Intrinsic free energy of the conformational transition of the KcsA signature peptide from conducting to nonconducting state

Ilja V. Khavrutskii*, Mikolai Fajer and J. Andrew McCammon*

Howard Hughes Medical Institute, Center for Theoretical Biological Physics,
Department of Chemistry and Biochemistry, University of California San Diego
La Jolla, California 92093-0365

ikhavru@mccammon.ucsd.edu, jmccammon@ucsd.edu

Model setup

The model of the wild type ion selectivity signature peptide was derived from the KcsA structures obtained in the presence of Tl^+ ions in place of K^+ ions. There are two such crystallographic structures: 1R3J that corresponds to high Tl^+ concentration, and 1R3K that corresponds to low Tl^+ concentration.[1] We extracted the following sequence from the monomer: $THR^{72}(1)$-$ALA^{73}(2)$-$THR^{74}(3)$-$THR^{75}(4)$-$VAL^{76}(5)$-$GLY^{77}(6)$-$TYR^{78}(7)$-$GLY^{79}(8)$. Superscript numbering corresponds to the original KcsA numbering in the PDB files, whereas numbers in parenthesis correspond to the model. In a single letter aminoacid code the sequence is T72-A73-T74-T75-V76-G77-Y78-G79 or TATTVGYG for short.

The corresponding double ala-mutant was generated by replacing the two glycine residues namely GLY77 and GLY79 with alanine, resulting in THR(1)-ALA(2)-THR(3)-THR(4)-VAL(5)-ALA(6)-TYR(7)-ALA(8) sequence or TATTVAYA for short.

Because the peptide model was derived by truncating the KcsA peptide we mended the ends of the chain with the standard neutral capping groups, namely Acetyl and N-methyl amide for the N- and C-termini, respectively. In the following we use the standard definitions of atom types from the CHARMM22 forcefield.

Methods
Molecular Dynamics

We use CHARMM22 TIP3P water model (CTIP3P) and peptide parameters.[2] Initial simulations were performed using GBMV implicit solvent model to expedite path optimization. The following parameters of the GBMV model were used (see charm c32b1 documentation for explanation of the parameters):

Table 1. GBMV parameters

GBMV	
BETA	-20

EPSILON	80
DN	1.0
WATR	1.4
GEOM	
TOL	1e-8
BUFR	0.5
MEM	10
CUTA	20
HSX1	-0.125
HSX2	0.25
ALFRQ	1
EMP	1.5
P4	0.0
P6	8.0
P3	0.70
ONX	1.9
OFFX	2.1
WTYP	2
NPHI	38
SHIFT	-0.102
SLOPE	0.9085
CORR	1

With the GBMV implicit solvent model we employed the following MD protocol. The electrostatic and vdW non-bonded interactions were truncated by switching functions between 12 and 13 Å. Langevin dynamics (LD) was performed with leap frog integrator using a 1.5 fs time step at 298 K and with a friction coefficient of 10 ps^{-1} for all heavy atoms. All bonds involving hydrogen atoms were constrained using SHAKE[3,4] with tolerance of 10^{-8} Å. Each evolution step involved 5,000 equilibration steps and 25,000 production steps. Coordinates from production runs were recorded every 50 steps for subsequent averaging.

The results of the GBMV simulations have been refined using explicit CTIP3P solvent model. The explicit water simulations were performed in the NPT ensemble using truncated octahedron box. For the wild type peptide the water box contained 1155 water molecules, whereas for the mutant the box contained 1222 water molecules. The nearest image distance was approximately 36 and 37 Å for the wild type and mutant peptides, respectively. The temperature was maintained with the Nose-Hoover thermostat at 298 K using a thermal piston of mass 250 (kcal/mol)•ps,[5] whereas the pressure was maintained at 1 atm by the Langevin piston method with the piston mass of 250 amu and Langevin collision frequency of 20 ps^{-1}.[6] Electrostatic interactions were computed using particle mesh Ewald method[7] with 12 Å real space cutoff. The vdW non-bonded interactions were truncated by switching functions between 10 and 12 Å. The covalent bonds between the hydrogen and the heavy atom were constrained using the SHAKE[4,8] algorithm with a tolerance in the bond length deviations of 10^{-10} Å. The MD integration step size was 2 fs. For the evolution runs we performed short MD runs including 5,000 steps of equilibration

and 25,000 steps of production. The solute coordinates from the production runs were recorded for subsequent averaging every 50 MD steps.

Transition Path Ensemble Optimization with ggaHFB

Initial path for optimization was prepared as follows. Starting from the X-ray structure corresponding to high Tl^+ concentrations (PDB code 1R3J), we gradually rotate its atoms into pII-like state producing three more states. Specifically, we rotate all the atoms of residues 1 through 5 along with the HN atom of residue 6 by -35.0 degrees about the N-CA bond of residue 6 ($\varphi(6)$ angle). In addition, we rotate atoms of residues 7 through 8 along with atom O of residue 6 by -20.0 degrees about the CA-C bond of residue 6 ($\psi(6)$ angle). Overall we perform three such rotations to produce intermediates int1, int2 and int3.

We then take the low Tl^+ concentration structure (PDB code 1R3K) as yet another intermediate to seed the path produced by the rotation described above. This structure introduces the VAL sidechain rotation. The reason this structure cannot be used as the product is that its backbone atoms collapse during the free energy optimization back to the high Tl^+ concentration structure, while leaving the VAL sidechain in the conformation found in 1R3K structure. Therefore, the 1r3k structure is placed right after the 1r3j but before the int1. Thus, the following sequence of structures 1r3j, 1r3k, int1, int2, int3 are fed to the ggaHFB interpolation procedure to generate 12 beads using the Fourier series truncation parameter of 4.

In constructing the Reactive Coordinate Space (RCS), we setup the following subspaces or levels:

Table 2. RCS vectors for different levels

Level	Atom types	Comments
Level1	C, N, CA, NT, CY, CAY and CAT	these atoms form the base of the polypeptide chain
Level2	O, CB and OY.	one bond away from the base level1
Level3	CG, CG1, CG2 and OG1	two bonds away from the base
Level4	CD1 and CD2.	three bonds away
Level5	CE1 and CE2.	four bonds away
Level6	CZ	five bonds away
Level7	OH	six bonds away

Only heavy (non hydrogen) atoms of the peptide are included into the RCS.

Using these levels we define RCS1 and RCS2 as follows. RCS1 combines levels 1 through 7, whereas RCS2 only includes the base level 1.

Throughout the path optimization we used uniform step size parameter beta. Slightly different protocols were used for the wild type and mutant optimization with the details provided in the following tables.

Table 3. The optimization protocol for the wild type.

Beads	ggaHFB Steps	Force constant	Beta	Trunc
GBMV Implicit	water			
12	1-100	10.0	0.000	8
	101-199	5.0	0.004	8
23	200-201	1.0	0.000	18
	202-499	5.0	0.005	18
	500-599	10.0	0.002	18
CTIP3P Explicit	water			
23	600-609	10.0	0.000	18
	610-1099	10.0	0.002-0.004	18
45	1100-1199	10.0	0.0035	42
89	1200-1209	10.0	0.0035	86

Table 2. The optimization protocol for the mutant.

Beads	ggaHFB Steps	Force constant	Beta	Trunc
GBMV Implicit	water			
12	1-9	10.0	0.000	8
	10-20	10.0	0.001	8
	21-199	10.0	0.002	8
23	200-209	1.0	0.000	18
	210-599	5.0	0.005	18
	600-799	10.0	0.003	18
CTIP3P Explicit	water			
23	800-804	10.0	0.000	18
	805-1199	10.0	0.002-0.004	18
45	1200-1299	10.0	0.0035	42
89	1300-1309	10.0	0.0035	86

PMF Integration with ggaHFB

Once the optimization was completed as determined by cessation of the changes in the coordinates of the RCS atoms a collection procedure was initiated. Note, that we only performed path optimizations on the RCS1 free energy surface and not RCS2. However, we used the subset of the atoms from the optimized RCS1 transition path ensemble as a reference to compute the PMF for the RCS2 level.

The final PMFs were collected using the averaged positions of the RCS atoms from 1 ns long batches of MD simulations using the final 89–bead path optimized at the RCS1 level as a reference. To restrain the atoms to the reference positions we used the force constant of $10.0 \, \text{kcal} \cdot \text{g}^{-1} \cdot \text{Å}^{-2}$. To compute the RCS1 level PMFs only five batches were necessary to achieve high convergence, whereas for the RCS2 level we needed 13 batches. The averages were combined into the final cumulative averages over the whole simulation time using the standard procedure described earlier.

During the final integration procedure we used the Fourier series truncation parameter of 88 with 1024 quadrature points for the reversible work line integral.

Comments on the PMF Validity

Keeping the RCS1 path and releasing the sidechains for the RCS2 PMF calculations substantially expands the SCS space. Unfortunately, proper averaging over all possible configurations of the sidechains necessitates overcoming barriers as high as 5-8 kcal/mol as we have seen on the V76 example. Therefore, a complete averaging cannot be achieved for the RCS2 PMF within the limited simulation time of regular MD. We could improve sampling by using parallel tempering or replica exchange MD,[9] but that would increase the cost of the computed PMFs by a factor equal to the number of replicas. Therefore, in this paper we limit ourselves to regular MD simulations. We still get apparently well-converged PMFs without having properly sampled alternative configurations of the sidechains. As expected the RCS2 PMFs take significantly longer to converge to the accuracy comparable to that of the RCS1 PMFs. Nevertheless, the RCS2 ensemble contains the discontinuities in the regions of the V76 transitions inherited from the preceding RCS1 path even after 13 ns. In particular, the sharp peaks corresponding to the V76 rotations at RCS1 level collapse at the RCS2 level creating the discontinuities in the SCS space. Strictly speaking such discontinuities invalidate the portion of the RCS2 PMFs in that region. Therefore, these PMFs should only be considered as tentative until more extensive sampling of the sidechain conformations is achieved. Work along those lines is currently in progress in our lab.

Nevertheless, the portion of the PMF underlying the important backbone transition from the α-strand to the pII state remains virtually unchanged upon going from RCS1 to RCS2 level. Thus, for the wild type peptide the forward activation barrier is 5.9 kcal/mol and the pII state is still less stable than the α-strand by slightly smaller 1.7 kcal/mol. Restoring the conducting state requires overcoming a slightly higher barrier of 4.2 kcal/mol. For the mutant we find the forward barrier of 0.9 kcal/mol and the relative pII state stabilization energy of 7.0 kcal/mol that makes the reverse barrier increase slightly to 7.9 kcal/mol. The differences between RCS1 and RCS2 PMFs are quite small in the backbone transition region. Thus, we conclude that structural details of the sidechains have very little effect on this functionally important backbone transition making it extremely robust.

The PMFs for the wild type peptide indicate that the free energy landscape is funneled toward the α-strand, whereas the mutation changes direction of transition opposing the α-

strand formation. Changes in the PMFs due to mutation are consistent with the Hammond postulate.[10] Further examination of the optimized transition path ensembles reveals that formation of the α-strand is coupled with the hydrophobic collapse between the V76 and Y78 residues. Such hydrophobic interactions are often considered the driving force in protein folding. Nevertheless, the strength of this particular interaction is not sufficient to stabilize the α-strand in the mutant.

Minimum Adiabatic Potential Energy Transition Path Optimization with ggaHFB
To optimize the paths in gas phase using the bare CHARMM22 forcefield we start from the minimum free energy transition path ensembles optimized on RCS1 surface in water. These paths have 89 beads. We first rebuild all the hydrogen positions by potential energy optimization with the fixed RCS1 atoms for all the beads. Because tyrosine and threonine OH groups have two and three rotameric states, we have to initialize their dihedral angles Cε1-Cς-Oη-Hη and Cα-Cβ-Oγ-Hγ, respectively. Thus, we assign 180, 180, -30 and 0 degrees for the T72, T74, T75 and Y78, dihedral angles, correspondingly. Adding tyrosine and threonine hydrogen atoms from the OH groups to the RCS1 we create the new reactive coordinate space RCS1h that is sufficient to integrate the adiabatic potential energy along the path.

Initially we performed ggaHFB optimization using 89 beads. With 89 beads we used 76 basis functions, the force constant of 20.0 $kcal \cdot g^{-1} \cdot Å^{-2}$ and step size parameter of 0.0025. We did not attempt to find the optimal optimization conditions and executed on the order of 2000 ggaHFB optimization steps. For each bead we used the mass weighted harmonic restraints in Cartesian coordinates, and tandem Steepest Descent/Adaptive Basis Newton Raphson optimizer with up to 200/1000 optimization steps. Because bead optimizations are very fast they were done for each bead consecutively within a single CHARMM input script, which was run on a single CPU. The optimization was set to exit once the average gradient change was less than 10^{-5} $kcal \cdot mol^{-1} \cdot Å^{-1}$. Throughout this work we truncated the electrostatic interactions with 16 Å cutoff, switching the interactions off between 16 and 18 Å. The non-bonded list cutoff was 21 Å.

As the ggaHFB optimization continues the path length increases and the OH groups rotate to different optimal positions. Soon the forces between the beads become discontinuous and adiabatic potential energy profile integration via the generalized line integral formalism no longer gives correct result. Despite that the optimization can be continued further and the progress of the path optimization can be followed by the path RMSD or by plotting a two-dimensional projection of the path onto the reactant and product vectors (not shown).

We started the optimization with 89 beads and then subsequently increased the number of beads to 177, 353, and finally 705 beads until the energy profile could be integrated precisely again. With 177 beads we performed on the order of 1000 ggaHFB optimization steps using up to 168 basis functions and the step size parameter of 0.0025. With 353 beads we performed on the order of 100 ggaHFB steps with up to 324 basis functions and we also changed the force constant of the harmonic restraint from 20.0 to 40.0 $kcal \cdot g^{-1} \cdot Å^{-2}$, and turned of the steepest descent optimization boost by setting the step size

parameter to 0.0000. Finally with 705 beads we performed on the order of 100 ggaHFB steps with up to 695 basis functions keeping the rest of the parameters the same as in the case of 353 beads.

Computing the Final Adiabatic Potential Energy Profile with ggaHFB
In the end of optimization we performed a single ggaHFB step with 705 beads and 705 basis functions (for the highest accuracy) in the Fourier series, using the force constant of $40.0 \ \text{kcal} \cdot \text{g}^{-1} \cdot \text{Å}^{-2}$. To evaluate the generalized line integral we used 2820 quadrature points. Finally, we have computed the exact energies along the structures at each of the 2820 quadrature points. The highest deviations must be observed at the progress variable value of 1.0. We find the accumulated errors with respect to the exact CHARMM22 energies of 0.07 and 0.12 kcal/mol for the wild type and mutant adiabatic energy profiles, respectively. These numbers could be improved even further with additional beads, because the onsets of the deviations between the exact and the ggaHFB profiles appear at the sharp peaks.

Energy Profiling of the Minimum Free-energy Transition Path Ensembles

Gas Phase Energy

We used Density Functional Theory model, namely B3LYP with 6-31G(d) basis set to compute the single point Quantum Mechanical energies along the path. To do that we optimized all the degrees of freedom orthogonal to RCS1 except for the four dihedral angles mentioned above that were kept restrained at the 180, 180, -30 and 0 degrees for the T72, T74, T75 and Y78, respectively using the CHARMM22 Molecular Mechanical forcefield. The B3LYP/6-31G(d) energies were then compared to the corresponding CHARMM22 energies.

Implicit Solvent Energy

To compare the Quantum Mechanical and Molecular Mechanical energies in water, we employed Polarizable Continuum Model. For the QM-PCM model we used B3LYP/6-31G(d) with scrf=(pcm,solvent=water,read) keywords in the Gaussian G03 input file with additional parameters pcmdoc, radii=uaks, scfvac, ofac=0.8, rmin=0.5. For the MM we used GBMV implicit solvent model with the exact same parameters as has been described above. Prior to computing single point QM-PCM energies we optimized all the degrees of freedom orthogonal to the RCS1 except the four dihedral angles mentioned above at the MM-GBMV level.

Conclusion
The present work demonstrates the utility of the novel ggaHFB method in studying complex processes on multidimensional free energy surfaces on the example of the important functional transition of the selectivity filter of KcsA ion channel. Other important questions that require exploring multidimensional free-energy surfaces can

now be addressed. Work is now in progress in our lab to verify the hypothesis for the ion selectivity (see Figure 1) within the tetrameric KcsA channel.

Figures

Figure 1S. An illustration of the ion selectivity hypothesis
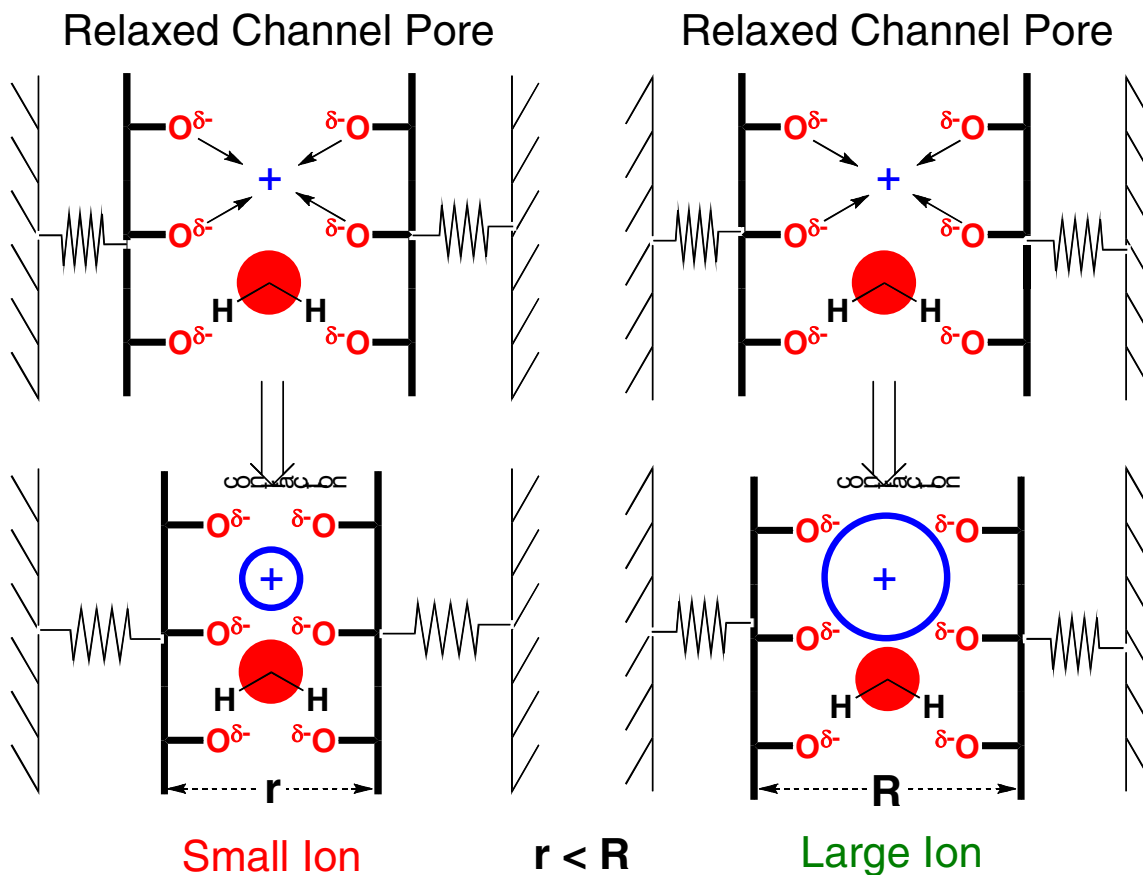
Thick black lines represent the signature peptide of the selectivity filter connected with springs to the outer barrel of the channel. In this model, smaller ions will **impede co-translating water passage** to a greater degree than larger ions by contracting the carbonyl rings of the channel. The ion sizes are not drawn to scale, but exaggerated to demonstrate the point.

References

(1)   Zhou, Y.; MacKinnon, R. *J. Mol. Biol.* **2003**, *333*, 965-975.
(2)   MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586-3616; MacKerell, A. D., Jr.; Brooks, B. R.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *The Encyclopedia of Computational Chemistry*; Schleyer, P.

v. R., Schreiner, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P., Henry F. Schaefer, I., Eds.; John Wiley & Sons: Chichester, 1998; Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935; Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954-9960; Mark, P.; Nilsson, L. *J. Comp. Chem.* **2002**, *23*, 1211-1219.

(3)   Ryckaert, J.-P.; Ciccotti , G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327-341.

(4)   Lazaridis, T.; Tobias, D. J.; Brooks, C. L., III; Paulaitis, M. E. *J. Chem. Phys.* **1991**, *95*, 7612-7625; Tobias, D. J.; Brooks III, C. L. *J. Chem. Phys.* **1988**, *89*, 5115-5127.

(5)   Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697; Nose, S. *J. Chem. Phys.* **1984**, *81*, 511-519.

(6)   Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384-2393; Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

(7)   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577-8593; Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(8)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.

(9)   Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140-150; Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607 - 2609; Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141-151; Roitberg, A. E.; Okur, A.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 2415-2418.

(10)   Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334-338.