PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

Paper

Elucidating Factors Important for Monovalent Cation Selectivity in Enzymes: *E. coli* β-Galactosidase as a Model

Robert W. Wheatley,[a] Douglas H. Juers,[b] Bogdan B. Lev,[c,d] Reuben E. Huber[a] and Sergei Yu. Noskov*[c]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Many enzymes require a specific monovalent cation (M^+), that is either Na^+ or K^+, for optimal activity. While high selectivity M^+ sites in transport proteins have been extensively studied, enzyme M^+ binding sites generally have lower selectivity and are less characterized. Here we study the M^+ binding site of the model enzyme *E. coli* ß-galactosidase, which is about 10 fold selective for Na^+ over K^+. Combining data from X-ray crystallography and computational models, we find the electrostatic environment predominates in defining the Na^+ selectivity. In this lower selectivity site rather subtle influences on the electrostatic environment become significant, including the induced polarization effects of the M^+ on the coordinating ligands and the effect of second coordination shell residues on the charge distribution of the primary ligands. This work expands the knowledge of ion selectivity in proteins to denote novel mechanisms important for the selectivity of M^+ sites in enzymes.

Introduction

Na^+ and K^+ are essential ions in biological systems. While these two monovalent cations (designated as M^+) are chemically similar, proteins frequently have a robust specificity for one or the other. Transport proteins such as the sodium-coupled secondary leucine transporter (LeuT) or the bacterial potassium or sodium channels can be up to 1000 fold selective[1, 2] and the factors critical for the high selectivity in these sites have been investigated extensively.[3-10] In addition to transport proteins, some enzymes require Na^+ or K^+ for optimal activity. Generally, these monovalent cations act in structural roles and/or are involved in substrate binding by making the binding pocket more electrostatically attractive to negatively charged ligands.[11] Examples of these cations stabilizing enzyme structure include Na^+ in thrombin[12] and in tryptophan synthase.[13] In their substrate binding roles, the major biological function of these cations appears to be counteracting the negative charge on substrate phosphate groups, often in conjunction with a divalent cation. A K^+ ion serves this role in pyruvate kinase[14] and fructose-1,6-biphosphatase.[15] In these enzymes, substitution with the competing M^+ causes a large reduction or even a complete loss of enzymatic activity, requiring selectivity to be maintained in a site dynamic enough to allow for substrate binding and release. While the M^+ selectivity in enzymes is generally lower than for transport proteins (10 to 100 fold selectivity), the exact mechanisms that govern site specificity for a given ion and its coupling to enzyme activity are yet to be widely studied. Additionally, knowledge of these mechanisms can lead to the ability to program the ion specificity of enzymes, an application of great pragmatic value.[16]

To provide additional insights on cation selectivity mechanisms, we investigated a well-known enzyme—*E. coli* ß-galactosidase (*lacZ*, E.C 3.2.1.23). Over 60 years ago, ß-galactosidase was one of the first enzymes shown to have activity stimulated by a monovalent metal ion.[17] The enzyme was also of central importance for the pioneering study of gene regulation.[18] ß-Galactosidase is a retaining glycosidase with 1023 amino acids[19, 20] per each identical subunit of the tetrameric protein. While the natural substrate is lactose, the enzyme hydrolyzes and transgalactosylates a variety of ß-D-galactopyranosides. It is a model enzyme for glycosidases that act on disaccharides and is one of the most commonly used enzymes in molecular biology. Industrially, ß-galactosidases are used in the hydrolysis of lactose in dairy products[21] and in the synthesis of galactooligosaccharides used as food additives.[22] In addition to the monovalent cation, ß-galactosidase activity also requires a divalent cation, either Mg^{2+} or Mn^{2+}. This divalent ion does not directly interact with the substrate but is located adjacent to the catalytic Glu 461.[23]

ß-Galactosidase binds either Na^+ or K^+ but is about 10 times more selective for Na^+.[24] This selectivity for Na^+ is atypical for M^+ containing enzymes. A recent survey[11] only identified two other enzymes, both class II aldolases, e.g. *E. coli* fructose-1,6-bisphosphate aldolase[25], that utilize a M^+ in substrate binding but prefer Na^+ over K^+. These enzymes are not, however, representative of all class II aldolases. Yeast fructose-1,6-bisphosphate aldolase activity is stimulated by K^+ [26] while Na^+ has little effect.[27]

ß-Galactosidase is a 5 domain protein (Fig. 1a).[28, 29] Two important catalytic residues (Glu 461 and Glu 537) are contained

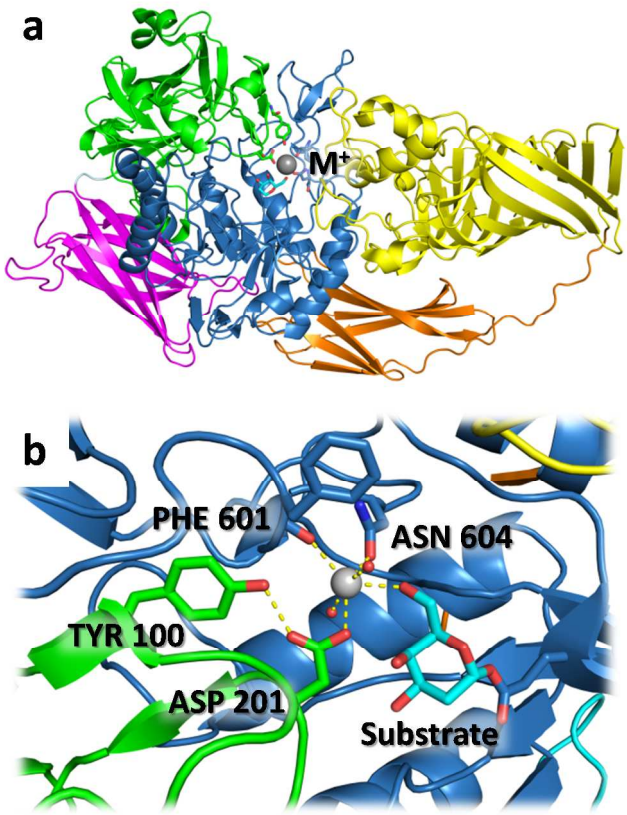

Fig. 1 The M^+ binding site of ß-galactosidase (lacZ). (a) The M^+ binding site shown in relationship to the domain structure of ß-galactosidase. Domain 1 through 5 are colored in sequence green, magenta, blue, orange, and yellow. The M^+ is shown as a grey sphere. (b) Detailed view of the M^+ binding site, showing the interaction of the M^+ (grey sphere) with the protein, water molecules (red spheres), and a substrate analog (cyan). The protein residues are colored by domain, with the same color scheme as the top figure.

in the central TIM barrel domain (domain 3, blue). Additional structures from domain 1 (green), domain 5 (yellow), and an extended loop from the TIM barrel itself modify the substrate binding site. The domain 5 structure has a role in glucose binding, while the extended TIM barrel loop and domain 1 stucture interact with galactose and the M^+. The M^+ binding site is composed of side chain oxygens from Asp 201 and Asn 604 and the backbone carbonyl oxygen of Phe 601 (Fig. 1b). Asp 201 is part of domain 1 while Phe 601 and Asn 604 are part of the extended loop from the TIM barrel. This M^+ site has a low number—only three—of protein ligands compared to the more extensively studied transport proteins.

The M^+ interacts directly with the substrate, coordinating the galactose O6 oxygen.[30] Substrates that lack an O6 hydroxyl group display a 0.9–1.4 kcal/mol decrease in binding free energy.[24] The interaction of the M^+ with a neutral hydroxyl substrate group, instead of a charged group, is the only example of this mode of substrate binding presently confirmed by x-ray crystallography. (Initial structures of coenzyme B_{12}-dependent diol dehyratase contained a K^+ coordinating two substrate hydroxyls.[31] However, subsequent studies[32, 33] suggested that this atom is actually Ca^{2+} in the diol dehyratase and related enzymes).[34] The interaction of the M^+ with a hydroxyl group is not, however, unprecedented, as serine and threonine residues are often found in M^+ binding sites, such as that of pyruvate kinase.[35] Unlike some enzymes, ß-galactosidase is active with either Na^+ or K^+ bound although the M^+ bound to ß-galactosidase has large and complex effects on the enzyme. The Na^+ enzyme generally has a greater substrate affinity than the K^+ enzyme, while the metals have been found to affect the rate of each of the hydrolytic half reactions differently.[24, 36] Work to explain these differences is ongoing.

In this study, a combination of structural and computational techniques is used to examine the factors that influence the selectivity of *E. coli* ß-galactosidase for Na^+ and K^+. A high-resolution crystal structure of K^+ ligated ß-galactosidase was obtained and forms the basis for comparisons with the Na^+ bound enzyme and for an understanding of the chemistry that governs selective Na^+ binding. Recently developed computational protocols for a reduced binding site representation allowed us to isolate the roles played by the structural rigidity of the protein scaffold and the electrostatic environment, including local polarization effects. Finally, the biological significance of the M^+ site selectivity is considered.

Experimental and Computational Methods

Buffers

Crystallization Buffer: 100 mM Bis-Tris (pH 6.5), 100 mM NaCl, 200 mM $MgCl_2$, 10 mM DTT, and PEG 8000 (10% w/w).
Mother Liquor: 100 mM Bis-Tris (pH 6.5), 100 mM KCl, 200 mM $MgCl_2$, 10 mM DTT, and PEG 8000 (10% w/w).

Crystallography

Standard protocols were used to express and purify β-galactosidase.[29] Aliquots were frozen in liquid nitrogen and stored at -70 °C until needed. Crystals were grown at 15 °C by hanging drop vapor diffusion over 1 mL of Crystallization Buffer in 24 well plates (Hampton Research).[29] Diffraction quality crystals were obtained by microseeding with starting crystals suspended in Crystallization Buffer. Different dilutions of seed crystals were utilized; the dilution required was determined empirically. The hanging drops (10 µL) consisted of equal volumes of protein and seed solution. The Na^+ was replaced with K^+ via serial dilution of the crystal into mother liquor over several hours, with estimated final concentrations of 100 mM K^+ and < 16 nM Na^+. For cryoprotection the crystal was equilibrated via serial soaks of about 1 hour in ~5% steps to a solution composed of 30% DMSO/70% mother liquor. Diffraction intensities were collected from single crystals at 100 K at ALS Beamline 5.0.2. Data were processed with MOSFLM[37] and scaled with SCALA.[38] A previously determined structure (1DP0) was used as the starting model for refinement. The final model was obtained by iterative cycles of model building with O[39] and Coot[40] and refinement with TNT[41] and Refmac.[42] Anisotropic B-factors were used for the active site K^+ ions, which showed oblong electron density ($2F_o$-F_c). Occupancies of all metal ions (K^+ and Mg^{2+}) and bound DMSO molecules were then refined with Phenix.[43] Coordinates and structure factors were deposited in the Protein Data Bank (4TTG).

Molecular Dynamics

Simulations with additive force-fields (CHARMM-27):

Molecular dynamic (MD) simulations were based on the Na^+ replete β-galactosidase structure 1DP0[29]. Upon hydration, the system consisted of 340 544 atoms and was approximately 205 × 152 × 106 Å in size. Simulations were performed with additive force-fields (CHARMM-27), with the Lennard–Jones parameters for Na^+ and K^+ adjusted to yield the experimental solvation free energies in bulk water[5]. Periodic boundary conditions in a NPT ensemble were utilized. The Nose-Hoover thermostat was used to maintain a pressure of 1 atm and a temperature of 300 K. The SHAKE algorithm was used to maintain the length of hydrogen atom bonds. Electrostatic interactions were calculated using the particle-mesh Ewald (PME) summation method[44] with a coupling parameter of 0.34 and a 6th-order spline for mesh interpolation. Non-bonded pair lists were maintained out to 16 Å. A real-space cut-off of 12 Å was used, with Lennard–Jones (LJ) interactions truncated from 10 to 12 Å via an atom-based energy switch algorithm with long-range LJ corrections applied.[45, 46] The leap-frog algorithm was used to integrate equations of motions. A 1.9 ns production run was performed with an integration time of 1 fs. This run was used to obtain the positional fluctuations and the water structure needed for construction of the reduced binding sites models.

Simulations with Drude polarizable force-fields:

MD simulations with polarizable force fields utilized the Drude polarizable force-field developed by the groups of MacKerell and Roux. The latest protein force-field compatible with the M^+ parameters was used[47-50]. The positions of auxiliary Drude particles (attached to all non-hydrogen atoms) were propagated via an extended Lagrangian formalism through the assignment of a small mass (0.4 amu) at low temperature (1 K) using a separate thermostat, known as a dual-thermostat method. The Velocity-Verlet2 integrator and Langevin Dynamics[48, 49] were employed. Minimizations were performed initially on only the Drude particles with the atoms constrained using the Steepest Descent (SD) optimizer. The remaining MD parameters were based on those used with the CHARMM-27 force field.

Quantum Mechanics Optimizations

QM Reduced Models:

Quantum mechanics calculations were used to estimate the optimal M^+ position in model systems representing the M^+ binding site. The calculations were performed with Gaussian 09[51] employing Density Functional Theory with the B3LYP functional. Geometry optimizations were performed with a 6-311+G(2df,2p) basis set. The model systems were based on the Na^+ replete β-galactosidase structure 1DP0 and the K^+ replete structure 4TTG. In addition to the metal, the system included the side chains of Tyr 100, Asp 201, Asn 604, and Trp 568. To obtain a covalently closed system, the Cβ of these residues were converted to methyl groups. The system also included the Phe 601 – Cys 602 peptide, with the Cα of each of these residues converted to methyl groups. Hydrogen atoms were added to the structures and their positions were optimized. With the positions of the remaining atoms fixed, the optimal M^+ location was then calculated.

QM-MM models:

QM/MM minimizations were performed on systems derived from both the Na^+ (1DP0) and K^+ (4TTG) replete β-galactosidase structures. The hydrated systems contained a single protein chain and consisted of approximately 35 000 atoms in a 90 × 90 × 140 Å system. The system was equilibrated for at least 1 ns using CHARMM27 force fields as described above. Additionally, the protein backbone, the M^+, and the side chains of Tyr 100, Asp 201, Trp 568 and Asn 604 were restrained to the experimental conformation with a harmonic energy of 10 kcal/mol/Å^2. For QM/MM minimizations, SCCDFTTB methods[52, 53] were employed with the recently developed DFTB3 functional for biological ligands and monovalent cations[54, 55]. The QM/MM decomposition was introduced with a link atom approach and implemented the "Divided Frontier Charge" scheme to minimize potential artefacts of the decomposition.[56] The QM sub-system consisted of the M^+, the side chains of Tyr 100, Asp 201, Asn 604, and Try 568, the main chain atoms (side chains excluded) from Gln 600 C to Cys 602 C, and the waters within 4 Å of the M^+. Molecular mechanics atoms were represented with the CHARMM-27 force field and the TIP3P water model. Protein atoms further than 10 Å from the QM region were fixed in place and water molecules further than 10 Å from the M^+ were deleted from the system. The system was exhaustively minimized with infinite non-bonded interaction cut-offs and a SCF convergence criteria set to 10^{-7}.

Reduced Representation of the Metal Binding Site

MD simulations were performed with reduced models of the ß-galactosidase M^+ binding site in order to dissect the effects of the protein structure and electrostatic environment on ion selectivity. Such reduced site representations have a long history in the study of ion selectivity and can range from utilizing a single minimized structure[57-59] to an ensemble of structures encompassing the conformational space of the binding site[7, 8, 60-65]. The parametric variation of different renstraints within the reduced model allows dissection of structural and energetic factors on the emergent selectivity of the site. The former is almost impossible to investigate with MD simulations of complete protein systems.

We followed the protocol of Yu, Noskov and Roux,[66, 67] one of several like-spirited approaches developed over the last 30 years.[68-70] This protocol has been validated for several systems and displays formidable predictive power.[71] This protocol models the protein structure of the M^+ binding site with two forces. The first force (ΔW_c^{site}) represents the general architectural confinement provided by the protein. Specifically, it represents the smallest spherical volume encompassing the dynamic fluctuations of each atom as measured from all atom MD simulations of the protein. When only this restraint is imposed the ligands in the M^+ binding site act like a confined microdroplet and are able to move sufficiently to accommodate different size ions. Within this regime, it is the local interactions that provides selectivity. The second force (ΔW_g^{site}) represents the case where the protein imposes a precise geometry on the coordinating ligands. In the limiting case, this force reduces the binding site to a rigid pocket and site volume dominates selectivity. This force is imposed by means of the Lagrange multiplier λ_g. Numerous simulations exhaustively sampled both the confining forces,

The reduced system was based on Tyr 100, Asp 201, Trp 568, Phe 601, Cys 602, and Asn 604. Ala was substituted for Phe 601

and Cys 602, as the backbone peptide between these residues, not the side chains, interacts with the M^+. Tyr 100 and Trp 568 do not directly interact with the M^+, but were included as they formed hydrogen bonds with Asp 201 and Asn 604, respectively. Nine water molecules (from the equilibrated complete protein system) were also included and confined within a 6 Å sphere of the M^+ site using a harmonic energy of 20 kcal/mol/Å^2. The size of the hydration sphere was selected to encompass two solvation shells of K^+ [48, 72], the larger of the two ions being studied. The M^+ was maintained at the centre of the system with a harmonic energy of 5 kcal/mol/Å^2. Coordinates were derived from the Na^+ replete crystal structure 1DP0[29]. Architectural confinement restraints (ΔW_c^{site}) were calculated from the last 1 ns of the complete protein MD simulation described above. Before this calculation, each frame of the equilibration simulation was aligned using only the residues selected for the model system. This alignment step insured that the restraints represented only local fluctuations in the binding site and not larger scale movements, such as that between entire domains, which also occurred. Reduced system simulations were performed with the CHARMM-27 force field (TIP3P water) and with the polarizable Drude force field (SWM4-NDP water). In the latter case, the polarizable force field model was used for all atoms except those of Trp 568.

Free Energy Perturbation

Free Energy Perturbation (FEP) simulations for ion selectivity were performed using the CHARMM PERTurb[52] module to evaluate the relative free energy of selectivity between pairs of ions.

$$\Delta\Delta G_{Na,K} = \left[G_{Na}^{site} - G_{Na}^{water}\right] - \left[G_{K}^{site} - G_{K}^{water}\right] = \Delta G_{Na/K}^{site} - \Delta G_{Na/K}^{water}$$

$\Delta G_{Na/K}^{water}$ describes the free energy difference between two ions in bulk water and $\Delta G_{Na/K}^{site}$ corresponds to the free energy difference between two ions in the protein binding site. $\Delta\Delta G_{Na,K}$ is the free energy difference between the two ions in the binding site and the two ions in the bulk water. Thus, $\Delta\Delta G_{Na,K}$ is calculated from two independent FEP simulations, a protein system and a bulk water system. Each of the FEPs simulations in one direction consisted of 11 λ-windows. The sampling time for each λ-window was set to 1 ns. The integration time step was set to 1 fs. The WHAM (weighted histogram analysis method) was used to compute the free energy differences, and the WHAM tolerance was set to 0.00001.

The bulk water system consisted of a solvent box containing 512 water molecules (either TIP3P or SWM4) and, fixed at the centre, a single ion atom. Simulation parameters were unchanged from the protein systems. The $\Delta G_{Na/K}^{water}$ was 18.12 kcal/mol for the CHARMM-27 force-field and 17.39 kcal/mol for the polarizable Drude force field. The computed free energy difference between Na^+ and K^+ in water corresponds closely to the previously published data for CHARMM-27 and Drude force-fields.[5, 47, 48]

Results and Discussion

β-Galactosidase has a dissociation constant (K_d) for Na^+ of 0.36 ± 0.09 mM while the K_d for K^+ is 10.8 times higher, 3.9 ± 0.6 mM.[24] The selectivity between Na^+ and K^+ in ß-galalactosidase is at the lower end of the typical range (10 to 100 fold) of other enzymes and is much less than the 1000 fold selectivity commonly reported for Na^+/K^+ membrane channels. The selection of Na^+ or K^+ by proteins is ultimately due to the difference in free energies of the cations in a totally solvated system versus a protein bound system. The free energies of each are large numbers and selection is due to a relatively small difference between these large numbers. Experimental and theoretical investigations with ion channels have found the factors that govern Na^+ or K^+ selectivity can be dissected into three, albeit at times interrelated, categories: the site coordination number, the site volume, and the local electrostatic environment.[59]

Table 1 Statistics for crystallographic data collection and refinement of the K^+ structure. Data in parentheses refer to the highest resolution shell.

Data Collection Statistics	
Space Group	$P2_12_12_1$
Unit Cell	
a, b, c (Å)	149.3, 168.1, 200.3
α, β, γ (°)	90, 90, 90
Resolution Range (Å)	56.453–1.60 (1.64–1.60)
Unique Reflections	604,015 (39,123)
Completeness (%)	92.9 (81.6)
Redundancy	2.9 (2.5)
R_{meas} (%)[a]	8.9 (30.5)
Average($I/\sigma(I)$)	9.5 (3.3)
Refinement Statistics	
Resolution Range (Å)	55.81–1.60 (1.64–1.60)
Number of reflections	595,237 (37,696)
R_{free} set size	10,033
Number of atoms	
Total	37,642
Protein	32,632
Solvent (HOH/ions)	4482
Heterogen (DMSO)	528
R (%)	0.1598
R_{free} (%)	0.1880
Average B-factor (Å^2)	16.1
RMS bond (Å)	0.020
RMS angles (°)	1.96
PDB ID	**4TTG**

[a] R_{meas} was calculated by SCALA.[38]

X-Ray Structure

To investigate how these factors apply to the Na^+ selectivity of β-galactosidase, the structure of the enzyme complexed with K^+ was solved. Statistics for data collection and refinement are presented in Table 1. This structure was compared (Fig. 2) with the structure of the enzyme complexed with Na^+ (1DP0).[29] Both structures were of high quality (Na^+ resolution 1.7 Å, R=15.7; K^+ resolution 1.6 Å, R=16.0) allowing detailed analysis. The two M^+ ions bind in essentially the same location. Both ions are coordinated by 3 protein ligands—the side chain oxygens of Asp 201 and Asn 604, and the main chain carbonyl of Phe 601. Overall, there are no major conformational differences between the two structures and the RMSD between the structures is only 0.12 Å. However, structural differences are observed at the M^+ binding site. Compared to Na^+, the larger K^+ causes an increase of 0.19 Å in the average coordination distance to the three protein

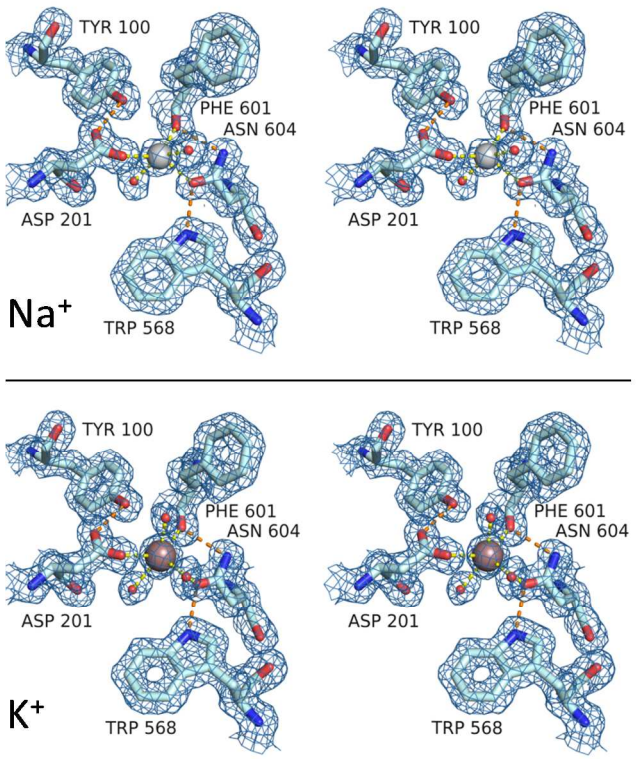

Fig. 2 Comparison of sodium and potassium complexed β-galactosidase. Metal coordination interactions are shown in yellow. Hydrogen bonds involved in metal binding site stabilization are shown in orange. The 2Fo-Fc electron density maps are contoured at 1 standard deviation. (Fo=Observed amplitudes, Fc=Calculated amplitudes.)

ligands—from 2.38 Å to 2.57 Å (Table 2). Specifically, in the K^+structure the side chains of Asp 201 and Asn 604 move a small amount but the position of the main chain atoms in these residues is largely unchanged. The Phe 601 main chain carbonyl is slightly displaced, requiring a slight movement of the protein backbone but this change does not propagate to neighbouring amino acids. The first shell of waters coordinated by the M^+ is different between the Na^+ and K^+ replete enzymes. The Na^+ enzyme has 2 water ligands (5 ligands total) forming a square pyramidal coordination structure while K^+ has 3 water ligands (6 ligands total) forming a trigonal biprismatic structure. In both structures, a water molecule is coordinated to the face of the metal opposite the active site (below Asp 201 in Fig. 2). However, on the opposite side of the metal, facing towards the substrate-binding site, there is one water in the Na^+ structure compared to two in K^+ structure. The location of each of these two waters in the K^+ structure is different from that of the one water in the Na^+ structure. The average water to M^+ distance increases by 0.51 Å in the K^+ compared to the Na^+ structure, from 2.26 Å to 2.77 Å. With both Na^+ and K^+, the amino acids coordinating the metals have temperature factors less than the average of the entire protein.

Selectivity and Coordination Number

The crystal structures showed that the Na^+ and K^+ bound at the active site have different coordination numbers. When bound to proteins, the general trend is that a higher coordination number is found with the larger K^+ ion. In a survey of protein structures, Na^+ was found to have from 3 to 7 ligands with an average of 5, while K^+ had 4 to 9 ligands, with an average of 6.[5] This difference is due to reduced steric repulsion allowing a larger number of ligands to more easily pack around the larger ion. In simulations of the K^+ selective pore in the KcsA channel, the observed selectivity for K^+ was attributed to the eight coordinating oxygen atoms. Decreasing the number of ligands in the binding site corresponded with increasing selectivity for Na^+,[62] showing the coordination environment can influence selectivity. However, the β-galactosidase M^+ site has only 3 protein ligands, leaving a large part of the coordination sphere filled by mobile water. This design appears incompatible with a single defined coordination environment and, indeed, the crystal structures show that the coordination number can change between 5 and 6 ligands. Additionally, this site composition allows both Na^+ and K^+ ions to achieve their average coordination number for a protein environment. Overall, the above observations make it unlikely that the coordination environment is important for M^+ selectivity.

Selectivity and M^+ Site Volume

Based on small molecule crystal structures and high-resolution protein structures, Na^+ to ligand distances normally have a range of approximately 2.2 to 2.6 Å, with an average of around 2.4 Å. K^+ to ligand distances are larger, with a range of approximately 2.6 to 3.0 Å and an average of 2.8 Å.[73, 74] In β-galactosidase the Na^+ to protein distances are around 2.4 Å, or close to the average (Table 2). However, the distance from K^+ to the three protein ligands averages around 2.6 Å, less than the average and close to the minimum normally observed. When we deliberately increased the coordination distance by increasing constraints during refinement, signals in the resulting F_o-F_c electron density map (which represents differences between the model and the data) indicated the new model was inconsistent with the crystallographic data. These observations suggest that constraints on the volume of the binding site by the three protein ligands may be important for selecting the smaller Na^+ over the larger K^+. A snug fit model has been used to explain selectivity in small molecules such as the cyclic membrane carrier valinomycin.[75] In addition, simulations of some protein Na^+ binding sites, such as the NA2 Na^+ site of the membrane transporter LeuT, show that structural constraints are necessary to account for the observed Na^+ selectivity.[5]

In this case, however, the M^+ binding site of β-galactosidase has only three protein ligands, arranged in a ring surrounding the metal (Fig. 2). The remaining ligands are water molecules, which are relatively mobile compared to the covalently linked protein ligands. In the crystal structures the M^+ is found in the plane defined by the three protein ligands, however, there appears to be little to constrain the M^+ to this plane. Thus K^+ should be able to move out of this plane and to more optimal coordination distances if being in the plane was energetically unfavourable. Both experimental and computational evidence supports this reasoning. First, β-galactosidase has been crystallized with numerous compounds representing various stages along the reaction pathway. Alignments of these structures, an example of which is shown in Fig. 3, indicate that the Na^+ can move relative

Table 2 Monovalent Cation Coordination Distances. Distances (Å) for each of the four subunits of the tetramer are reported. Database (DB) numbers are from,[74] except for the K^+ to Asn distance, which is not reported in this reference. In this case the K^+ to carbonyl distance was used. Distances in boldface deviate from the database value by more than one σ. Ch A refers to chain A of the tetramer, etc.

	Ch A	Ch B	Ch C	Ch D	Average	DB	Deviation
Sodium structure							
Asp 201 OD2	2.39	2.33	2.36	2.31	2.35	2.41	-0.06
Phe 601 O	2.37	2.39	2.40	2.36	2.38	2.38	0.00
Asn 604 OD1	2.42	2.24	2.31	2.43	2.40	2.38	0.02
				Avg	2.38	2.39	-0.01
Water 1	**2.26**	**2.28**	**2.29**	**2.14**	**2.24**	2.41	-0.17
Water 2	**2.28**	**2.25**	**2.20**	2.37	**2.27**	2.41	-0.14
Potassium Structure							
Asp 201 OD2	**2.60**	**2.59**	**2.62**	**2.53**	**2.59**	2.81	-0.24
Phe 601 O	**2.55**	**2.51**	**2.54**	**2.48**	**2.52**	2.74	-0.19
Asn 604 OD1	**2.60**	**2.62**	**2.61**	**2.57**	**2.60**	2.74	-0.15
				Avg	**2.57**	2.76	-0.19
Water 1	**2.45**	**2.53**	**2.59**	**2.36**	**2.48**	2.81	-0.40
Water 2	2.82	2.75	**2.70**	2.94	2.80	2.81	-0.04
Water 3	**3.05**	**3.03**	**2.93**	**3.07**	**3.02**	2.81	0.21

to the ligand ring. Note that the Na^+ moves twice as far (0.8 Å) as the protein ligands (0.4 Å). Similar ranges of positions are found with K^+ bound structures (D. H. Juers, unpublished observations). Second, QM/MM systems that modelled the M^+ binding site and surrounding water show that, after minimization, the optimal K^+ position is again in the plane of the ligand ring. The coordination distances were also less than optimal, averaging 2.67 Å. A Na^+ replete system was found to be similar to the crystal structure, with a 2.37 Å average coordination distance. As it was possible the hydrogen bond network from the solvent in the QM/MM system was forcing the K^+ into a strained coordination site, full QM models of an unhydrated binding site were also analysed. These models retained the binding site geometry from the crystal structures but allowed the M^+ to move. Again, the M^+ was found to stay essentially in the plane of the ring. The optimal Na^+ coordination distance was calculated as 2.43 Å, and the optimal K^+ distance was 2.63 Å. When a model was constructed with K^+ substituted into the Na^+ binding site (starting coordination distances 2.40 Å), the metal was indeed found to move away from the plane of the ring. In this case, an average coordination distance of 2.56 Å was obtained. All together, this analysis suggests the K^+ coordination distances, while less than average, reflect the chemical environment of the β-galactosidase binding site and do not impose a significant energy penalty for K^+ binding. The smaller than average coordination distances observed in the K^+ crystal structure only correlated with, and did not cause, the Na^+ selectivity.

Selectivity and the Local Electrostatic Environment

The electrostatic environment is the final factor identified as being important for Na^+/K^+ selectivity. The presence of a negative charge has been shown to favour selection of the smaller of two cations,[76] as this produces a shorter, and hence stronger, electrostatic bond. Simulations with simple models have suggested that the replacement of ester groups with carboxyl groups changes the selectivity from K^+ to Na^+ in membrane channels.[4] The presence of a carboxylate group in the LeuT NA1 Na^+ binding site accounts for this site's selectivity.[5] In β-galactosidase, the carboxylate side chain of Asp 201 places a formal negative charge in the binding site. The data from the crystal structures, which eliminated coordination number and site volume as selectivity factors, indicates, albeit in a negative manner, that the electrostatic environment is key to Na^+ selectivity in β-galactosidase.

To provide further evidence that the electrostatic environment accounts for the M^+ selectivity, computational studies were employed. As discussed in the Methods section, such a goal was most readily achieved with a molecular dynamics protocol that employs a reduced model of the ion binding site (Fig. 4) to isolate electrostatic and structural factors involved in selectivity.[66] This

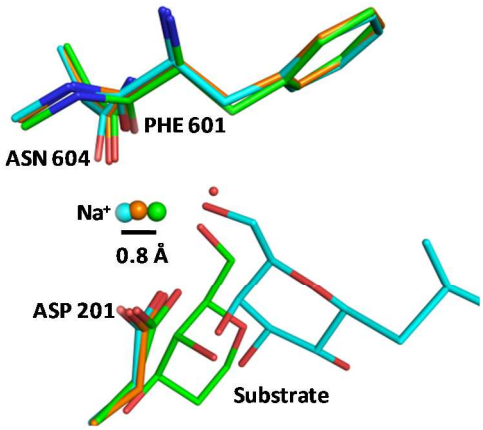

Fig. 3 The dynamic nature of the M^+ binding site. A comparison of the Na^+ position in three structures representing various phases of the ß-galactosidase reaction. Orange: unligated with substrate site water as red sphere (1DP0). Blue: IPTG inhibitor representing substrate in a precatalytic complex (1JYX). Green: galactosylated enzyme representing the covalent intermediate (1JZ4).

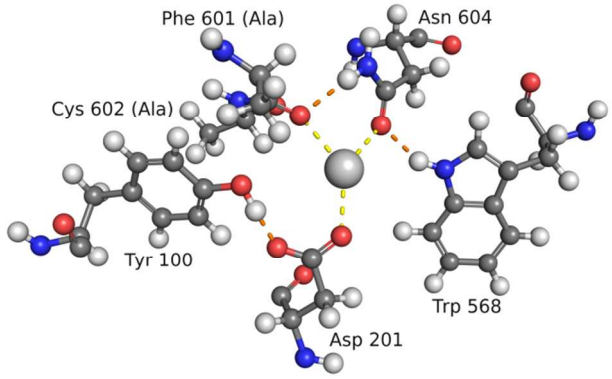

Fig. 4 Reduced system used for molecular dynamic modeling. Metal coordination interactions are shown in yellow. Hydrogen bonds involved in metal binding site stabilization are shown in orange. For clarity, the 9 water molecules included in the system have been omitted from this diagram. Note the side chains of Phe 601 and Cys 602 were modeled as alanine residues.

protocol investigates the relationship between two extremes. The first is a confined microdroplet model in which the atoms are relatively free to move and the general charge environment provides selectivity. The second is a rigid pocket model (imposed by means of the Lagrange multiplier λ_g) where site volume dominates selectivity. By performing simulations which exhaustively sample the two confining forces, all feasible effects of the protein structure on the binding site are calculated. The results from FEP simulations utilizing the Drude polarizable force-field are shown in Fig. 5a. $\Delta\Delta G_{Na,K}$ values greater than 0 indicate the system is selective for Na^+. The 10.8 fold selectivity of β-galactosidase corresponds to a $\Delta\Delta G_{Na,K}$ of +1.4 kcal/mol. Significantly, the system was selective for Na^+, even at high RMSD values. The high RMSD regime represents the confined microdroplet extreme and shows that, for this binding site, the electrostatic environment alone is sufficient to result in Na^+ selectivity. As expected, when additional structural forces were applied (λ_g force was very strong), the selectivity for Na^+ increased further. However, this structural effect was only observed when the λ_g force held the model atoms in position with sub-angstrom precision and the atomic motion (RMSD) was much less than a biologically relevant ~0.5 Å (positional fluctuations calculated from all atom simulations). Thus, the rigid pocket regime did not appear to be biologically significant, a finding consistent with the analysis of the crystal structure. Also importantly, when the RMSD was damped to the biologically realistic ~0.5 Å, the system quite closely replicated (within 0.7 kcal/mol) experimental Na^+ selectivity, indicating the system was a valid model of the binding site. Boot-strapping and block-averaging analysis for WHAM estimates of free energies differences between two monovalent cations ($\Delta\Delta G_{M+/X+}$) have generally found uncertainties to be within 0.5 to 0.75 kcal/mol. In summary, these calculations indicate that for the β-galactosidase M^+ binding site, selectivity is primarily governed by the electrostatic environment.

The Drude polarizable force fields explicitly account for electronic polarizability of the binding pocket. Results, however, are drastically different for FEP computations performed with less sophisticated classical force fields, in which the electric field

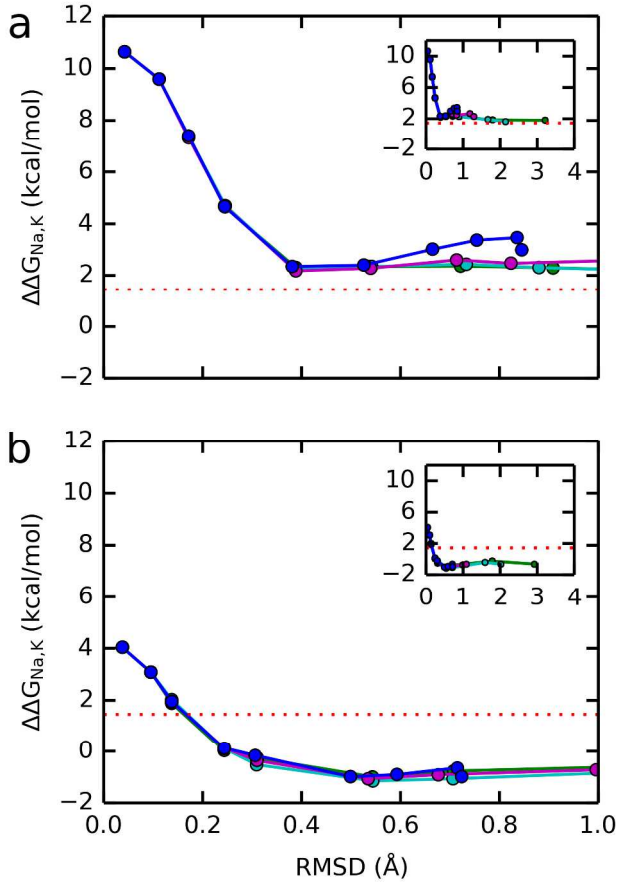

Fig. 5 Free energy perturbations with the model system. (a) Polarizable Drude force field. (b) Classical CHARMM-27 force field. For each data series, ΔW_c^{site} was held to a constant value (green 0.01, cyan 0.1, magenta 1, blue 10 kcal) while ΔW_a^{site} was varied from 0.005 to 500 kcal. The dotted line indicates the experimental Na^+ selectivity.

distribution on the protein atoms remain constant regardless of the environment. Specifically, FEP simulations performed with the CHARMM-27 force field were under most conditions selective for K^+, and underestimate the free energy difference by approximately 2.5 kcal/mol at the biologically relevant ~0.5 Å RMSD (Fig. 5b). This difference significantly exceeds the expected 0.5 to 0.75 kcal/mol resolution of FEP simulations. The finding that the CHARMM-27 force-field does not represent this binding site showed that even though the binding site contained a carboxylate group (Asp 201), the presence of this charge alone was not sufficient to produce a Na^+ selective system. The positive charge on the M^+ would also be expected to induce an increased negative charge on the binding site ligands. It was only by including this effect in the computational system, using the polarizable force field, that the experimentally observed preference for Na^+ was reproduced. Related QM calculations also found that M^+ binding has a significant effect on the partial atomic charges of the M^+ ligands (Supplementary Table 1). In summary, local polarization effects are critically important in the β-galactosidase binding site and this work shows that in moderately selective systems these effects can become significant.

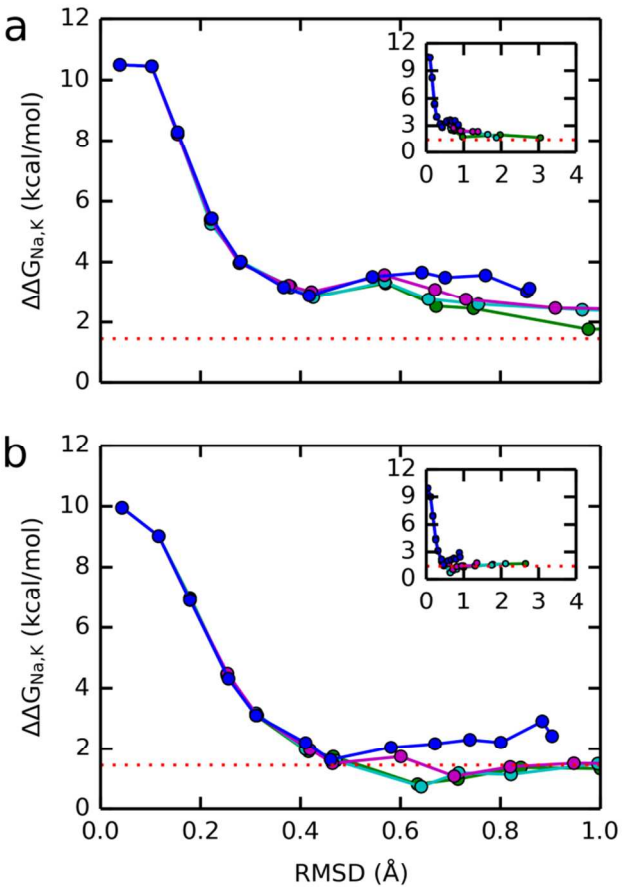

Fig. 6 Free energy perturbations with modified systems (polarizable Drude force field). (a) Tyr 100 deleted. (b) Trp568 deleted. For each data series, ΔW_c^{site} was held to a constant value (green 0.01, cyan 0.1, magenta 1, blue 10 kcal) while ΔW_a^{site} was varied from 0.005 to 500 kcal. The dotted line indicates the experimental Na^+ selectivity.

Secondary Coordination Shell Interactions

An interesting aspect of Asp 201 is that while one carboxyl oxygen interacts with the M^+, the second carboxyl oxygen forms a hydrogen bond with Tyr 100. When Tyr 100 was deleted from the model system (Fig. 6a) the system actually became more selective for Na^+, with an $\Delta\Delta G_{Na,K}$ increase of approximately 1.1 kcal/mol around biologically relevant conditions. Significantly, the crystal structures show the hydrogen bond between Tyr 100 and Asp 201 averages 2.47 Å in the four units of the Na^+ tetramer and 2.53 Å over the subunits of the K^+ tetramer. This bond is quite short, consistent with the presence of a shared proton. Hence this interaction would reduce the negative charge on the Asp 201 carboxylate group and in turn reduce the selectivity for Na^+. QM calculations (data not shown) using the model system described above but with Tyr 100 deleted also found such an electrostatic effect occurs. This data shows that residues outside the first coordination shell can be important for selectivity.

In addition to its effect on selectivity, the Tyr–Asp hydrogen bond is important for maintaining the geometry of the binding site. Specifically, the bond orients Asp 201 so that only one carboxyl oxygen interacts with the M^+. Using a whole protein model, both molecular dynamic and QM/MM simulations show that when Ala is substituted for Tyr 100, the Asp 201 side chain will move so both carboxyl oxygens are approximately equidistant from the M^+ (see Supplementary Fig. 1). It also should be noted that the geometric restraints of the model system were necessary for maintaining the correct geometry of the Tyr 100 – Asp 201 interaction. When the restraints were relaxed (RMSD of the system much greater than biological), the Asp carboxyl group would move away from Tyr 100 and towards the M^+. This problem was found with both the classical and the Drude polarizable force fields and shows the complexity of modelling even simple interactions between amino acids with current force fields.

The role of Asp 201 in M^+ binding has been previously investigated experimentally by substitution with Asn.[24] It was expected that this substitution would increase the selectively for K^+ by eliminating the negative charge in the binding site. However, it was found that the substituted enzyme bound *both* Na^+ and K^+ poorly. With the Asn substituted enzyme, preserving the polarity of the hydrogen bond with Tyr 100 would require the Asn side chain carboxamide to be oriented with the oxygen towards Tyr 100. In turn the carboxamide nitrogen would be oriented towards the ion-binding site, an orientation detrimental for ion binding.

Trp 568 is the other second shell residue that interacts with a M^+ coordinating ligand, forming a hydrogen bond with the side chain oxygen of Asn 604. In this case, the hydrogen bond is close to typical distances (average of 2.89 Å in the Na^+ crystal and 2.87 Å in the K^+ crystal) and would be expected to have only minor effects on the binding site charge. When Trp 568 was deleted from the model system (Fig. 6b) selectivity remained within the computational precision of the experimental value. However, it was found that, as restraints were relaxed, the RMSD of the M^+ ligating Asn OD1 atom increased rapidly compared to the initial system with Trp 568 included (data not shown). Thus Trp 568 functions in stabilizing the position ofAsn 604 within the binding site.

Conclusion

The fact the β-galactosidase selectivity mechanism is based on the electrostatic environment has consequences for enzyme function. In β-galactosidase the M^+ binding site has a small number—only three—of protein ligands. Such an arrangement keeps a large face of the M^+ exposed for substrate binding (Fig. 1). Additionally, the design of the binding site allows the M^+ to move during the reaction (Fig. 3). Thus, a selectivity mechanism based on the electrostatic environment allows for a highly dynamic metal binding site. Mechanisms based on coordination number or site volume require more ligands around the metal and would not be as compatible with such requirements. As enzyme sites where the M^+ interacts with the substrate would have, to some extent, a dynamic nature, M^+ sites with selectivity based on the electrostatic environment would have a broad biological utility.

The relativity low selectivity between K^+ and Na^+ in *E. coli* β-galactosidase may also be important, providing a novel mechanism for optimizing enzyme activity. In the energized (actively metabolizing) *E. coli* cytoplasm there is generally a tenfold lower concentration of Na^+ than K^+. Since ß-galactosidase binds Na^+ about 10 fold better than K^+, under conditions in the actively metabolizing cell cytoplasm, the ratio of Na^+:K^+ bound

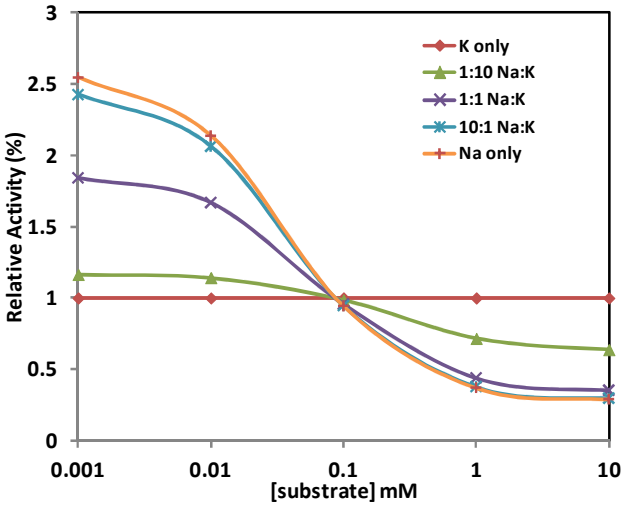

Fig. 7 Effect of the ratio of Na^+:K^+ in the medium on the β-galactosidase reaction rate at various concentrations of substrate. From the ratio of cations in the medium, the ratio of Na^+:K^+ bound to the enzyme was calculated using the Kd values for each cation. Activity is calculated from the kinetic data for the standard model substrate pNPG[24] and is relative to the rate calculated for a medium containing only K^+.

to ß-galactosidase would be close to 1:1. Under these conditions (medium Na^+:K^+ ratio of 1:10) and high substrate concentrations, Fig. 7 shows the mixed enzyme population is about 70% more active than the enzyme with Na^+ alone. However, the low intracellular Na^+ concentration is maintained by active transport, since compared to the cytoplasm most environments contain much more Na^+ than K^+. Sea water contains 480 mM Na^+ and only 10 mM K^+. In de-energized (non-metabolizing) *E. coli*, the Na^+ gradient gradually collapses[77] and the result is that non-metabolizing cells have a higher intracellular Na^+ concentration than metabolizing cells. Fig. 7 shows that, under conditions of low substrate concentrations and a 10:1 Na^+:K^+ medium ratio, β-galactosidase activity is close to twice that which would be found with the Na^+:K^+ ratio of energized cells. This calculation suggests that under starvation conditions, the M^+ selectivity could provide a mechanism which optimizes β-galactosidase for low nutrient concentrations.

In conclusion, Na^+ selectivity in the *E. coli* ß-galactosidase M^+ binding site was not due to the coordination environment or geometric constraints. Instead, the electrostatic environment was found to determine the preference for Na^+ over K^+. Factors affecting the electrostatic environment included the induced polarization effects of the M^+ and the interaction of second shell residues (Tyr 100). Finally, this work suggests how selectivity mechanisms can be important for enzymes function.

Notes

We would like to thank Qiang Cui and Maximillian Kubilius for their help with Na^+ and K^+ parameters for SCCDFTTB computations. Useful discussions with Chunfeng Zhao and Benoit Roux are acknowledged. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant program (RGPIN 340946-07) and a Resource Allocation Award from the Compute Canada (Westgrid). SN thanks an Alberta Innovates Technology Futures (AITF) New Faculty award. BL was supported by the AITF graduate scholarship.

[a] *Division of Biochemistry, Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4, Canada.*
[b] *Department of Physics, Whitman College, Walla Walla, Washington, 99362, USA.*
[c] *Centre for Molecular Simulation and Department of Biological Sciences, BI-447, University of Calgary,2500 University Drive NW, Calgary, AB, T3A 2T3, Canada. E-mail: snoskov@ucalgary.ca*
[d] *Current address: School of Applied Sciences, Health Innovations Research Institute, RMIT University, GPO Box 2476, Melbourne VIC 3001, Australia*
*Sergei Noskov is the corresponding author for this manuscript.

References

1. A. Alam and Y. Jiang, *Nat. Struct. Mol. Biol.*, 2009, **16**, 35-41.
2. S. W. Lockless, M. Zhou and R. MacKinnon, *PLoS Biol.*, 2007, **5**, e121.
3. B. Roux, S. Berneche, B. Egwolf, B. Lev, S. Y. Noskov, C. N. Rowley and H. Yu, *J. Gen. Physiol.*, 2011, **137**, 415-426.
4. S. Y. Noskov and B. Roux, *J. Gen. Physiol.*, 2007, **129**, 135-143.
5. S. Y. Noskov and B. Roux, *J. Mol. Biol.*, 2008, **377**, 804-818.
6. B. Corry and M. Thomas, *J. Am. Chem. Soc.*, 2012, **134**, 1840-1846.
7. M. Thomas, D. Jayatilaka and B. Corry, *PLoS Comput. Biol.*, 2013, **9**, e1002914.
8. D. L. Bostick and C. L. Brooks, 3rd, *J. Am. Chem. Soc.*, 2010, **132**, 13185-13187.
9. I. Kim and T. W. Allen, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 17963-17968.
10. N. Chakrabarti, C. Ing, J. Payandeh, N. Zheng, W. A. Catterall and R. Pomes, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 11331-11336.
11. M. J. Page and E. Di Cera, *Physiol. Rev.*, 2006, **86**, 1049-1092.
12. A. O. Pineda, C. J. Carrell, L. A. Bush, S. Prasad, S. Caccia, Z.-W. Chen, F. S. Mathews and E. Di Cera, *J. Biol. Chem.*, 2004, **279**, 31842-31853.
13. S. Rhee, K. D. Parris, S. A. Ahmed, E. W. Miles and D. R. Davies, *Biochemistry*, 1996, **35**, 4211-4221.
14. T. M. Larsen, L. T. Laughlin, H. M. Holden, I. Rayment and G. H. Reed, *Biochemistry*, 1994, **33**, 6301-6309.
15. V. Villeret, S. Huang, H. J. Fromm and W. N. Lipscomb, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 8916-8920.
16. S. Rana, N. Pozzi, L. A. Pelc and E. Di Cera, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 5221-5225.
17. J. Lederberg, *J. Bacteriol.*, 1950, **60**, 381-392.
18. F. Jacob and J. Monod, *J. Mol. Biol.*, 1961, **3**, 318-356.
19. A. Fowler and I. Zabin, *J. Biol. Chem.*, 1978, **253**, 5521-5525.
20. A. Kalnins, K. Otto, U. Ruther and B. Muller-Hill, *EMBO J.*, 1983, **2**, 593-597.
21. V. Gekas and M. López-Leiva, *Process Biochem.*, 1985, **20**, 2-12.
22. A. Gosling, G. W. Stevens, A. R. Barber, S. E. Kentish and S. L. Gras, *Food Chem.*, 2010, **121**, 307-318.
23. S. Lo, M. L. Dugdale, N. Jeerh, T. Ku, N. J. Roth and R. E. Huber, *Prot. J.*, 2010, **29**, 26-31.
24. J. Xu, M. A. A. McRae, S. Harron, B. Rob and R. E. Huber, *Biochem. Cell Biol.*, 2004, **82**, 275-284.
25. D. R. Hall, G. A. Leonard, C. D. Reed, C. I. Watt, A. Berry and W. N. Hunter, *J. Mol. Biol.*, 1999, **287**, 383-394.
26. R. D. Kobes, R. T. Simpson, B. L. Vallee and W. J. Rutter, *Biochemistry*, 1969, **8**, 585-588.
27. O. C. Richards and W. J. Rutter, *J. Biol. Chem.*, 1961, **236**, 3177-3184.
28. R. H. Jacobson, X. J. Zhang, R. F. DuBose and B. W. Matthews, *Nature*, 1994, **369**, 761-766.

29. D. H. Juers, R. H. Wigley, X. Zhang, R. E. Huber, D. E. Tronrud and B. W. Matthews, *Protein Sci.*, 2000, **9**, 1685-1699.
30. D. H. Juers, T. D. Heightman, A. Vasella, J. D. McCarter, L. Mackenzie, S. G. Withers and B. W. Matthews, *Biochemistry*, 2001, **40**, 14781-14794.
31. J. Masuda, N. Shibata, Y. Morimoto, T. Toraya and N. Yasuoka, *Structure*, 2000, **8**, 775-788.
32. T. Kamachi, M. Takahata, T. Toraya and K. Yoshizawa, *J. Phys. Chem. B*, 2009, **113**, 8435-8438.
33. T. Toraya, S. Honda and K. Mori, *Biochemistry*, 2010, **49**, 7210-7217.
34. M. Yamanishi, M. Yunoki, T. Tobimatsu, H. Sato, J. Matsui, A. Dokiya, Y. Iuchi, K. Oe, K. Suto, N. Shibata, Y. Morimoto, N. Yasuoka and T. Toraya, *Eur. J. Biochem.*, 2002, **269**, 4484–4494.
35. M. S. Jurica, A. Mesecar, P. J. Heath, W. Shi, T. Nowak and B. L. Stoddard, *Structure*, **6**, 195-210.
36. K. Wallenfels and O. P. Malhotra, in *Advances in Carbohydrate Chemistry*, ed. L. W. Melville, Academic Press, 1962, vol. Volume 16, pp. 239-298.
37. A. W. Leslie and H. Powell, in *Evolving Methods for Macromolecular Crystallography*, eds. R. Read and J. Sussman, Springer Netherlands, 2007, vol. 245, ch. 4, pp. 41-51.
38. Collaborative Computational Project Number 4, *Acta Crystallogr.*, 1994, **D50**, 760-763.
39. T. A. Jones, J. Y. Zou, S. W. Cowan and M. Kjeldgaard, *Acta Crystallogr.*, 1991, **A47**, 110-119.
40. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr.*, 2010, **D66**, 486-501.
41. D. E. Tronrud, *Methods Enzymol.*, 1997, **Volume 277**, 306-319.
42. G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Crystallogr.*, 1997, **D53**, 240-255.
43. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr.*, 2010, **D66**, 213-221.
44. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
45. X. Wu and B. R. Brooks, *J. Chem. Phys.*, 2008, **129**, 154115.
46. P. Lagüe, R. W. Pastor and B. R. Brooks, *J. Phys. Chem. B*, 2003, **108**, 363-368.
47. G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux and A. D. MacKerell Jr, *Chem. Phys. Lett.*, 2006, **418**, 245-249.
48. G. Lamoureux and B. Roux, *J. Phys. Chem. B*, 2006, **110**, 3308-3322.
49. P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux and A. D. MacKerell, *J. Chem. Theory Comput.*, 2013, **9**, 5430-5449.
50. H. Li, Ngo, V. A., Da Silva, M. C., Callahan, K. M., Salahub, D. R., Roux, B., & Noskov, S. Y., *J. Phys. Chem. B*, 2015. In Press.
51. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT, 2009.
52. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545-1614.
53. Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**, 569-585.
54. M. Kubillus, Kubar, T., Gaus, M., Řezáč, J., & Elstner, M., *J. Chem. Theory Comput.*, 2015, **11**, 332–342.
55. M. Gaus, A. Goez and M. Elstner, *J. Chem. Theory Comput.*, 2013, **9**, 338-354.
56. P. H. Konig, M. Hoffmann, T. Frauenheim and Q. Cui, *J. Phys. Chem. B*, 2005, **109**, 9082-9095.
57. T. Dudev and C. Lim, *Chem. Rev.*, 2003, **103**, 773-787.
58. T. Dudev and C. Lim, *Accounts Chem Res*, 2014, **47**, 3580-3587.
59. T. Dudev and C. Lim, *Chem. Rev.*, 2013, **114**, 538-556.
60. P. D. Dixit, S. Merchant and D. Asthagiri, *Biophys. J.*, 2009, **96**, 2138-2145.
61. D. Asthagiri, L. R. Pratt and M. E. Paulaitis, *J. Chem. Phys.*, 2006, **125**.
62. M. Thomas, D. Jayatilaka and B. Corry, *Biophys. J.*, 2007, **93**, 2635–2643.
63. D. L. Bostick, K. Arora and C. L. Brooks, *Biophys. J.*, 2009, **96**, 3887-3896.
64. D. Asthagiri, P. D. Dixit, S. Merchant, M. E. Paulaitis, L. R. Pratt, S. B. Rempe and S. Varma, *Chem. Phys. Lett.*, 2010, **485**, 1-7.
65. D. M. Rogers and S. B. Rempe, *J. Phys. Chem. B*, 2011, **115**, 9116-9129.
66. H. Yu, S. Y. Noskov and B. Roux, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20329-20334.
67. H. B. Yu, S. Y. Noskov and B. Roux, *J. Phys. Chem. B*, 2009, **113**, 8725-8730.
68. D. Boda, J. Giri, D. Henderson, B. Eisenberg and D. Gillespie, *J. Chem. Phys.*, 2011, **134**.
69. D. Boda, M. Valisk, B. Eisenberg, W. Nonner, D. Henderson and D. Gillespie, *Biophys. J.*, 2007, 609a-609a.
70. A. Malasics, D. Gillespie, W. Nonner, D. Henderson, B. Eisenberg and D. Boda, *Bba-Biomembranes*, 2009, **1788**, 2471-2480.
71. B. Lev and S. Y. Noskov, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2397-2404.
72. B. Lev, B. Roux and S. Y. Noskov, *J. Chem. Theory Comput.*, 2013, **9**, 4165-4175.
73. M. M. Harding, *Acta Crystallogr.*, 2002, **D58**, 872-874.
74. M. M. Harding, *Acta Crystallogr.*, 2006, **D62**, 678-682.
75. B. Dietrich, *J. Chem. Educ.*, 1985, **62**, 954-964.
76. G. Eisenman, *Biophys. J.*, 1962, **2**, 259 - 323.
77. A. M. Castle, R. M. Macnab and R. G. Shulman, *J. Biol. Chem.*, 1986, **261**, 3288-3294.

