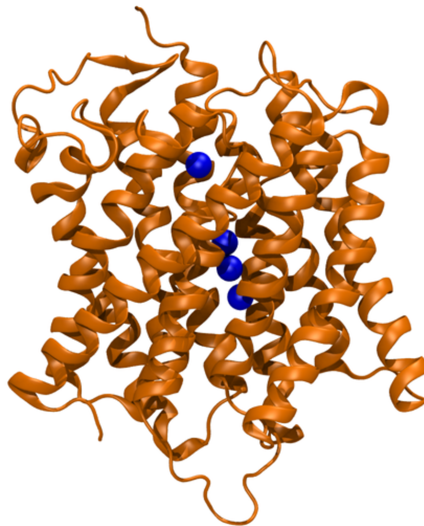
University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group

Forcing substrates through channels



James Gumbart

January 2011

A current version of this tutorial is available at
http://www.ks.uiuc.edu/Training/Tutorials/
Join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Contents

1 Introduction **3**
1.1 Required programs 5
1.2 Getting started 5

2 Exploring the pathway with SMD **6**
2.1 Setting up and running the SMD simulation 6
2.2 Analysis of the SMD simulation 9

3 Determining the potential of mean force with ABF **11**
3.1 Setting up and running the ABF simulation 11
3.2 Analysis of the ABF simulations 13

1 Introduction

Living cells are defined in part by the possession of a membrane barrier surrounding their perimeter, thereby delineating the border between the inside and the outside. The existence of this barrier, as well as the proteins that mediate exchange of ions and other substrates across it, leads immediately to the question of how transport across the membrane, whether it be active or passive, is achieved and, in particular, how its done selectively. Although some common mechanisms exist, each protein has evolved specific features to optimally transport its particular substrate. The transport process is characterized typically by a substrate pathway along with associated conformational changes in the protein. These changes are often small or even non-existent for passive channels, but can be quite large for active transporters. The substrate pathway can also be characterized by the substrate's free energy along it, known as the potential of mean force (PMF), which dictates the speed of transport and the selectivity.

While X-ray crystallography and other experimental structural techniques can provide snapshots of membrane proteins, even in multiple states, they cannot display the dynamics connecting those states. Computational methods, namely molecular dynamics (MD) simulations, provide a means of animating the static structures, allowing for, e.g., the visualization of an entire permeation event of a substrate through a channel or transporter. However, the relevant time scales for many transport processes is beyond that afforded by MD, which is typically limited to a few microseconds currently. As such, methods to accelerate the process of interest in a simulation have been developed. One example is steered molecular dynamics (SMD) (1, 2), a method in which forces are applied to part of the system, e.g., the substrate, to drive it along a predefined direction. Commonly implemented using a constant force or a constant velocity, this method is useful for exploring possible permeation pathways through a membrane protein. However, because the forces imposed are usually orders of magnitude greater than would be experienced in a living system, interpretation of the results is often limited to a qualitative description of the possible behavior of the substrate and protein. For a quantitative picture of transport, one must turn to more advanced methods. For example, the results of multiple SMD simulations combined through application of Jarzynski's equality provides a means of recovering an equilibrium PMF from non-equilibrium events (3, 4, 5, 6, 7). Alternatively, adaptive biasing forces (ABF) can generate quasi-equilibrium trajectories from which the PMF can be deduced (8, 9, 10, 11).

Numerous channels and transporters have been studied using MD-based methods. SMD has been used to address substrate binding and/or transport in the water channel aquaporin (12, 13, 14, 15), the ADP/ATP carrier AAC (16), neurotransmitter transporters LeuT (17), the sugar transporter LacY (18, 19), and many others. In addition, SMD has been used to probe the conformational changes underlying function for many membrane proteins, such as the vitamin B_{12} transporter BtuB (20) and the protein-conducting channel SecY (21, 22, 23). ABF has been used to find the free-energy profile for, e.g., glycerol in the channel GlpF (24), ADP in AAC (25), and ions in the nicotinic acetylcholine and glycine

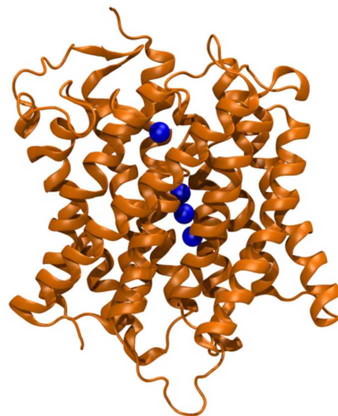


Figure 1: Crystal structure of AmtB. The top of the figure is the extracellular side of the channel and the bottom is the cytoplasmic side. The four crystallographically resolved NH_3/NH_4^+ molecules, denoted Am1 to Am4 from top to bottom, are also seen in blue.

receptors (26). Both methods, SMD and ABF, are clearly well established, but still require care to ensure their proper application.

In this tutorial, applications of SMD and ABF to the ammonium transporter AmtB will be explored (see Fig. 1). AmtB has been well studied by computational methods, including those used to calculate free energies (27, 28, 29, 30, 31, 32). SMD will be used first to gain an approximate knowledge of the permeation pathway and the barriers along it. Then the PMF for ammonia in the central region of the channel will be calculated using ABF. The appropriate choice of parameters and potential difficulties will also be discussed.

The reader of this tutorial is assumed to be familiar with the use of NAMD to perform "standard" calculations, including energy minimization and MD simulations. General documentation, tutorials and templates of NAMD configuration files are available from the Documentation section of the NAMD web page.

Many of the simulations in this tutorial can, out of necessity, take a significant amount of time to run on a single processor. If this time becomes prohibitive to completing the tutorial, example output is provided throughout to enable the reader to still carry out the desired analysis.

The investigations of AmtB described below will utilize the 1.35-Å structure solved in 2004 (33). Prior to carrying out SMD and ABF simulations, a membrane-water system containing AmtB should be built and equilibrated.

1.1 Required programs

1 VMD:

Available at http://www.ks.uiuc.edu/Research/vmd/ (for all platforms)

2 NAMD (ver. 2.7b2 or higher):

Available at http://www.ks.uiuc.edu/Research/namd/ (for all platforms)

3 Plotting program:

The free program xmgrace (unix) is recommended. It can be downloaded from http://plasma-gate.weizmann.ac.il/Grace/. Other useful graphing programs are Mathematica, http://www.wolfram.com/, Matlab, http://www.mathworks.com/, and gnuplot, http://www.gnuplot.info/. For Windows, Microsoft Excel can also be used.

1.2 Getting started

1 A fully built and equilibrated system is required for the tutorial. The necessary files are provided in `channel-tutorial-files/common/`, or can be downloaded from http://www.ks.uiuc.edu/Training/Tutorials/science/channel/. Example input files for running the simulation as well as output are also provided.

2 To build the system from scratch using the tools within VMD, begin by downloading the structure of AmtB from the Protein Data Bank at http://www.pdb.org/ (PDB code: `1U7G`). Then follow the **Membrane Protein Tutorial** to prepare and equilibrate your system.

2 Exploring the pathway with SMD

Before investing computer time in relatively lengthy free-energy calculations, SMD simulations will be used to gain a qualitative picture of the permeation pathway for ammonia in the AmtB channel. The crystallized ammonium ion on the extracellular side of the channel, mutated to a neutral ammonia molecule, will be pulled toward the cytoplasmic side at a constant velocity. You will then analyze the resulting output.

2.1 Setting up and running the SMD simulation

1 The equilibrated system will be used as the starting point for the constant velocity SMD simulations. Set up a configuration file for a standard equilibrium simulation, which will be modified accordingly.

2 To use the SMD feature in NAMD, the following parameters need to be set in the configuration file:

```
SMD on
SMDFile smd01.ref
SMDk 5
SMDVel .00001 ;# .00001 A/timestep = 10 A/ns (1 fs timesteps used)
SMDDir 0 0 -1
SMDOutputFreq 100
```

The parameter `SMDVel` defines a pulling speed of 10 Å/ns, which is fast enough to traverse the channel, defined by `SMDDir` to be in the $-z$ direction, in an ∼5-ns simulation. The file, `smd01.ref`, is almost identical to the PDB of the entire system and controls to which atoms the SMD force will be applied. A non-zero value in the occupancy column of this file indicates forces will be applied, a "0" not. For pulling a single ammonia molecule, only the nitrogen of the ammonia will be set to 1. Finally, `SMDk` specifies the force constant (in kcal/molÅ^2) for the spring connecting the SMD atom to the imaginary atom moving at constant velocity.

Setting the speed. The pulling velocity should be as slow as possible in order to maximize sampling of the channel environment. Choose a velocity that allows you to complete the simulation in a reasonable amount of time given the computer power available.

Setting the force constant. The force constant affects how well the constant velocity of the pulled atom is maintained. It should be high enough to ensure that the potential in the channel is accurately measured but not so high that the measurement is dominated by noise; this is also known as the "stiff-spring approximation" (6). The appropriate value will depend on the size of the barriers encountered, although a good rule-of-thumb is that the deviation due to thermal noise, i.e., $\Delta z = \sqrt{\frac{k_B T}{k}}$, is at least less than 0.5 Å. See Fig. 2 for a comparison of the resulting force and distance profiles using different force constants.

3 To prevent the protein, or even the entire system, from drifting under the applied force, restraints must be applied to counterbalance it. However, it is important that these restraints do not limit any potential conformational changes of the protein during the simulation. The following parameters apply positional restraints to a limited number of atoms:

```
constraints on
consexp 2
consref rest6.ref
conskfile rest6.ref
conskcol O
selectConstraints on
selectConstrX off
selectConstrY off
selectConstrZ on
```

Of particular note is the restriction to apply restraints in only the z direction, which corresponds to the SMD direction. Similar to the `SMDFile` above, the `conskfile` specifies which atoms to apply restraints to as well as the force constant for those harmonic restraints; in this case, the C_α atoms of residues 7, 97, 149, 225, 333, and 348, which are at the periphery of the extracellular side of AmtB, are chosen with a force constant of 5.0 kcal/mol·Å^2 The file in `consref` specifies the positions to which those atoms should be restrained, and many times is identical to the input PDB file.

Alternative restraints. To restrain the center-of-mass of a large group of atoms in a single dimension (x, y, or z), SMD can also be used. Set the `SMDFile` to one containing the atoms to restrain, the `SMDDir` to `0 0 1`, and `SMDVel` to `0`. Because of the way SMD is formulated, non-axial directions cannot be used in this manner. Also, should SMD be used for restraints, an alternative method will be needed to apply the steering forces, such as `tclForces`.

4 In some cases, the molecule being pulled may find it easier to move laterally outside of the channel before continuing translocation than to cross a large barrier. To limit such a possibility, restraints in the xy plane can be applied. However, these restraints should only affect the molecule if it

tries to move outside the channel, and otherwise be zero. One way to implement such a restraint in NAMD is through `tclBCforces`. Additionally, the `colvars` (short for collective variables) module is useful for defining restraints, as well as for calculating free energies (see below).

In the NAMD configuration file, an external file containing the colvars setup is referenced:

```
colvars on
colvarsConfig restrain.in
```

In `restrain.in`, you will find the following:

```
colvarsTrajFrequency 1000
colvarsRestartFrequency 1000

colvar {
   name restrain01

   width 0.5

   lowerboundary 0.0
   upperboundary 8.0

   lowerwallconstant 100.0
   upperwallconstant 100.0

   distanceXY {
     main {
       atomnumbers { <list of ammonia atom(s)> }
     }
     ref {
       atomnumbers { <list of protein atoms> }
     }
     axis ( 0.0, 0.0, 1.0 )
   }
}
```

The `distanceXY` collective variable defines a restraint that, in this case, limits the motion of the ammonia molecule to a cylinder within 8 Å of the central axis of the protein.

Atom numbering. In `colvars`, as in NAMD, atom numbering begins from 1, whereas in VMD, it begins from 0. An easy way to get the 1-based atom numbers in VMD is to use the keyword `serial` instead of `index`.

5 Run the simulation. For a 5-ns simulation, it may take 1-3 days to complete, depending on available computational resources. Sample output is provided.

2.2 Analysis of the SMD simulation

1 For a constant-velocity simulation, it is particularly useful to look at the force as a function of position in the channel. To extract this information from the log file, tools such as awk or grep (Linux) can be used. Alternatively, one can write a script in VMD to extract the required information and output it to a new file. For the SMD simulation run here, the content of the script reads as follows:

```
### Open the log file for reading and the output .dat file for writing
set file [open "AmtB-SMD01.log" r]
set output [open "FvsP.dat" w]

### Loop over all lines of the log file
while { [gets $file line] != -1 } {
### Determine if a line contains SMD output.  If so, write the
### current $z$ position followed by the force along $z$ scaled by
### by the direction of pulling (0,0,-1)

   if {[string range $line 0 3] == "SMD "} {
     puts $output "[expr [lindex $line 4]] [expr -1*[lindex $line 7]]"
   }
}
### Close the log file and the output file
close $file
close $output
```

Run the script in VMD and plot the resulting output in, e.g., Grace or Excel (see Fig. 2). Right now, the data appears very noisy. To clean it up, calculate a running average with a window size of 500. This can be done in Grace by going to the menu **Data** → **Transformations** → **Running Averages** or in Excel by typing in an adjacent column of the first row `AVERAGE($B1:$B500)`, and then using the "fill down" function (location depends on Excel version) to fill the remaining cells.

2 Examine the simulation trajectory in VMD and try to correlate the force peaks in the plot with specific events. For example, the first two large peaks correspond to breaching the first two hydrophobic gates in AmtB, formed by Phe107 and Phe215. The three internal binding sites for NH_3 are also revealed by this plot. Compare them to the location of those found in the crystal structure. Finally, the other gate formed by Tyr32 and Phe31 can also be identified. If you have not completed your own simulation yet, an example trajectory with frames taken every 10 ps for the first SMD simulation is provided (`AmtB-SMD01-10ps.dcd`).

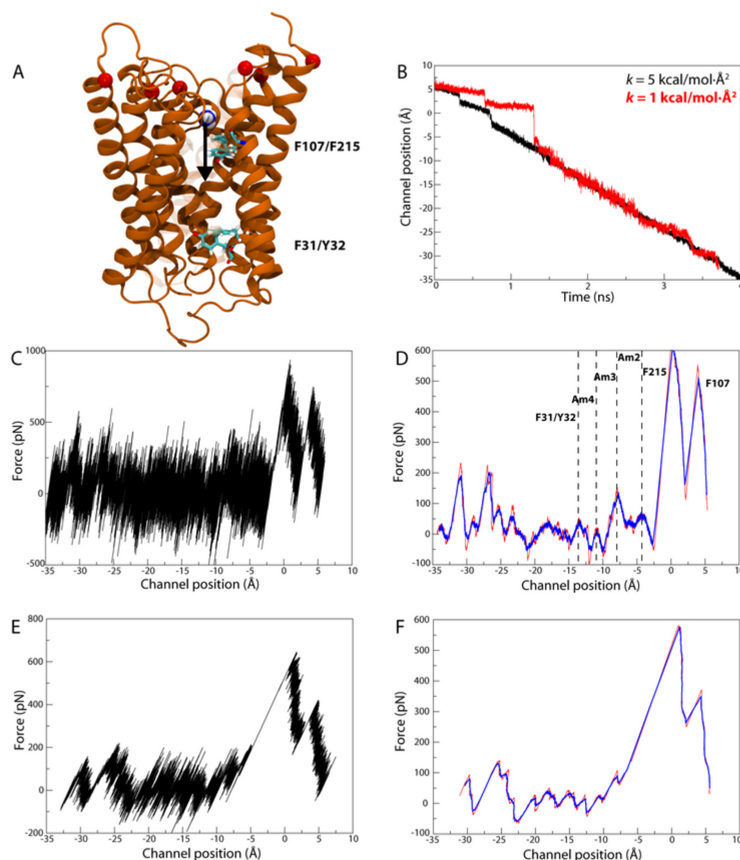


Figure 2: SMD as applied to NH_3 in AmtB. (A) Simulation setup. The NH_3 molecule begins above the first gate and is pulled in the z direction toward the second gate and beyond. Red spheres near the top indicate residues restrained to negate any net motion arising from the SMD forces. (B) Position of NH_3 during the SMD simulations for two different values of the SMD spring constant k. Notice how a larger k (black) enforces better the constant velocity of the pulled atom than the smaller k (red). (C) Force vs. position for SMD with $k = 5$ kcal/mol·Å^2. (D) Running average of force vs. position with averages taken over 0.5-ns (light, red line) and 1-ns (heavy, blue line) windows. Force peaks are correlated with specific channel features. (E,F) Same as (C,D) but with $k = 1$ kcal/mol·Å^2.

3 Determining the potential of mean force with ABF

Now with a basic understanding of the NH_3 permeation pathway and the barriers along it, a more quantitative characterization can be carried out. We will use the adaptive biasing forces method in NAMD (10, 34) to determine the potential of mean force (PMF) along the channel axis.

3.1 Setting up and running the ABF simulation

1 First, an appropriate reaction coordinate must be defined. While an obvious choice would be just to use the absolute position of NH_3 along the channel axis, this does not account for any fluctuations of the channel in z. A more apt choice is to relate the position of the molecule with that of the channel's center of mass; this also has the benefit of negating the need for additional restraints. Next, appropriate bounds for this coordinate must be defined. Using VMD, load the equilibrated PDB file for the AmtB system. Make two atom selections, one for NH_3 (i.e., "resname AMM1 and noh") and one for the protein C_α atoms (i.e., "protein and name CA"). You can get the distance along z separating them with the `tcl` command

```
expr [lindex [measure center $AMM1sel] 2] - [lindex [measure center $protCAsel] 2]
```

which is, approximately, the beginning of your reaction coordinate. We will use the two hydrophobic gates to more explicitly define the range, i.e., residues 107 and 215 on one side and 31 and 32 on the other. Again, using the `tcl` command above, but with $upperCAsel and $lowerCAsel instead of $AMM1sel, find the relevant region of the channel.

2 To enhance the efficiency of the ABF algorithm, the span of the reaction coordinate will be subdivided into equally spaced windows. The range found above is approximately [-13,5], while the NH_3 is initially positioned 9 Å above the center of the protein. Therefore, we will use four windows, each 5-Å wide, over the range [-13,7]. Create four directories, one for each window. Using the trajectory from the SMD simulation, find frames in which NH_3 is somewhere in each of the windows, i.e., between 2 and 7 Å above the protein's center of mass and so on. Write out a PDB for the entire system for each representative frame named, e.g., `win1start.pdb`. These will be the starting points for the four individual ABF simulations.

3 The NAMD configuration file used for the SMD simulations can be used again here with a few minor edits. First, the keyword `coordinates` should reference the appropriate starting frame for a given window. Remove any reference to SMD and constraints, but leave that for colvars.

4 As in the SMD section above, colvars will be used for the ABF calculations. The colvars input is similar, with a few important differences:

```
colvarsTrajFrequency 1000
colvarsRestartFrequency 1000

colvar {
   name Translocation

   width 0.1

   lowerboundary --z1--
   upperboundary --z2--

   lowerwallconstant 100.0
   upperwallconstant 100.0

   distanceZ {
     main {
       atomnumbers { <list of ammonia atom(s)> }
     }
     ref {
       atomnumbers { <list of protein atoms> }
     }
     axis ( 0.0, 0.0, 1.0 )
   }
}

abf {
   colvars Translocation
   fullSamples 1000
}
```

In this case, instead of `distanceXY` used to restrain the off-axis movement, the `distanceZ` variable is used. The width of 0.1 Å defines the bin size and is sufficiently small to generate a smooth PMF. The biasing forces and PMF calculation are invoked by the `abf` block. Create a colvars input file for each window, defining the lower and upper boundaries appropriately.

5 Run the simulation for each window. To restart an ABF simulation, in the subsequent configuration files, add the line

```
colvarsInput <inputname>.restart.colvars.state
```

which references the output of the previous run.

Judging convergence. It is difficult *a priori* to define the amount of sampling required to obtain a converged PMF. One way to assess convergence is to plot sequential PMFs and observe any changes as the simulation progresses. After two or more runs for a given window, plot the last two PMFs and compare. You can also plot the counts, which measure the number of times each bin has been sampled. When the PMF is converged, NH_3 should diffuse freely along the reaction coordinate, as the biasing forces cancel out any barriers and wells in the free energy that are encountered. Therefore, the counts should also increase uniformly for all bins after convergence. As another way to judge convergence, you can plot the forces stored in the `*.grad` files for all windows, checking for continuity across the window boundaries.

3.2 Analysis of the ABF simulations

Because some windows may require significant sampling to obtain a reasonably accurate PMF, you may also use the provided output for further analysis.

1 To re-combine the output from the separate windows in a single PMF, we will again use ABF. Create a new directory called **Merge** and place in there the last **grad** and **count** files from each window. Also copy a NAMD configuration file and a colvars input file.

2 In the configuration file, set it to `run` for 0 steps. Ensure that it still properly references PDB, PSF, and restart files. It is not important which input files you use for this simulation.

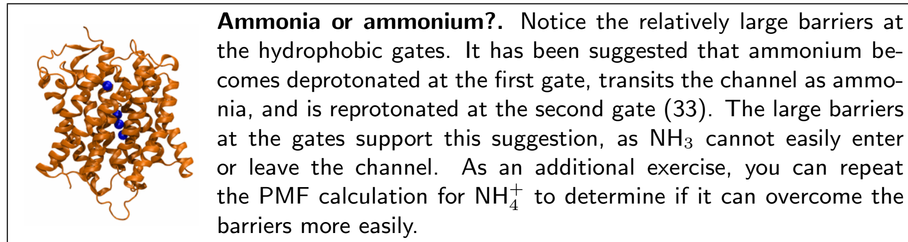
3 You will need to modify the colvars input file in two ways. First, change the upper and lower boundaries so that they encompass the entire range, i.e., -13 and 7. Then, add the following line to the **abf** block:

```
inputPrefix <win1name> <win2name> <win3name> <win4name>
```

where `win_name` refers to the prefix for the output files of that window.

4 Run the new simulation. NAMD will read in the gradients and counts from the provided output of the individual windows and generate a new set of outputs, including a new PMF, covering the entire reaction coordinate.

5 Plot the final PMF. Through comparison to the simulation, try to correlate the various extrema to specific locations in the channel. The minima should correspond to the locations of the crystallized ammonia molecules in the PDB `1U7G`, as shown in Fig. 3.



Ammonia or ammonium?. Notice the relatively large barriers at the hydrophobic gates. It has been suggested that ammonium becomes deprotonated at the first gate, transits the channel as ammonia, and is reprotonated at the second gate (33). The large barriers at the gates support this suggestion, as NH_3 cannot easily enter or leave the channel. As an additional exercise, you can repeat the PMF calculation for NH_4^+ to determine if it can overcome the barriers more easily.

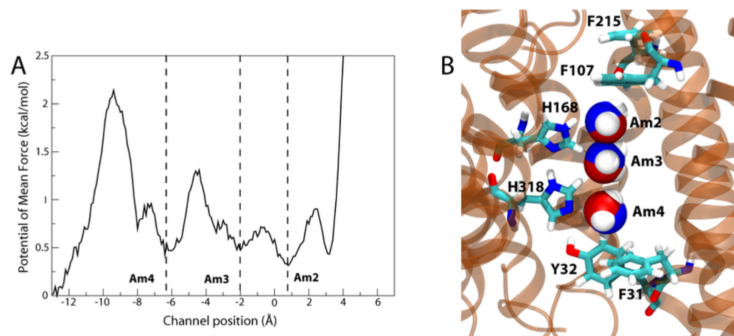


Figure 3: ABF as applied to NH_3 in AmtB. (A) PMF as a function of NH_3 position relative to the protein's center. (B) Close-up on the interior of the channel. The locations of the three minima in the PMF are indicated by the blue and white NH_3 molecules, with the darker red spheres showing the crystallized positions. Notice the near perfect overlap. Key residues at the gates and in the channel are also highlighted.

Acknowledgements

Development of this tutorial was supported by the National Institutes of Health (P41-RR005969 - Resource for Macromolecular Modeling and Bioinformatics).

References

[1] Izrailev S, et al. (1997) Molecular dynamics study of unbinding of the avidin-biotin complex Biophys J 72:1568–1581

[2] Sotomayor M, Schulten K (2007) Single-molecule experiments in vitro and in silico Science 316:1144–1148

[3] Jarzynski C (1997) Nonequilibrium equality for free energy differences Phys Rev Lett 78:2690–2693

[4] Jarzynski C (1997) Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach Phys Rev E 56:5018–5035

[5] Hummer G, Szabo A (2001) Free energy reconstruction from nonequilibrium single-molecule pulling experiments Proc Natl Acad Sci USA 98:3658–3661

[6] Park S, et al. (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality J Chem Phys 119:3559–3566

[7] Park S, Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulations J Chem Phys 120:5946–5961

[8] Darve E, Pohorille A (2001) Calculating free energies using average force J Chem Phys 115:9169–9183

[9] Hénin J, Chipot C (2004) Overcoming free energy barriers using unconstrained molecular dynamics simulations J Chem Phys 121:2904–2914

[10] Darve E, Rodríguez-Gómez D, Pohorille A (2008) Adaptive biasing force method for scalar and vector free energy calculations J Chem Phys 128:144120

[11] Hénin J, et al. (2010) Exploring multidimensional free energy landscapes using time-dependent biases on collective variables J Chem Theor Comp 6:35–47

[12] Jensen MØ, et al. (2002) Energetics of glycerol conduction through aquaglyceroporin GlpF. Proc Natl Acad Sci USA 99:6731–6736

[13] Ilan B, et al. (2004) The mechanism of proton exclusion in aquaporin channels Proteins: Struct, Func, Bioinf 55:223–228

[14] Wang Y, Schulten K, Tajkhorshid E (2005) What makes an aquaporin a glycerol channel: A comparative study of AqpZ and GlpF Structure 13:1107–1118

[15] Chen H, et al. (2007) Charge delocalization in proton channels. I. The aquaporin channels and proton blockage Biophys J 92:46–60

[16] Wang Y, Tajkhorshid E (2008) Electrostatic funneling of substrate in mitochondrial inner membrane carriers Proc Natl Acad Sci USA 105:9598–9603

[17] Celik L, Schiott B, Tajkhorshid E (2008) Substrate binding and formation of an occluded state in the leucine transporter Biophys J 94:1600–1612

[18] Yin Y, et al. (2006) Sugar binding and protein conformational changes in lactose permease Biophys J 91:3972–3985

[19] Jensen MØ, et al. (2007) Sugar transport across lactose permease probed by steered molecular dynamics Biophys J 93:92–102

[20] Gumbart J, Wiener MC, Tajkhorshid E (2007) Mechanics of force propagation in TonB-dependent outer membrane transport Biophys J 93:496–504

[21] Gumbart J, Schulten K (2006) Molecular dynamics studies of the archaeal translocon Biophys J 90:2356–2367

[22] Gumbart J, Schulten K (2007) Structural determinants of lateral gate opening in the protein translocon Biochemistry 46:11147–11157

[23] Gumbart J, Schulten K (2008) The roles of pore ring and plug in the SecY protein-conducting channel J Gen Physiol 132:709–719

[24] Henin J, et al. (2008) Diffusion of glycerol through *Escherichia coli* aquaglyceroporin GlpF Biophys J 94:832–839

[25] Dehez F, Pebay-Peyroula E, Chipot C (2008) Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel J Am Chem Soc 130:12725–12733

[26] Ivanov I, et al. (2007) Barriers to ion translocation in cationic and anionic receptors from the cys-loop family J Am Chem Soc 129:8217–8224

[27] Lin Y, Cao Z, Mo Y (2006) Molecular dynamics simulations on the *Escherichia coli* ammonia channel protein AmtB: Mechanism of ammonia/ammonium transport J Am Chem Soc 128:10876–10884

[28] Luzhkov VB, et al. (2006) Computational study of the binding affinity and selectivity of the bacterial ammonium transporter AmtB Biochemistry 45:10807–10814

[29] Nygaard TP, et al. (2006) Ammonium recruitment and ammonia transport by *E. coli* ammonia channel AmtB Biophys J 91:4401–4412

[30] Lamoureux G, Klein ML, Bernèche S (2007) A stable water chain in the hydrophobic pore of the AmtB ammonium transporter Biophys J 92:L82–L84

[31] Bostick DL, III CLB (2007) Deprotonation by dehydration: The origin of ammonium sensing in the AmtB channel PLoS Comput Biol 3:e22

[32] Yang H, et al. (2007) Detailed mechanism for AmtB conducting NH_4^+/NH_3: Molecular dynamics simulations Biophys J 92:877–885

[33] Khademi S, et al. (2004) Mechanism of ammonia transport by Amt/MEP/Rh: Structure of AmtB at 1.35 Å Science :1587–1594

[34] Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD J Comp Chem 26:1781–1802