Two mechanisms of ion selectivity in protein binding sites

Haibo Yu[a,1], Sergei Yu. Noskov[b], and Benoît Roux[a,2]

[a]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637; and [b]Institute for Biocomplexity and Informatics, Department of Biological Sciences, University of Calgary, Calgary, AB, Canada T2N 1N4

Edited* by William F. DeGrado, Penn, Philadelphia, PA, and approved September 23, 2010 (received for review May 21, 2010)

A theoretical framework is presented to clarify the molecular determinants of ion selectivity in protein binding sites. The relative free energy of a bound ion is expressed in terms of the main coordinating ligands coupled to an effective potential of mean force representing the influence of the rest of the protein. The latter is separated into two main contributions. The first includes all the forces keeping the ion and the coordinating ligands confined to a microscopic subvolume but does not prevent the ligands from adapting to a smaller or larger ion. The second regroups all the remaining forces that control the precise geometry of the coordinating ligands best adapted to a given ion. The theoretical framework makes it possible to delineate two important limiting cases. In the limit where the geometric forces are dominant (rigid binding site), ion selectivity is controlled by the ion-ligand interactions within the matching cavity size according to the familiar "snug-fit" mechanism of host-guest chemistry. In the limit where the geometric forces are negligible, the ion and ligands behave as a "confined microdroplet" that is free to fluctuate and adapt to ions of different sizes. In this case, ion selectivity is set by the interplay between ion-ligand and ligand-ligand interactions and is controlled by the number and the chemical type of ion-coordinating ligands. The framework is illustrated by considering the ion-selective binding sites in the KcsA channel and the LeuT transporter.

computations | free energy | ion coordination | KcsA | LeuT

The binding of small ions is fundamental to the structure and function of biological systems. Ions are involved in the folding of proteins and nucleic acids, enzyme catalysis, and in numerous cellular signaling processes. Monovalent cations such as Na^+ and K^+ play an important role in the homeostasis and electric activity of living cells, modulating biomolecular dynamics and stability through both specific and nonspecific interactions (1, 2). The importance of those small cations is most strikingly examplified by their implication in a wide variety of membrane transport proteins, e.g., ion channels (3, 4), transporters (5–8), and ATP-driven pumps (9–11).

Understanding the microscopic mechanisms by which ions associate with high specificity to protein binding sites is a question that has fascinated scientists for decades. Because K^+ and Na^+ ions are strongly bound to water molecules in bulk solutions, the protein must provide coordinating groups compensating the loss of hydration. Selectivity arises when this energetic compensation is more favorable for one type of ion than for another, relative to the difference in the hydration free energy between two ions. The 1,000-fold discrimination between K^+ and Na^+ achieved by some membrane proteins seems particularly remarkable because these two monovalent cations are so similar, differing only slightly in their atomic radii (by ~0.38 Å). The simplest and most appealing explanation of ion selectivity relies primarily on the protein structure, whereby implicitly assuming that the protein binding site is providing a cavity of the appropriate size for one ion, but is unable (for structural reasons) to adapt to an ion of a different size. This view has been widely used to discuss the ion selectivity of channels and membrane transporters (12). For example, the determination of the three-dimensional structure of the KcsA K^+ channel at atomic resolution using X-ray crystallography reveals a series of K^+ binding sites along the narrow pore, that seem perfectly adapted to provide an optimal coordination for K^+ but not for Na^+ (3). This view, which relies on a precise control of the protein structure at the sub-ångström level, is in close correspondence with the classical concepts invoked in "host-guest" chemistry (13, 14). In the field of permeation and ion channels, this structural explanation of selectivity has been traditionally called the "snug-fit" mechanism (15).

The availability of high resolution X-ray structures of membrane proteins with a highly selective ion-binding site offers a great opportunity for advancements in our understanding of ion selectivity using advanced computational methods. Despite limitations due to their approximating nature, current atomic computational models appear to be sufficiently accurate to gain mechanistic insights about ion selectivity. This statement is supported by the results from free energy perturbation molecular dynamics (FEP/MD) simulations carried out on a number of systems and are all consistent with experimental observations (16–23): the KcsA channel and valinomycin are highly selective for K^+ (16, 17), the NaK channel is permissive for both Na^+ and K^+ (18), and the two binding sites of the leucine transporter LeuT are highly selective for Na^+ (22).

In the case of the KcsA K^+ channel, computational studies have led to the following paradox: the narrow pore is highly selective for K^+ over Na^+ according to all-atom FEP/MD simulations, even though it is inherently too flexible to satisfy the requirements of the classic explanation based on a fixed cavity size (16, 17). This counterintuitive result led to the suggestion that robust selectivity in a flexible structure could arise depending on the number and type of ligands coordinating the ion, without the need to enforce the protein geometry to sub-ångström precision (16–18). Several studies confirmed that selectivity can be explained on the basis of local interactions (24–26), but there is a lack of consensus about the significance of the spatial restriction imposed on the ligands by the protein structure surrounding the selective binding site (26, 27). Although these ideas seem conceptually simple, assessing the exact role of the local and nonlocal forces exerted on the ion and coordinating ligands by the surrounding protein scaffold is challenging because it is difficult to identify the underlying physicochemical basis for the computed selectivity from large-scale FEP/MD simulations.

Our goal with the present analysis is to gain deeper understanding of the structural and energetic factors governing ion selectivity by introducing a statistical mechanical framework that

Author contributions: H.Y., S.Y.N., and B.R. designed research; H.Y., S.Y.N., and B.R. performed research; H.Y., S.Y.N., and B.R. analyzed data; and H.Y., S.Y.N., and B.R. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]Present address: School of Chemistry, University of Wollongong, Wollongong, NSW 2522, Australia.

[2]To whom correspondence should be addressed. E-mail: roux@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1007150107/-/DCSupplemental.

enables a stricter definition of the microscopic elements entering the construction of reduced binding site models in proteins. The analysis sheds new light on the existence of two distinct limiting mechanisms giving rise to ion selectivity in proteins. The theoretical framework is elaborated in the next section, and then illustrated with an application to the ion-selective binding sites of the KcsA channel and the LeuT transporter.

Theoretical Developments

Statistical Mechanical Reduction. A selective ion-binding site in a protein in equilibrium with bulk solvent is considered. Assuming that ionic species i and j are present in solution, binding selectivity is governed by the relative free-energy $\Delta\Delta G_{ij} = \Delta G^{\text{site}}_{ij} - \Delta G^{\text{bulk}}_{ij}$, where $\Delta G^{\text{bulk}}_{ij} = [G^{\text{bulk}}_i - G^{\text{bulk}}_j]$ is the free energy difference between ion i and j in the bulk solvent, and $\Delta G^{\text{site}}_{ij} = [G^{\text{site}}_i - G^{\text{site}}_j]$ is the free energy difference between ion i and j in the binding site. The relative free energy of ion i and j in the binding site can be written in terms of configurational integrals,

$$e^{-\beta\Delta G^{\text{site}}_{ij}} = \frac{\int_{\text{site}} d\mathbf{R} e^{-\beta U_i(\mathbf{R})}}{\int_{\text{site}} d\mathbf{R} e^{-\beta U_j(\mathbf{R})}}, \qquad [1]$$

where $\mathbf{R}$ represents all degrees of freedom in the system, $\beta = 1/k_{\text{B}}T$, and U_i and U_j represent the total potential energy with ion i or j, respectively.

While Eq. **1** can be employed to carry out free energy computations from all-atom MD simulations, it remains difficult to unambiguously identify direct cause-and-effect in complex systems. For example, the very strong interactions of the ion with the protein ligands are balanced by multiple long-range interactions arising from the packing of atoms around the binding site that confers stability to its three-dimensional configuration. Those long-range interactions may, or may not, be sensitive to the type of ion bound in the site. To make progress, we seek to reduce the complexity of Eq. **1** by integrating out all of the degrees of freedom that are not immediately relevant to the local interactions. To this end, we distinguish a reduced subsystem comprising only the bound ion and the n most important ligands that are directly participating in the binding site. By ligands we mean to designate the molecular groups coordinating the ion directly, together with the nearest covalently attached atoms. Typical reduced subsystems are illustrated in Fig. 1 for the KcsA channel and the LeuT transporter. Letting $\mathbf{X}$ represent the coordinates of the flexible reduced subsystem, and $\mathbf{Y}$ the remaining coordinates, Eq. **1** can be reexpressed as,

$$\begin{aligned} e^{-\beta\Delta G^{\text{site}}_{ij}} &= \frac{\int_{\text{site}} d\mathbf{X} d\mathbf{Y} e^{-\beta U_i(\mathbf{X},\mathbf{Y})}}{\int_{\text{site}} d\mathbf{X} d\mathbf{Y} e^{-\beta U_j(\mathbf{X},\mathbf{Y})}} \\ &= \frac{\int_{\text{site}} d\mathbf{X} e^{-\beta[U^{\text{il}}_i(\mathbf{X}) + U^{\text{ll}}(\mathbf{X}) + \Delta W^{\text{site}}(\mathbf{X})]}}{\int_{\text{site}} d\mathbf{X} e^{-\beta[U^{\text{il}}_j(\mathbf{X}) + U^{\text{ll}}(\mathbf{X}) + \Delta W^{\text{site}}(\mathbf{X})]}}, \end{aligned} \qquad [2]$$

where U^{il} and U^{ll} represent the ion-ligand and ligand-ligand interactions, respectively, and $\Delta W^{\text{site}}(\mathbf{X})$ is an effective potential of mean force (PMF) that incorporates all the influence of the rest of the system (protein, membrane, and solvent).

It is important to note that the statistical mechanical reduction of the subsystem is achieved by introducing a fixed assignment of the participating molecular moieties; no spatial restriction is actually associated with the choice of degrees of freedom $\mathbf{X}$ appearing in Eq. **2**. This present development is in contrast with alternate statistical mechanical reductions that rely on a *fixed* spatial boundary (e.g., based on a distance criterion) to identify and separate a system into an inner and an outer region (28). In that sense, the separation between the "inner" ($\mathbf{X}$) and "outer" ($\mathbf{Y}$) degrees of freedom in Eq. **2** is based on a flexible boundary. Similar ideas were used to formulate the properties of bulk solvent in terms of a reduced subsystem containing a fixed number of solvent molecules (29).

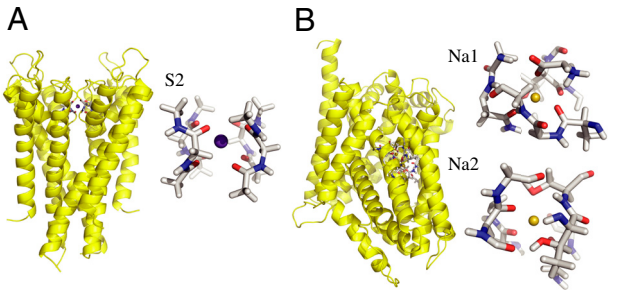


Fig. 1. (*A*). The full KcsA structure (PDB code: 1K4C) is shown in yellow ribbon, and the central K^+-selective S2 binding site of the KcsA channel formed by the backbone carbonyl of Val76 and Gly77 chosen to construct the reduced reduced model consisting of four diglycine molecules. (*B*). The full LeuT structure (PDB code: 2A69) is shown in yellow ribbon, and the two Na^+-selective binding sites Na1 and Na2 with the protein atoms incorporated into the reduced model.

Confinement and Geometric Contributions. By definition, the effective ΔW^{site} in Eq. **2** rigorously incorporates all influence from the three-dimensional protein structure on the ion and coordinating ligands. In other words, ΔW^{site} is formally associated with all "architectural" forces arising from the protein scaffold. It can be safely assumed that, for a fixed configuration $\mathbf{X}$ for the ion and its coordinating ligands in the reduced subsystem, $\Delta W^{\text{site}}(\mathbf{X})$ does not depend explicitly on the ion type. This assumption is reasonable because the bound ion in the fixed configuration $\mathbf{X}$ is surrounded by the coordinating ligands and is not in direct contact with the remaining atoms of the outer region. The outer atoms are some distance away from the bound ions and their interactions with the ion are mainly electrostatic. Those long-range interactions are nearly the same as long as an ion of the same charge is present in the reduced subsystem. For example, computations indicate that ΔW^{site} for Na^+ and K^+ in the S2 binding site of the KcsA channel differ by less than 0.2 kcal/mol.

At a conceptual level, it is useful to separate the architectural forces acting on the reduced subsystem into two distinct contributions, which we will call "confinement" and "geometric," $\Delta W^{\text{site}} = \Delta W^{\text{site}}_c + \Delta W^{\text{site}}_g$. The first contribution, ΔW^{site}_c, accounts for the generic effect of the protein structure surrounding the reduced subsystem imposing an upper limit on the fluctuations of the bound ion and its coordinating ligands. Without the surrounding protein structure, the atoms of the reduced subsystem would move away from one another and the binding site would not maintain its integrity. By virtue of the surrounding protein structure, the instantaneous atomic fluctuations of the atoms within the reduced subsystem are contained and never exceed some maximum value. Bounded fluctuations is the generic effect of architectural confinement.

Multiple mathematical forms of such a generic confinement are possible. For the sake of simplicity, for each atom k in the reduced subsystem we define the architectural confinement as the smallest spherical volume V_k encompassing all the dynamic excursions of that atom, independently of the type of ion that is bound in the site. Each spherical volume V_k is parametrically defined by the position of the center $\bar{\mathbf{r}}_k$ and radius R_k; those parameters can be extracted from MD simulations of the full system. Accordingly, the architectural confinement contribution is defined as,

$$e^{-\beta\Delta W^{\text{site}}_c(\mathbf{X})} = \Pi_k H(|\mathbf{r}_k - \bar{\mathbf{r}}_k| - R_k) \equiv H_c(\mathbf{X}), \qquad [3]$$

where $\mathbf{r}_k$ is the position of the kth atom, and the H's are Heaviside step-functions. Additional order parameters (e.g., distance, angles) can also be included in a similar way. This type of

construction defines a minimal default model with probability distribution,

$$\rho_0(\mathbf{X}) = H_c(\mathbf{X})e^{-\beta[U_i^{\mathrm{il}}(\mathbf{X})+U^{\mathrm{ll}}(\mathbf{X})]} \quad [4]$$

that incorporates the generic effect of architectural confinement by the surrounding protein structure in an idealized fashion: each atom k of the reduced subsystem can fluctuate and make arbitrary excursions, as long as the fluctuations remain bounded. If the confinement function H_c is chosen to match the upper bound of all atomic fluctuations, whether an ion of type i or j is bound, H_c is independent of ion type by construction. We will refer to the situation embodied by $\rho_0(\mathbf{X})$ defined in Eq. **4** as the *confined microdroplet* model. In order to serve as a useful reference default model in the following developments, it is important that the idealized reduced model based on $\rho_0(\mathbf{X})$ provides a reasonably good "mimic system" able to reproduce the qualitative physical behavior observed in the all-atom system.

It is interesting to note that there is a certain analogy between Eq. **3** and how a hard-core radius might be defined in terms of the shortest distance observed between any two particles in the liquid over multiple configurations (30). The concept of the hard-core repulsion makes it possible to treat the thermodynamic effects of weak interactions such as van der Waals as a separate perturbation (30). Here, a similar idea is used as an idealization of architectural confinement to define a convenient minimal default model, upon which additional features can be built.

The second contribution, $\Delta W_g^{\mathrm{site}}$, accounts for the remaining architectural forces from the surrounding protein that enforce a precise geometry to the atoms of the inner region. For example, the geometrical architectural forces control the magnitude of the fluctuations of each atom k within the generic confinement volume V_k. An optimal model for $\Delta W_g^{\mathrm{site}}$ should recapitulate the ensemble properties of the system extracted from all-atom MD simulations. But it is also important to avoid introducing unwarranted assumptions into the functional form of $\Delta W_g^{\mathrm{site}}$. To achieve this goal, we adopt a cross entropy method used previously by Hummer et al. (31) in studies of the hydrophobic solvation, and write

$$\eta = -\int d\mathbf{X}\rho(\mathbf{X})\ln[\rho(\mathbf{X})/\rho_0(\mathbf{X})], \quad [5]$$

where $\rho(\mathbf{X})$ is the probability distribution of the system. We seek to maximize the relative entropy η in Eq. **5** under the constraint that $\rho(\mathbf{X})$ is normalized, and that the quadratic fluctuations of the atoms in the reduced subsystem must match the value extracted from all-atom MD simulations. The constrained optimization problem is solved by introducing the Lagrange multiplier λ_g to satisfy the condition on the normalization and the overall fluctuations. The resulting normalized distribution is,

$$\rho(\mathbf{X};\lambda_g) = \frac{H_c^{\mathrm{site}}(\mathbf{X})e^{-\beta[U_i^{\mathrm{il}}(\mathbf{X})+U^{\mathrm{ll}}(\mathbf{X})+\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g)]}}{\int d\mathbf{X}H_c^{\mathrm{site}}(\mathbf{X})e^{-\beta[U_i^{\mathrm{il}}(\mathbf{X})+U^{\mathrm{ll}}(\mathbf{X})+\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g)]}}, \quad [6]$$

where

$$\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g) = \lambda_g\sum_k(\mathbf{r}_k - \bar{\mathbf{r}}_k)^2. \quad [7]$$

In principle, it would be possible to build models that satisfy higher order moments of the atomic fluctuations following the same procedure (see Section A in *SI Appendix*).

The Lagrange multiplier λ_g has the dimension of a harmonic spring constant (kcal/mol/Å^2), and its optimal value should be determined self-consistently by performing averages with the distribution $\rho(\mathbf{X};\lambda_g)$ to match the reference values extracted from all-atom MD with ion i and j. Two extreme qualitative behaviors are expected: (*i*) if the binding site is very stiff, the root-mean-square (rms) fluctuations of the reduced subsystem are small and λ_g is expected to be very large; (*ii*) if the binding site is flexible, the rms fluctuations of the reduced subsystem are large and λ_g is expected to be small. When $\lambda_g = 0$, the distribution $\rho(\mathbf{X};\lambda_g)$ returns to the generic confined microdroplet model $\rho_0(\mathbf{X})$ of Eq. **4**.

Relative Free Energies. The relative binding free energy of ions i and j is,

$$e^{-\beta\Delta G_{ij}^{\mathrm{site}}} = \frac{\int_{\mathrm{site}} d\mathbf{X}H_c(\mathbf{X})e^{-\beta[U_i^{\mathrm{il}}(\mathbf{X})+U^{\mathrm{ll}}(\mathbf{X})+\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g)]}}{\int_{\mathrm{site}} d\mathbf{X}H_c(\mathbf{X})e^{-\beta[U_j^{\mathrm{il}}(\mathbf{X})+U^{\mathrm{ll}}(\mathbf{X})+\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g)]}}, \quad [8]$$

where H_c given by Eq. **3** is the generic confinement chosen to match the upper bound of all atomic fluctuations regardless of the bound ion type, and $\Delta W_g^{\mathrm{site}}(\mathbf{X};\lambda_g)$, given by Eq. **7**, is a harmonic potential with force constant λ_g adjusted to optimally match the atomic rms fluctuations from the all-atom simulations. By construction, the only term that depends on ion type i or j in Eq. **8** is the ion-ligand interaction energies $U_i^{\mathrm{il}}(\mathbf{X})$ and $U_j^{\mathrm{il}}(\mathbf{X})$ from the reduced subsystem.

It is useful to derive some simplified expressions for $\Delta G_{ij}^{\mathrm{site}}$. To this end, let us define the 1-ion self free energy G_i^{site},

$$e^{-\beta G_i^{\mathrm{site}}} = \frac{\int_{\mathrm{site}} d\mathbf{X}H_c^{\mathrm{site}}(\mathbf{X})e^{-\beta W_i(\mathbf{X})}}{\int_{\mathrm{site}} d\mathbf{X}H_c^{\mathrm{site}}(\mathbf{X})}, \quad [9]$$

where $W_i = U_i^{\mathrm{il}} + U^{\mathrm{ll}} + \Delta W_g^{\mathrm{site}}(\lambda_g)$. The self free energy may be expressed as,

$$e^{\beta G_i^{\mathrm{site}}} = \int dw\mathscr{P}_i(w)e^{\beta w}, \quad [10]$$

where the density of state $\mathscr{P}_i(w)$ is defined as,

$$\mathscr{P}_i(w) = \frac{\int_{\mathrm{site}} d\mathbf{X}\delta[W_i(\mathbf{X}) - w]H_c^{\mathrm{site}}(\mathbf{X})e^{-\beta W_i(\mathbf{X})}}{\int_{\mathrm{site}} d\mathbf{X}H_c^{\mathrm{site}}(\mathbf{X})e^{-\beta W_i(\mathbf{X})}}. \quad [11]$$

The self free energy G_i^{site} in Eq. **10** can be developed as cumulant expansion, yielding $G_i^{\mathrm{site}} \approx \langle w\rangle_{(i)} + \langle(w - \langle w\rangle_i)^2\rangle_{(i)}/2k_{\mathrm{B}}T$. To lowest order, the entropy difference between ion of type i and j is negligible and the quadratic fluctuations are not expected to depend strongly upon the ion type. In this case, the relative free energy of ion i and j in the binding site is dominated by,

$$\Delta G_{ij}^{\mathrm{site}} \approx \langle w\rangle_{(i)} - \langle w\rangle_{(j)} \approx [\langle U_i^{\mathrm{il}} + U^{\mathrm{ll}}\rangle_{(i)} - \langle U_j^{\mathrm{il}} + U^{\mathrm{ll}}\rangle_{(j)}] + [\langle\Delta W_g^{\mathrm{site}}\rangle_{(i)} - \langle\Delta W_g^{\mathrm{site}}\rangle_{(j)}]. \quad [12]$$

The first and second terms in **[12]**, $\langle U_i^{\mathrm{il}}\rangle_{(i)} - \langle U_j^{\mathrm{il}}\rangle_{(j)}$ and $\langle U^{\mathrm{ll}}\rangle_{(i)} - \langle U^{\mathrm{ll}}\rangle_{(j)}$, correspond to the difference in the average ion-ligand and ligand-ligand interactions, respectively. The last term, $\langle\Delta W_g^{\mathrm{site}}\rangle_{(i)} - \langle\Delta W_g^{\mathrm{site}}\rangle_{(j)}$, corresponds to the difference in the geometric architectural contribution to the PMF. The latter depends implicitly on λ_g, and it is identically equal to zero if $\lambda_g = 0$.

Results and Discussion

Two Distinct Physical Limits. The formal separation of ΔW^{site} makes it possible to delineate two important limiting cases. The first corresponds to the situation where the geometric structural forces associated with ΔW_g are very strong. As a consequence, the configuration of the binding site is very stiff, its geometry is strictly enforced, and the average ion-ligand interaction energy dominates the relative free-energy difference,

$$\lim_{\lambda_g \to \infty} \Delta G_{ij}^{\text{site}} \approx \langle U_i^{\text{il}} \rangle_{(i)} - \langle U_j^{\text{il}} \rangle_{(j)}, \quad [13]$$

where the brackets with subscripts represent thermal averages for ion i and j. This expression is consistent with the concept of a fixed cavity size as in host-guest chemistry (12), traditionally called the snug-fit mechanism of selectivity in the ion channel literature (15). In this case, the selectivity is predominantly controlled by the ion-ligand interactions relative to the hydration free energy. The surrounding protein structure is very stiff and precisely dictates the configurational geometry of the ligands that is best adapted to an ion of a given size.

A second limiting case corresponds to the situation where the geometric structural forces are negligible. This situation is represented by the generic reference model defined by Eq. **4**. The ion and the ligands are allowed to fluctuate and freely adapt to coordinate an ion of a different size, but are otherwise restricted to remain within a small spatial region strictly enforced by the function H_c^{site}. The relative free energy is dominated by

$$\lim_{\lambda_g \to 0} \Delta G_{ij}^{\text{site}} \approx \langle U_i^{\text{il}} + U^{\text{ll}} \rangle_{(i)} - \langle U_j^{\text{il}} + U^{\text{ll}} \rangle_{(j)}. \quad [14]$$

This situation corresponds to the confined microdroplet limit, representing a flexible binding site (17). In this limit, the probability of the configurations is dictated by the strong ion-ligand interactions causing an induced-fit coordination of the ion (i.e., even an ion of the "wrong" size ends up being well coordinated). Structurally, the ion-ligand system is allowed to fluctuate and displays some local liquid-like features allowing a finite width for the first peak in the ion-ligand radial distribution function (17).

For Na^+ and K^+, the difference in ion-ligand interaction energy, $\langle U_{\text{Na}}^{\text{il}} \rangle_{(\text{Na})} - \langle U_{\text{K}}^{\text{il}} \rangle_{(\text{K})}$, is always a large negative number (favorable), whereas the corresponding difference in ligand-ligand interaction, $\langle U^{\text{ll}} \rangle_{(\text{Na})} - \langle U^{\text{ll}} \rangle_{(\text{K})}$, is more typically a positive number (unfavorable) (32). Thermodynamically, the mean ion-ligand interaction does not give rise to selectivity in the confined microdroplet limit, in contrast with the host-guest mechanism with fixed cavity radius as described by [**13**]. The lack of selectivity from the mean ion-ligand interaction has been documented very clearly by Asthagiri et al. (24). In the confined microdroplet, both the favorable (negative) ion-ligand and unfavorable (positive) ligand-ligand interactions contribute and compete, in a nontrivial manner, to control selectivity (17). There are some similarities with Eisenman's classic concept of field-strength, in which structurally featureless binding sites are selective by virtue of the chemical type of ion-coordinating ligands (33). This view is also closely related to the model proposed to explain the selectivity of Ca^{2+} channels in which the ion and the negatively charged carboxylate groups of the protein are allowed to dynamically fluctuate within a confined subvolume (34).

Spatial confinement is an essential component of the context that gives rise to selectivity in the limit where $\lambda_g \to 0$ in [**14**]. Factors that lead to a structural disruption of that context can result in a loss of selectivity of a binding site. Interestingly, a search through the structural database reveals that a common feature of ion-binding sites in proteins is the existence of a shell of nonpolar hydrophobic groups surrounding the ion-coordinating ligands (35). Presumably, this serves to provide a region of low dielectric to protect the confinement and enhance the local electrostatic interactions. While confinement is ultimately ensured by the three-dimensional fold of the protein, with all its *nonlocal* long-range complexities, it is important to realize that the character of ΔW_c^{site} does not need to be complex to support selectivity. Nontrivial free-energy patterns can emerge from the *local* interactions between the confined ion and the ligand, even in the context of a relatively generic and featureless ΔW_c^{site} (32). Therefore, in the confined microdroplet limit, the protein structure plays a critical role to support selectivity, albeit, not by enforcing a specific cavity size with a ligand geometry to sub-ångström precision (17).

The K^+ S2 Binding Site in KcsA. To illustrate the concepts presented above, we first consider the S2 binding site of the KcsA channel, which is the most selective when the narrow pore is in its conductive conformation (16, 17). The pore of the KcsA can adopt several conformations (3, 36), but the conductive state observed in crystal structure obtained at high K^+ concentration that we consider here is the conformation relevant for ion selectivity (37). Following the formal developments of Eqs. **3**, **6**, and **7**, we construct a reduced computational model of the S2 site of KcsA (Fig. 1*A*). The reduced model comprises four diglycine peptides corresponding to the backbone of Val76 and Gly77 of the four subunits in the KcsA channel, for a total of eight ion-coordinating carbonyl ligands. By construction, ΔW_c^{site} is designed to prevent fluctuations of the atoms that would exceed the largest excursion observed during an all-atom MD simulation of KcsA in a fully solvated membrane. The dynamics of the binding site produced by the reduced model with only the confinement is qualitatively similar to that observed from an all-atom MD simulation of KcsA with bound ions in a fully solvated membrane (see Movie S1 and Movie S2 animations provided in *SI Appendix*). This microdroplet model does not prevent collapse of the ion-coordinating ligand onto Na^+. The ΔW_g^{site} is designed to enforce the optimal coordinating geometry for K^+. Analysis indicates that the value of λ_g matching the atomic fluctuations from the all-atom MD is on the order of 0.5–1.0 kcal/mol/Å^2.

In Fig. 2, the results of free-energy calculations of $\Delta\Delta G_{\text{Na,K}}$ are shown as a function of the strength of the geometric forces. In the confined microdroplet limit, the site is selective for K^+ over Na^+ by about 6 kcal/mol. As the geometric structural force is increased via the Lagrange multiplier λ_g, the fluctuations progressively decrease and selectivity increases. Roughly two distinct regimes can be identified. When the architecture is enforced by a force constant λ_g of more than 10–100 kcal/mol/Å^2, the ion-ligand interactions start to make an increasingly important contribution to $\Delta\Delta G_{\text{Na,K}}$ while the ligand-ligand interactions decrease progressively down to zero. This behavior corresponds to the snug-fit limit. Selectivity starts to be strictly dictated by the cavity size only when the structural distortion of the coordinating ligand by the binding of Na^+ is causing a sufficiently large energy penalty, i.e., when $\lambda_g \geq 100$ kcal/mol/Å^2. At such values of λ_g, the rms fluctuations of the ligands become smaller than about 0.2–0.3 Å, corresponding to the size difference between Na^+ and K^+. At the opposite limit, when $\lambda_g \leq 10$ kcal/mol/Å^2, both

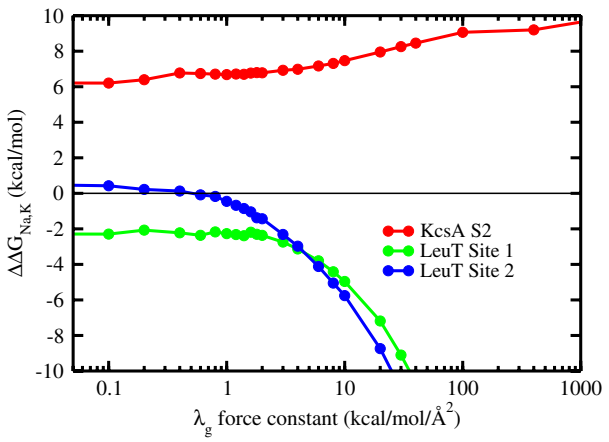


Fig. 2. Ion selectivity $\Delta\Delta G_{\text{Na,K}}$ as a function of the geometric force constant λ_g calculated from FEP/MD simulations of the reduced models of KcsA and LeuT binding sites.

the ion-ligand and ligand-ligand interactions contribute significantly to $\Delta\Delta G_{Na,K}$. This behavior corresponds to the confined microdroplet regime. The behavior of the reduced model in this regime is not very sensitive to the specific radius of the confinement volume (see Section B in *SI Appendix*). It is particularly noteworthy that the selectivity for K^+ over Na^+ is robust down to $\lambda_g = 0$ kcal/mol/Å^2, where the atomic fluctuations are more than twice as large as the size difference between Na^+ and K^+. This observation implies that no strain energy contribution from the geometric structural potential is needed to maintain selectivity in the confined microdroplet limit. This result is consistent with the study of Asthagiri et al. (24), in which a small variance in selectivity was observed when the restraining potential was varied from weak to strong.

In the intermediate regime, where geometric structural forces are too weak to prevent adaptation of the ligands to the small ion, there is an induced fit at the cost of a structural strain energy arising from the deformation of the ligands from their ideal geometry. Using a perturbative argument, this strain energy corresponds approximately to $\langle \Delta W_g^{site} \rangle_{(i)} - \langle \Delta W_g^{site} \rangle_{(j)}$. An appropriate name for this situation might be the "strained-fit" regime. Nevertheless, as observed from Fig. 3, the magnitude of both $\langle \Delta U^{il} \rangle$ and $\langle \Delta U^{ll} \rangle$ remains considerable as long as the binding site remains sufficiently flexible (i.e., for λ_g smaller than 100 kcal/mol/Å^2). The importance of both $\langle \Delta U^{il} \rangle$ and $\langle \Delta U^{ll} \rangle$ can also be displayed by using configurations extracted from all-atom MD simulations of the KcsA channel in a fully hydrated membrane, although the results converge with more difficulty because the analysis involves taking the difference between averages of large fluctuating quantities. From a structural point of view, whether the binding site is extremely flexible or rigid, near ideal K^+-oxygen coordination distances are invariably observed whenever an instantaneous configuration is energy-minimized.

The Na^+ Binding Sites in LeuT. To further appreciate the difference between the snug-fit and confined microdroplet limits, it is helpful to compare the above results with those obtained on LeuT, a bacterial homologue of Na^+-dependent neurotransmitter transporters. LeuT possesses two ion-binding sites, called Na1 and Na2 (Fig. 1*B*), which are highly selective for Na^+ over K^+ and only weakly selective for Na^+ over Li^+ (5, 22, 23). The site Na1 comprises six ligands, including a negatively charged carboxylate from the bound leucine substrate, while the site Na2 comprises five neutral ligands. As shown in Fig. 2, the calculated selectivity for a reduced model of the Na1 site retains some selectivity comparable to that from all-atom simulations even in the confined microdroplet limit (22). In contrast, the reduced model of the site Na2 does not display any selectivity in the absence of additional structural forces to stabilize an optimal cavity size for Na^+. These observations are not very sensitive to the particular radius of the confinement volume (see Section B in *SI Appendix*). The mechanism giving rise to Na^+ selectivity in these two sites in LeuT corresponds naturally to two different limits. In Na1, robust selectivity arises in the confined microdroplet by virtue of the strong electrostatic field arising from the negatively charged carboxylate group-the latter is a classic Eisenman high field-strength ligand that favors the smaller of the two cations (33). Selectivity in the site Na1 thus reflects the local thermodynamics determined by the composition of the confined microdroplet. In Na2, selectivity arises by virtue of geometric forces. Because the coordinating ligands forming the Na2 binding site are taken from backbone and side-chain atoms of nearest neighboring residues (e.g., m, $m-1$, and $m+1$) along the polypeptide chain, the protein exploits the local stiff degrees of freedom from the covalent structure to establish a cavity size that is robust and best-adapted to Na^+. This situation is reminiscent of K^+-selective valinomycin, a small cyclic ionophore made rigid by virtue of its covalent and hydrogen bonding structure (38, 39).

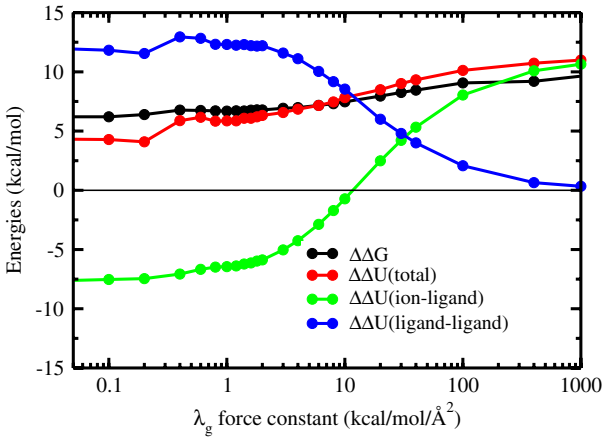


Fig. 3. Decomposition of the various energetic terms to the selectivity in the reduced model of the S2 binding site of KcsA. The ion selectivity, total energy, ion-ligand interaction energy U^{il}, and ligand-ligand interaction energy U^{ll} in the reduced model consisting of four diglycine molecules as a function of the strength of the geometric structural forces.

Conclusion

A statistical mechanical framework was developed to express the relative free energy of bound ions in terms of local interactions and the effect of the surrounding protein architecture. The latter appears as an effective potential acting on the local subsystem comprising the ion and the most relevant ligands. The framework helps to shed new light on the main elements giving rise to ion selectivity in protein binding sites. Separating the influence of architecture conceptually into confinement and geometric contributions, two important limiting cases were highlighted. In the limit where the geometric structural forces are dominant, selectivity is controlled by the ion-ligand interactions determined by the cavity size according to the familiar snug-fit mechanism of host-guest chemistry. In the limit where the geometric structural forces are negligible, the ligands are free to fluctuate and locally adapt to an ion of a different size. In this "confined microdroplet" limit, the free energies arise from the local thermodynamics of ion-ligand and ligand-ligand interactions and selectivity is controlled by both the number and chemical type of ion-coordinating ligands. This mechanism relies on a balance of ion-ligand and ligand-ligand interactions and its capacity to produce robust selectivity under well defined conditions is unambiguous.

The snug-fit and the confined microdroplet limits set in place two conceptually extreme physical views of ion selectivity. Real systems are expected to take on a mixed character of these two idealized extremes, being closer to one or the other. Some binding sites can be robustly selective in the confined microdroplet limit, and other binding sites cannot be selective without additional protein forces stabilizing an optimal ligand configuration. In the case of the site S2 of KcsA, the analysis of computational models shows that selectivity can be realized over a wide range of conditions of structural flexibility. Because flexibility of the pore is critical to keep the free-energy barriers low for rapid ion conduction (16), it seems reasonable that the selectivity of the binding sites in K^+ channels would need to be closer to a confined microdroplet mechanism. In the case of LeuT, the computational models shows that the site Na1 is inherently selective for Na^+ because of the negatively charged ligand provided by the leucine substrate, while the site Na2 becomes selective only by virtue of the local geometric forces.

In all three cases examined here, the architectural stiffness provided by the surrounding protein structure serves to enhance

rather than oppose the selectivity that is already emerging in the confined microdroplet regime. The latter corresponds to an inherent trend that is robustly set by the number and type of ion-coordinating ligands. Additional computations with reduced models show that reverting such trend is possible, but only by making the local structure extremely stiff (see Section C in *SI Appendix*), leading to conditions that may be difficult to realistically achieve for most flexible biological macromolecules. This analysis suggests that stabilizing the local geometry to build upon the inherent selectivity set by the number and type of ion-coordinating ligands is probably one of the key *design principles* of ion-selective binding sites in proteins and other biological macromolecules. It will be interesting to examine the effect of selective binding of Zn^{2+}, Mg^{2+}, and Ca^{2+} on the folding and stabilization of RNA and DNA structures in the context of the present framework (40, 41).

Materials and Methods

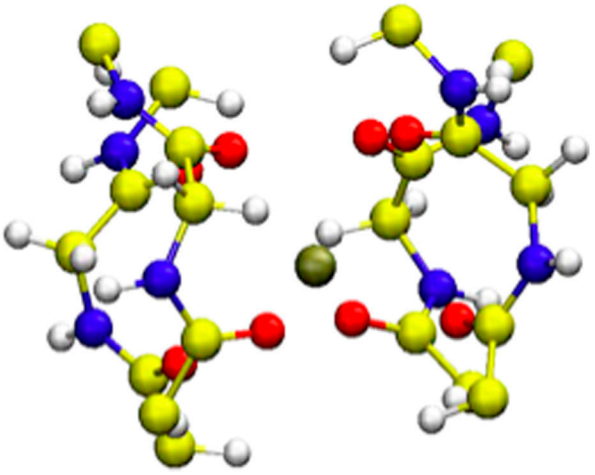
A reduced model of the K^+-selective S2 binding site of KcsA was constructed from four diglycine molecules (Fig. 1*A*). The resulting reduced model is similar to a model previously used by Asthagiri et al. (24). In addition, reduced models of the two Na^+-selective binding sites Na1 and Na2 in the LeuT transporter were constructed by including the molecular groups within 5 Å around the bound ion (Fig. 1*B*). For both systems, the range of confinement of the protein atoms in the reduced model (R_k in Eq. **3**) was extracted from over 1 ns well equilibrated all-atom MD simulations of the complete protein in a fully solvated lipid bilayer with either Na^+ or K^+ bound at the binding sites of interest (16, 17, 22). For the S2 site of the KcsA channel, the variables for the confinements include 52 Cartesian atomic positions and 16 distances between carbonyl oxygen atoms. For the binding sites of LeuT, the variables for the confinements include 43 and 37 Cartesian atomic positions for the sites Na1 and Na2, respectively. All details about the confinement used for KcsA and LeuT are given in Section D of *SI Appendix*. Geometric structural forces were introduced in the form of a harmonic restraining potential with respect to the configurations optimized from the X-ray structure of KcsA (3) and LeuT (5). Selectivity for K^+ and Na^+ was calculated from more than 4 ns FEP/MD simulations of the reduced model, postprocessed using the weighted histogram analysis method (42, 43). The MD simulations were performed with CHARMM c35a1 and PARAM27 force field (44, 45). The simulation methodology and PARAM27 force field parameters have been described previously (16–18). Additional computations accounting explicitly for induced polarization indicate that the conclusions are unchanged.

ACKNOWLEDGMENTS. This work was supported by the National Institute of Health via Grant GM-62342 (to H.Y. and B.R.) and the Canadian Institute for Health Research (CIHR) 200804MOP-186232 (to S.Y.N.). S.Y.N. is CIHR New Investigator and an Alberta Heritage Foundation Medical Scholar.

1. Hille B (2001) *Ion channels of excitable membranes* (Sinauer Associates, Sunderland, MA), 3rd ed.
2. Page MJ, Di Cera E (2006) Role of Na^+ and K^+ in enzyme function. *Physiol Rev* 86:1049–1092.
3. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R (2001) Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 Å resolution. *Nature* 414:43–48.
4. Zhou Y, MacKinnon R (2003) The occupancy of ions in the K^+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J Mol Biol* 333:965–975.
5. Yamashita A, Singh S, Kawate T, Jin Y, Gouaux E (2005) Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* 437:215–223.
6. Abramson J, et al. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301:610–615.
7. Boudker O, Ryan RM, Yernool D, Shimamoto K, Gouaux E (2007) Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 445:387–393.
8. Faham S, et al. (2008) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science* 321:810–814.
9. Murata T, Yamato I, Kakinuma Y, Leslie AG, Walker JE (2005) Structure of the rotor of the V-Type Na^+-ATPase from Enterococcus hirae. *Science* 308:654–659.
10. Morth JP, et al. (2007) Crystal structure of the sodium-potassium pump. *Nature* 450:1043–1049.
11. Shinoda T, Ogawa H, Cornelius F, Toyoshima C (2009) Crystal structure of the sodium-potassium pump at 2.4 Å resolution. *Nature* 459:446–450.
12. Gouaux E, Mackinnon R (2005) Principles of selective ion transport in channels and pumps. *Science* 310:1461–1465.
13. Pedersen C, Frensdor H (1972) Macrocyclic polyethers and their complexes. *Angew Chem Int Ed* 1:16–25.
14. Dietrich B (1985) Coordination chemistry of alkali and alkaline-earth cations with macrocyclic ligands. *J Chem Educ* 62:954–964.
15. Armstrong C, Bezanilla F (1972) Negative conductance caused by entry of sodium and cesium ions into potassium channels of squid axons. *J Gen Physiol* 60:588–608.
16. Bernèche S, Roux B (2001) Energetics of ion conduction through the K^+ channel. *Nature* 414:473–77.
17. Noskov SY, Bernèche S, Roux B (2004) Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* 431:830–834.
18. Noskov SY, Roux B (2007) Importance of hydration and dynamics on the selectivity of the KcsA and NaK channels. *J Gen Physiol* 129:135–143.
19. Neyton J, Miller C (1988) Potassium blocks barium permeation through a calcium-activated potassium channel. *J Gen Physiol* 92:549–567.
20. Neyton J, Miller C (1988) Discrete Ba^{2+} block as a probe of ion occupancy and pore structure in the high-conductance Ca^{2+}-activated K^+ channel. *J Gen Physiol* 92:569–586.
21. Vergara C, Alvarez O, Latorre R (1999) Localization of the K^+ lock-In and the Ba^{2+} binding sites in a voltage-gated calcium-modulated channel. Implications for survival of K^+ permeability. *J Gen Physiol* 114:365–376.
22. Noskov SY, Roux B (2008) Control of ion selectivity in LeuT: two Na^+ binding sites with two different mechanisms. *J Mol Biol* 377:804–818.
23. Krishnamurthy H, Piscitelli CL, Gouaux E (2009) Unlocking the molecular secrets of sodium-coupled transporters. *Nature* 459:347–355.
24. Asthagiri D, Pratt L, Paulaitis M (2006) Role of fluctuations in a snug-fit mechanism of KcsA channel selectivity. *J Chem Phys* 125:24701.
25. Varma S, Rempe S (2007) Tuning ion coordination architectures to enable selective partitioning. *Biophys J* 93:1093–1099.
26. Bostick DL, Brooks CL (2007) Selectivity in K^+ channels is due to topological control of the permeant ion's coordinated state. *Proc Natl Acad Sci USA* 104:9260–9265.
27. Yu HB, Noskov SY, Roux B (2009) Hydration number, topological control, and ion selectivity. *J Phys Chem B* 113:8725–8730.
28. Asthagiri D, et al. (2010) Ion selectivity from local configurations of ligands in solutions and ion channels. *Chem Phys Lett* 485:1–7.
29. Beglov D, Roux B (1994) Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J Chem Phys* 100:9050–9063.
30. Weeks JD, Chandler D, Andersen HC (1971) Role of repulsive forces in determining equilibrium structure of simple liquids. *J Chem Phys* 54:5237.
31. Hummer G, Garde S, Garca AE, Pohorille A, Pratt LR (1996) An information theory model of hydrophobic interactions. *Proc Natl Acad Sci USA* 93:8951–8955.
32. Roux B (2010) Exploring the ion selectivity properties of a large number of simplified binding site models. *Biophys J* 98:2877–2885.
33. Eisenman G (1962) Cation selective electrodes and their mode of operation. *Biophys J* 2 (supp 2):259–323.
34. Boda D, et al. (2009) Ionic selectivity in L-type calcium channels by electrostatics and hard-core repulsion. *J Gen Physiol* 133:497–509.
35. Yamashita M, Wesson L, Eisenman G, Eisenberg D (1990) Where metal ions bind in proteins. *Proc Natl Acad Sci USA* 87:5648–5652.
36. Cordero-Morales J, et al. (2006) Molecular determinants of gating at the potassium-channel selectivity filter. *Nat Struct Mol Biol* 13:311–318.
37. Valiyaveetil F, Leonetti M, Muir T, Mackinnon R (2006) Ion selectivity in a semisynthetic K^+ channel locked in the conductive conformation. *Science* 314:1004–1007.
38. Åqvist J, Alvarez O, Eisenman G (1992) Ion-selective properties of a small ionophore in methanol studied by free energy perturbation simulations. *J Phys Chem* 96:10019–10025.
39. Varma S, Sabo D, Rempe S (2008) K^+/Na^+ selectivity in K channels and valinomycin: over-coordination versus cavity-size constraints. *J Mol Biol* 376:13–22.
40. Draper DE, Grilley D, Soto AM (2005) Ions and RNA folding. *Annu Rev Bioph Biom* 34:221–243.
41. Woodson SA (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* 9:104–109.
42. Kumar S, Bouzida D, Swendsen R, Kollman P, Rosenberg J (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13:1011–1021.
43. Souaille M, Roux B (2001) Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computers Physics Communications* 135:40–57.
44. Brooks BR, et al. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614.
45. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.

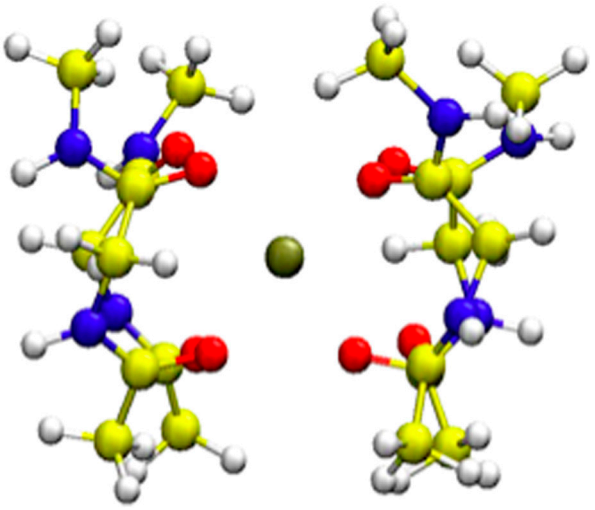
Supporting Information

Yu et al. 10.1073/pnas.1007150107



Movie S1. The S2 site in an all-atom molecular dynamics simulation of KcsA in a fully solvated membrane.

Movie S1 (MPG)



Movie S2. The S2 site in a molecular dynamics simulation of the reduced model enforced with the generic confinement potential according to the distribution of $\rho_0(X)$ in Eq. **4**.

Movie S2 (MPG)

Supplementary Information for "Two mechanisms of ion selectivity in protein binding sites" by Yu, Noskov and Roux

A) Additional theoretical developments

The atomic fluctuations are obtained from all-atom MD simulations of the solvated protein. The confinement function $H_{\rm c}(\mathbf{X})$, then is constructed with Heaviside step-functions to match the upper bounds of fluctuations, whether an ion of type i or j is bound. It is particularly important that $H_{\rm c}$ be independent of ion type. This implies that imposing the confinement on the all-atom system costs no free energy because $H_{\rm c}$ only forbids configurations that are never observed in the first place. The relative free energy of ion i and j in the binding site is not affected by the confinement,

$$\begin{aligned} e^{-\beta\Delta G_{ij}^{\rm site}} &= \frac{\int_{\rm site} d\mathbf{R}\ e^{-\beta U_i(\mathbf{R})}}{\int_{\rm site} d\mathbf{R}\ e^{-\beta U_j(\mathbf{R})}} \\ &= \frac{\int_{\rm site} d\mathbf{R}\ H_{\rm c}(\mathbf{X})\ e^{-\beta U_i(\mathbf{R})}}{\int_{\rm site} d\mathbf{R}\ H_{\rm c}(\mathbf{X})\ e^{-\beta U_j(\mathbf{R})}} \\ &= \frac{\int_{\rm site} d\mathbf{X}\, d\mathbf{Y}\ H_{\rm c}(\mathbf{X})\ e^{-\beta U_i(\mathbf{X},\mathbf{Y})}}{\int_{\rm site} d\mathbf{X}\, d\mathbf{Y}\ H_{\rm c}(\mathbf{X})\ e^{-\beta U_j(\mathbf{X},\mathbf{Y})}} \\ &= \frac{\int_{\rm site} d\mathbf{X}\ H_{\rm c}(\mathbf{X})\ e^{-\beta[U_i^{\rm il}(\mathbf{X})+U^{\rm ll}(\mathbf{X})+\Delta W^{\rm site}(\mathbf{X})]}}{\int_{\rm site} d\mathbf{X}\ H_{\rm c}(\mathbf{X})\ e^{-\beta[U_j^{\rm il}(\mathbf{X})+U^{\rm ll}(\mathbf{X})+\Delta W^{\rm site}(\mathbf{X})]}} \end{aligned} \tag{1}$$

Comparing with the Eq. [2] in the main text shows that

$$e^{-\beta\Delta W^{\rm site}(\mathbf{X})} \equiv H_{\rm c}(\mathbf{X}) e^{-\beta\Delta W^{\rm site}(\mathbf{X})} \tag{2}$$

This means that the confinement imposed by $H_{\rm c}(\mathbf{X})$ can be treated as being implicitly part of the function $\Delta W^{\rm site}(\mathbf{X})$. By construction, the step-function $H_{\rm c}$ is equal to zero or one, and $H_{\rm c}^2 = H_{\rm c}$. The remainder $\Delta W_{\rm g}^{\rm site}$ can thus safely be defined as,

$$e^{-\beta\Delta W^{\rm site}(\mathbf{X})} \equiv H_{\rm c}(\mathbf{X}) e^{-\beta\Delta W_{\rm g}^{\rm site}(\mathbf{X})} \tag{3}$$

Similar operations are carried out to define the so-called "cavity potential function" in hard-sphere liquids (see [1]). The present construction based on Eq. (3) defines a minimal default model with probability distribution,

$$\rho_0(\mathbf{X}) = H_{\rm c}(\mathbf{X})\ e^{-\beta[U_i^{\rm il}(\mathbf{X})+U^{\rm ll}(\mathbf{X})]} \tag{4}$$

that incorporates the generic effect of architectural confinement by the surounding protein structure in an idealized fashion: each atom k of the reduced subsystem can fluctuate and make arbitrary excursions, as long as it remains confined within its spherical volume V_k. To determine $\Delta W_{\rm g}^{\rm site}$, we seek to maximize the cross-entropy η [2]

$$\eta = -\int d\mathbf{X}\ \rho(\mathbf{X})\ \ln\left[\rho(\mathbf{X})/\rho_0(\mathbf{X})\right] \tag{5}$$

under the constraint that the probability distribution $\rho(\mathbf{X})$ is normalized, and that the atomic fluctuations of the all atoms in the reduced subsystem match the fluctuations extracted from all-atom MD simulations. Denoting the set of coordinates $\mathbf{X}$ as $\{x_1, x_2, \ldots, x_n\}$, the average moments

extracted from the all-atom system can be written as

$$F^{(i_1,i_2,\ldots i_n)} = \left\langle (x_1)^{i_1} \ldots (x_n)^{i_n} \right\rangle \tag{6}$$

Then, the constrained optimization problem is solved by introducing a complete set of Lagrange multiplier $\lambda_{\rm g}^{(i_1,i_2,\ldots i_n)}$ to satisfy the condition on the normalization and reproduce the atomic fluctuation moments $F^{(i_1,i_2,\ldots i_n)}$,

$$\eta = -\int d\mathbf{X}\; \rho(\mathbf{X})\; \ln\left[\rho(\mathbf{X})/\rho_0(\mathbf{X})\right] + \sum_{\rm all}\; \lambda_{\rm g}^{(i_1,i_2,\ldots i_n)}\; \left[\int d\mathbf{X}\; \rho(\mathbf{X})\left((x_1)^{i_1} \ldots (x_n)^{i_n}\right) - F^{(i_1,i_2,\ldots i_n)}\right] \tag{7}$$

where the sum runs to include the atomic fluctuations *to all order*. The normalization condition enters via the term $F^{(0,0,\ldots,0)} = 1$ corresponding to the zeroth moment with $i_1 = 0, i_2 = 0, \ldots, i_n = 0$. The resulting normalized distribution can be formally written as,

$$\rho(\mathbf{X}) = \frac{H_{\rm c}^{\rm site}(\mathbf{X})\; e^{-\beta[U_i^{\rm il}(\mathbf{X}) + U^{\rm ll}(\mathbf{X}) + \Delta W_{\rm g}^{\rm site}(\mathbf{X})]}}{\int d\mathbf{X}\; H_{\rm c}^{\rm site}(\mathbf{X})\; e^{-\beta[U_i^{\rm il}(\mathbf{X}) + U^{\rm ll}(\mathbf{X}) + \Delta W_{\rm g}^{\rm site}(\mathbf{X})]}} \tag{8}$$

where

$$\Delta W_{\rm g}^{\rm site}(\mathbf{X}) = \sum_{\rm all}\; \lambda_{\rm g}^{(i_1,i_2,\ldots i_n)}\left[(x_1)^{i_1} \ldots (x_n)^{i_n}\right] \tag{9}$$

By virtue of the confinement condition implying that $\rho(\mathbf{X}) = 0$ when any of the variable $\mathbf{X}_k$ lies outside the microscopic volumes V_k, all moments $F^{(i_1,i_2,\ldots i_n)}$ are finite and bounded. For this reason, the distribution $\rho(\mathbf{X})$ is uniquely determined by $\rho_0(\mathbf{X})$ and the series of moments $F^{(i_1,i_2,\ldots i_n)}$. In principle, the above procedure can be carried out to match the average moments extracted from the all-atom MD exactly to all order, thus yielding the exact form of $\Delta W_{\rm g}^{\rm site}$. However, it is sufficient to consider only the quadratic fluctuations for the sake of the analysis presented in the article aimed at delineating the two dominant mechanisms of ion selectivity.

Ultimately, it is important that the reference model based on $\rho_0(\mathbf{X})$ provides a good starting point to reproduce the physical behavior of the exact all-atom system. This way, the reduced system with $\lambda_{\rm g} = 0$ is already providing a reasonable "mimic system". To illustrate the character of the mimic system, two movie animations of the KcsA binding site S2 as Supplemental Information. The first movie, **kcsa-all-atom.mpg**, shows the dynamics of the binding site taken from an all-atom MD simulation of KcsA in a fully solvated membrane. The second movie, **kcsa-reduced-model.mpg**, shows the dynamics of the reduced model with only the confinement according to the distribution $\rho_0(\mathbf{X})$. The behavior of the active site is very similar in both cases, and only careful observation reveals that the fluctuations are slightly larger in the reduced model simulations (as expected).

B) Generic uniform confinement

The trends observed in the reduced systems are not highly sensitive to the confinement radius R_k. To illustrate this, the relative free energy of K^+ and Na^+ in the KcsA and LeuT binding sites was recomputed with FEP/MD assuming a uniform generic confinement. In these computations, the geometric contribution, $\lambda_{\rm g}$, is set to zero. The results of FEP/MD computations are shown in Fig. 1. The results from FEP/MD show that the main trend holds for a wide range of confinement radius. It should be noted that imposing a confinement radius smaller than 0.5 Å is akin to assume that the

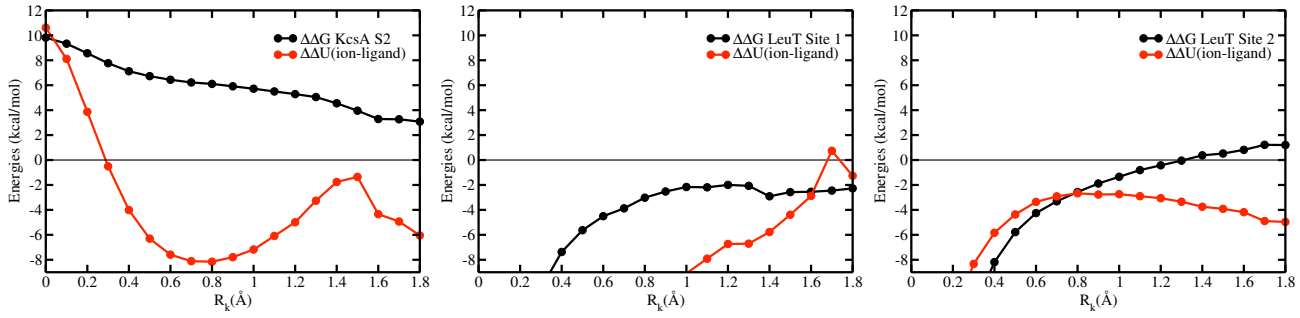


Figure 1: Effect of changing the confinement radius in the reduced binding site models of KcsA and LeuT.

binding site is very rigid and that the snug-fit limit is reproduced with R_k approaching 0.0 Å. This analysis indicates that, in the absence of addition information from all-atom MD, meaningful trends can be extracted from reduced models with a generic confinement radius of 1.0 to 1.4 Å. Such values correspond to typical thermal fluctuations extracted from all-atom MD simulations.

C) Reduced systems with adapted optimal geometry

The relative free energies calculated in the KcsA and LeuT reduced systems with λ_g set to zero correspond to the confined microdroplet limit. Those free energies set the natural trend arising solely from the number and the type of ligands coordinating the ion. An interesting question is whether it is possible to revert the natural trend if the architectural forces (i.e., $\Delta W_g^{\rm site}$) were reinforcing a coordination geometry adapted to best-fit K^+ or Na^+. To address this question, putative K^+- and Na^+-adapted geometries must be constructed for LeuT and KcsA, respectively. The K^+-adapted optimal geometries of each LeuT reduced sites were constructed by performing energy minimization in the presence of a K^+ ion in the binding site. This yields the optimal geometry $\overline{\mathbf{r}}_k[{\rm K-adapted}]$ for those two binding sites. The Na^+-adapted optimal geometry of the KcsA reduced model was constructed in the same way with energy minimization in the presence of a bound Na^+. This yields the optimal geometry $\overline{\mathbf{r}}_k[{\rm Na-adapted}]$. Then, the relative free energy was recomputed as a function of the force constant λ_g using FEP/MD simulations. The results are shown in Fig. 2. It is clear that reverting the natural tendency of a reduced model in the confined microdroplet limit may be difficult. A fairly stiff restoring force λ_g on the order of 50-100 kcal/mol/Å^2 is required to dictate an optimal geometry and produce selectivity for Na^+ over K^+ from the ligands of the KcsA S2 site. A similar trend is observed for the site Na1 of LeuT. On the other hand, the selectivity of the site Na2 of LeuT can be swayed to favor Na^+ or K^+ by virtue of the geometric architectural forces. This is understandable because this site displays no selectivity in the confined microdroplet limit. It is "selectivity-neutral" by virtue of the number and type of ligands.

These results suggest the following "design principles" for a general ion-selective binding site. Let us consider a situation where one is interested in designing a selective ion binding site at one location in a macromolecular structure. It is assumed that it is possible to modify chemically the local structure such that N ligands are made available to coordinate the desired ion. The key question is to how to best choose the number and type of ligands such that this putative binding site will have the desired selectivity. To answer this question, one could construct a reduced model of the binding

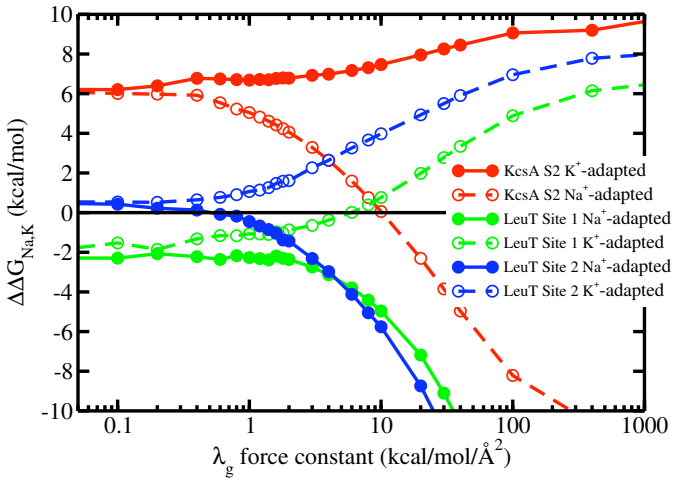


Figure 2: Selectivity in reduced models in which the optimal geometry is adapted to best-fit K^+ or Na^+. The curves with solid lines correspond to the results shown in the article. The curves with dashed lines correspond to putative KcsA and LeuT sites stabilizing an optimal geometry (via λ_g) adapted by energy minimization for Na+ or K^+, respectively. Except for the site Na2 of LeuT, the onset of selectivity is established only for large values of λ_g, indicating that a fairly stiff geometry is required to counter-balance the natural trend observed in the confined microdroplet limit ($\lambda_g = 0$).

site with some generic confinements, and compute the relative free energy of K^+ and Na^+ using FEP/MD simulations. These calculations will display the natural selectivity of the reduced binding site in the confined microdroplet limit, which is predominantly set by the number and the type of ligands. If there is a hint of the desired selectivity, or no selectivity at all, it should be possible to stabilize the optimal coordination geometry for the desired ion and enhance the selectivity. If, on the other hand, the wrong trend is displayed in the confined microdroplet, it will be difficult to correct for this by trying to stiffen some optimal coordinating geometry for the desired ion. The simplest course of action is to go back and choose a different number and type of ligands to achieve the desired selectivity.

D) Confinement of reduced models

The S2 binding site of KcsA

The confinement potential utilized for the S2 binding site of KcsA comprises two components. The first component is defined as the maximum displacement of the atomic positions relative to a fixed point in space. The parameters are given in a CHARMM coordinate format, with column X, Y, Z being the position of the center $\bar{\mathbf{r}}_k$ and the W array being radius R_k in Eq. [3] in the main text, respectively. The W entries equal to zero imply that no confinement is applied to this atom.

```
* S2 KcsA
*
   77
    1    1 GLY  N     -4.64887  -0.83509   9.50473 B    77     1.11800
    2    1 GLY  HN    -5.13162  -0.96595   9.31059 B    77     1.50200
    3    1 GLY  CA    -4.12878  -1.23317  10.78173 B    77     1.08400
    4    1 GLY  HA1   -3.40890  -1.73482  10.80684 B    77     1.33100
    5    1 GLY  HA2   -4.93703  -1.57129  11.42373 B    77     1.18300
    6    1 GLY  C     -3.91184   0.08970  11.94418 B    77     1.12900
    7    1 GLY  O     -2.94688   0.10042  11.71373 B    77     1.55800
    8    1 GLY  NT    -4.84581   0.39329  12.52523 B    77     1.09100
    9    1 GLY  HNT   -5.75408   0.15978  12.25791 B    77     1.18200
   10    1 GLY  CAT   -4.89088   1.48798  13.47304 B    77     1.07200
   11    1 GLY  HT1   -4.57916   2.44754  12.99660 B    77     0.00000
   12    1 GLY  HT2   -4.21093   1.29591  14.33593 B    77     0.00000
   13    1 GLY  HT3   -5.92118   1.62012  13.87340 B    77     0.00000
   14    1 GLY  CAY   -4.00278   0.28112   7.37116 B    77     1.05700
   15    1 GLY  HY1   -3.24541   0.93240   6.88598 B    77     0.00000
   16    1 GLY  HY2   -4.12294  -0.64513   6.77327 B    77     0.00000
   17    1 GLY  HY3   -4.96939   0.82090   7.42553 B    77     0.00000
   18    1 GLY  CY    -3.52841  -0.07013   8.76067 B    77     1.08700
   19    1 GLY  OY    -2.52505  -0.11182   9.09946 B    77     1.22600
   20    2 GLY  N     -2.19409   4.94613   8.96996 C    77     1.02000
   21    2 GLY  HN    -2.24598   6.08285   8.43493 C    77     1.37800
   22    2 GLY  CA    -2.70641   5.01697  10.05355 C    77     1.10500
   23    2 GLY  HA1   -3.46214   4.32659  10.13897 C    77     1.29700
   24    2 GLY  HA2   -3.28208   5.83456  10.11106 C    77     1.29900
   25    2 GLY  C     -1.76919   4.86712  11.23659 C    77     1.09900
   26    2 GLY  O     -1.75490   4.09548  11.28954 C    77     1.48800
   27    2 GLY  NT    -1.31894   5.93100  11.93388 C    77     1.12300
   28    2 GLY  HNT   -1.67602   6.80704  11.65500 C    77     1.16100
   29    2 GLY  CAT   -0.24696   5.99186  12.87711 C    77     1.09000
   30    2 GLY  HT1    0.64864   5.45525  12.48882 C    77     0.00000
   31    2 GLY  HT2   -0.54832   5.50931  13.83445 C    77     0.00000
   32    2 GLY  HT3    0.03669   7.04651  13.08899 C    77     0.00000
   33    2 GLY  CAY   -0.81216   4.39049   6.82920 C    77     1.07100
   34    2 GLY  HY1    0.16495   3.86935   6.74337 C    77     0.00000
   35    2 GLY  HY2   -1.51476   3.96081   6.08616 C    77     0.00000
   36    2 GLY  HY3   -0.67103   5.47043   6.63011 C    77     0.00000
   37    2 GLY  CY    -1.36706   4.18442   8.21649 C    77     1.06000
   38    2 GLY  OY    -1.22304   3.41142   8.45477 C    77     1.40000
   39    3 GLY  N      4.03509   2.59256   9.62796 D    77     1.03200
   40    3 GLY  HN     4.50148   2.88034   9.53588 D    77     1.49700
   41    3 GLY  CA     3.34242   3.57992  10.39441 D    77     1.20800
   42    3 GLY  HA1    2.69244   4.04598  10.34533 D    77     1.62300
   43    3 GLY  HA2    4.12096   3.85002  10.70243 D    77     1.60000
   44    3 GLY  C      3.09656   2.50590  11.85192 D    77     1.04700
   45    3 GLY  O      2.07879   2.30212  11.82239 D    77     1.35700
   46    3 GLY  NT     3.86380   2.16979  12.72009 D    77     1.11400
   47    3 GLY  HNT    4.43620   2.70630  12.30309 D    77     1.33200
   48    3 GLY  CAT    3.68712   1.15789  13.69221 D    77     1.16900
   49    3 GLY  HT1    3.35986   0.20112  13.22228 D    77     0.00000
   50    3 GLY  HT2    2.91221   1.45859  14.43426 D    77     0.00000
   51    3 GLY  HT3    4.63800   0.97166  14.23921 D    77     0.00000
   52    3 GLY  CAY    3.69380   1.32209   7.54516 D    77     1.05300
   53    3 GLY  HY1    2.95043   0.67782   7.03029 D    77     0.00000
   54    3 GLY  HY2    4.05174   2.10118   6.84300 D    77     0.00000
   55    3 GLY  HY3    4.54526   0.69880   7.88397 D    77     0.00000
   56    3 GLY  CY     3.02610   1.97702   8.74189 D    77     1.07000
   57    3 GLY  OY     1.88018   1.92906   8.87224 D    77     1.34400
   58    4 GLY  N      1.32823  -2.78946  10.07423 A    77     1.08300
   59    4 GLY  HN     1.43283  -3.71749  10.03683 A    77     1.42200
   60    4 GLY  CA     1.90688  -2.63456  11.09797 A    77     1.19500
   61    4 GLY  HA1    2.33895  -1.91769  10.96887 A    77     1.48800
   62    4 GLY  HA2    2.37273  -3.23512  11.56719 A    77     1.57800
   63    4 GLY  C      0.86306  -2.34886  12.24334 A    77     1.09500
   64    4 GLY  O      0.81213  -1.46278  12.17049 A    77     1.51900
   65    4 GLY  NT     0.38985  -2.90084  12.77926 A    77     0.99800
   66    4 GLY  HNT    0.86825  -4.00821  12.92656 A    77     1.31400
   67    4 GLY  CAT   -0.76324  -2.97653  13.94003 A    77     1.07200
   68    4 GLY  HT1   -1.49106  -2.14257  13.81008 A    77     0.00000
   69    4 GLY  HT2   -0.30895  -2.88181  14.95179 A    77     0.00000
   70    4 GLY  HT3   -1.32163  -3.93705  13.88415 A    77     0.00000
   71    4 GLY  CAY    0.40915  -2.98550   7.87001 A    77     1.01500
   72    4 GLY  HY1   -0.68424  -2.79282   7.79145 A    77     0.00000
   73    4 GLY  HY2    0.90315  -2.54174   6.97738 A    77     0.00000
```

```
  74    4 GLY  HY3    0.58314  -4.08192   7.87008 A    77     0.00000
  75    4 GLY  CY     0.92835  -2.36978   9.07829 A    77     1.02000
  76    4 GLY  OY     0.79681  -1.39207   9.09792 A    77     1.28500
  77    5 POT  POT   -0.29568   0.95990  10.39158 ION1 1      0.00000
```

The second component is defined as the maximum atom-atom distance for selected pairs of atoms.

```
Max    -atom1--  -atom2--
2.60 - A  77  O  B  77  O
5.85 - A  77  O  B  77  O
2.60 - B  77  O  C  77  O
5.85 - B  77  O  C  77  O
2.60 - C  77  O  D  77  O
5.85 - C  77  O  D  77  O
2.60 - D  77  O  A  77  O
5.85 - D  77  O  A  77  O
2.55 - A  77  OY B  77  OY
5.30 - A  77  OY B  77  OY
2.55 - B  77  OY C  77  OY
5.30 - B  77  OY C  77  OY
2.55 - C  77  OY D  77  OY
5.30 - C  77  OY D  77  OY
2.55 - D  77  OY A  77  OY
5.30 - D  77  OY A  77  OY
4.00 - A  77  O  C  77  O
6.85 - A  77  O  C  77  O
4.00 - B  77  O  D  77  O
6.85 - B  77  O  D  77  O
4.15 - A  77  OY C  77  OY
6.05 - A  77  OY C  77  OY
4.15 - B  77  OY D  77  OY
6.05 - B  77  OY D  77  OY
2.60 - A  77  O  A  77  OY
4.35 - A  77  O  A  77  OY
2.60 - B  77  O  B  77  OY
4.35 - B  77  O  B  77  OY
2.60 - C  77  O  C  77  OY
4.35 - C  77  O  C  77  OY
2.60 - D  77  O  D  77  OY
4.35 - D  77  O  D  77  OY
```

The Na1 and Na2 binding sites of LeuT

The confinement potential is defined as the maximum displacement of the atomic positions. The parameters are given in a CHARMM coordinate format, with column x, y, z being the position of the center $\overline{\mathbf{r}}_k$ and the wmain being radius R_k in Eq. 3 in the main text, respectively. The W entries equal to zero imply that no confinement is applied to this atom.

Site Na1:

```
* Na1 LeuT
*
   81
    1    1 ALA  N     -4.48081  -4.64420 -11.79484 PEPA 22     1.46700
    2    1 ALA  HN    -4.89939  -4.98648 -12.63667 PEPA 22     0.00000
    3    1 ALA  CA    -5.40563  -4.89222 -10.74875 PEPA 22     1.39700
    4    1 ALA  HA    -5.62682  -3.88769 -10.24458 PEPA 22     0.00000
    5    1 ALA  CB    -6.69048  -5.46111 -11.47168 PEPA 22     1.39500
    6    1 ALA  HB1   -7.61508  -5.55596 -10.84547 PEPA 22     0.00000
    7    1 ALA  HB2   -6.58102  -6.32203 -11.98911 PEPA 22     0.00000
    8    1 ALA  HB3   -7.07498  -4.75454 -12.25364 PEPA 22     0.00000
    9    1 ALA  C     -4.86612  -5.66815  -9.60906 PEPA 22     1.22700
   10    1 ALA  O     -4.78615  -5.13223  -8.47641 PEPA 22     1.45300
   11    2 VAL  N     -4.53507  -6.93257  -9.78260 PEPA 23     1.33300
   12    2 VAL  HN    -4.30741  -7.19782 -10.68002 PEPA 23     0.00000
   13    2 VAL  CA    -3.99151  -7.79372  -8.77376 PEPA 23     1.25500
   14    2 VAL  HA    -4.74773  -7.84537  -7.97203 PEPA 23     0.00000
   15    2 VAL  CG2   -5.28338  -9.79101  -9.64792 PEPA 23     1.50300
   16    2 VAL  HG21  -6.08104  -9.57368  -8.86190 PEPA 23     0.00000
   17    2 VAL  HG22  -5.33442 -10.93680  -9.77930 PEPA 23     0.00000
   18    2 VAL  HG23  -5.57129  -9.23543 -10.52472 PEPA 23     0.00000
   19    2 VAL  C     -2.63555  -7.28323  -8.22651 PEPA 23     1.24500
   20    2 VAL  O     -1.63616  -7.15601  -8.95264 PEPA 23     1.45300
   21    3 GLY  N     -2.52937  -7.08193  -6.90302 PEPA 24     1.31900
   22    3 GLY  HN    -3.33184  -7.24492  -6.29610 PEPA 24     0.00000
   23    3 GLY  CA    -1.37813  -6.56942  -6.26564 PEPA 24     1.26800
   24    3 GLY  HA1   -1.09824  -5.52959  -6.62736 PEPA 24     0.00000
   25    3 GLY  HA2   -0.61090  -7.26881  -6.49779 PEPA 24     0.00000
   26    3 GLY  C     -1.54909  -6.67550  -4.76397 PEPA 24     1.28500
   27    3 GLY  O     -2.34566  -7.48950  -4.23121 PEPA 24     1.64400
   28    4 GLY  N     -3.01325  -4.72510  -2.71252 PEPA 26     1.39200
   29    4 GLY  HN    -2.76725  -4.49326  -3.60216 PEPA 26     0.00000
   30    4 GLY  CA    -4.33893  -4.34292  -2.25966 PEPA 26     1.54900
   31    4 GLY  HA1   -4.70701  -3.81496  -3.16597 PEPA 26     0.00000
   32    4 GLY  HA2   -4.29150  -3.73550  -1.37946 PEPA 26     0.00000
   33    4 GLY  C     -5.27464  -5.58255  -2.02276 PEPA 26     1.58700
   34    4 GLY  O     -6.13720  -5.59557  -1.18917 PEPA 26     1.87200
   35    5 ASN  N     -4.97838  -6.64670  -2.77703 PEPA 27     1.32700
   36    5 ASN  HN    -4.27007  -6.54229  -3.50105 PEPA 27     0.00000
   37    5 ASN  CA    -5.70771  -7.92277  -2.63846 PEPA 27     1.31600
   38    5 ASN  HA    -6.74807  -7.69983  -2.53151 PEPA 27     0.00000
   39    5 ASN  CB    -5.51978  -8.79483  -3.81436 PEPA 27     1.31200
   40    5 ASN  HB1   -5.97581  -9.77648  -3.67018 PEPA 27     0.00000
   41    5 ASN  HB2   -4.43633  -8.98068  -3.97342 PEPA 27     0.00000
   42    5 ASN  CG    -5.84466  -8.16210  -5.13499 PEPA 27     1.35100
   43    5 ASN  OD1   -5.05583  -7.39220  -5.65381 PEPA 27     1.36300
   44    5 ASN  ND2   -7.04811  -8.49758  -5.65121 PEPA 27     2.02900
   45    5 ASN  HD21  -7.18784  -8.21696  -6.59782 PEPA 27     0.00000
   46    5 ASN  HD22  -7.74238  -8.95615  -5.14990 PEPA 27     0.00000
   47    6 PHE  C     -7.07425  -0.47374  -5.14500 PEPA 251    1.37200
   48    6 PHE  O     -6.11063  -0.59362  -5.95852 PEPA 251    1.44700
   49    7 THR  N     -7.80608  -1.56003  -4.85277 PEPA 252    1.32900
   50    7 THR  HN    -8.57516  -1.40809  -4.22665 PEPA 252    0.00000
   51    7 THR  CA    -7.62203  -2.91768  -5.40950 PEPA 252    1.26100
   52    7 THR  HA    -6.59691  -3.04106  -5.35871 PEPA 252    0.00000
   53    7 THR  CB    -8.37455  -4.03279  -4.73924 PEPA 252    1.35900
   54    7 THR  HB    -8.20211  -4.03243  -3.64557 PEPA 252    0.00000
   55    7 THR  OG1   -7.83232  -5.34314  -5.09241 PEPA 252    1.71100
   56    7 THR  HG1   -8.30261  -5.94141  -4.51355 PEPA 252    0.00000
   57    7 THR  CG2   -9.94075  -3.99243  -4.97911 PEPA 252    1.39100
   58    7 THR  HG21 -10.40829  -4.59541  -4.20398 PEPA 252    0.00000
   59    7 THR  HG22 -10.23948  -4.47044  -5.93823 PEPA 252    0.00000
   60    7 THR  HG23 -10.31515  -2.93996  -4.96457 PEPA 252    0.00000
   61    7 THR  C     -7.89688  -3.06788  -6.95122 PEPA 252    1.17800
   62    7 THR  O     -7.03304  -3.58694  -7.60694 PEPA 252    1.22200
   63    8 LEU  N     -8.96861  -2.45400  -7.45039 PEPA 253    1.20800
   64    8 LEU  HN    -9.63466  -1.94032  -6.92111 PEPA 253    0.00000
   65    8 LEU  CA    -9.15805  -2.57887  -8.93708 PEPA 253    1.23800
   66    8 LEU  HA    -8.86834  -3.58906  -9.26383 PEPA 253    0.00000
   67    9 ASN  CG    -8.80016  -7.32268  -8.26319 PEPA 284    1.38500
   68    9 ASN  OD1   -7.54947  -7.42695  -8.32864 PEPA 284    1.53700
   69    9 ASN  ND2   -9.44924  -6.83353  -7.17989 PEPA 284    1.74700
   70    9 ASN  HD21  -9.00469  -6.19215  -6.56059 PEPA 284    0.00000
   71    9 ASN  HD22 -10.46213  -6.90326  -7.08779 PEPA 284    0.00000
   72   10 SOD  SOD   -5.94865  -5.26686  -6.33815 ION4 1      0.00000
   73   11 LEU  N     -4.30479  -2.17549  -7.97037 ION3 1      1.18700
```

```
   74   11 LEU  HT1   -4.15557  -1.77182  -8.89003 ION3 1      0.00000
   75   11 LEU  HT2   -4.98310  -1.67557  -7.36414 ION3 1      0.00000
   76   11 LEU  HT3   -4.60807  -3.20786  -8.00681 ION3 1      0.00000
   77   11 LEU  CA    -3.06718  -2.31030  -7.17436 ION3 1      1.17500
   78   11 LEU  HA    -2.22624  -2.73718  -7.81270 ION3 1      0.00000
   79   11 LEU  C     -3.26218  -3.23866  -5.93652 ION3 1      1.17200
   80   11 LEU  OT1   -2.25587  -3.44439  -5.18847 ION3 1      1.35000
   81   11 LEU  OT2   -4.38742  -3.77343  -5.68337 ION3 1      1.41200
```

Site Na2:

```
* Na2 LeuT
*
   65
    1    1 GLY  N     -2.70469  -7.59600 -14.07503 PEPA 20     1.46700
    2    1 GLY  HN    -2.67925  -7.71740 -15.08395 PEPA 20     0.00000
    3    1 GLY  CA    -1.47138  -7.36084 -13.44406 PEPA 20     1.63000
    4    1 GLY  HA1   -0.79086  -7.58176 -14.26876 PEPA 20     0.00000
    5    1 GLY  HA2   -1.29001  -8.03109 -12.63918 PEPA 20     0.00000
    6    1 GLY  C     -1.46963  -5.97171 -12.80873 PEPA 20     1.44600
    7    1 GLY  O     -0.93499  -5.87891 -11.71060 PEPA 20     1.72300
    8    2 ASN  N     -2.10679  -4.97874 -13.49629 PEPA 21     1.49000
    9    2 ASN  HN    -2.34657  -5.19075 -14.47332 PEPA 21     0.00000
   10    2 ASN  CA    -2.53088  -3.68098 -12.93543 PEPA 21     1.33700
   11    2 ASN  HA    -1.65980  -3.19620 -12.51839 PEPA 21     0.00000
   12    2 ASN  C     -3.38067  -3.88233 -11.74850 PEPA 21     1.33100
   13    2 ASN  O     -3.10501  -3.45553 -10.64251 PEPA 21     1.53100
   14    3 VAL  N     -4.53507  -6.93257  -9.78260 PEPA 23     1.33300
   15    3 VAL  HN    -4.30741  -7.19782 -10.68002 PEPA 23     0.00000
   16    3 VAL  CA    -3.99151  -7.79372  -8.77376 PEPA 23     1.25500
   17    3 VAL  HA    -4.74773  -7.84537  -7.97203 PEPA 23     0.00000
   18    3 VAL  C     -2.63555  -7.28323  -8.22651 PEPA 23     1.24500
   19    3 VAL  O     -1.63616  -7.15601  -8.95264 PEPA 23     1.45300
   20    4 GLY  N     -2.52937  -7.08193  -6.90302 PEPA 24     1.25900
   21    4 GLY  HN    -3.33184  -7.24492  -6.29610 PEPA 24     0.00000
   22    4 GLY  CA    -1.37813  -6.56942  -6.26564 PEPA 24     1.26800
   23    4 GLY  HA1   -1.09824  -5.52959  -6.62736 PEPA 24     0.00000
   24    4 GLY  HA2   -0.61090  -7.26881  -6.49779 PEPA 24     0.00000
   25    5 PHE  C      3.18381 -10.08145  -8.52876 PEPA 348    1.36900
   26    5 PHE  O      2.66014  -9.50800  -9.46280 PEPA 348    1.76400
   27    6 ALA  N      2.63165  -9.98630  -7.34322 PEPA 349    1.29000
   28    6 ALA  HN     2.95139 -10.50527  -6.51650 PEPA 349    0.00000
   29    6 ALA  CA     1.53597  -9.03579  -7.08634 PEPA 349    1.35200
   30    6 ALA  HA     0.72761  -9.15026  -7.76912 PEPA 349    0.00000
   31    6 ALA  C      1.97237  -7.57624  -7.02444 PEPA 349    1.35000
   32    6 ALA  O      1.44730  -6.74717  -7.76167 PEPA 349    1.41000
   33    7 GLY  N      3.14844  -7.25539  -6.35976 PEPA 350    1.27700
   34    7 GLY  HN     3.62060  -7.86093  -5.72853 PEPA 350    0.00000
   35    7 GLY  CA     3.74937  -5.85491  -6.35587 PEPA 350    1.32500
   36    7 GLY  HA1    4.63974  -5.94889  -5.73604 PEPA 350    0.00000
   37    7 GLY  HA2    3.01859  -5.18945  -5.89470 PEPA 350    0.00000
   38    7 GLY  C      4.23305  -5.32599  -7.64543 PEPA 350    1.20000
   39    7 GLY  O      3.75561  -4.29937  -8.13330 PEPA 350    1.48500
   40    8 THR  N      4.24286  -6.04901 -10.80201 PEPA 352    1.45000
   41    8 THR  HN     3.98349  -6.81134 -10.31085 PEPA 352    0.00000
   42    8 THR  CA     3.21738  -5.81417 -11.94058 PEPA 352    1.40700
   43    8 THR  HA     3.81246  -5.62687 -12.82292 PEPA 352    0.00000
   44    8 THR  CB     2.28819  -6.97491 -12.21944 PEPA 352    1.56400
   45    8 THR  HB     1.59339  -6.61860 -13.00626 PEPA 352    0.00000
   46    8 THR  OG1    1.65027  -7.28013 -11.01703 PEPA 352    1.75700
   47    8 THR  HG1    1.77542  -8.17691 -10.81317 PEPA 352    0.00000
   48    8 THR  CG2    2.93520  -8.33104 -12.64670 PEPA 352    1.73800
   49    8 THR  HG21   3.63517  -8.14206 -13.48752 PEPA 352    0.00000
   50    8 THR  HG22   2.12243  -9.01878 -12.95886 PEPA 352    0.00000
   51    8 THR  HG23   3.39500  -8.85948 -11.72117 PEPA 352    0.00000
   52    8 THR  C      2.41007  -4.55826 -11.76266 PEPA 352    1.34300
   53    8 THR  O      1.82346  -4.03134 -12.74644 PEPA 352    1.55700
   54    9 SER  N      2.28004  -4.10196 -10.54542 PEPA 353    1.36100
   55    9 SER  HN     2.88165  -4.38930  -9.82269 PEPA 353    0.00000
   56    9 SER  CA     1.50059  -2.87821 -10.24503 PEPA 353    1.25800
   57    9 SER  HA     0.67411  -2.81515 -10.95711 PEPA 353    0.00000
   58    9 SER  CB     0.74021  -2.85277  -8.90192 PEPA 353    1.60200
   59    9 SER  HB1   -0.00626  -2.11724  -8.76097 PEPA 353    0.00000
   60    9 SER  HB2    1.67041  -2.53991  -8.27787 PEPA 353    0.00000
   61    9 SER  OG     0.23326  -4.08988  -8.34446 PEPA 353    1.59100
   62    9 SER  HG1    0.55670  -3.96559  -7.41534 PEPA 353    0.00000
   63    9 SER  C      2.30616  -1.60645 -10.46262 PEPA 353    1.24600
   64    9 SER  O      1.79656  -0.56759 -10.92516 PEPA 353    1.43300
   65   10 SOD  SOD    0.31275  -5.96473  -9.71533 ION1 1      0.00000
```

References

[1] Hansen, J & McDonald, I. (1986) *Theory of Simple Liquids, 2nd Edition.* (Academic Press, London).

[2] Hummer, G, Garde, S, Garcia, A. E, Pohorille, A, & Pratt, L. R. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8951–8955.