Chemical Visualization of Human Pathogens: the Retroviral Capsids

Juan R. Perilla*†‡, Boon Chong Goh*†‡, John E. Stone*, and Klaus Schulten*†
*Beckman Institute, University of Illinois at Urbana-Champaign
†Department of Physics, University of Illinois at Urbana-Champaign
‡Co-first author

***Abstract*—**

Retroviruses are pathogens characterized by their ability to incorporate viral DNA into a host cell's genome. Retroviruses like Rous Sarcoma Virus (RSV) infect cells during mitosis, when the chromatin is exposed to the cytoplasm. Conversely, the genus of lentiviruses, like the human immunodeficiency virus (HIV), have evolved to infect non-dividing cells [1]. Despite infecting cells at different stages of their life cycles, RSV and HIV share a similar late stage replication cycle that is highly dependent on the group actin polyprotein precursor (Gag), which contains the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins. Both HIV's CA and Gag are considered unexploited targets for pharmaceutical intervention. We describe the techniques that were used to build, simulate, analyze and visualize the structures of both Gag and CA, and we discuss scientific visualization needs that spurred development of an interactive GPU-accelerated ray tracing engine and the use of remote visualization technologies.

I. Introduction

Retroviruses are parasites that pose a major health threat to humans (e.g., HIV and human T-cell leukemia virus), as well as other animals (e.g., chickens (RSV), monkeys (M-PMV), mice (MLV), etc). In the case of HIV, numerous treatments have been developed, but the virus adapts quickly to antiviral drugs such that new compounds need to be refined continuously. HIV infects the human cell by inserting its genes, packaged into a capsid, into the cell's nucleus, and thus taking control over the cell's machinery. Once the cell is infected, the virus replicates and is released into the host's bloodstream.

As a virus that infects non-dividing cells, which have their genome protected behind a nuclear membrane, HIV has to take advantage of natural cell responses to induce cooperation of the host to reach inside the nucleus. Some species of monkeys are immune to HIV by spoiling this cooperation with the capsid, attacking the capsid instead. Likewise, pharmacological interventions seek to attack the HIV capsid to prevent cooperation with the cellular machinery, or simply by breaking the capsid apart. Such interventions, however, require knowledge of the chemical and physical properties of the capsid. In an accompanying movie we show the HIV capsid, a colossal structure, containing about 1,300 proteins with altogether 4 million atoms. Although the capsid proteins are all identical, they nevertheless arrange themselves into a largely asymmetric structure. The large size and lack of symmetry posed a huge challenge to resolving the chemical structure of the HIV capsid [2]. Embedding the structure into a computational model that also included physiological solvent led to a 64-million-atom simulation, the largest such simulation achieved to date [3]. While the solution of the HIV capsid structure was a great achievement, it was just the starting point for the main research agenda, which is developing new antiviral drugs. Computational biologists presently simulate how the capsid exploits cellular proteins to be guided to the cell's nucleus and how antiviral drugs interfere with that process. The currently identified drug candidates, while effective against the capsid, are unfortunately also highly toxic to the patient. An incredibly precise picture of the interactions of the capsid with the cell and drugs has emerged from the simulation, which can help guide future drug development. Such research, however, requires the most extreme computer power available today.

Newly released viruses are unable to infect healthy cells until they reach a mature state. An alternative strategy of preventing virus spread is therefore, to lock the viral particles in their immature, non-infectious state. However, to make the immature virus an attractive target for structure-based drug development one needs to know its chemical structure. Unfortunately, the complexity and size of the viral particle – an incomplete hexagonal shell close to 100 nm in size – have prevented the experimental determination of the atomic-detail chemical structure of the virus. Nonetheless, we have recently determined an atomistic structure of the immature retroviral lattice for Rous sarcoma virus [4] as presented in the accompanying movie. The multi-domain RSV model reveals novel features of the packing and dynamics of the immature capsid protein with implications for the maturation process and confirms the stabilizing roles of the so-called upstream and downstream domains of the immature RSV. Since RSV is a close cousin of HIV, the RSV structure may lead to a better experimentation and targeting of HIV.

II. Computational Challenges

Computer simulations enable researchers to study the mechanics of virus processes at atomic resolution, probing spatial and temporal resolutions that are currently inaccessible to experimental methods alone [3]. Petascale supercomputers provide the computing power necessary to simulate very large biomolecular systems such as the HIV virus in all-atom detail [2]. These studies require tremendous computational capabilities for both the simulations as well as the analysis and visualization of the results.

The immature RSV hexamer model (Figure 1 right) was derived through a combination of state-of-the-art modeling techniques, including, cryo-EM-guided homology modeling, and large-scale molecular dynamics simulations. Chemical structures of RSV capsid proteins, including the p10 peptide,

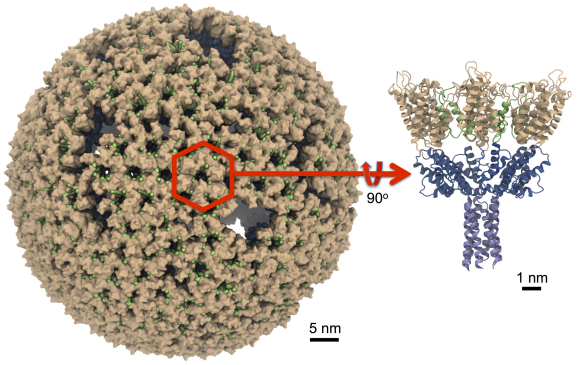

Fig. 1. The immature retroviral capsid for Rous sarcoma virus [4] (left) are composed of hundreds of identical hexamer models (right). The immature RSV hexamer model consists of p10 peptide (green), NTD (cream), CTD (violet-blue) and SP-NC (blue-gray).

N-terminal domain (NTD), and C-terminal domain (CTD) [5], [6], were docked in a cryo-EM density map of an immature retroviral lattice [7]. To complete the construction of an immature RSV model, a crucial domain downstream of the capsid protein, namely spacer peptide-nucleocapsid (SP-NC), must be incorporated into the model. However, structural knowledge of this domain is scarce as experimentalists struggle to elucidate the structure of the SP-NC domain at high resolution. Therefore, the SP-NC domain has to be modeled using purely computational techniques. The SP-NC domain was previously proposed to be a six-helix bundle (6HB) and a 6HB model was proposed whose structure is based on an ion channel [8]. The tertiary arrangement of the 6HB model had to be optimized such that a ring of salt bridges can be formed to stabilize the 6HB model [4]. To refine the structure of the immature RSV hexamer model, an enhanced sampling technique named replica exchange MD simulation, a feature available in NAMD [9], was employed to simultaneously simulate $>$100 copies of a 0.6-million-atom system for a total simulation time of 5 μs. Such massive calculation require 5,000 Cray XK7 GPU-accelerated compute nodes and can only be completed using petascale supercomputers.

The scaffold of an entire immature RSV capsid model (Figure 1 left) was constructed using the positional information derived from the Thomson problem [10]. Four-hundred negatively-charged particles were placed on the surface of a sphere of radius 60 nm, and subsequently equilibrated by Coulombic repulsion. Once equilibrated, particles were deleted manually to mimic the pattern of the defects as observed in the cryo-electron tomogram of authentic immature RSV capsids [11]. The all-atom model of the immature RSV hexamer model was subsequently mapped to the final positions of the remaining particles. The hexamers were rotated along the azimuthal direction to minimize steric clashes, the resulting model is showed in the accompanying movie and in Fig. 1. Visualizing the incomplete spherical shell of the immature RSV capsid allows us to inspect the possible mechanisms for enzymatic reactions that initiate the maturation process.

The model of the mature HIV-1 capsid [2] required an ingenious combination of computational and experimental techniques. The hexamer-hexamer interactions were derived using the molecular dynamics flexible fitting (MDFF) method [12]. This method incorporates the electronic density obtained by the microscope, with a quantum-mechanically derived atomic force-field [13], allowing determination of the location of individual atoms. Due to the lack of experimental information regarding the interactions between pentamers and hexamers, the model was derived entirely by computational modeling through long-time scale simulations. Remarkably, the 1.5 μs of simulation obtained for the pentamer-of-hexamers – a 1.5-million-atom system – required the continuous use of 400 Cray XK7 nodes for four months [2].

With the building blocks of a capsid at hand, namely the hexamer-of-hexamers and pentamer-of-hexamers, it was essential to determine a realistic scaffold of an HIV-1 capsid. Thus, to sample the conformational space spanned by mature retroviral capsids, the spiral algorithm was employed [14]. The algorithm treats the capsid as a cubic planar graph, analogous to fullerenes, enabling the search of candidate conformations by shuffling of pentagons [15]. As illustrated in the movie, determination of the 3-D geometry of each graph, was then performed by embedding of the cubic graph in two dimensions employing the Tutte algorithm [16], followed by mapping of the 2-D graph into a sphere of arbitrary radius [15]. Since the connectivity of the edges is dictated by the graph, a global geometry optimization is performed by treating each edge as a Carbon atom and employing a molecular forcefield. After the candidate structures are generated, the cross-correlation between each model and cryo-electron tomography of authentic capsids was calculated to measure the likelihood of each model [2]. The scaffold was then used to place hexamers-of-hexamers, as well as pentamers-of-hexamers, yielding a complete model of a retroviral capsid. Furthermore, the model was further optimized by MD simulations and validated in living human cells [2]. The HIV capsid is the largest biomolecular dynamics simulation – a 64-million-atom system – performed to date [3]. The simulation required the sustained use of NAMD on 8,000 Cray XK7 nodes for a period of two months, in order to obtain enough sampling.

Parallel MD simulations were performed using NAMD [9], running on the GPU-accelerated Cray XK7 compute nodes of Titan at ORNL and Blue Waters at NCSA. NAMD and VMD use the Cray XK7 GPUs for acceleration of large-size and long-timescale simulation, analysis, and visualization tasks [17]–[21], and they provide built-in scripting and plug-in interfaces allowing them to be extended and customized for the research tasks described above.

III. Visualization Techniques

All of the renderings of the all-atom virus capsids, molecular dynamics simulations, and experimental cryo-EM densities were produced using VMD [22]. The visualizations encompass a wide variety of molecular representations. In several overview renderings of the virus capsids, capsid subunits are drawn using “QuickSurf”, a fast GPU-accelerated molecular surface representation provided in VMD [20], [23], as shown in the viral overview renderings shown in Figs. 1 and 2. The QuickSurf representation allows the structure to be depicted with smoothly variable detail, enabling atomic-scale features to be clearly shown or abstracted via smoothing. Renderings of the pentameric and hexameric lattice scaffolding used to construct initial all-atom virus capsid models were produced using the VMD scripting interfaces [24].

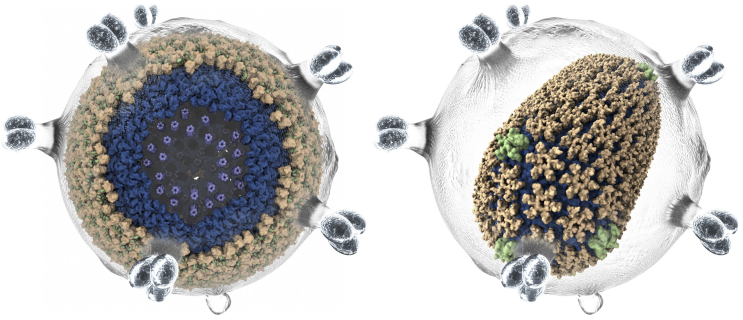

Fig. 2. VMD overview rendering of the immature retroviral lattice for Rous sarcoma virus [4] (left) and the mature HIV capsid (right) composited with a conceptual rendering of the viral envelope.

During fitting of the all-atom virus capsid to experimental cryo-EM density, individual protein domains are shown in so-called "ribbon" or "cartoon" representations in combination with the target experimental cryo-EM density map drawn with a transparent material that allows the interior protein backbone structure to be clearly seen. The combination of multiple superimposed molecular representations makes it possible to see the improving fit between the experimental cryo-EM density map and the all-atom virus structure as fitting progresses.

Colors were selected to clearly delineate key components of the immature and mature virus capsid structures, highlighting for example, the distinction between the large number of hexameric subunits and the 12 pentameric subunits in the mature HIV capsid. Different colors were also selected for latch proteins, and N-terminal and C-terminal subunits. Shadows and broad-area lighting techniques such such as ambient occlusion help provide the viewer with visual cues that elucidate pores, crevices, pockets, and cavities in molecular surfaces. The use of ambient occlusion lighting and carefully selected surface parameters clarifies the boundaries of individual hexamer and pentamer subunits, and provides shading that helps differentiate the N-terminal and C-terminal subunits.

The VMD `ViewChangeRender` and `VMDMovie` plugins were used to develop, manage, and render all of the individual movie segments. Movie segments were produced using keyframes composed from user-specified view orientations, with associated simulation trajectory frames, molecular representations, and lighting, shading, and material properties.

IV. Interactive GPU-Accelerated Ray Tracing

Built-in parallel ray tracing (RT) engines in VMD eliminate the need for intermediate disk I/O that would be harmful for performance [25], and they provide advanced lighting and shading capabilities unavailable in most rasterization-based visualization tools, including ambient occlusion lighting and shadows, depth-of-field focal blur, reflection and high quality transparency, stereoscopic cameras, and panoramic projections All of the renderings produced with VMD used the built-in TachyonL-OptiX GPU-accelerated RT engine [20], a lightweight adaptation of the Tachyon RT engine [19], [26] based on OptiX [27] and CUDA [28].

A fully-interactive version of the TachyonL-OptiX GPU-accelerated RT engine allows the user to view the molecular scene in exactly the way it will appear in final rendering, but achieving high interactivity through a progressive-refinement RT algorithm that updates the display continuously as stochastic samples accumulated. When the user changes the view orientation or global lighting or shading parameters, VMD clears the RT accumulation buffer and stochastic sampling starts anew. Large scenes with complex shading can be interactively rendered with faster display or sample convergence rates using multiple GPUs in parallel.

To facilitate accurate and rapid previewing of HIV capsid structures, VMD interactive RT was used extensively on a high performance remote visualization server with four NVIDIA Tesla K80 GPUs. Remote visualization allowed researchers to exploit the interactive RT feature of VMD while traveling, using conventional laptop computers that would otherwise be entirely incapable of such high-quality interactive rendering. The remote visualization server provided GPU-accelerated H.264 video streaming via NICE DCV and the NVIDIA GRID and NVENC video SDKs, encoding directly from the GPU framebuffer to achieve high frame rates and reduced latency by minimizing PCIe transfers between the GPUs and the host.

V. Final Rendering and Post Production Editing

Final rendering of movie frames was completed on the remote visualization server described above and on the Blue Waters Cray XK7 nodes [19], [20], [29]. Movie frames were rendered in 16:9 aspect ratio at HD 1920×1080 resolution, with 12 antialiasing samples per pixel and 12 ambient occlusion shadow rays per hit, and two directional lights; transmission rays and shadow filtering were performed for transparent geometry. The final movie was composed from individual movie segments, still images, artwork, and audio narration components that were assembled using the Adobe After Effects nonlinear video editing software. Narration audio was acquired and composed using Adobe Audition. All audio tracks were normalized to +0.1 db and each voice track was individually equalized and processed with a noise reduction filter. Music was composed and performed by: http://www.bensound.com.

VI. Conclusions

Ongoing advances in experimental imaging and all-atom modeling of virus structures promise new insights into the basis of viral infection. The processes of structure determination, simulation, analysis, and visualization of large viruses constantly pose new methodological, software, and hardware challenges due to their size and complexity. By harassing the power of the state-of-the-art supercomputers and cryo-electron microscopy, we are now able to derive atomic models of the mature and immature retroviral capsids. These capsid models will serve as a platform to investigate the delicate interactions of the capsid with cellular host factors and drug molecules.

The HIV modeling and visualization tasks described herein are only possible with the use of petascale supercomputers, and represent the current threshold for computational virology. As such, extensive and routine use of the capabilities of the Blue Waters and Titan supercomputers was required, and high performance remote visualization was employed to accommodate a mix of highly parallel, and highly interactive modeling and visualization tasks. In the future we expect that high-interactivity remote visualization will be a feature of all

high performance computing centers, eliminating the need for transfer of large datasets and enabling interactive modeling and visualization tools to be easily and effectively used.

ACKNOWLEDGMENTS

This research is part of the Blue Waters sustained-petascale computing project supported by NSF awards OCI-0725070 and ACI-1238993, the state of Illinois, and "The Computational Microscope" NSF PRAC awards OCI-0832673 and ACI-1440026. This research used the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, supported by DOE Contract DE-AC05-00OR22725. The authors acknowledge support from NSF rapid response research award ACI-1524703, and NIH grants 9P41GM104601 and 5R01GM098243-02. The authors acknowledge Dr. Jodi Hadden for narrating the introduction of the movie.

REFERENCES

[1] Laura Hilditch and Greg J Towers. A model for cofactor use during HIV-1 reverse transcription and nuclear entry. *Curr. Opin. Virol.*, 4:32–36, 2014.

[2] Gongpu Zhao, Juan R. Perilla, Ernest L. Yufenyuy, Xin Meng, Bo Chen, Jiying Ning, Jinwoo Ahn, Angela M. Gronenborn, Klaus Schulten, Christopher Aiken, and Peijun Zhang. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, 497:643–646, 2013.

[3] Juan R. Perilla, Boon Chong Goh, C. Keith Cassidy, Bo Liu, Rafael C. Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[4] Boon Chong Goh, Juan R. Perilla, Matthew R. England, Katrina J. Heyrana, Rebecca C. Craven, and Klaus Schulten. Atomic modeling of an immature retroviral lattice using molecular dynamics and mutagenesis. *Structure*, 23:1414–1425, 2015.

[5] Narayanasamy Nandhagopal, Alan A. Simpson, Marc C. Johnson, Adam B. Francisco, Gisela W. Schatz, Michael G. Rossmann, and Volker M. Vogt. Dimeric rous sarcoma virus capsid protein structure relevant to immature Gag assembly. *J. Mol. Biol.*, 335:275–282, 2004.

[6] R Campos-Olivas, J L Newman, and M F Summers. Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses. *J. Mol. Biol.*, 296(2):633–49, 2000.

[7] Tanmay A M Bharat, Norman E Davey, Pavel Ulbrich, James D Riches, Alex de Marco, Michaela Rumlova, Carsten Sachse, Tomas Ruml, and John A G Briggs. Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy. *Nature*, 487(7407):385–9, July 2012.

[8] Di L Bush, Eric B Monroe, Gregory J Bedwell, Peter E Prevelige, Judith M Phillips, and Volker M Vogt. Higher order structure of the Rous sarcoma virus SP assembly domain. *J. Virol.*, 88(March):5617–5629, 2014.

[9] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comp. Chem.*, 26:1781–1802, 2005.

[10] J J Thomson. XXIV. on the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Philosophical Magazine Series 6*, 7(39):237–265, 1904.

[11] Alex de Marco, Norman E Davey, Pavel Ulbrich, Judith M Phillips, Vanda Lux, James D Riches, Tibor Fuzik, Tomas Ruml, Hans-Georg Kräusslich, Volker M Vogt, and John A G Briggs. Conserved and variable features of Gag structure and arrangement in immature retrovirus particles. *J. Virol.*, 84(22):11729–36, 2010.

[12] Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, 16:673–683, 2008.

[13] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theor. Comp.*, 8(9):3257–3273, 2012.

[14] David E Manolopoulos and Patrick W Fowler. Molecular graphs, point groups, and fullerenes. *The Journal of chemical physics*, 96(10):7603–7614, 1992.

[15] Peter Schwerdtfeger, Lukas Wirz, and James Avery. Program fullerene: a software package for constructing and analyzing structures of regular fullerenes. *Journal of computational chemistry*, 34(17):1508–1526, 2013.

[16] William T Tutte. How to draw a graph. *Proc. London Math. Soc*, 13(3):743–768, 1963.

[17] John E. Stone, James C. Phillips, Peter L. Freddolino, David J. Hardy, Leonardo G. Trabuco, and Klaus Schulten. Accelerating molecular modeling applications with graphics processors. *J. Comp. Chem.*, 28:2618–2640, 2007.

[18] James C. Phillips, John E. Stone, and Klaus Schulten. Adapting a message-driven parallel application to GPU-accelerated clusters. In *SC '08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, Piscataway, NJ, USA, 2008. IEEE Press.

[19] John E. Stone, Barry Isralewitz, and Klaus Schulten. Early experiences scaling VMD molecular visualization and analysis jobs on Blue Waters. In *Extreme Scaling Workshop (XSW), 2013*, pages 43–50, Aug. 2013.

[20] John E. Stone, Kirby L. Vandivort, and Klaus Schulten. GPU-accelerated molecular visualization on petascale supercomputing platforms. In *Proceedings of the 8th International Workshop on Ultrascale Visualization*, UltraVis '13, pages 6:1–6:8, New York, NY, USA, 2013. ACM.

[21] John E. Stone, Ryan McGreevy, Barry Isralewitz, and Klaus Schulten. GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. *Faraday Discuss.*, 169:265–283, 2014.

[22] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[23] Michael Krone, John E. Stone, Thomas Ertl, and Klaus Schulten. Fast visualization of Gaussian density surfaces for molecular dynamics and particle system trajectories. In *EuroVis - Short Papers 2012*, pages 67–71, 2012.

[24] James C. Phillips, John E. Stone, Kirby L. Vandivort, Timothy G. Armstrong, Justin M. Wozniak, Michael Wilde, and Klaus Schulten. Petascale Tcl with NAMD, VMD, and Swift/T. In *SC'14 workshop on High Performance Technical Computing in Dynamic Languages*, SC '14, pages 6–17. IEEE Press, 2014.

[25] John Stone and Mark Underwood. Rendering of numerical flow simulations using MPI. In *Second MPI Developer's Conference*, pages 138–141. IEEE Computer Society Technical Committee on Distributed Processing, IEEE Computer Society Press, 1996.

[26] John E. Stone. *An Efficient Library for Parallel Ray Tracing and Animation*. Master's thesis, Computer Science Department, University of Missouri-Rolla, April 1998.

[27] Steven G. Parker, James Bigler, Andreas Dietrich, Heiko Friedrich, Jared Hoberock, David Luebke, David McAllister, Morgan McGuire, Keith Morley, Austin Robison, and Martin Stich. OptiX: a general purpose ray tracing engine. In *ACM SIGGRAPH 2010 papers*, SIGGRAPH '10, pages 66:1–66:13, New York, NY, USA, 2010. ACM.

[28] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with CUDA. *ACM Queue*, 6(2):40–53, 2008.

[29] Melih Sener, John E. Stone, Angela Barragan, Abhishek Singharoy, Ivan Teo, Kirby L. Vandivort, Barry Isralewitz, Bo Liu, Boon Chong Goh, James C. Phillips, Lena F. Kourkoutis, C. Neil Hunter, and Klaus Schulten. Visualization of energy conversion processes in a light harvesting organelle at atomic detail. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '14. IEEE Press, 2014.