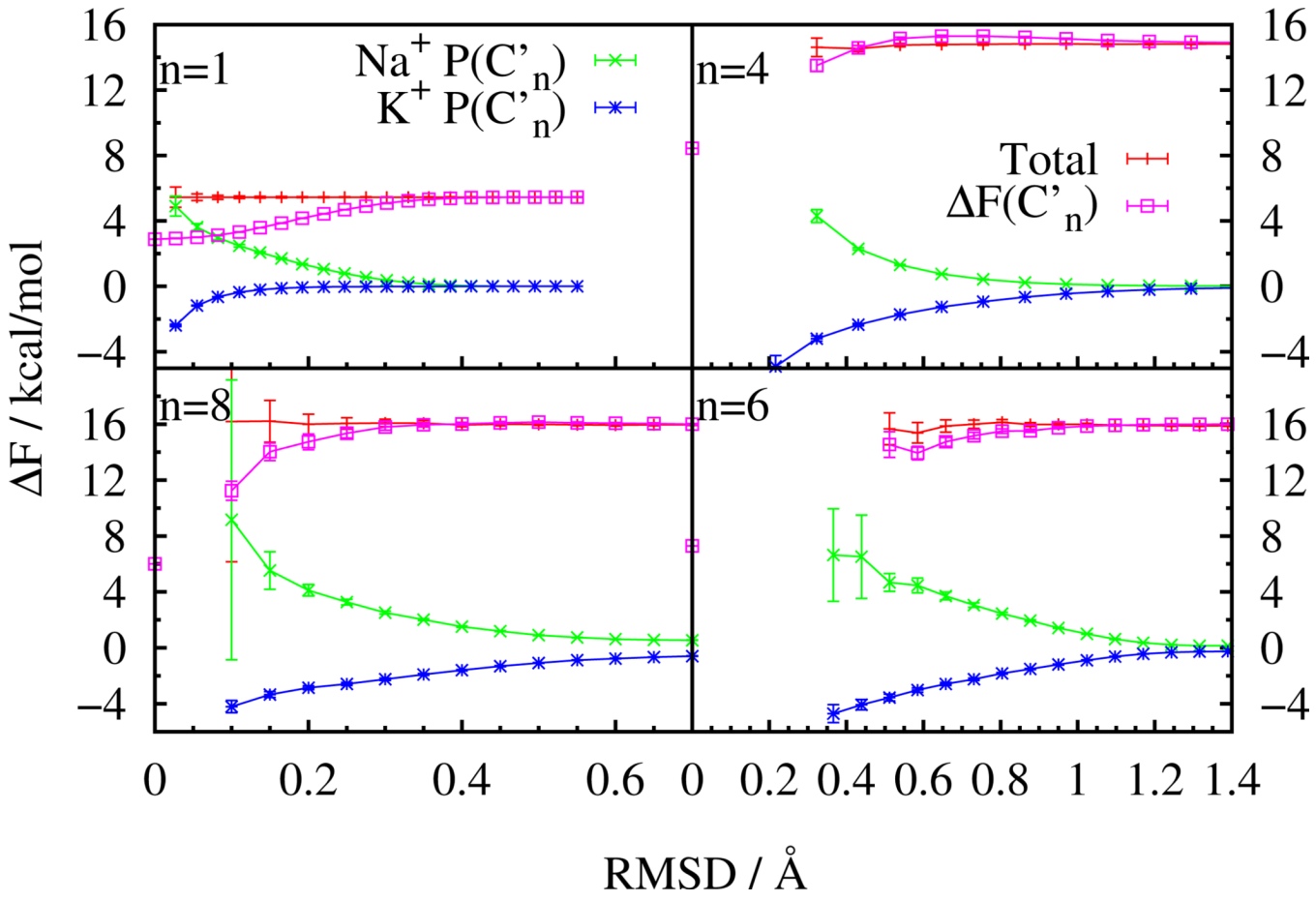

Figure 4.

Sodium to potassium mutation free energies computed for individual coordination states, C_n, as a function of distance (RMSD) from a K^+-*n* water minimum energy structure. The total selectivity is calculated from the sum of free energies for placing a constraint on Na^+ ($F_{Na^+} \rightarrow D_{Na^+}$, green x-s), mutating the ion given C_n ($D_{Na^+} \rightarrow D_{K^+}$, magenta boxes), and removing the constraint from K^+ ($D_{K^+} \rightarrow F_{K^+}$, blue stars). At zero RMSD (far left of each plot), $\Delta F(C_n^{'})=\Delta E$ is shown. These contributions sum to a constant for each *n*, the un-constrained $Na^+ \rightarrow K^+$ mutation free energy (red). In bulk water, the mutation free energy is 17.3 kcal/mol.[30]

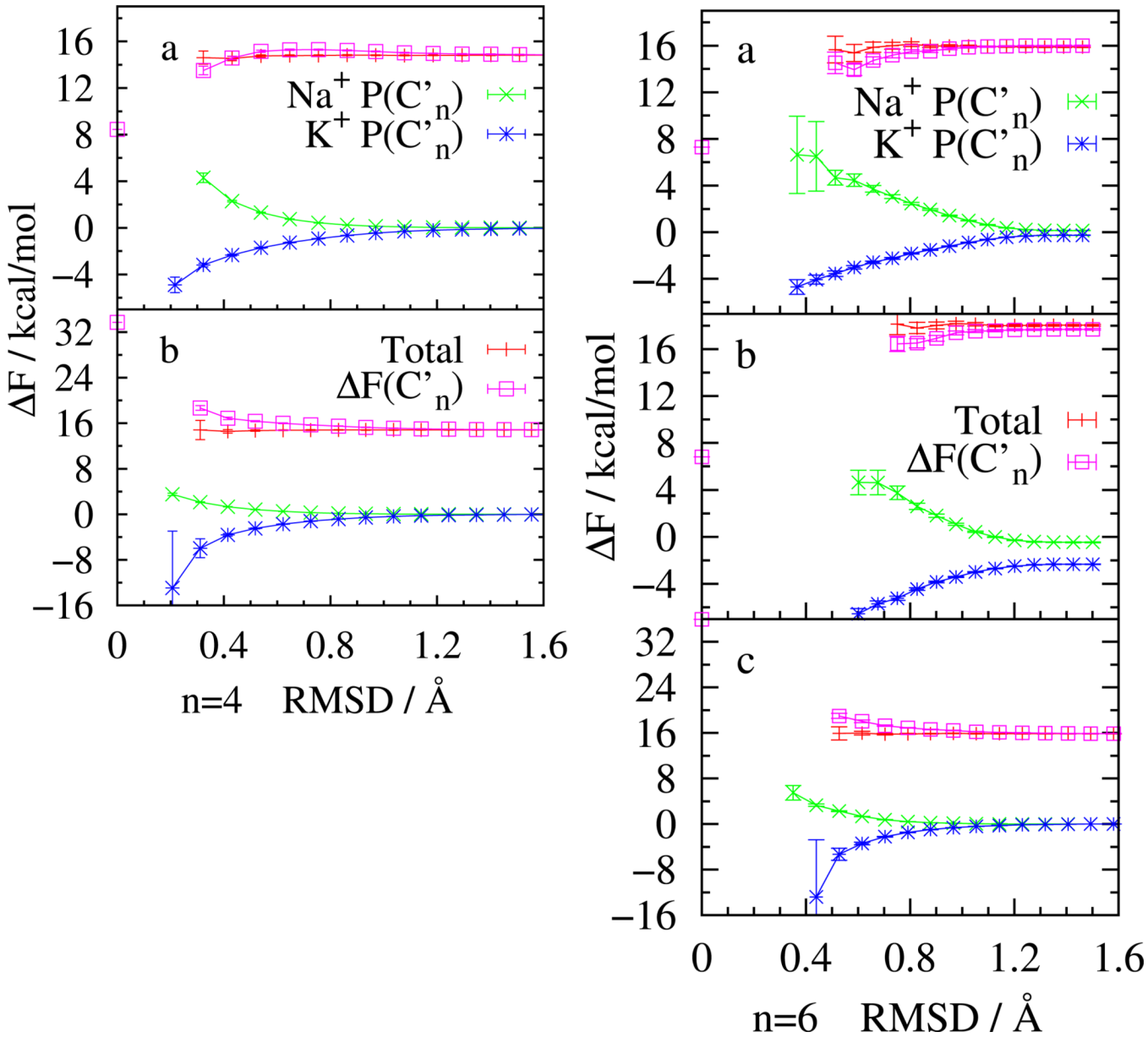

Figure 5.

Comparison of selectivity components for alternate choices of minimum energy geometry at fixed number of water ligands, *n*. For $n = 4$ (left), panel a is constructed using RMSD from a K^+ minimum energy structure, while b is constructed from a Na^+ minimum. For $n = 6$ (right), a and b are based on K^+ structures, and c is from the most highly occupied Na^+ minimum.

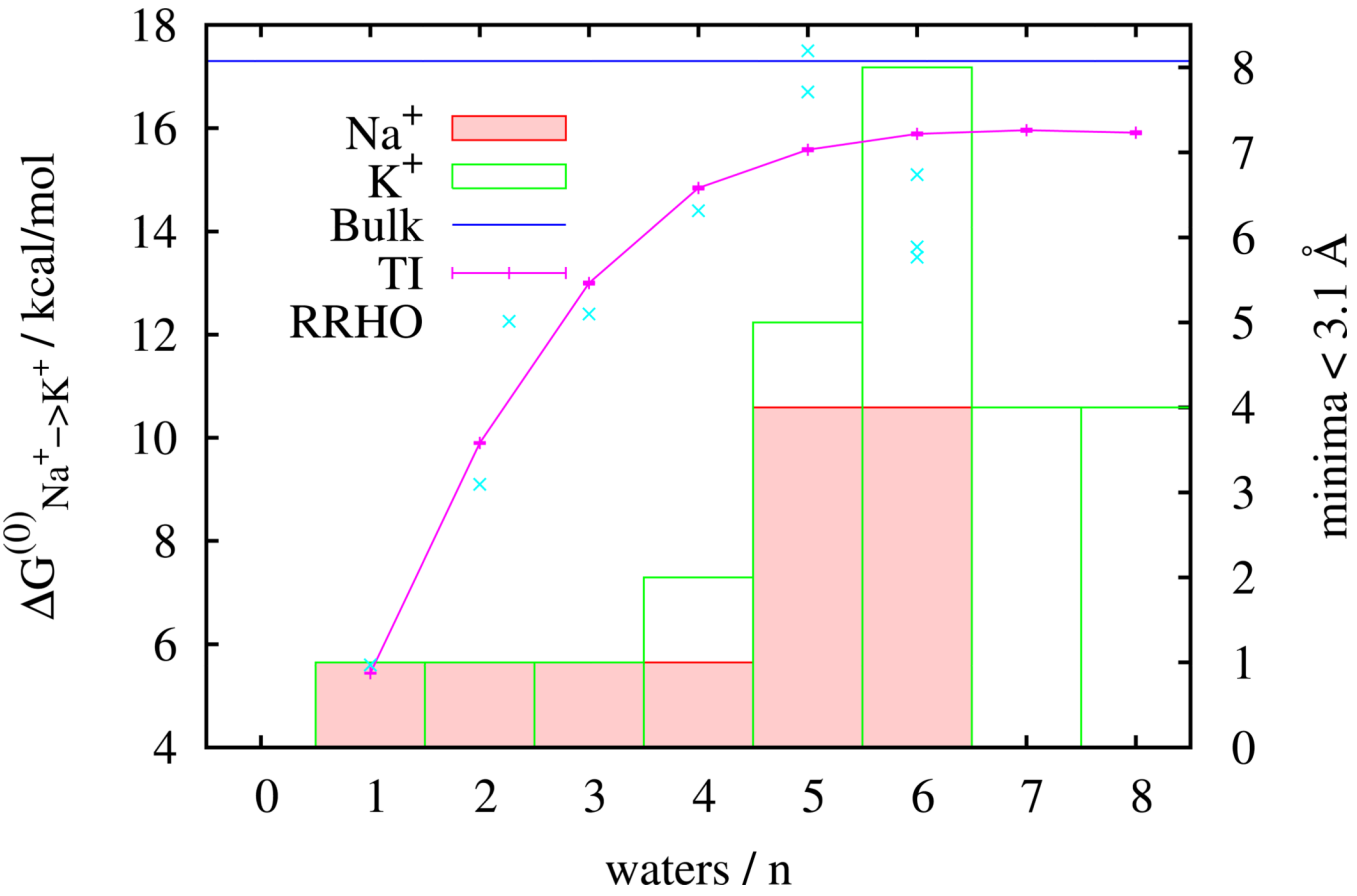

Figure 6.

Ion selectivities for constant water number, *n*, and wall force (Eq. 4) computed with the AMOEBA force field using TI and the RRHO approximation (left scale). The RRHO approximation to the free energy difference is plotted for Na^+ clusters with at least 5% percent occupancy at 298.15 K, subtracted from the free energy of the K^+ cluster with nearest RMSD minimum energy structure. The number of minimum energy structures within the 3.1Å constraint is plotted on the right scale.