microtubule would not be possible. These molecular systems are composed by several dozens of millions of atoms and must be simulated for long periods of time (microseconds). These projects are examples of how **Blue Waters enables bold, new projects** that push the limits of what can be done with scientific computing. In our case, that means expanding molecular dynamics simulation capabilities from simulating just a few proteins to simulating full organelles.

PUBLICATIONS AND DATA SETS

Goh, B. C., et al., Atomic modeling of an immature retroviral lattice using molecular dynamics and mutagenesis. *Structure*, 23 (2015), pp. 1414-1425.

Fraiberg, M., CheY's acetylation sites responsible for generating clockwise flagellar rotation in Escherichia coli. *Mol. Microbiol.*, 95:2 (2015), pp. 231-44.

Sisneros, R., C. Malone, A. Nonaka, and S. Woosley Chemical visualization of human pathogens: the Retroviral Capsids. *SC15 Exhibition, Huston, TX: High Performance Computing, Networking Storage and Analysis.*

Lu, M., et al., Dynamic allostery governs cyclophylin A-HIV capsid interplay. *PNAS*, 112: 47 (2015), pp. 14617-14622.

Cassidy, C. K., et al., CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. eLife, (2015), 10.7554/*eLife*.08419.

Stone, J. E., Atomic detail visualization of photosynthetic membranes with GPU-accelerated ray tracing. *Parallel Comput.* 55 (2016), pp. 17-27.

Liu, C., et al., Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.*, 7 (2016), pp. 10714-10724.

Stone, J. E., Sherman, W. R. and Schulten K., Immersive molecular visualization with omnidirectional stereoscopic ray tracing and remote rendering. IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW) (IEEE Computer Society, Chicago, Illinois, May 23-27, 2016).

Stone, J. E., et. al. , High performance molecular visualization: In-situ and parallel rendering with EGL. IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW) (IEEE Computer Society, Chicago, Illinois, May 23-27, 2016).

Perilla, J.R., et. al., All-Atom Molecular Dynamics of Virus Capsids as Drug Targets. *J. Phys. Chem. Lett.* 7 (20126), pp. 1836–1844.

Heyrana, K., et al., Contributions of charged residues in structurally dynamic capsid surface loops to Rous sarcoma virus assembly. *J. Virol*, 90:12 (2016), pp. 5700-5714.

Stone, J. E., Early experiences porting the NAMD and VMD molecular simulation and analysis software to GPU-accelerated OpenPOWER platforms. International Workshop on OpenPOWER for HPC (ISC High Performance conference, Frankfurt, Germany, June 23, 2016).

THE RECYCLING MACHINERY OF THE CELL

Allocation: Illinois/500 Knh
PI: Klaus Schulten[1]
Co-PIs: Till Rudack[1] , Lela Vukovic[2]
Collaborator: Wolfgang Baumeister[3]

[1]University of Illinois at Urbana Champaign
[2]University of Texas at El Paso
[3]Max Plank Institute of Biochemistry

EXECUTIVE SUMMARY

While waste recycling became popular in our daily life more recently, living cells have mastered recycling of their protein content since their very beginning. Recycling of unneeded protein molecules in cells is performed by a molecular machine called 26S proteasome, which cuts these proteins into smaller pieces for reuse as building blocks for new proteins. Proteins that need to be recycled are labeled by tags made of poly-ubiquitin protein chains. The 26S proteasome machine recognizes and binds to these tags, pulls the tagged protein close, then unwinds it, and finally, cuts it into pieces.

Despite its substantial role in the cell's life cycle, the proteasome's atomic structure and function remain elusive. Employing a combination of computational techniques implemented using nanoscale molecular dynamics (NAMD) along with cryo-electron microscopy (EM) data, we obtained an atomic structure of the human 26S proteasome and investigated the mechanism underlying substrate recruitment and unfolding.

FIGURE 1: The recycling system of the cell. The 26S proteasome is the key player of the human protein recycling system. The first structure of the human 26S proteasome obtained through integrative modeling utilizing Blue Waters will lead to breakthroughs in understanding its detailed function and will play a pivotal role in the development of the 26S proteasome as a drug target for molecular disease therapies.

INTRODUCTION

Recycling of proteins by degradation is vital for a variety of essential cellular processes, including protein quality control, cell cycle regulation, adaptive immune response, and apoptosis. The 26S proteasome is responsible for the vast majority of regulated intracellular protein degradation and is an important drug target for multiple diseases, including cancer, neurodegenerative diseases, and immuno-inflammatory disorders. The 26S proteasome is an adenosine triphosphate (ATP) hydrolysis driven 2.5 MDa molecular machine that recruits, unfolds, and degrades poly-ubiquitin tagged proteins through a complex interaction clockwork of over 60 known protein subunits (Fig. 1).

Despite its substantial role in the cell's life cycle, the proteasome's atomic structure and function remained elusive. However, recent developments in hybrid experimental methods based on the revolutionary advance of electron microscopy, together with improvements in real space refinement methods [1], have led to previously unimaginable

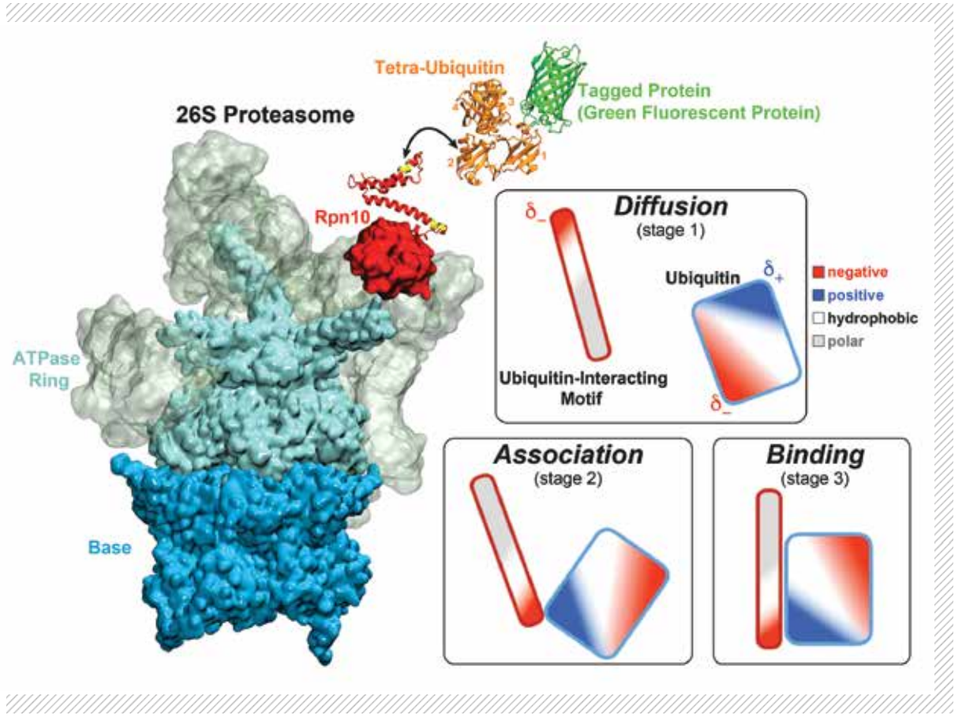

FIGURE 2: Recognition of the tetra-ubiquitin tag by the 26S proteasome in three stages: In stage 1 of the recognition process conserved complementary electrostatic patterns of Rpn10 and ubiquitins guide protein association; stage 2 induces refolding of Rpn10 and tetra-ubiquitin tag; stage 3 facilitates formation of hydrophobic contacts between the tag and Rpn10.

structural information on large macromolecular complexes [2].

In this report, we present the **first** atomic structure of the human 26S proteasome with bound nucleotides based on a 3.9 Å resolution cryo-EM density. It was obtained by integrative modeling combining molecular dynamics flexible fitting (MDFF) with *de novo* structure prediction algorithms [3]. Furthermore, utilizing molecular dynamics (MD), steered molecular dynamics (SMD) simulations, and generalized simulated annealing (GSA) techniques, we revealed the underlying mechanism of poly-ubiquitin tagged protein recognition and recruitment through the 26S proteasome [4].

METHODS & RESULTS

To build an accurate atomic model of the human 26S proteasome we followed the strategy established for large macromolecular complexes [1,2]. We first built comparative models of the human 26S proteasome subunits based on the yeast 26S proteasome and extended structurally-unresolved segments with *de novo* modeling. Next, we refined the structure according to the density in real space by combining MDFF with *de novo* structure prediction as well as Monte Carlo-based backbone and side chain rotamer search algorithms in an iterative manner. The features observed in the resulting structure are important for coordinating the proteasomal subunits during substrate processing. One of the key novel features is that in the predominant state of the heterohexameric ATPase motor of the proteasome there is one adenosine diphosphate (ADP) and five ATP-bound. The structure of the ADP-bound subunit is distinct from the other five subunits most notably in the pore loop region, which is known to be a key region interacting with the substrate during unfolding. Detailed investigations of the unfolding process by MD simulations and path sampling techniques are now possible and planned in future.

Furthermore, we employed MD, SMD, and GSA simulations on Blue Waters, utilizing QwikMD [5], our new intuitive "point and click" graphical interface connecting visual molecule dynamics (VMD) and NAMD. QwikMD significantly simplifies the setup, execution, and analysis of NAMD simulations from laptop computers up to supercomputers. Our simulations elucidated the process of how ubiquitin-tagged proteins are recognized by the partially disordered flexible arm of the ubiquitin receptor Rpn10 of the 26S proteasome (Fig. 2). Rpn10 consists of a globular domain that binds to the 26S proteasome and a flexible arm that contains ubiquitin-interacting motifs (UIM). Previous studies identified that ubiquitin binds the UIMs through hydrophobic interactions; however, poly-ubiquitin mostly assumes closed forms in solution, in which the hydrophobic patches are protected. Our results reveal that hydrophobic UIMs of Rpn10, which bind to hydrophobic patches of ubiquitin, are likewise mostly protected prior to binding. Therefore, refolding of both tetra-ubiquitin and Rpn10 to reveal hydrophobic patches is a necessary step of the binding process. Since it is unlikely that protected hydrophobic patches can initiate refolding and efficiently search for each other, our results indicate that electrostatic interactions are responsible for the initial stage of the mutual recognition between UIMs and Rpn10.

The combination of MDFF with Monte Carlo-based rotamer search algorithms enabled us to obtain a structural model based on high-resolution (< 4.0 Å) cryo-EM densities and obtain new insights into substrate recognition and unfolding. The obtained atomic structure permits further computational studies of proteasomal function and will serve as a starting point for future structure-guided drug discovery that will assist to develop the proteasome as a drug target.

WHY BLUE WATERS

Investigating the functional processes of large protein machinery, such as the proteasome (3.5 M atoms), which occur on the millisecond timescale, is **only possible on petascale** computing resources, such as Blue Waters.

De novo structure prediction and Monte Carlo-based sampling algorithms are only efficient if thousands of models are predicted. Employing these algorithms for the numerous subunits of the proteasome is a well-suited task for the large-scale parallel architecture of Blue Waters.

NEXT GENERATION WORK

We would like to investigate how the motor action of ATP hydrolysis in the proteasome is driving the subunit reorganization during the functional cycle of substrate processing, which requires extensive sampling of long trajectories.

In order to exploit the proteasome's role as a drug target, high-throughput screening of thousands of drugs that possibly interact with the proteasome might be feasible by the year 2020 and would require pre-exascale computing for such a complex macromolecular machinery as the 26S proteasome.

PUBLICATIONS AND DATA SETS

Schweitzer, A., et al., The structure of the human 26S proteasome at a resolution of 3.9 Å. *PNAS*, (2016).

Zhang Y., et al., Recognition of poly-ubiquitins by the proteasome through protein re-folding guided by electrostatic and hydrophobic interactions. *J. Phys. Chem. B*, (2016). DOI: 10.1021/acs.jpcb.6b01327