Molecular Dynamics Simulations of Biomolecular Complexes

Jeff Wereszczynski
Illinois Institute of Technology

1 Summary of Proposed Research

We are requesting continued use of XSEDE resources to support four independent projects that each utilize conventional and enhanced sampling molecular dynamics (MD) simulations. Specifically, these project are:

1. The nucleosome core particle (NCP) consists of ∼147 base pairs of DNA wrapped around a histone octamer and is the fundamental unit of DNA packaging in eukaryotes. The substitution of variant proteins to the octamer has been linked to epigentic regulation *in vivo*, such as the increased presence of H2A.B in transcriptionally active regions of DNA. While several studies have highlighted potential mechanisms, a fully atomistic understanding of H2A.B-regulated NCP destabilization has not been described. Here, we propose a combination of conventional and free energy MD simulations to study the mechanisms of epigenetic activation by H2A.B substitution.

2. The chromatosome is an extension of the NCP which binds an additional ∼20 base pairs of DNA due to the presence of a linker histone. Experiments have shown that linker histones can bind to DNA in either an on- or off-dyad locations, the selection of which is dictated by modifications to the linker histone sequence. However, the molecular mechanisms that drive the binding mode, as well as how they are influenced by specific linker histone residues, remains poorly understood. Here, we will perform a series of conventional MD simulations of three different linker histones bound in on- and off-dyad locations to determine how specific residues influence the structure and energetics of the chromatosome.

3. Chromatin fibers consist of repeating NCP units and form the building blocks of chromosomes within the nucleus of eukaryotic cells. Therefore, the structure and dynamics of poly-NCP arrays directly influences several processes, including gene expression, DNA replication, and DNA repair. In this project, we aim to study the relationship between structure, function, and dynamics in poly-NCP arrays. More concretely, we will perform fully atomistic MD simulations of an octa-NCP construct in a compact state. Analysis will focus on the stability and conformational dynamics of the chromatin fiber, along with the importance of stacking interactions between adjacent nucleosomes and inter-NCP contacts formed by the histone tails. Furthermore, we will focus on identifying allosteric networks within poly-NCP arrays.

4. Pili are protein structures that protrude from bacterial cell walls and function to enhance their virulence. They are bio-synthesized by the C class sortase proteins, which was the subject of our previous work. Structurally, pili are polymers built by covalently linking various pilin subunits. However, there is little detail in the literature about the molecular structure and dynamics of these building blocks or how their properties are potentially modified due to polymerization. Here, we propose to implement free energy MD simulations on the dimeric form of *Streptococcus pneumoniae* and *Corynebacterium diphtheriae* pilins. Simulations will be carried out both in the presence and absence of the experimentally determined intra-domain isopeptide bonds that have been proposed to impart mechanical strength to pili.

2 Background and Specific Goals

2.1 Simulations of the H2A.B Variant Nucleosome (NIH 1R15GM114758)

Eukaryotic cells store their genetic material into chromosomes through a hierarchical packaging scheme whose fundamental unit is the nucleosome core particle (NCP) [1]. The NCP wraps ∼147 base pairs of DNA around a symmetric histone octamer formed by two (H2A-H2B) dimers flanking an $(H3\text{-}H4)_2$ tetramer [2]. *In vivo*, NCPs act as epigenetic governors of transcription and alter gene expression through a variety of chemical changes [3, 4]. For example, substitution of canonical histones with a structurally similar but sequentially deviant protein, a histone "variant", can comprehensively modulate gene expression. Our previous work was focused on a transcription repressor, the macroH2A variant, in both homogeneous (both dimers replaced) [5] and heterogeneous (one canonical dimer and one variant dimer) nucleosomes. In contrast, Barr body deficient H2A (H2A.B) is associated with transcriptionally active regions of DNA[6], but little is known about how the structure and dynamics of H2A.B containing NCPs affect their function.

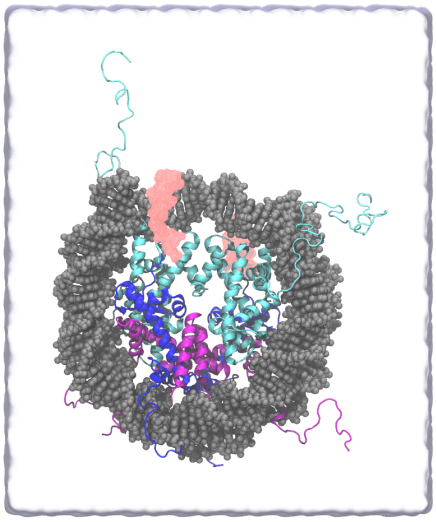

Figure 1: The initial conformation of the H2A.B NCP system. Two H2A.B(blue)-H2B(purple) dimers flank the $(H3\text{-}H4)_2$ tetramer (cyan) and wrap 147 base pairs of DNA (grey). Red sections indicate the truncated H2A C-terminus.

Like other H2A variants, the histone fold of H2A.B is believed to follow the structural motif of three alpha helices connected by short loops and flanked by flexible N- and C-terminal tails. However, the sequence of this fold is only ∼50% identical to the canonical protein, making H2A.B the most sequentially divergent of the major variants [3]. Indeed, these sequence variations alter important chemical properties of the NCP, such as reducing the total number of lysine residues and eliminating the acidic patch believed to be crucial in internucleosomal interactions. Strikingly, the C-terminal sequence of H2A.B is ∼20 amino acids shorter than the canonical protein, reducing the sequence responsible for forming dimer-tetramer interactions by ∼25%. Furthermore, these C-terminal residues are located near the DNA entry/exit site in canonical H2A (Figure 1). Thus, the lack of C-terminal sequence may destabilize the NCP by degrading both protein-protein and protein-DNA interactions and result in an open, more accessible structure [7–9].

In the simulations proposed here, we will comprehensively study the effects of H2A.B substitution on NCP dynamics and stability through a combination of conventional and free energy MD simulations. Results from conventional simulations will be compared against our previous work to determine the differences in dynamics between H2A.B and canonical nucleosomes in the hundreds of nanoseconds timescale, and solution ensembles will be compared against experimental scattering profiles [7, 8]. Furthermore, free energy calculations will be conducted for both the canonical and H2A.B NCPs to quantify the disparity in DNA binding capabilities. The success of these free energy calculations will serve to guide unwinding simulations on future NCP systems, such as other variant NCPs, the chromatosome, and post-translationally modified NCPs. These simulations, along with our previous work, will provide further understanding of the interaction between structure, dynamics, and function in variant-containing nucleosomes.

2.2 Simulations of the Chromatosome (NSF 1552743)

The chromatosome is an extension of the NCP that contains the same structural foundations with an additional ∼20 base pairs of DNA (∼167 total) accompanied by a highly-conserved linker histone (Figure 2). Colloquially known as histone H1, this nuclear protein plays a crucial role in the condensation of nucleosome chains into higher order structures [10], as well as other cellular functions [11] such as gene expression [12, 13], heterochromatin genetic activity [14], and cell differentiation [15, 16], among many others [17–19]. Linker histones have also been found to rapidly exchange locations throughout the chromosome [20, 21]. In avian species, H1 is replaced by the H5 isoform.

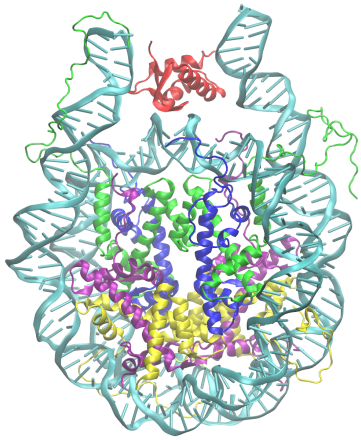

Figure 2: Structure of the Chromatosome (initial PDB ID 4QLC). The structure is color-coded as follows: Histone H2A (purple), Histone H2B (yellow), Histone H3 (green), Histone H4 (blue), and Histone H5 (red). Solvated system size is ∼270,000 atoms

Although H1 and H5 have highly similar primary and tertiary structures, recent experiments have shown that they exhibit different binding modes on the nucleosome. Histone H5 binds centered on the dyad axis in the "on-dyad" location, whereas histone H1 binds in a skewed mode relative to the dyad axis, or off-dyad. In a recent study, Zhou *et al* examined the binding modes of wild type globular histone H1 (GH1), H5 (GH5), and a GH5 penta-mutant that had five residues switched from their GH5 to GH1 sequences (GH5_pMu) [22]. Results showed that these small number of mutations were able to shift the binding state from the H5 on-dyad location to the H1 off-dyad location. This suggests that these five residues may be the key determinants for linker histone binding modes on chromatosomes. However, the molecular mechanisms that orchestrate this conformational shift, the role of specific linker histone residues, and how residue modifications affects this process is poorly understood. Moreover, the extent to which these phenomenon alter greater chromatin dynamics is also unclear.

Here, we will study the atomic scale mechanisms that dictate linker histone binding modes through a series of conventional MD simulations of different chromatosomes containing GH1, GH5, and GH5_pMu. We hypothesize that each linker histone is capable of binding in both the on- and off-dyad locations, and that differences in the histones shift the free energies of the states. Therefore, we will initiate simulations of each system from both the on- and off-dyad locations. Analysis will focus on the structure and energetics of specific contacts that stabilize these two binding modes, with an emphasis on how the mutated residues in GH5_pMu create the experimentally observed shift in binding.

2.3 Simulations of the Canonical Octanucleosome (NSF 1552743)

In vitro, nucleosomes form long "beads on a string" like structures in which many nucleosomes are periodically positioned on a long DNA molecule. In its most compact form, poly-NCP arrays condense into structures that are composed of tetra-NCP repeats with DNA that zigzags between the individual units [23]. These tetra-NCP structures likely form the basis of chromatin fibers and thus represent a second level of gene compaction [1, 24–26]. What the relationship is between the structure, dynamics, and function of these constructs, and how these properties are modulated by chromatin remodeling factors, are questions of deep scientific interest that have yet to be fully understood.

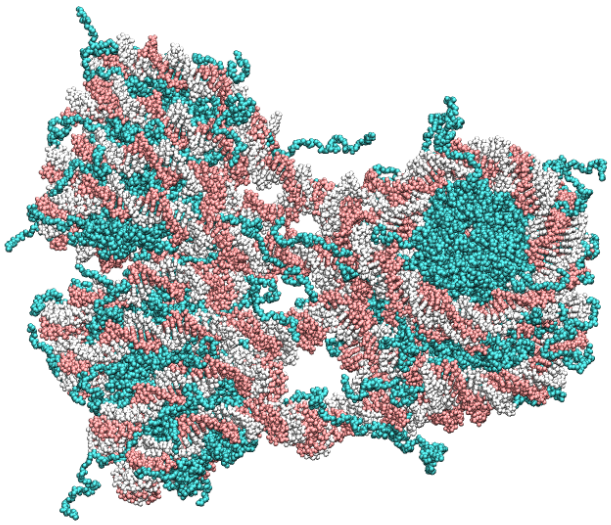

Figure 3: The octanucleosome system to be simulated here, based on EMDataBank ID emd_2600.ccp4 and PDB ID 1KX5. When fully solvated the system size is $\sim$ 2,800,000 atoms

In previous work on mono-NCPs, we showed that chromatin remodeling factors influence not only the NCP structure, but also the dynamic networks that extend throughout the complex [5]. We hypothesize that the regulation of allosteric networks is a general feature of chromatin remodeling factors, and that histone variants and PTMs modulate both intra- and inter-NCP dynamics as a method of influencing chromatin stability far from their physical location. If true, these networks would likely propagate through the inter-nucleosomal DNA and histone tails, and they would link locations that affect nucleosome stability, such as the DNA entry and exit points, as well as the locations most heavily influenced by remodeling factors, such as the L1 loops and the acidic patch.

In this project, we will begin to test the hypothesis that allosteric networks link regions *between* nucleosomes that are known to be important for chromatin fiber regulation. To do this, we will perform a set of conventional MD simulations of an octa-NCP system (Figure 3). Analysis will focus on determining the strength and linchpins of inter-NCP and inter-tetraNCP dynamic networks with a combination of methods that we previously applied to mono-NCP systems [5]. In future years, we will build on these results by simulating poly-NCP constructs that include histone variants and PTMs, as well as systems in other, non-compact, conformations. We also aim to use our full atomistic simulations to validate existing Coarse Grained models that have been used to simulate mono-NCPs [27] and their reliability to be used in simulating larger poly-NCP to model realistic chromatin fibers.

2.4 Simulations of Pilin Subunits (NIH 1R35GM119647)

Pili, which are elongated protein structures that reside on the cell surface of bacteria, have been shown to increase bacterial virulence. They are formed by sortase C (SrtC) proteins through the polymerizing of pilin sub-units [28]. The main pilin monomer of *Streptococcus pneumoniae* (RrgB) is made up of four domains as shown in Figure 4. The second and third domains are arranged side to side, while the first and last units lie on top and below these middle units. In the case of *Corynebacterium diphtheriae*, its main building block (SpaA) is a smaller three-domain protein with domains that are placed on top of each other. Interestingly, each of the domains in both proteins contain isopeptide bonds that have been proposed to provide additional mechanical stability.

Previously, we examined the lid opening process of SrtC, and our results predicted that it likely requires the action of an external factor to function due to the high calculated free energies required for the lid opening process; a result that was later confirmed by NMR dynamics experiments carried out by our collaborators [29]. We further went on to examine the energetics of pilin sorting signals binding to SrtC (Progress Report, section 1.2). For the present project, we aim to understand the structural dynamics of the *Streptococcus pneumoniae* and *Corynebacterium diphtheriae* pilin monomers, explain how dimerization affects overall dynamics, and elucidate the molecular reasons for the intra-domain isopeptide bonds observed in these pilin molecules.

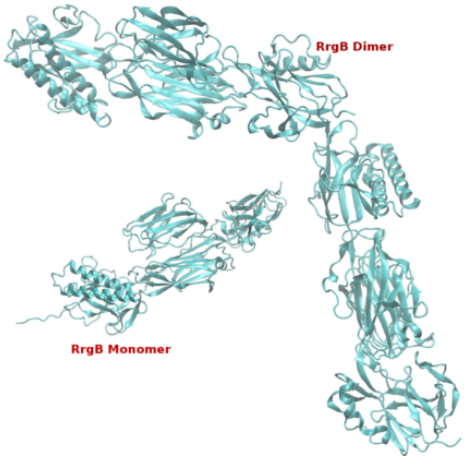

Figure 4: Monomer and dimer of the *Streptococcus pneumoniae* RrgB pilin

3 Proposed Simulations and System Details

Here, we describe the specific details of the simulations we will perform in each project. These simulations will use either NAMD, GROMACS, or AMBER [30–32], and GPU acceleration will be used whenever feasible [33]. All simulations with explicit solvent environments will be performed with periodic boundary conditions, with non-bonded interactions truncated at 10 Å, and long-range electrostatics treated with the particle mesh Ewald (PME) method [34]. Hydrogen containing bonds will be constrained using SHAKE [35] in AMBER and NAMD, and P-LINCS will restrain similar bonds in GROMACS simulations [36]. All simulations are to be conducted with a 2 fs timestep unless otherwise specified.

3.1 Simulations of the H2A.B Nucleosome

To better understand the structure-dynamics-function relationship in H2A.B containing NCPs, we are requesting XSEDE resources for conducting two complementary arrays of simulations. In the first, we will conduct conventional MD (cMD) simulations on the H2A.B NCP. In the second, we will calculate the free energy of unwinding of canonical and H2A.B NCPs via umbrella sampling (US) calculations. In both sets of simulations, the protein and nucleic acid forcefield parameters will come from FF14SB, the most recent AMBER fixed point-charge forcefield [37, 38], and ions will be paramaterized according to Joung and Cheatham [39].

Table 1: A summary of the total requested resources for simulations relating to the H2A.B NCP. Conventional simulations have a benchmark speed of **300 SUs/ns on Comet** with NAMD, whereas umbrella sampling calculations were benchmarked at **5 SUs/ns on XStream** with AMBER.

NCP Type	Simulation Method	System Size (atoms)	Simulation Length	Number of Simulations	Comet SUs	XStream SUs
H2A.B	cMD	241,730	250 ns	3	225,000	———
H2A.B	US	24,509	250 ns	36	———	45,000
Canonical	US	25,172	250 ns	36	———	45,000
				Total:	**225,000**	**90,000**

3.1.1 Conventional Molecular Dynamics

Since no crystal structure of the H2A.B NCP exists, an initial conformation was constructed from the 1.9 Å resolution X-ray structure of the canonical NCP (PDB: 1KX5) [40]. The canonical H2A sequence was truncated and mutated to the sequence of H2A.B in VMD and then neutralized by the addition of Cl^- ions. The neutralized H2A.B NCP was then solvated in a box of TIP3P water[41] at a salt concentration of ∼150 mM NaCl, and the box was extended 10 Å from the solute, yielding a system size of 241,730 atoms.

Simulations will be performed in the NPT ensemble at 300 K and 1 atm using the NAMD engine and a 2 fs timestep. Long range correlations, DNA-protein interactions and binding energies, and dimer-tetramer interactions will be analyzed from the resulting trajectories and will be compared against our previous results for the canonical NCP[5]. Furthermore, the H2A.B solution of states will be compared to experimental scattering data [7, 8] using our recently developed in-house protocol (see Progress Report, Section 1.4).

To determine the amount of necessary simulation time, we direct to our previous work on canonical and macroH2A variant NCPs in which our simulations required 50-60 ns in order for the extended histone tails to equilibrate and contract [5]. Therefore, in order to obtain ∼200 ns of production data, a total of 250 ns per simulation must be conducted. Three separate simulations will be employed to ensure meaningful statistics in observed dynamic behavior. Thus, a total of 750 ns of cMD will be conducted for the H2A.B NCP. To conduct these simulations, we request **225,000 SUs** on Comet, in accordance with our benchmarking data. Additionally, we are requesting **0.42 TB** of storage space on Data Oasis for this portion of the project.

3.1.2 Umbrella Sampling of DNA Unwinding

To determine the simulation protocol for our umbrella sampling calculations, we reference a recent coarse-grained simulation of DNA unwinding from the standard NCP [42]. The reported free energy profile suggests that the unwrapping of DNA occurs in an asymmetric pathway in which only one end of DNA unwinds, in agreement with other studies [43, 44]. To generate their PMF, Zhang *et al.* coupled harmonic restraints to the radius of gyration (R_g) of all DNA base pairs over 12 windows spanning a range of 45-72.5 Å, with centers evenly spaced every 2.5 Å. However, sufficient sampling for a similar protocol over such a wide range of R_g values at the atomistic level is currently unfeasible. Instead, we will utilize the asymmetric unwinding behavior to conduct two separate umbrella sampling calculations per NCP system, one for each DNA end. The reaction coordinate will be defined by the R_g of 73 base pairs extending from the dyad axis in either direction, and harmonic biases will be introduced to 18 windows spread over a range of R_g = 42.5-59.5 Å. Note that we are increasing the number of windows relative to the study done by Zhang *et al.*, as we expect that our all-atom model will have smaller fluctuations, and thus require a higher density of windows, then their course grain model. Probability densities from all windows will be converted into a 1-dimensional potential of mean force (PMF) using WHAM [45].

The accuracy of any free energy calculation relies heavily on the convergence of each simulation window. Therefore, we propose that each window will be run for ∼250 ns, in accordance to our cMD protocol. To increase computational performance, we will employ a hydrogen mass repartitioning protocol to increase our timestep two-fold ($\Delta t = 4$ fs) [46], and we will treat the solvent environment implicitly, which has been shown to provide good interpretations of nucleosome dynamics [47]. In

doing so, we will reduce each system to ∼25,000 atoms to be simulated at 300 K in the NVT ensemble with all pair-wise interactions calculated directly.

We will utilize the efficient GPU performance of the recently developed pmemd.CUDA (v16) for simulating each window. From our benchmarks on XStream, simulating the proposed windows will require 45,000 SUs per NCP type, so we request a total of **90,000 XStream SUs** to conduct our umbrella sampling calculations. Furthermore, we request **0.52 TB** of storage space on Data Oasis for this portion of the project.

3.2 Simulations of the Chromatosome

To determine the structural and energetic factors that dictate the binding states of linker histones, conventional MD simulations of chromatosomes with GH1, GH5, and the GH5_pMu linker histones will be conducted. Each simulation will be initiated in both the on- and off-dyad conformations to determine the how seemingly minor structural differences shift the equilibrium of these states. Simulations of the on-dyad state will be initiated from the 4QLC crystal structure of the H5 chromatosome system, with missing loops and tails modelled using Modeller 9.17 [48]. Currently there is no atomic-resolution structure of a chromatosome containing an off-dyad linker histone. Therefore, these simulations will be based on the cryogenic electron microscopy (cryo-EM) micrograph of a chromatin fiber which contains histone H1 in the off-dyad binding mode (EMDataBank ID emd_2600.ccp4) [49]. Using this electron density map, we will perform molecular dynamics flexible fitting (MDFF) calculations to fit linker histones to the off-dyad position [50]. These simulations will be performed with local resources.

Simulation parameters will be similar to those for our H2A.B simulations, including the use of an NPT ensemble at 300 K and 1 atm with the NAMD engine and a 2 fs timestep. Protein and nucleic acid parameters will be taken from the AMBER ff14SB force field, and TIP3P waters will be used. As in our previously performed simulations of the canonical and macroH2A-containing NCP, as well as the proposed H2A.B simulations, each simulation will be performed for 250 ns three times. Analysis will focus on the complex stability, as well as the binding energetics and dynamics of the different linker histones.

Using current resources allocated by XSEDE, we have already begun — and expect to finish — three simulations of the chromatosome containing GH5 in the on-dyad binding mode. Therefore, we are requesting resources for a total of 15 separate 250 ns simulations to complete three trajectories of the remaining five systems. In accordance to our benchmarking data, we are requesting **1,312,500 Comet hours** (see Table 2). Additionally, we are requesting **2.40 TB** of storage to store our trajectory files.

Table 2: A summary of the total requested resources for simulations relating to the chromatosome. A benchmark speed of **350 SUs/ns** was measured on Comet.

Linker Histone (Conformation)	Simulation Method	System Size (atoms)	Simulation Length (ns)	Number of Simulations	Comet SUs
GH1 (On-Dyad)	cMD	270,000	250	3	262,500
GH1 (Off-Dyad)	cMD	270,000	250	3	262,500
GH5 (Off-Dyad)	cMD	270,000	250	3	262,500
GH5_pMu (On-Dyad)	cMD	270,000	250	3	262,500
GH5_pMu (Off-Dyad)	cMD	270,000	250	3	262,500
				Total:	**1,312,500**

Table 3: Simulations we are requesting XSEDE resources for in the octanucleosome project. A conversion rate of **2997 SUs/ns** was measured on Stampede (see Benchmark data).

System	System Size (Atoms)	Simulation Time	Number of Simulations	Stampede SUs
OctaNCP	2,800,000	500 ns	2	3,000,000
OctaNCP	2,800,000	250 ns	1	750,000
			Total:	**3,750,000**

3.3 Simulations of the Canonical Octanucleosome

To begin our study of inter-NCP allosteric networks, we are performing cMD simulations of a compact octanucleosome comprised of canonical histones. The original structure for these simulations is based on the 11 Å resolution cryo-EM micrograph of a 30 nm chromatin fiber reconstituted on 12 x 177 base pairs (EMDataBank ID emd_2600.ccp4) [49]. Individual NCPs are based on the 1.9 Å crystal structure by Davey *et al.* [40] (PDB ID: 1KX5). Each tetra-NCP is based on the crystal structure by Schalch *et al.*[23] (PDB ID: 1ZBB). We use the AMBER14SB force field to describe the proteins and the DNA. The system is solvated with TIP3P water molecules, and counterions have been added to neutralize the system and simulate a 150 mM NaCl environment [37, 51]. The resulting system has ∼2,800,000 atoms. We use Gromacs 5.0.6 to run these simulations. Pressure is controlled using the Parrinello-Rahman barostat [52, 53] while temperature is controlled using the Nosé-Hoover thermostat [54, 55].

To map allosteric effects in the NCP, we will utilize multiple techniques. Residue-residue correlations will be calculated by utilizing mutual information methods [56, 57], while allosteric networks will be mapped with the Weighted Implementation of Sub-optimal Pathways approach [58]. In addition, the importance of individual residues to dynamic networks throughout the structure will be computed with edge-betweenness centrality calculations [59]. Finally, we will determine how packing into chromatin fibers alters the dynamics of a single NCP by computing the Kullback-Leibler divergence of dihedral angle populations between these simulations and our previous monoNCP simulations [60].

To estimate the required simulation times for convergence of these calculations, we refer to our previous NCP study where we performed three independent 250 ns simulations for each construct. Given the system size of the octaNCP, dynamics is expected to be slowed by packing. We estimate that simulations twice as long, or 500 ns in length, will be required to achieve an acceptable level of sampling. We have already performed 250 ns of cMD of a canonical octanucleosome. Preliminary results show that simulation times of ∼150 ns are required to reach equilibrium. In addition, it is important that we perform these simulations multiple times to delineate which dynamical properties are robust. Therefore, we intend to perform three 500 ns simulations. Based on the timing in the benchmark section, we therefore request **3,750,000 Stampede hours** (see Table 3). In addition, we are requesting **3.96 TB** of storage to save our trajectory files (see scaling document).

3.4 Umbrella Sampling Molecular Dynamics Simulations of Pilin Subunits

Using in house resources, we implemented three conventional and three accelerated MD simulations of the monomer and dimer forms of RrgB and SpaA. The results suggest that their structures are more dynamic in the absence of the intra-domain isopeptide bonds. In some of the calculations, a kink appears in the dimer with the angle between monomers approaching approximately 110°. Such structural changes were not seen in the simulations of dimers with an isopeptide bond, which could imply that isopeptide bonds are vital to pili structural integrity. However, it could also be

Table 4: Simulations we are requesting XSEDE resources for umbrella sampling of pilin proteins. A conversion rate of **60 SUs/ns** was used for the simulations (see Benchmark data).

System	Number of Simulations	Length	Comet SUs
RrgB Dimer	100 ns × 31 windows	3.1 µs	186,000
RrgB Dimer with Iso	100 ns × 31 windows	3.1 µs	186,000
SpaA Dimer	100 ns × 31 windows	3.1 µs	186,000
SpaA Dimer with Iso	100 ns × 31 windows	3.1 µs	186,000
		Total:	**744,000**

that the calculations implemented for structures with an isopeptide bond are yet to find this kinked conformation. Clarifying this requires the use of free energy calculations. Therefore, we propose to implement umbrella sampling calculations to further extend the aMD and cMD simulations we have already performed and provide detailed molecular level information about the role of isopeptide bonds in pilin molecules.

The octahedral box containing a solvated RrgB dimer is made up of 987,650 particles. The SpaA system has a slightly smaller total number of particles due to its smaller size. To implement the umbrella sampling calculations, a dihedral collective variable will be used. The vertices of this dihedral will be made from the center of mass of groups of atoms in the first and last sub-units of RrgB and SpaA. These calculations will therefore make use of the non-GPU accelerated version of AMBER, since such groupings are yet to be implemented in the GPU accelerated module. The dihedral angle will be varied from 90° to 180° in 3° increments, leading to thirty windows per system. Each window will be run for 100 ns, which means a total time of 3.1 µs per system, and windows will be combined to create a PMF using WHAM [45]. A 100 ns simulation time was chosen since we observed equilibration times of between 20 to 40 ns in the aforementioned cMD calculations, which will allow 60 to 80 ns of sampling time for collecting statistics. Calculations will be carried out three times for RrgB and SpaA dimers in the presence and absence of isopeptide bonds, yielding a total of 12.0 µs simulation time for this project. Therefore, in accordance with our benchmarking data, we request **744,000 Comet SUs** for our free energy calculations of pilin structures. We are also requesting **14.9 TB** of storage for these calculations.

4 Research Team and Supporting Grants

1. Samuel Bowerman (graduate student): Sam will perform the free energy simulations relating to the canonical and H2A.B variant nucleosomes. He is supported by the PI's R15 award (NIH 1R15GM114758).

2. Havva Kohestani (graduate student): Havva will perform the conventional MD simulations of the H2A.B NCP. She is supported by the PI's R15 award (NIH 1R15GM114758).

3. Emmanuel Naziga (postdoc): Emmanuel will conduct the simulations of pilin subunits. He is supported by the PI's MIRA award (NIH 1R35GM119647).

4. Francisco Rodríguez Ropero (postdoc): Francisco will conduct the simulations of the octanucleosome. He is supported by the PI's CAREER award (NSF 1552743).

5. Dustin Woods (graduate student): Dustin will perform the simulations of the chromatosome. He is supported by the PI's CAREER award (NSF 1552743).

5 Summary

There will be six individuals working on the proposed research program: one PI, two postdocs, and three graduate students. Our total request of **4,500,000 SUs on Stampede, 2,281,500 SUs on Comet, 90,000 SUs on XStream, 18.24 TB of storage on Data Oasis, and 3.96 TB of storage on Ranch** is summarized in Table 5.

Project	Stampede (SUs)	Comet (SUs)	XStream (SUs)	Data Oasis (TB)	Ranch (TB)
H2A.B NCPs	—	225,000	90,000	0.94	—
Chromatosome	—	1,312,500	—	2.40	—
OctaNCP	3,750,000	—	—	—	3.96
Pilin subunits	—	744,000	—	14.90	—
Total	3,750,000	2,281,500	90,000	18.24	3.96

Table 5: Total XSEDE Allocation Request. We are requesting hours on three XSEDE resources: Stampede, Comet, and XStream.

6 Additional Considerations

HPC Facilities in the Wereszczynski Lab Our lab has exclusive access to our own computing cluster that has 12 compute nodes, each of which is comprised of 16 CPU cores running at 2.6 GHz, 64 GB of RAM, and four NVIDIA 1080 GTX GPU cards. The cluster also contains a high-speed infiniband interconnect, a head node, and ∼52 TB of storage.

Other HPC Allocations Our lab was recently awarded an allocation on Anton 2 to study the diffusion properties of modified lipopolysaccharide bilayers. There is no overlap between the projects discussed in this proposal and the ones to be performed on Anton 2.

Molecular Dynamics Simulations of Biomolecular Complexes Progress Report

Jeff Wereszczynski
Illinois Institute of Technology

1.1 Simulations of PKCα Bilayer Binding (Status: Published [61])

Protein Kinase Cα (PKCα) is a member of a large family of serine/threonine kinases that is involved in regulating a wide array of cellular functions such as proliferation, differentiation, and motility. It is activated by binding of its C2 and C1 domains to the inner leaflet of the cell through a process that involves the recognition of diacylglycerol and phosphatidylserine molecules. Here, we examined the first step of this activation process, the docking of the C2 domain to the bilayer, through a series of conventional and free energy MD simulations of the C2 domain interacting with three model bilayers: pure POPC, POPC/POPS/PIP2, and POPC/POPS/PIP2/POG. Results showed that POPS and PIP2 are critical for C2 domain docking, which occurs in both perpendicular and parallel orientations. Our results shed new insights into the process of C2 bilayer binding and suggest new mechanisms for the roles of different phospholipids in the PKCα activation process.

1.2 Simulations of SortaseC Enzymes (Status: Published [29] and In Preparation)

Previously, we detailed progress from equilibrium dynamics and lid opening of the sortase C-1 (SrtC-1) enzyme. We found that lid opening (to access the enzyme active site) does not occur freely in wild type SrtC-1 but is observed in several mutant forms including one (GSlid) that had the entire lid-containing region (residues 55-69) replaced with a 15 residue glycine-serine linker and represented a model with a completely open lid. These findings were confirmed using NMR data from our collaborators. We further investigated the free energy barrier that must be overcome to open the lid via umbrella sampling calculations. These calculations demonstrated that energies $\sim$ 20 kcal/mol was required to open the lid in the wild type, while significantly lower values ($\sim$ 1.5-12 kcal/mol) were necessary for the mutants. As detailed in our previous proposal, we continued this project by examining the energetics of the SrtC proteins in complex with the substrate sorting signals. Results from thermodynamic integration calculations (Table 1) show that SrtC-1 and SrtC-3 prefer the RrgB (IPQTG) sorting signal compared to RrgA (YPRTG) and RrgC (VPDTG), although in some cases the energy differences are rather small. This could imply that, as has been proposed in the literature, these enzymes could act on different substrates depending on cellular conditions. SrtC-2 has a strong preference for RrgA compared to RrgB while SrtC-3 prefers RrgB to the two other pilins.

Table 1: Relative free energies (kcal/mol) for transforming the RrgB sorting signal to the RrgA and RrgC sorting signals while in complex with srtC proteins as obtained from thermodynamic integration calculations.

System	Transformation	Free Energy
SrtC-1	IPQTG $\mapsto$ VPDTG	0.67
	IPQTG $\mapsto$ YPRTG	1.60
SrtC-2	IPQTG $\mapsto$ VPDTG	0.52
	IPQTG $\mapsto$ YPRTG	-3.90
SrtC-3	IPQTG $\mapsto$ VPDTG	1.99
	IPQTG $\mapsto$ YPRTG	2.88

1.3 Simulations of Antimicrobial Peptides (Status: Submitted)

Using our previous allocation, we conducted MD simulations to study the role of arginine in antimicrobial peptide/bilayer association, the results of which have recently been submitted for publication. Antimicrobial peptides (AMPs) are key components of the innate immune response in many organisms, where they display a broad spectrum of antimicrobial activity against bacteria, viruses, and parasites [62–64]. Most AMPs are short peptides containing a high proportion of cationic residues [62, 65, 66] that interact more strongly with the anionic membranes typical of prokaryotic organisms than the net neutral eukaryotic membranes. AMPs containing arginine are more active than those containing lysine, and switching arginine for lysine decreases the antimicrobial activity, membrane disruptive ability, and cytotoxicity for many AMPs [67–72].

To examine the source of this disparity, we conducted comparative simulations of a model prokaryotic membrane system (POPC/POPG) interacting with two different mutants of a prototypical AMP, KR-12: one with all lysines mutated to arginines (R-KR12) and one with all arginines mutated to lysines (K-KR12). Both peptides approach the bilayer in an analogous fashion and display similar preferences for hydrogen bonding with the anionic POPGs. However, R-KR12 binds tighter to the bilayer than K-KR12 and forms significantly more hydrogen bonds with the bilayer, leading to considerably longer interaction times. Further simulations with methylated R-KR12 mutants reveal that it is the strength of these peptide-bilayer hydrogen bonds, rather than the number of bonds formed, that is responsible for the extensive interaction seen in the R-KR12 system.

Additionally, replica exchange umbrella sampling simulations were performed to compare the native state of both peptides in solution and when inserted into the bilayer, sampling along two dimension: alpha, a measure of how helical the peptide is, and the peptide's end-to-end distance. The resultant potentials of mean force (PMFs) reveal that both peptides are disordered in solution but form an amphipathic α-helix when inserted into the bilayer (Figure 1). Overall, these results highlight the role of charge and hydrogen bond strength in peptide bilayer insertion, and offer potential insights for designing more potent therapeutic analogues in the future.

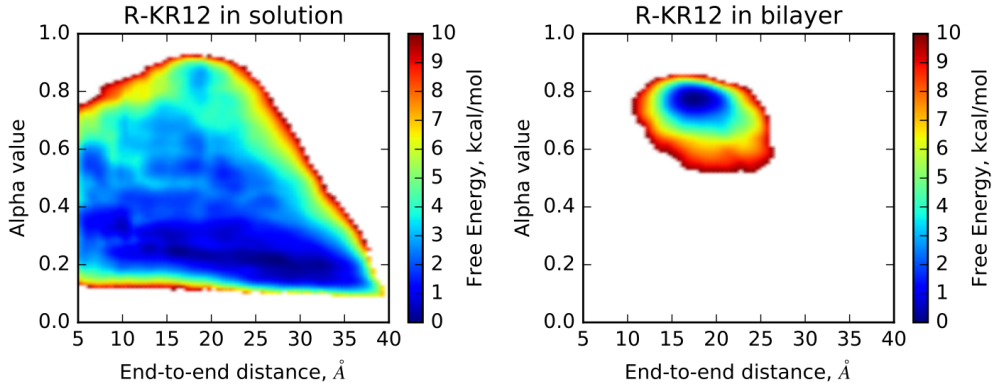

Figure 1: Two-dimensional PMFs for R-KR12 free in solution (left) and inserted in the POPC/POPG bilayer (right). Similar energy landscapes were observed for K-KR12. While both peptides are unfolded in solution, insertion into the bilayer promoted folding into an alpha helical state.

1.4 Bayesian SAXS Refinement of aMD Trajectories (Status: In preparation)

Resources from our previous allocation were used to develop an iterative Bayesian method for fitting accelerated MD (aMD) trajectories to empirical small angle X-ray scattering (SAXS) data. The fitting routine will be implemented as a module in SASSIE, a publicly available SAXS-analysis web

server [73], and a manuscript outlining it will be submitted to the *JCTC* in January 2017. We will apply this technique to a variety of systems in the future, including H2A.B, and we summarize our results for the tri-ubiquitin test case below.

Ubiquitin is a prominent cellular signaling protein for a variety of biological functions, and different ubiquitin oligomers can be formed by a covalent isopeptide linkage between the C-terminus of one monomer to one of eight possible sites on another monomer. In our study, we produced cMD and aMD trajectories for five linkages (K6, K11, K29, K48, and K63). Our fitting protocol identified that different solution ensembles were defined by different linkage types, in agreement with previous results [74–77]. Furthermore, we observed that aMD trajectories typically converged to higher model quality faster than their cMD counterparts (Figure 2). The improved performance of aMD trajectories was related to aMD boosting providing not only larger maximum variations in structural RMSD but also reducing the correlation time by lowering the energy barriers between states.

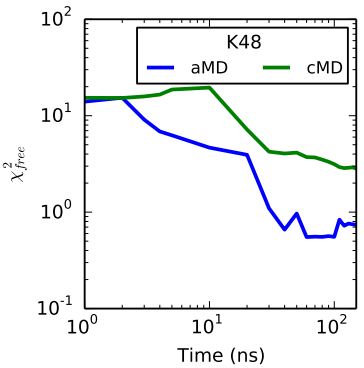

Figure 2: Model goodness-of-fit (χ^2_{free}) vs simulated time for the K48-linked ubiquitin trimer for both aMD (blue) and cMD (green) trajectories.

1.5 Simulations of the canonical octanucleosome (Status: Ongoing)

Using resources from our previous allocation, we have recently obtained 250 ns of one production run for the octanucleosome system. Root Mean Square Deviations (RMSD) were calculated for each nucleosome core particle (NCP) as a function of simulation time (Figure 3). In all eight NCPs, we observe that the RMSD values converge after ∼150 ns. Therefore, we conclude that 150 ns of simulation is necessary for equilibration in current and future octanucleosome trajectories. To achieve a sufficient sampling to map allosteric effects in the octanucleosome system, we estimate that 500 ns per simulation will be needed.

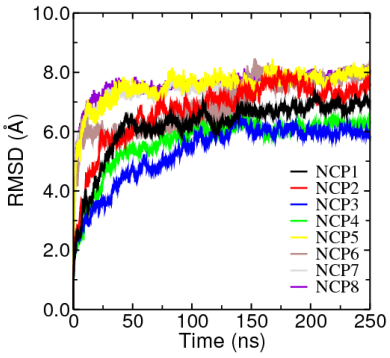

Figure 3: Root mean square deviation for each NCP in the octanucleosome system. Only alpha carbons were considered.

1.6 Simulations of the Canonical-macroH2A Hybrid NCP (Status: Ongoing)

Another portion of our previous XSEDE allocation was dedicated to simulating a hybrid NCP in which one dimer possessed the canonical H2A histone and the other contained the macroH2A variant. Three separate simulations of 250 ns were conducted, and the resulting trajectories were analyzed analogously to our published study of homogeneous macroH2A-like NCPs [5, 78]. We previously observed that the lysine sidechain in the macroH2A L1 loop consistently forms a direct interaction with a neighboring DNA phosphate, but we find this interaction in less than 2% of hybrid NCP frames. Instead, the lysine preferentially interacts with the canonical histone, most notably through hydrogen bonding with either the sidechain of N38 in the L1 loop (24.4% of frames) or the backbone of R35 in the α1 helix (17.6% of frames). Interestingly, the MMGBSA-calculated DNA binding energy of the hybrid NCP is significantly more favorable than the canonical molecule ($\Delta\Delta G$ = -34.1 ± 7.6 kcal/mol). Since -0.7 kcal/mol can be attributed directly to H2A sequence deviations, the majority of favorability may be associated with differences in local dynamics created through long range interactions with the macroH2A sequence. We are presently working to identify and analyze potential pathways.

References

[1] K. Luger, M. L. Dechassa, and D. J. Tremethick. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews Molecular Cell Biology*, 13(7):436–447, 2012.

[2] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. Crystal structure of the nucleosome core particle at 2.8 å resolution. *Nature*, 389(6648):251–260, 1997.

[3] Clemens Bönisch and Sandra B Hake. Histone h2a variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Research*, 40(21):10719–10741, 2012.

[4] Gregory D Bowman and Michael G Poirier. Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews*, 115(6):2274–2295, 2014.

[5] S. Bowerman and J. Wereszczynski. Effects of macroH2A and H2A. Z on nucleosome structure and dynamics as elucidated by molecular dynamics simulations. *Biophysical Journal*, 110(2):327–337, 2016.

[6] Michael Y Tolstorukov, Joseph A Goldman, Cristele Gilbert, Vasily Ogryzko, Robert E Kingston, and Peter J Park. Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Molecular Cell*, 47(4):596–607, 2012.

[7] Masaaki Sugiyama, Yasuhiro Arimura, Kazuyoshi Shirayama, Risa Fujita, Yojiro Oba, Nobuhiro Sato, Rintaro Inoue, Takashi Oda, Mamoru Sato, Richard K Heenan, et al. Distinct features of the histone core structure in nucleosomes containing the histone H2A.B variant. *Biophysical Journal*, 106(10):2206–2213, 2014.

[8] Yasuhiro Arimura, Hiroshi Kimura, Takashi Oda, Koichi Sato, Akihisa Osakabe, Hiroaki Tachiwana, Yuko Sato, Yasuha Kinugasa, Tsuyoshi Ikura, Masaaki Sugiyama, et al. Structural basis of a nucleosome containing histone H2A.B/H2A.Bbd that transiently associates with reorganized chromatin. *Scientific Reports*, 3:3510, 2013.

[9] Yunhe Bao, Kasey Konesky, Young-Jun Park, Simona Rosu, Pamela N Dyer, Danny Rangasamy, David J Tremethick, Paul J Laybourn, and Karolin Luger. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *The EMBO Journal*, 23(16):3314–3324, 2004.

[10] Thomas J Maresca, Benjamin S Freedman, and Rebecca Heald. Histone h1 is essential for mitotic chromosome architecture and segregation in xenopus laevis egg extracts. *Journal of Cell Biology*, 169(6):859–869, 2005.

[11] Sonja P Hergeth and Robert Schneider. The h1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Reports*, 16(11):1439–1453, 2015.

[12] Yuhong Fan, Tatiana Nikitina, Jie Zhao, Tomara J Fleury, Riddhi Bhattacharyya, Eric E Bouhassira, Arnold Stein, Christopher J Woodcock, and Skoultchi. Histone h1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, 123(7):1199–1212, 2005.

[13] Xuetong Shen and Martin A Gorovsky. Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell*, 86(3):475–483, 1996.

[14] Xingwu Lu, Sandeep N Wontakal, Harsh Kavi, Byung Ju Kim, Paloma M Guzzardo, Alexander V Emelyanov, Na Xu, Gregory J Hannon, Jiri Zavadil, Dmitry V Fyodorov, and Arthur I Skoultchi. Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science*, 340(6128):78–81, 2013.

[15] Hansol Lee, Raymond Habas, and Cory Abate-Shen. Msx1 cooperates with histone H1b for inhibition of transcription and myogenesis. *Science*, 304(5677):1675–1678, 2004.

[16] Yu Zhang, Dilaware Khan, Julia Delling, and Edda Tobiasch. Mechanisms underlying the osteo- and adipo-differentiation of human mesenchymal stem cells. *The Scientific World Journal*, 2012:1–14, 2012.

[17] Maria A Christophorou, Gonalo Castelo-Branco, Richard P Halley-Stott, and et. al. Citrullination regulates pluripotency and histone h1 binding to chromatin. *Nature*, 507(7490):104–108, 2014.

[18] Tina Thorslund, Anita Ripplinger, Saskia Hoffmann, and et. al. Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature*, 527(7578):389–393, 2014.

[19] Anita Konishi, Shigeomi Shimizu, Junko Hirota, and et. al. Involvement of histone H1.2 in apoptosis induced by DNA double-strand breaks. *Cell*, 114(6):673–688, 2003.

[20] Melody A Lever, John P H Th'ng, Xuejun Sun, and Michael J Hendzel. Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, 408(6814):873–876, 2000.

[21] Tom Misteli, Akash Gunjan, Robert Hock, Michael Bustin, and David T Brown. Dynamic binding of histone H1 to chromatin in living cells. *Nature*, 408(6814):877–881, 2000.

[22] Bing-Rui Zhou, Hanqiao Feng, Rodolfo Ghirlando, Shipeng Li, Charles D SChwieters, and Yawen Bai. A small number of residues can determine if linker histones are bound on or off dyad in the chromatosome. *Journal of Molecular Biology*, 428(20):3948–3959, 2016.

[23] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141, 2005.

[24] D. J. Tremethick. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*, 128(4):651–654, 2007.

[25] M. P. Scheffer, M. Eltsov, and A. S. Frangakis. Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 108(41):16992–16997, 2011.

[26] K. Maeshima, R. Imai, S. Tamura, and T. Nozaki. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123(3):225–237, 2014.

[27] B. Zhang, W. Zheng, G. A. Papoian, and P. G. Wolynes. Exploring the Free Energy Landscape of Nucleosomes. *Journal of the American Chemical Society*, 138:8126–8133, 2016.

[28] F. Neiers, C. Madhurantakam, S. Falker, C. Manzano, A. Dessen, S. Normark, B. Henriques-Normark, and A. Achour. Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. *J. Mol. Biol.*, 393(3):704–716, 2009.

[29] A. W. Jacobitz, E. B. Naziga, S. W. Yi, S. A. McConnell, R. Peterson, M. E. Jung, R. T. Clubb, and J. Wereszczynski. The "Lid" in the Streptococcus pneumoniae SrtC1 Sortase

Adopts a Rigid Structure that Regulates Substrate Access to the Active Site. *Journal of Physical Chemistry B*, 120(33):8302–8312, 2016.

[30] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781–1802, 2005.

[31] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[32] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26:1668–1688, 2005.

[33] Andreas W Gotz, Mark J Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C Walker. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *Journal of chemical theory and computation*, 8(5):1542–1555, 2012.

[34] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald - an N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98:10089–10092, 1993.

[35] V. Krautler, W. F. Van Gunsteren, and P. H. Hunenberger. A fast SHAKE: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, 22:501–508, 2001.

[36] Berk Hess. P-lincs: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4(1):116–122, 2008.

[37] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[38] Ivan Ivani, Pablo D Dans, Agnes Noy, Alberto Pérez, Ignacio Faustino, Adam Hospital, Jürgen Walther, Pau Andrio, Ramon Goñi, Alexandra Balaceanu, et al. Parmbsc1: a refined force field for DNA simulations. *Nature Methods*, 13(1):55–58, 2016.

[39] In Suk Joung and Thomas E Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30):9020–9041, 2008.

[40] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *Journal of Molecular Biology*, 319(5):1097–1113, 2002.

[41] Daniel J Price and Charles L Brooks III. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics*, 121(20):10096–10103, 2004.

[42] Bin Zhang, Weihua Zheng, Garegin A Papoian, and Peter G Wolynes. Exploring the free energy landscape of nucleosomes. *Journal of the American Chemical Society*, 138(26):8126–8133, 2016.

[43] Joshua Lequieu, Andrés Córdoba, David C Schwartz, and Juan J de Pablo. Tension-dependent free energies of nucleosome unwrapping. *ACS Central Science*, 2(9):660–666, 2016.

[44] Thuy TM Ngo, Qiucen Zhang, Ruobo Zhou, Jaya G Yodh, and Taekjip Ha. Asymmetric unwrapping of nucleosomes under tension directed by dna local flexibility. *Cell*, 160(6):1135–1144, 2015.

[45] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[46] Chad W Hopkins, Scott Le Grand, Ross C Walker, and Adrian E Roitberg. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of Chemical Theory and Computation*, 11(4):1864–1874, 2015.

[47] Ramu Anandakrishnan, Aleksander Drozdetski, Ross C Walker, and Alexey V Onufriev. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophysical Journal*, 108(5):1153–1164, 2015.

[48] Andrej Sali and Tom L Blundell. Comparative protein modelling by satisfaction of spacial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.

[49] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li. Cryo-em Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science*, 344(6182):376–380, 2014.

[50] L. G. Trabuco, E. Villa, E. Schreiner, C. B. Harrison, and K. Schulten. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods*, 49(2):174–180, 2009.

[51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79:926–935, 1983.

[52] M. Parrinello and A. Rahman. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics. *Journal of Applied Physics*, (52):7182–7190, 1981.

[53] S. Nosé and M. Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, (50):1055–1076, 1983.

[54] S. Nosé. A Molecular Dynamics Method for Simulations in the Canonical ensemble. *Molecular Physics*, (52):255–268, 1984.

[55] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, (31):16951697, 1985.

[56] G. Scarabelli and B. J. Grant. Kinesin-5 allosteric inhibitors uncouple the dynamics of nucleotide, microtubule, and neck-linker binding sites. *Biophysical Journal*, 107(9):2204–2213, 2014.

[57] O. F. Lange and H. Grubmuller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, 2006.

[58] A. T. Van Wart, J. Durrant, L. Votapka, and R. E. Amaro. Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *Journal of Chemical Theory and Computation*, 10(2):511–517, 2014.

[59] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[60] C. L. McClendon, L. Hua, A. Barreiro, and M. P. Jacobson. Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. *Journal of Chemical Theory and Computation*, 8(6):2115–2126, 2012.

[61] M. Alwarawrah and J. Wereszczynski. Investigation of the Effect of Bilayer Composition on PKC-C2 Domain Docking Using Molecular Dynamics Simulations. *Journal of Physical Chemistry B*, 2016.

[62] M. Zasloff. Antimicrobial peptides of multicellular organisms. *Nature*, 415(6870):389–395, 2002.

[63] D. M. Bowdish, D. J. Davidson, Y. E. Lau, K. Lee, M. G. Scott, and R. E. Hancock. Impact of LL-37 on anti-infective immunity. *Journal of Leukocyte Biology*, 77(4):451–459, 2005.

[64] R. E. Hancock and H. G. Sahl. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nature Biotechnology*, 24(12):1551–1557, 2006.

[65] Y. Shai. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochimica et Biophysica Acta - Biomembranes*, 1462(1-2):55–70, 1999.

[66] P. Wadhwani, R. F. Epand, N. Heidenreich, J. Burck, A. S. Ulrich, and R. M. Epand. Membrane-active peptides and the clustering of anionic lipids. *Biophysical Journal*, 103(2):265–274, 2012.

[67] R. A. Llenado, C. S. Weeks, M. J. Cocco, and A. J. Ouellette. Electropositive charge in alpha-defensin bactericidal activity: functional effects of Lys-for-Arg substitutions vary with the peptide primary structure. *Infection and Immunity*, 77(11):5035–5043, 2009.

[68] I. Nakase, S. Okumura, S. Katayama, H. Hirose, S. Pujals, H. Yamaguchi, S. Arakawa, S. Shimizu, and S. Futaki. Transformation of an antimicrobial peptide into a plasma membrane-permeable, mitochondria-targeted peptide via the substitution of lysine with arginine. *Chemical Communications (Cambridge, England)*, 48(90):11097–11099, 2012.

[69] S. T. Yang, S. Y. Shin, C. W. Lee, Y. C. Kim, K. S. Hahm, and J. I. Kim. Selective cytotoxicity following Arg-to-Lys substitution in tritrpticin adopting a unique amphipathic turn structure. *FEBS Letters*, 540(1-3):229–233, 2003.

[70] L. T. Nguyen, L. de Boer, S. A. Zaat, and H. J. Vogel. Investigating the cationic side chains of the antimicrobial peptide tritrpticin: hydrogen bonding properties govern its membrane-disruptive activities. *Biochimica et Biophysica Acta - Biomembranes*, 1808(9):2297–2303, 2011.

[71] A. Bonucci, E. Balducci, M. Martinelli, and R. Pogni. Human neutrophil peptide 1 variants bearing arginine modified cationic side chains: effects on membrane partitioning. *Biophysical Chemistry*, 190-191:32–40, 2014.

[72] B. Mishra, R. F. Epand, R. M. Epand, and G. Wang. Structural location determines functional roles of the basic amino acids of KR-12, the smallest antimicrobial peptide from human cathelicidin LL-37. *RSC Advances*, 3(42):19560–19571, 2013.

[73] Stephen J. Perkins, David W. Wright, Hailiang Zhang, Emre H. Brookes, Jianhan Chen, Thomas C. Irving, Susan Krueger, David J. Barlow, Karen J. Edler, David J. Scott, Nicholas J.

Terrill, Stephen M. King, Paul D. Butler, and Joseph E. Curtis. Atomistic modelling of scattering data in the Collaborative Computational Project for Small Angle Scattering (CCP-SAS). *Journal of Applied Crystallography*, 49(6):1861–1875, Dec 2016.

[74] Ranjani Varadan, Michael Assfalg, Aydin Haririnia, Shahri Raasi, Cecile Pickart, and David Fushman. Solution conformation of lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of Biological Chemistry*, 279(8):7055–7063, 2004.

[75] Anja Bremm, Stefan MV Freund, and David Komander. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase cezanne. *Nature Structural & Molecular Biology*, 17(8):939–947, 2010.

[76] Yu Ye, Georg Blaser, Mathew H Horrocks, Maria J Ruedas-Rama, Shehu Ibrahim, Alexander A Zhukov, Angel Orte, David Klenerman, Sophie E Jackson, and David Komander. Ubiquitin chain conformation regulates recognition and activity of interacting proteins. *Nature*, 492(7428):266–270, 2012.

[77] Takeshi Tenno, Kenichiro Fujiwara, Hidehito Tochio, Kazuhiro Iwai, E. Hayato Morita, Hidenori Hayashi, Shigeo Murata, Hidekazu Hiroaki, Mamoru Sato, Keiji Tanaka, and Masahiro Shirakawa. Structural basis for distinct roles of lys63- and lys48-linked polyubiquitin chains. *Genes to Cells*, 9(10):865–875, 2004.

[78] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation*, 8(9):3314–3321, 2012.

Molecular Dynamics Simulations of Biomolecular Complexes

Jeff Wereszczynski
Illinois Institute of Technology

1 Benchmarks and Scaling

Here, we present benchmarking data for each of the projects detailed in the main section of our proposal. In Table 1 we summarize the computing portion of this section, whereas we summarize our storage request in Table 2.

To calculate the "simulation efficiency", we have assumed that simulations scale nearly linearly up to one node (X=16 CPUs on Stampede, 24 on Comet), and we have taken the ratio of the number of SUs per ns of a simulation at N number of processors relative to the simulation on one node. That is:

$$\text{Efficiency} = \frac{SUs\,(X\ CPUs)}{SUs\,(N\ CPUs)} \tag{1}$$

For each system we have chosen the number of CPUs we will use based upon a balance of computational efficiency and simulation speed. In general, we will maintain an efficiency above 80%.

Table 1: Overview of benchmarking data for all the systems discussed in this proposal.

System	Method	MD Package	# CPUs	Efficiency	SUs/ns	Simulation Time (ns)	Stampede SUs	Comet SUs	XStream SUs
H2A.B	cMD	NAMD	480	80%	300	750	———	225,000	———
H2A.B	US	AMBER	1	100%	5	9,000	———	———	45,000
NCP	US	AMBER	1	100%	5	9,000	———	———	45,000
Chromatosome	cMD	NAMD	288	80%	350	3,750	———	1,312,500	———
OctaNCP	cMD	GROMACS	1024	80%	3,000	1,500	3,750,000	———	———
Pilin subunits	US	AMBER	72	82%	60	12,400	———	744,000	———
						Total:	**3,750,000**	**2,281,500**	**90,000**

Table 2: Overview of storage requirements for all the systems discussed in this proposal.

System	Method	Storage/ns (MB)	Simulation Time (ns)	Data Oasis Storage (TB)	Ranch Storage (TB)
H2A.B	cMD	560	750	0.42	—
H2A.B	US	28.9	9,000	0.26	—
NCP	US	28.9	9,000	0.26	—
Chromatosome	cMD	640	3,750	2.40	—
OctaNCP	cMD	2640	1,500	—	3.96
Pilin subunits	US	1229	12,400	14.90	—
			Total:	**18.24**	**3.96**

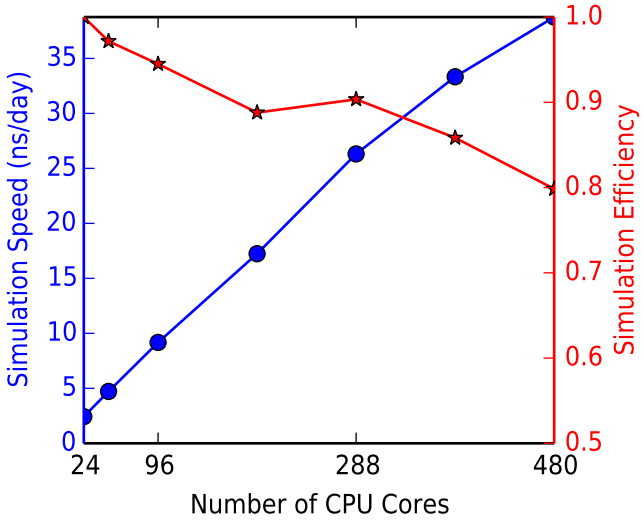

Figure 1: Scaling data for cMD simulations of the H2A.B NCP on Comet.

1.1 Simulations of H2A.B Nucleosomes

We request a total of **225,000 SUs on Comet** (3 simulations × 250 ns per simulation × 300 SUs per ns) and **90,000 SUs on XStream** (2 systems × 36 windows per system × 250 ns per window × 5.0 SUs per ns) for simulations relating to H2A.B variant NCPs. For the storage needs of this project, we request 940 GB on Oasis. The individual computational costs are broken down below.

1.1.1 Conventional Molecular Dynamics

Based on the benchmarking data presented in Figure 1, we found that the cMD simulations of the H2A.B NCP scales efficiently to 20 nodes. This results in a computational cost of 297.7 SUs per ns of simulation, which we round to 300 SUs per ns. In addition, saving every 5 ps results in a trajectory file size of 560 MB per ns of simulation. Therefore, we request 420 GB (560 MB per ns × 250 ns per simulation × 3 simulations) of storage on Oasis for the cMD simulations.

1.1.2 Umbrella Sampling Calculations

The current implementation of the free energy module in pmemd.CUDA can only utilize one GPU per simulation, and our benchmarking runs have shown that a single window will run at a speed of 5.0 ns per day. Therefore, each simulation window will cost 4.8 SUs per ns of simulation on XStream, which we round to 5.0 SUs per ns. Because the simulation timestep of each window is double that of our cMD simulations, storing trajectory information at the same frequency results in a 10 ps per frame resolution and a trajectory file size of 28.9 MB per ns of simulation. Therefore, we request a total of 520 GB (28.9 MB per ns × 250 ns per window × 36 windows per system × 2 systems) of storage space on Oasis for the proposed umbrella sampling calculations.

1.2 Simulations of the Chromatosome

Based on the benchmarks acquired from Comet (Figure 2), simulations of the chromatosome scale well to 288 CPUs at 80% efficiency and results in a simulation speed of 20 ns/day. Thus, each simulation will require 345.6 SUs per ns of simulation, which we round to 350 SUs per ns. Therefore,

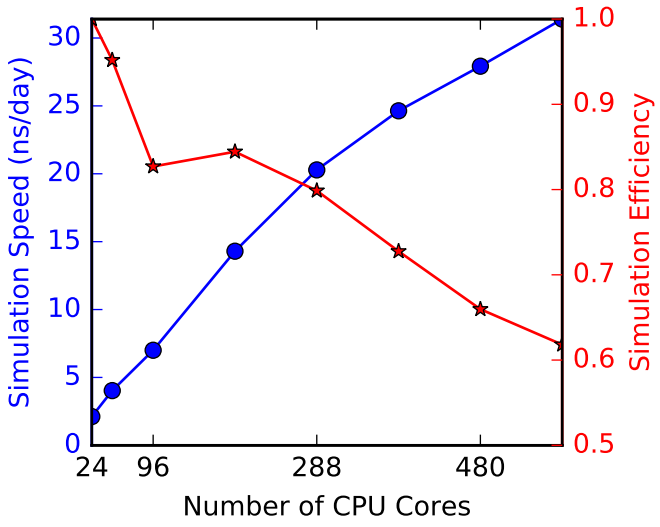

Figure 2: Scaling data for simulations of the chromatosome.

we are requesting **1,312,500 SUs on Comet** (5 systems × 3 simulations per system × 250 ns per simulation × 350 SUs per ns). Saving trajectories every 5 ps results in file sizes of 640 MB per ns of simulation; therefore, we are also requesting 2.40 TB of total storage space (640 MB per ns × 250 ns per simulation × 3 simulations per system × 5 systems).

1.3 Simulations of the Canonical Octanucleosome

Based on benchmarking data presented in Figure 3, we found that octaNCP simulations scale well to 1024 CPUs with a simulation efficiency of 80.0%. This results in a computational cost of 2997 SUs per ns of simulation, which we have rounded to 3000 SUs per ns for the remainder of this proposal. In addition, saving coordinates every 5 ps results in a trajectory file size of 2640 MB per ns of simulation.

In total, this project requests **3,750,000 SUs on Stampede** (1 system × 500 ns / simulation × 2 simulations / system × 3000 SUs / ns + 1 system × 250 ns / simulation × 1 simulations / system × 3000 SUs / ns). We are also requesting 3.96 TB of space on Ranch to backup our trajectory files (2640 MB/ns × 1500 ns).

1.3.1 Simulations of Pilin Subunits

Umbrella sampling (US) calculations will be performed with the CPU version of PMEMD. Benchmarking shows that these calculations only scale well to 72 CPUs on Comet (82.2% efficiency), resulting in a computational cost of 57.6 SUs per ns, which we round to 60 SUs per ns. We are therefore requesting **744,000 SUs on Comet** (60 SUs/ns × 3100 ns /system × 4 systems) for the pilin project. Saving snapshot every 10 ps requires 1229 MB per ns, giving a total storage requirement of approximately 14.9 TB.

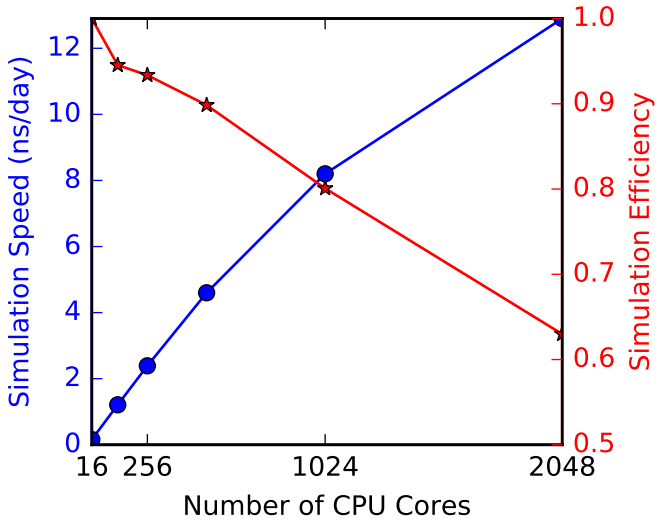

Figure 3: Scaling of Octanucleosome Simulations on Stampede

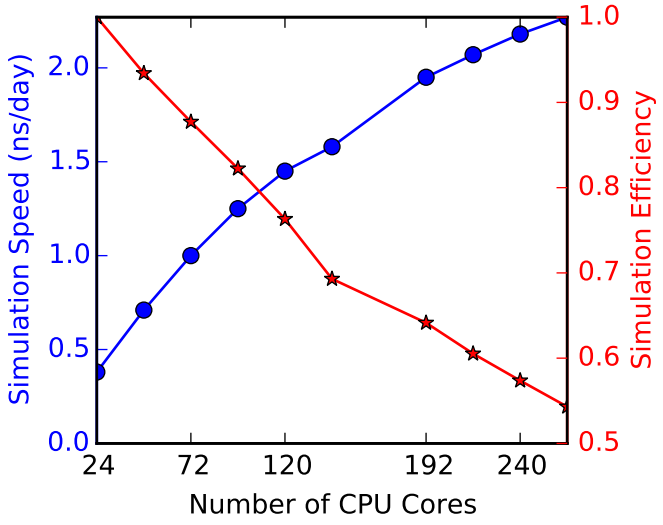

Figure 4: Benchmark for Pilin Simulations on Comet