

**The 26S Proteasome:
A computational study of the waste recycler of the cell**

Proposal for:
2014 Blue Waters allocation for UIUC Projects

PI: **Klaus Schulten**

Center for the Physics of Living Cells
University of Illinois at Urbana-Champaign
September 15, 2014

Project Abstract

Protein degradation is vital for a variety of essential cellular processes like apoptosis and transcription. Its malfunction is associated with severe diseases including cancer and neurodegenerative diseases. In eukaryotes, protein degradation is regulated by the ubiquitin-proteasome pathway, in which the 26S proteasome acts as its executive key player. The 26S proteasome is a 2.5 MDa multi-subunit molecular machine, which recruits, unfolds, and degrades tetra-ubiquitin tagged proteins. The recently obtained structure of the 26S proteasome provides a unique opportunity to explore for the first time its complex function and dynamics at atomic resolution through molecular dynamics (MD) simulations. Here, we propose to explore the first functional processes of the 26S proteasome degradation cycle, namely the recognition, recruitment and transport of a poly-ubiquitin tag. To overcome computational timescale limitations, we propose to employ MD simulations coupled with enhanced sampling methods, which can currently only be performed on a petascale machine such as Blue Waters. This project is a huge opportunity for computational biology at UIUC as it deals with one of the hottest areas of cell biology today, possibly Nobel Prize worthy, for which presently expertise and means (Blue Waters) reside only at UIUC.

Field of Science

The field of science of the proposed project is molecular cell biology.

Principal Investigator

Klaus Schulten is the Swanlund Professor of Physics at UIUC and a co-director of the NSF Physics Frontier Center for the Physics of Living Cells (CPLC). He is an expert in theoretical and computational biophysics and has directed the NIH-funded Biotechnological Research Center for Macromolecular Modeling and Bioinformatics for more than 20 years. The Center's molecular analysis and dynamics programs VMD and NAMD are used by over 260,000 registered users for research in many areas of biological science and are considered the fastest programs for simulations of large scale biological systems. Schulten's research papers have received over 70,000 citations, which is among the highest citation rates in all of computational biology.

Team Members

Dr. Till Rudack, *Alexander von Humboldt Postdoctoral Fellow*, trudack@ks.uiuc.edu

Dr. Lela Vukovic, *NSF Center for the Physics of Living Cells Postdoctoral Fellow*, lvukov1@ks.uiuc.edu

Yi Zhang, graduate student, yizhang@ks.uiuc.edu

Project Overview

Accurate and timely protein degradation is essential for maintenance of protein homeostasis in cells. Degradation in eukaryotes is regulated by the ubiquitin-proteasome pathway with 26S proteasome as a key player operating at its executive end [1]. In recent years, the 26S proteasome, also known as the waste recycler of the cell, has been identified as a possible drug target for various severe diseases [2]. For example, two approved anti cancer drugs are known to target the 26S proteasome [3, 4].

While the general function of the 26S proteasome is known to be recruiting, unfolding, and degrading tetra-ubiquitin tagged protein substrates, its detailed atomic mechanisms remain elusive. Recently, a near-atomic model of the complete 26S proteasome has been obtained, providing a structural context for understanding its function [5, 6]. However, complete understanding still requires identifying exact molecular processes performed by the 26S proteasome. MD simulations are uniquely suited to obtain dynamic insights into mechanisms of the 26S proteasome degradation cycle and close the last structural gaps, which are not accessible to structural experiments due to high flexibility or multi conformational structures.

Here, we propose to use MD simulations to examine in collaboration with our experimental coworkers the first step of the 26S proteasome degradation cycle, which is assumed to be driven by the highly flexible Rpn10 receptor [7]. The first step of this cycle entails recognition and recruitment of the tetra-ubiquitin tag on the substrate by Rpn10, followed by the transport of the tagged substrate by Rpn10 from solution to the active site of the de-ubiquitylating enzyme Rpn11. The relevant conformational changes of Rpn10, which are coupled to the events of recognition, binding and transport of tetra-ubiquitin, are expected to occur on timescales of several microsecond to milliseconds. These timescales are inaccessible to all-atom equilibrium MD simulations. However, study of conformational changes and binding events of flexible proteins on the stated timescales is well suited for enhanced sampling methods within the framework of MD simulations, such as generalized simulated annealing (GSAFold) [8] and replica exchange MD (REMD) [9] methods, both recently implemented in our group's NAMD software [10].

With the use of enhanced sampling MD, we will be able to: 1) obtain structures of flexible proteins, currently inaccessible to experiments, 2) for the first time, examine the physical plausibility of the hypothesis that flexible ubiquitin receptor Rpn10 delivers ubiquitin to the active site of de-ubiquitylating enzyme Rpn11; 3) determine how the 26S proteasome regulates flexible proteins to recognize and transport tagged proteins. The outcomes will be of high interest to experimentalists, since determining structure and dynamics of highly flexible proteins such as Rpn10 remains challenging.

Target Problem

Here, we propose to characterize with MD simulations how 26S proteasome recognizes, recruits, and transports the tetra-ubiquitin tag in the first step of the protein degradation cycle. This first step is thought to be primarily performed by Rpn10, which consists of a globular part and a long flexible arm, resembling a fishing rod, that contains a ubiquitin interacting motif (UIM) domain, resembling a fishing hook. While the globular part of Rpn10 binds stably to the proteasome complex [11, 12], the location, flexibility and dynamical behavior of the flexible Rpn10 arm within the 26S proteasome is unknown.

In principle, long MD simulations could be used to determine the location, flexibility and dynamical behavior of Rpn10 within the 26S proteasome,

in both free and ubiquitin-bound states. However, preliminary all-atom and coarse grained equilibrium MD simulations showed that the Rpn10 arm could interact and become trapped by the Rpn10 globular part on the timescale of 100 nanoseconds. The observed traps imply a rough free energy surface of Rpn10, containing multiple minima separated by high barriers. More trapped configurations of Rpn10 flexible arm are expected in the context of the complete proteasome. Therefore, in order to efficiently sample the whole conformational space of Rpn10 within the 26S proteasome complex, we propose to explore this complex with MD simulations combined with enhanced sampling methods GSAFold (later on referred as GSA) [8] and REMD [9], both recently implemented in NAMD [10].

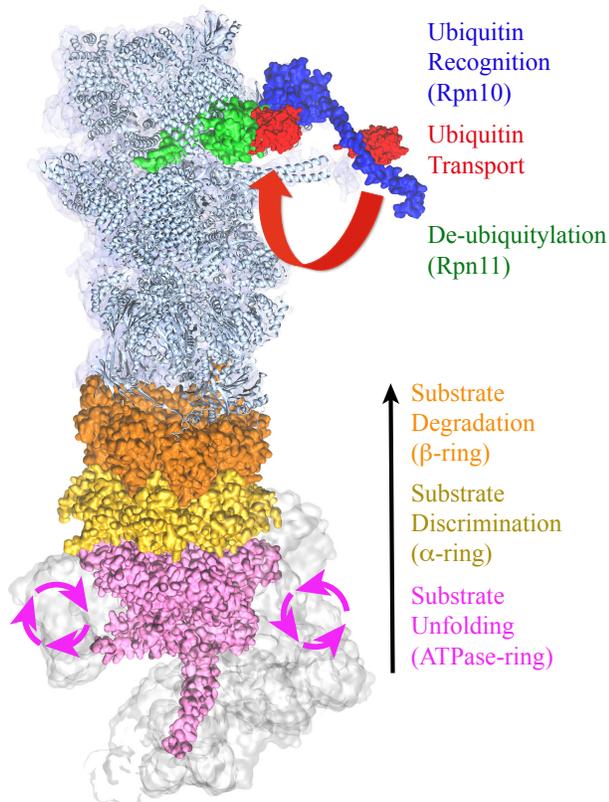


Figure 1: 26S Proteasome. The upper part depicts a ubiquitin (shown in red) recruitment by Rpn10 (dark blue) and transport of ubiquitin to the de-ubiquitylating protein Rpn11 (green) within the near-atomic structure of the 26S proteasome (light blue cartoon, PDB-ID: 4CR2) fitted by MDFF into a 8 Å cryo-EM density map (EMDB-ID: 2594). The 26S proteasome is rotationally symmetric; the bottom part shows the identical near-atomic structure in surface representation, highlighting the substrate processing. After recognition and cleavage of the tetra-ubiquitin tag by the specific subunits of the lid and the base (gray) the substrate is unfolded by the ATPase-ring (pink), checked by the α-ring (yellow), and degraded by the β-ring (orange).

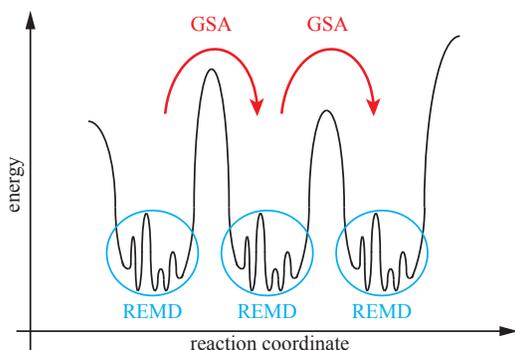


Figure 2: Sampling of the energy landscape. The combination of GSA and REMD enables the exploration of rough energy landscapes, with several local minima separated by high-energy barriers, which govern the biomolecular motion. By using the combination of GSA and REMD, the highly flexible multi conformational substates of Rpn10 become accessible to simulations.

The GSA method is uniquely suited for a broad exploration of rough free energy surfaces in large macromolecular complexes, as recently shown by our group for a cellulosome complex. While GSA simulations of partially restrained Rpn10 and 26S proteasome in implicit solvent will provide clusters of Rpn10 configurations, the obtained configurations need to be further explored by REMD simulations of fully flexible Rpn10 within the 26S proteasome in solvent. Flexibility of Rpn10 will be particularly important for examining how the bound tetra-ubiquitin modulates the conformational space available to Rpn10.

To overcome the computational time scale limit in REMD simulations and obtain a converged conformational ensemble of Rpn10 structures within the 26S proteasome, we propose to describe the system with the PACE hybrid-resolution model [13, 14], developed in our group. The PACE model, with proteins represented in a united-atom model and with solvent described in a coarse-grained solvent model, provides ~ 50 times acceleration in comparison with the all-atom model [13]. A combination of the PACE hybrid model and REMD was used recently with great success for examining self-assembly of flexible proteins into fibrils [15]. Since the combination of PACE and REMD lead to convergence in conformational sampling in [15], we expect similar convergence of Rpn10 configurations within the 26S proteasome.

Preliminary studies

Recently, first near-atomic structural models of 26S proteasomes in three different states (PDB identification codes: 4CR2, 4CR3, 4CR4) were derived from 8 Å-resolution cryo-EM density maps. These models were obtained by combining crystal structures of individual proteins (66 resolved proteins at present) and fitting them into density maps; the fitting procedure was performed with the help of the MDFF software within the NAMD program, developed by our lab [16, 17].

The obtained structural models require further improvements, such as addition of missing domains and flexible parts, to be complete and stable when probed in MD simulations. To perform the proposed study, it is crucial to obtain complete models of at least two 26S proteasome proteins, Rpn10 and Rpn11, which are known to bind to the polyubiquitin tag. We initially prepared complete structures of these proteins by combining available structural data on Rpn10 and Rpn11, preparing their homology models (where appropriate), and modeling in the missing parts with Modeller software [18] (Fig. 3). The prepared structures were relaxed in all-atom MD simulations (< 100 ns) and in hybrid reso-

lution 1-2 μ s MD simulations, until secondary structure convergence was observed. The completed and relaxed Rpn10 and Rpn11 structures have been incorporated by MDFF into the complete structural model of the 26S proteasome, thus providing a starting point structure for the proposed MD simulations.

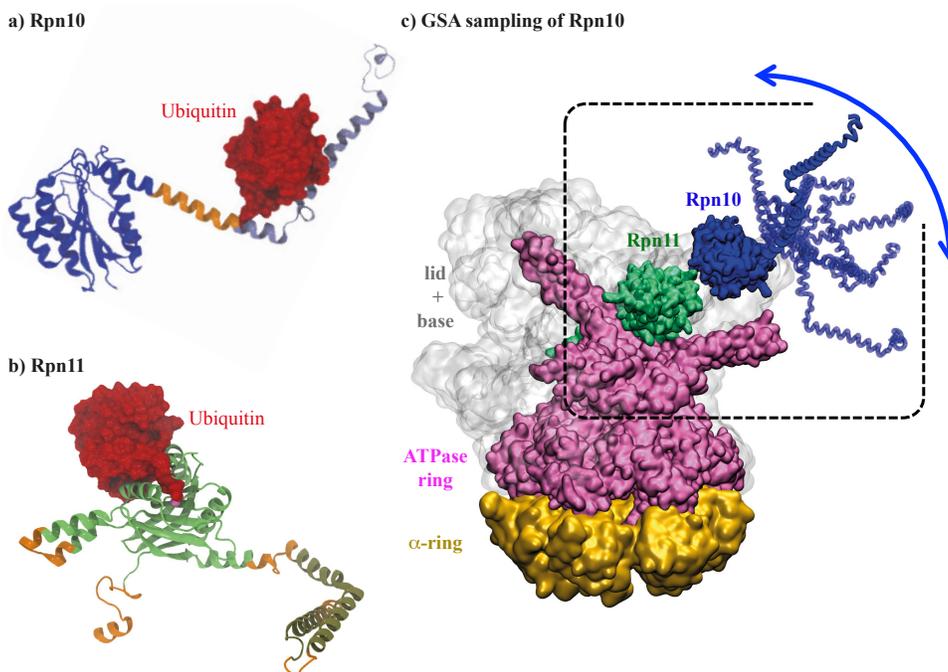


Figure 3: Simulation of the action of proteasome subunits Rpn10 and Rpn11. a) Complete structure of Rpn10. The globular part of Rpn10 (shown in dark blue) is a homology model based on the X-ray structure of Rpn10 of *S. pombe* (PDB ID 2X5N), the helical extension (light blue) that contains the ubiquitin interaction motif (UIM) is a homology model, based on its NMR structure for human Rpn10 with bound mono-ubiquitin (PDB ID: 2KDE). Experimentally not resolved parts (orange) were modeled with Modeller. **b) Complete structure of Rpn11.** The globular part of Rpn11 (light green) is based on the X-ray structure of the Rpn11/Rpn8 dimer in yeast (PDB ID: 4OCM). The helical N-terminus is based on PDB ID 3J47 (tan). Experimentally not resolved parts (orange) were modeled with Modeller. **c) Different conformations of Rpn10 obtained by GSA simulations.** Final structures of the preliminary GSA simulations of 32 configurations of isolated Rpn10 (dark blue) are overlaid with the complete 26S proteasome complex to give an impression of the configuration space covered by the flexible linker of the Rpn10.

To test if converged conformational ensembles could in principle be obtained by the proposed approach, we performed preliminary GSA simulations of isolated Rpn10 and REMD simulations of the Rpn10 flexible arm in solvent. A diverse set of structures of Rpn10 and its arm was generated already in short GSA (Fig. 3 c) and REMD simulations, indicating large flexibility of Rpn10 and the fact that the chosen simulation types are suitable to describe conformational space of Rpn10.

Description of Code(s)

NAMD (NANoscale Molecular Dynamics) is a parallel molecular dynamics code designed for high performance simulation of large biomolecular systems [19]. It employs the prioritized message-driven execution capabilities of the Charm++ parallel runtime system, allowing excellent parallel scaling on both massively parallel supercomputers

and commodity workstation clusters. Recognized as the leading software for running such simulations on large parallel machines, NAMD demonstrated scaling to thousands of processors in a 2002 paper that received a Gordon Bell Award. In 2012, a 100 million atom NAMD simulation was chosen as an acceptance benchmark for Blue Waters, and during the Blue Waters friendly user time, NAMD was also used to provide the first all-atom description of the 64 million atom HIV capsid, recently published in Nature [20].

A major advance in NAMD was support for petascale simulations of modestly sized systems through a new implementation of replica-exchange enhanced sampling methods including parallel tempering (temperature REMD), used in the proposed project. The NAMD 2.9 implementation was limited to Charm++ running on top of MPI, which was a particular limitation for Cray machines for which highly tuned Charm++ network layers had been developed. However, a new and completely backwards-compatible replica-exchange implementation into NAMD has been developed, which now directly supports these network layers as well as fine-grained Charm++ messaging.

The GSA method is an already established plugin in NAMD. GSA is highly scalable because each run is independently performed on one or a few nodes, so no communication among the nodes is needed. Therefore, GSA scales linearly with the number of investigated configurations. Different configurations are gained by different seeds for every run.

Experience, Readiness, Usage Plans and Funding Sources

Our software NAMD served as an acceptance test for the Blue Waters and is being used for our Petascale Computing Resource Allocation (PRAC) on Blue Waters. GSA and REMD methods are already implemented into the NAMD code. GSA was successfully applied on Blue Waters for the case of the cellulosome [21]. The development of NAMD on the Blue Waters is fully supported through PRAC by NCSA; no further support is needed.

Our estimate of Blue Waters usage per quarter is Q1: 40%, Q2: 10%, and Q3: 40% Q4: 10%. Within Q1 we will perform the GSA simulations. Within Q2 we will determine optimal simulation parameters for REMD, and perform production runs in Q3 and Q4.

Resources Required

The proposed simulations will be conducted using NAMD, a molecular dynamics program already successfully used on Blue Waters, as described above.

The first part of the proposed simulations will be conducted using GSAFold, a highly scalable simulated annealing method, implemented as a plugin into NAMD, which was already successfully used on Blue Waters [21]. Using the same parameters as in [21], benchmarks for 32 configurations of Rpn10 (4,126 atoms, 3,524 with fixed bonds and 602 with flexible bonds) with 60,000 steps required 21 hours on one node. We estimate 100,000 configurations as reasonable; the number of configurations is twice as large as used for the cellulosome [21]. However, in case of the cellulosome only two clusters of configurations were achieved, whereas benchmarks for Rpn10 with 320 configurations revealed already multiple configurations and a much wider configuration space. The re-

sults of the configuration space of the isolated Rpn10 will be used to enhance the simulation system by the parts of the proteasome that can be reached by the flexible part of the Rpn10. Two more simulations with and without ubiquitin will be performed within the context of the proteasome. As the additional parts are kept fixed, we assume a similar amount of computer time as for the isolated Rpn10. Hence, we will need 3 simulations of 100,000 configurations, which each will use 65,625 node-hours. In total we will use **196,875 node-hours** to conduct the GSA part of the study.

The second part of the proposed simulations will be conducted using REMD implemented into NAMD, which was already successfully used on Blue Waters. REMD simulations of two systems (each 180,000 beads) will be performed, containing the 26S proteasome regulatory particle by itself and with ubiquitin-bound Rpn10. The REMD predictor [22] suggested 475 replicas for the proposed system for a temperature range of 300 - 500K. A 5 ns test REMD simulation was performed for a 40,000 atom test system containing only the Rpn10 flexible arm, in order to determine optimal REMD parameters and identify a suitable temperature range that results in a versatile conformation ensemble of Rpn10. For this test system, we have been able to obtain a large ensemble of Rpn10 conformations and achieve a 40% replica-exchange rate with the suggested number of replicas, which is considered efficient for REMD simulations.

The performance of REMD simulations for the proposed proteasome system (180,000 beads) is determined based on the benchmark of 3.69 node-hours for 1 ns simulations for each of the 475 replicas, employing 4,750 Blue Water nodes. Based on our previous experience of simulating similar systems, we will perform 75 ns REMD simulations for each of the two systems to derive the average conformations of Rpn10 within the 26S proteasome. Therefore, we will need 2 simulations x 3.69 node-hours/ns/replica x 475 replicas x 75 ns = **262,910 node-hours**. It is also estimated that 5.0 TB of on-line disk storage will be required to house the simulation output.

In total we request **459,785 node-hours**.

Requested Start Date and Duration

The start date for the allocation for the proposed project is November 1, 2014. We expect to complete the proposed calculations within 9-12 months.

Bibliography

1. Förster, F., Unverdorben, P., Śledź, P., and Baumeister, W. “Unveiling the Long-Held Secrets of the 26S Proteasome” *Structure* 21, (2013): 1551–1562.
2. Dahlmann, B. “Role of Proteasomes in Disease” *BMC Biochemistry* 8, (2007): S3.
3. Adams, J. “Proteasome Inhibitors as New Anticancer Drugs” *Current Opinion in Oncology* 14, (2002).
4. Adams, J. “The Development of Proteasome Inhibitors as Anticancer Drugs” *Cancer Cell* 5, (2004): 417–421.
5. Śledź, P., Förster, F., and Baumeister, W. “Allosteric Effects in the Regulation of 26S Proteasome Activities” *Journal of molecular biology* 425, (2013): 1415–1423.
6. Unverdorben, P., Beck, F., Śledź, P., Schweitzer, A., Pfeifer, G., Plitzko, J. M., Baumeister, W., and Förster, F. “Deep Classification of a Large Cryo-EM Dataset Defines the Conformational Landscape of the 26S Proteasome” *Proceedings of the National Academy of Sciences* 111, (2014): 5544–5549.
7. Matyskiela, M. E., Bashore, C., Nogales, E., and Martin, A. “Complete Subunit Architecture of the Proteasome Regulatory Particle” *Nature* (2012).
8. Melo, M. C. R., Bernardi, R. C., Fernandes, T. V. A., and Pascutti, P. G. “GSAFold: A New Application of GSA to Protein Structure Prediction” *Proteins: Structure, Function, and Bioinformatics* 80, (2012): 2305–2310.
9. Sugita, Y. and Okamoto, Y. “Replica-Exchange Molecular Dynamics Method for Protein Folding” *Chemical Physics Letters* 314, (1999): 141–151.
10. Jiang, W., Phillips, J. C., Huang, L., Fajer, M., Meng, Y., Gumbart, J. C., Luo, Y., Schulten, K., and Roux, B. “Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD” *Computer Physics Communications* 185, (2014): 908–916.
11. Śledź, P., Unverdorben, P., Beck, F., Pfeifer, G., Schweitzer, A., Förster, F., and Baumeister, W. “Structure of the 26S Proteasome with ATP-S Bound Provides Insights into the Mechanism of Nucleotide-Dependent Substrate Translocation” *Proceedings of the National Academy of Sciences* 110, (2013): 7264–7269.
12. Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. “Molecular Architecture of the 26S Proteasome Holocomplex Determined by an Integrative Approach” *Proceedings of the National Academy of Sciences* 109, (2012): 1380–1387.

13. Han, W., Wan, C.-K., Jiang, F., and Wu, Y.-D. "PACE Force Field for Protein Simulations. 1. Full Parameterization of Version 1 and Verification" *Journal of Chemical Theory and Computation* 6, (2010): 3373–3389.
14. Han, W. and Schulten, K. "Further Optimization of a Hybrid United-Atom and Coarse-Grained Force Field for Folding Simulations: Improved Backbone Hydration and Interactions between Charged Side Chains" *Journal of Chemical Theory and Computation* 8, (2012): 4413-4424.
15. Han, W. and Schulten, K. "Fibril Elongation by A β 17-42: Kinetic Network Analysis of Hybrid-Resolution Molecular Dynamics Simulations" *Journal of the American Chemical Society* 136, (2014): 12450–12460.
16. Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. "Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics" *Structure* 16, (2008): 673–683.
17. Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., and Schulten, K. "Molecular Dynamics Flexible Fitting: A Practical Guide to Combine Cryo-Electron Microscopy and X-Ray Crystallography" *Methods* 49, (2009): 174–180.
18. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. "Comparative Protein Structure Modeling Using Modeller" *Current Protocols in Bioinformatics* (2002).
19. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. "Scalable Molecular Dynamics with NAMD." *Journal of computational chemistry* 26, (2005): 1781–802.
20. Zhao, G., Perilla, J. R., Yufenyuy, E. L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A. M., Schulten, K., Aiken, C., and Zhang, P. "Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics." *Nature* 497, (2013): 643–6.
21. Bernardi, R. C., Melo, M. C. R., and Schulten, K. "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems." *Biochimica et Biophysica Acta - General Subjects* (invited manuscript).
22. Patriksson, A. and Spoel, D. van der. "A Temperature Predictor for Parallel Tempering Simulations." *Physical chemistry chemical physics : PCCP* 10, (2008): 2073–7.