

# **CHARMM-GUI Drude Prepper for Molecular Dynamics Simulation Using the Classical Drude Polarizable Force Field**

Abhishek A. Kognole<sup>1†</sup>, Jumin Lee<sup>2†</sup>, Sang-Jun Park<sup>2†</sup>, Sunhwan Jo<sup>3†</sup>, Payal Chatterjee<sup>1</sup>, Justin A. Lemkul<sup>4</sup>, Jing Huang<sup>5</sup>, Alexander D. MacKerell, Jr.<sup>1\*</sup>, and Wonpil Im<sup>2\*</sup>

<sup>1</sup>Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA

<sup>2</sup>Departments of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University, PA, USA

<sup>3</sup>Leadership Computing Facility, Argonne National Laboratory, 9700 Cass Ave, Argonne, IL, USA

<sup>4</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

<sup>5</sup>Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: alex@outerbanks.umaryland.edu and wonpil@lehigh.edu

## ABSTRACT

Explicit treatment of electronic polarizability in empirical force fields (FFs) represents an extension over a traditional additive or pairwise FF and provides a more realistic model of the variations in electronic structure in condensed phase, macromolecular simulations. To facilitate utilization of the polarizable FF based on the classical Drude oscillator model, *Drude Prepper* has been developed in CHARMM-GUI. *Drude Prepper* ingests additive CHARMM protein structures file (PSF) and pre-equilibrated coordinates in CHARMM, PDB, or NAMD format, from which the molecular components of the system are identified. These include all residues and patches connecting those residues along with water, ions, and other solute molecules. This information is then used to construct the Drude FF-based PSF using molecular generation capabilities in CHARMM, followed by minimization and equilibration. In addition, inputs are generated for molecular dynamics (MD) simulations using CHARMM, GROMACS, NAMD, and OpenMM. Validation of the *Drude Prepper* protocol and inputs is performed through conversion and MD simulations of various heterogeneous systems that include proteins, nucleic acids, lipids, polysaccharides, and atomic ions using the aforementioned simulation packages. Stable simulations are obtained in all studied systems, including 5  $\mu$ s simulation of ubiquitin, verifying the integrity of the generated Drude PSFs. In addition, the ability of the Drude FF to model variations in electronic structure is shown through dipole moment analysis in selected systems. The capabilities and availability of *Drude Prepper* in CHARMM-GUI is anticipated to greatly facilitate the application of the Drude FF to a range of condensed phase, macromolecular systems.

## INTRODUCTION

Molecular simulations based on empirical force fields (FFs) are widespread and applied to a wide range of chemical, biological, and pharmacological systems. The increase in the number of systems subjected to molecular simulations is associated with a number of driving forces from both the experimental and computational sides. Experimentally, X-ray crystallography, NMR, and, more recently, cryogenic electron microscopy (cryo-EM) are producing an ever increasing number of 3-dimensional structures of a range of macromolecules that can be starting points for molecular simulations.<sup>1</sup> Moreover, solution techniques such as X-ray and neutron scattering, Förster resonance energy transfer (FRET), and other single-molecule experiments yield observables for which molecular simulations offer the potential of producing an atomic level interpretation.<sup>2-4</sup> In addition, the combination of quantum mechanical or machine learning models in conjunction with molecular mechanics (QM/MM or ML/MM) methods offers the ability to understand data from various kinetic experiments at an atomic and electronic level.<sup>5, 6</sup>

Computational drivers leading to the increased use of molecular simulations include both hardware and software contributions. Ever increasing accessibility to multicore CPU and GPU technologies,<sup>7</sup> as well as specialized simulation hardware<sup>8</sup> allow for accessibility to more molecular systems for extended simulation times by a larger number of researchers including experimentalists interested in applying simulations to accentuate their experimental portfolios. Software developments include programs used for performing simulations, tools for the preparation of simulation systems, algorithms to access larger timescales and ensembles, and the energy functions and FFs used in the molecular simulations. Simulation programs undergoing developments to take advantage of ongoing changes in hardware include GROMACS,<sup>9</sup> GROMOS,<sup>10</sup> CHARMM,<sup>11</sup> AMBER,<sup>12</sup> NAMD,<sup>13</sup> GENESIS<sup>14</sup>, TINKER,<sup>15</sup> OpenMM,<sup>16</sup> and LAMMPS,<sup>17</sup> among others. These programs include a range of advanced simulation algorithms including free energy methods and enhanced sampling approaches including temperature and Hamiltonian replica exchange methods and a wide collection of meta dynamics approaches, many of which are implemented in the PLUMED package.<sup>18</sup> In the area of simulation system preparation, most commercially available software packages, such as MOE (Chemical Computing Group), Maestro (Schrödinger), and Discovery Studio (BIOVIA) include system preparation capabilities, while there are academic tools such as CHARMM-GUI and MDWeb,<sup>19</sup> among others. In particular, CHARMM-GUI (<https://www.charmm-gui.org>), developed in the Im laboratory with contributions from a number of other groups, allows for the preparation of highly heterogeneous biomolecular systems and is recently expanding into non-traditional biological systems such as polymers and materials.<sup>20-26</sup>

Essential to the accuracy of molecular simulations is the quality of empirical FFs from which the energies and/or forces driving the sampling of conformational space are obtained. In other words, the FF accuracy, in the limit of full sampling of the partition function, dictates the quality and utility of the simulation outcomes. In the context of biological and pharmacological systems, a number of additive or pairwise FFs are available, including AMBER,<sup>27-31</sup> CHARMM,<sup>32-37</sup> OPLS,<sup>38-40</sup> and GROMOS,<sup>41-43</sup> among others. These FFs dominate the molecular simulation landscape due to their computational efficiency with a satisfactory level of accuracy. While these FFs continue to undergo improvements, there is evidence that inherent limitations in the underlying potential energy function, notably the additive approximation for the treatment of electrostatic interactions, ultimately place an upper limit in their accuracy<sup>44-47</sup> with recent examples being the treatment of disordered proteins.<sup>48, 49</sup>

To overcome these limitations, FFs that explicitly include the treatment of electronic polarization have been introduced and are now starting to be more widely used in application studies. Methods used to treat polarizability include induced dipole,<sup>50-53</sup> fluctuating charge,<sup>54-58</sup> and classical Drude oscillator (or charge-on-spring) models.<sup>59-61</sup> Of these, the AMOEBA<sup>52</sup> and Drude polarizable FFs

are the most comprehensive. The AMOEBA FF includes proteins,<sup>52</sup> nucleic acids,<sup>53</sup> and small molecules.<sup>62, 63</sup> The Drude FF, developed in the MacKerell laboratory in collaboration with Roux and coworkers, includes nucleic acids,<sup>64, 65</sup> lipids,<sup>66</sup> proteins,<sup>67</sup> carbohydrates,<sup>68-71</sup> ions,<sup>72, 73</sup> and organic molecules<sup>61</sup> with halogens<sup>74</sup> and methods to automate the prediction of electrostatic parameters using machine learning.<sup>75</sup> In addition to the explicit treatment of polarizability, the Drude FF also accounts for the anisotropic features of hydrogen bond acceptors via the inclusion of virtual “lone pair” particles and a non-diagonal polarizability tensor. The accuracy of nonbonded interactions is improved through shielded short-range electrostatic interactions via Thole scaling<sup>76</sup> and the use of atom pair-specific Lennard-Jones parameters (e.g., NBFIX in CHARMM nomenclature).<sup>77</sup> A particular advantage of the Drude model is the use of an explicit particle to treat the electronic degrees of freedom allowing for steric polarization effects to be modeled, a property that has been used to improve the treatment of Mg<sup>2+</sup> and halogens.<sup>74, 78</sup> The Drude model is computationally efficient, having only a 3-4 fold overhead over the additive model, such that molecular dynamics (MD) simulations up to and beyond 1  $\mu$ s are possible.<sup>61, 79</sup> This contrasts the AMOEBA polarizable FF, which is slower by over an order of magnitude.<sup>80, 81</sup>

In this work, we describe *Drude Prepper* (<https://www.charmm-gui.org/input/drude>) in CHARMM-GUI that provides molecular systems and input scripts for simulations using the Drude polarizable FF. *Drude Prepper* takes equilibrated systems prepared and simulated with the CHARMM additive FF as inputs and converts them for use with the Drude FF. Current capabilities detailed and validated in this paper include the necessary conversions of the coordinate and “protein structure files” (PSF) required for polarizable simulations, as well as inputs for MD simulations using CHARMM,<sup>82</sup> NAMD,<sup>83</sup> OpenMM,<sup>84</sup> and GROMACS,<sup>85</sup> all of which have been extended to treat the Drude FF and allow for MD simulations to be performed on a range of CPU and GPU platforms. It should be noted that the term “PSF” is historical in nature and that a PSF may be created for any class of molecule. The remainder of the paper describes the additive to Drude polarizable FF conversion algorithms and their application to a range of macromolecular systems. The latest versions of the CHARMM Drude polarizable FF can be obtained separately through MacKerell lab website ([http://mackerell.umaryland.edu/charmm\\_ff.shtml](http://mackerell.umaryland.edu/charmm_ff.shtml)). The Drude FF versions used for this study are summarized in Supporting Information **Table S1**.

## METHODS

### Conversion of Additive FF PSF to Drude FF PSF

The foundation of the generation of molecules in CHARMM is based on a combination of desired residues and patches of chemical entities to build all necessary atom connectivity information (i.e., PSF) required for the energy and force calculations, as well as other molecular manipulations. Residues are the fundamental chemical entities such as amino acids, nucleotides, monosaccharides, and lipids. The patch capability in CHARMM is a unique tool that manipulates these residues. For example, there are patches to link amino acid residues or nucleotides to create polypeptides or oligonucleotides, respectively, and patches allow for a nearly endless collection of chemical modifications, including disulfide bonds, changes in protonation states, and chemical modifications of amino acids or nucleotides. In addition, patches can be used as molecular modeling tools such as adding covalent bonds between protein residues and metals to assure that the chelation state of the metal is maintained throughout an MD simulation. This approach contrasts with other programs in which the final chemical entity must be input directly into the program. Indeed, the residue and patching capabilities in CHARMM allow for an infinite number of chemical connectivities to be readily created, with a prime example being the ability to create a wide range of poly- and oligosaccharides by simply listing the desired monosaccharides followed by the list of patches creating the various types of glycosidic linkages and chemical

modifications. Accordingly, several simulation packages including NAMD and OpenMM allow for CHARMM PSFs to be directly read, directly supplying the information required for molecular simulations with those programs.

The molecular generation tools were initially developed in the context of the CHARMM additive FF. Currently, the CHARMM all-atom additive FF contains over 2,200 residues and 300 patches, representing a near comprehensive collection of chemical coverage with emphasis on biological and pharmacological systems, although now extending significantly beyond that range. To take advantage of this, the Drude FF largely uses the same molecular generation strategy, importantly, maintaining the same residue and atom name nomenclature as in the additive FF except water molecules. Accordingly, it is possible and a good strategy to perform a CHARMM additive PSF to Drude PSF conversion by detecting the residues and patches used in an additive FF PSF and creating the corresponding Drude FF PSF.

*Drude Prepper* handles this conversion in two steps: 1) detecting the components and patches that comprise an input molecular system and 2) reassembling the components in the context of the Drude FF. The detection of system components is done based on the residue names in each segment (SEGID), e.g., a single polypeptide chain. For example, if all residues in a segment are amino acids, the segment is classified as a protein. DNA and RNA are detected in a similar way, but if the segment contains O<sub>2</sub> atoms, it is classified as RNA, and DNA otherwise. Default patches are then applied to create the respective biopolymers. Special patches applied to each segment are detected by checking for the existence of a certain atom in a residue. For example, an acetylated N-terminus is detected by confirming the presence of a CY atom at the N-terminal residue. In addition, a disulfide bond can be detected by searching for any SG atom bonded to the SG atom of another residue.

Handling carbohydrates is the most demanding task due to the wide variety of glycosidic linkages. To properly recognize the glycosidic linkages between sugar residues, *Drude Prepper* utilizes *Glycan Reader*<sup>86, 87</sup> that recognizes the monosaccharide types and linkages using the atomic coordinates and connectivity information in a PSF. The detected linkage information is passed in the form of a CHARMM stream file. Note that *Drude Prepper* only supports the sugars that are available in the Drude FF such as general furanoses, pyranoses, hexouronic acids, and other acid moieties.<sup>88</sup> Additional monosaccharides and linkage types, such as seven-, eight-, nine-carbon sugars, and N/O glycosidic linkages to proteins will be included as they are developed. More generally, *Drude Prepper* only covers those chemical entities explicitly in the Drude FF, although the coverage is currently quite extensive. On the other hand, there is currently no analog of the CHARMM General Force Field (CGenFF) and associated program,<sup>89</sup> although efforts towards a Drude General Force Field (DGnFF) are actively ongoing.

Water molecules are detected if a segment contains a residue named TIP3. All other components not described above are classified as “else”, and *Drude Prepper* issues an error if topologies and parameters for the residues in “else” are not available in the Drude FF. After all components are successfully recognized in the first step, they are reassembled based on the recognized residues and patches in the Drude FF. A mass of 0.4 AMU is subtracted from polarizable atoms (i.e., non-hydrogen atoms) and assigned to the Drude oscillators or particles. Lone pair particles are generated according to the topology of the residues in the Drude FF. While charge is distributed onto the lone pair, no mass is assigned to the lone pair particles as the forces acting on those particles are transformed back to their parent atom as required for propagation of a MD trajectory. The TIP3P<sup>90, 91</sup> water molecules from the additive PSF are converted to the SWM4-NDP<sup>92</sup> water model; a more accurate, but computationally more demanding water model SWM6 is also available.<sup>93</sup> In addition to creating the scripts to perform molecular simulations in CHARMM, *Drude Prepper* provides input scripts optimized for simulating the system with the Drude FF using NAMD, GROMACS, and OpenMM. Once the conversion from the additive to Drude FF PSF is complete

and the input scripts are created, *Drude Prepper* then performs initial system minimization using CHARMM. Details of these steps and the input options for the various programs follow.

## Drude System Preparation and Simulation Protocols

*Drude Prepper* creates a collection of CHARMM input and stream files for system setup, minimization, and simulation using the Drude FF. Brief descriptions of these input and stream files are included in **Table S2**. In addition, input scripts and supporting files for equilibration and production simulations using NAMD, OpenMM, and GROMACS can be found in each program directory from “download.tgz” that can be downloaded at the final step of *Drude Prepper*. Note that one can modify the inputs to adjust equilibration and production times and the trajectory saving steps for their specific needs. Detailed descriptions of the contents in these input files follow.

**CHARMM pre-equilibration with the additive FF:** First, all simulation systems are assumed to be built using *Solution Builder* or *Membrane Builder* modules in CHARMM-GUI and are equilibrated using the CHARMM additive FF and the equilibration inputs provided by CHARMM-GUI. We refer this additive FF based equilibration to as “pre-equilibration” in the context of preparing the simulation system for the Drude FF. Using the standard inputs generated by *Solution Builder* or *Membrane Builder*, the systems are minimized and equilibrated in the NVT (constant particle number, volume, and temperature) and NPT (constant particle number, pressure, and temperature) ensembles. System sizes are generated to have a minimum of 10 Å buffer of TIP3P water beyond the non-hydrogen atoms of the macromolecules in all directions. A force-switch smoothing function over 10 to 12 Å is applied for Lennard-Jones (LJ) interactions. The particle mesh Ewald (PME)<sup>94</sup> method is used for long-range electrostatic interactions, and the SHAKE<sup>95</sup> algorithm is utilized for fixing all bonds including hydrogen atoms. In this study, all pre-equilibration was performed at 303.15 K. The backbone and sidechain heavy atoms were harmonically restrained during this equilibration. The definition of backbone and sidechain heavy atoms are provided in **Table S3**. It needs to be emphasized that using a well pre-equilibrated system to initiate the Drude FF calculations is recommended, especially in cases where the starting structure is of low resolution, contains externally docked ligands, and/or have homology modeled sections, etc.

**CHARMM minimization:** Since the Drude particles are initially assigned the same coordinates as their parent atoms, the initial minimization requires special attention. *Drude Prepper* performs this minimization using CHARMM in two stages. First, only the Drude particles are minimized with a very strong harmonic restraint with a force constant of 100,000 kcal/mol/Å<sup>2</sup> applied to all real atoms including hydrogens. 100 steps of adopted basis Newton-Raphson (ABNR) are used in this stage. Next, the strong restraints on all real atoms are removed and only the non-hydrogen atoms (**Table S3**) are restrained with force constants of 1 kcal/mol/Å<sup>2</sup> (for backbone atoms) and 0.1 kcal/mol/Å<sup>2</sup> (sidechain atoms). The full system is minimized for 100 steps of steepest descent (SD) and another 100 steps of ABNR. Long-range electrostatic interactions are computed using PME<sup>94</sup> summation with a real space cutoff of 12 Å with the nonbond list maintained out to 16 Å in conjunction with an heuristic update, an Ewald splitting parameter ( $\kappa$ ) of 0.34 Å<sup>-1</sup>, a grid spacing of ~1.0 Å, and a sixth-order interpolation of the charge to the grid. The LJ interactions are truncated at 12 Å with a switching function<sup>96, 97</sup> over 10–12 Å. The isotropic long-range correction (LRC)<sup>98</sup> is added to the non-bonded interactions that assumes the radial distribution function is unity beyond the real space cutoff. Future versions of *Drude Prepper* will include the necessary scripts to perform simulations using the LJ-PME treatment of long-range LJ interactions.<sup>99</sup> The internal geometry of water molecules and covalent bonds involving hydrogen atoms are kept fixed using a separate SHAKE<sup>95</sup> algorithm called RATTLE/Roll. The minimized coordinates of the

system are used to generate the inputs for both equilibration and production in CHARMM or user-selected additional MD platform(s).

**CHARMM equilibration & production:** Performing a short MD equilibration is essential prior to production simulations. The MD equilibration involves a 100-ps simulation in the NPT ensemble using a small timestep of 0.5 fs. The simulation is performed with the velocity-Verlet integrator (VV2) developed specifically for the Drude model.<sup>82</sup> A dual Nose-Hoover<sup>100, 101</sup> thermostat is used, where the thermostat coupled to Drude particles is set to 1 K and the thermostat for real atoms is set to the target temperature. The friction coefficient is set to 10 ps<sup>-1</sup> for the temperature control and the pressure is controlled at 1 atm by a modified Anderson-Hoover barostat.<sup>102</sup> The nonbonded interaction options, the usage of RATTLE/Roll for water molecules, and positional harmonic restraints to the non-hydrogen atoms are the same as in the above CHARMM minimization. To avoid polarization catastrophe, a hard-wall constraint is used with a parent-Drude bond length cutoff of 0.2 Å.<sup>66, 72</sup> During the equilibration, a trajectory is saved every 2.5 ps for any analysis, and the final coordinate is used to initiate the production simulation. The input for the CHARMM production MD simulation is identical to that used for equilibration except that the integration timestep is set to 1 fs, coordinates are saved every 10 ps, and all positional harmonic restraints are removed.

Equilibration with OpenMM, GROMACS, or NAMD uses the same timestep and positional restraint protocol as in the above CHARMM equilibration with the differences being in the integrators, thermostats, and barostats as described below for the production MD inputs.

**NAMD:** The NAMD program<sup>83</sup> uses an extended Lagrangian method for integrating Drude particle positions in Langevin dynamics.<sup>82, 83</sup> For NPT simulation, the temperature is regulated by a dual thermostat<sup>82</sup> with Drude particles coupled to 1 K with a damping coefficient of 20 ps<sup>-1</sup> and real atoms coupled to the thermostat at the target temperature with a friction coefficient of 5 ps<sup>-1</sup>. The pressure is maintained at 1 atm using the Langevin piston method<sup>103</sup> with an oscillation period of 50 fs and a decay time of 25 fs. The timestep is set to 1 fs and coordinates are saved every 10 ps for production. PME<sup>94</sup> summation is used to compute long-range electrostatics with ~1 Å grid spacing and an interpolation order of 6. Non-bonded interactions are calculated with a real space cutoff of 12 Å and a switching distance is set to 10-12 Å for the LJ interactions and the isotropic LJ long-range correction is applied. Note that the force switching is turned off while using the Drude FF and a potential switch is used for the LJ interactions. The pair lists within 16 Å are updated every 10 steps. Constraints are added using SHAKE to keep the covalent bonds involving hydrogen atoms rigid. The Drude hardwall is set to 0.2 Å.

**OpenMM:** The OpenMM program<sup>16</sup> has GPU implementation<sup>84</sup> capability for the Drude FF. An extended Lagrangian dynamics simulation with velocity-Verlet integrator is implemented for treating the motion of atoms and the associated Drude particles. A dual thermostat with specific reference temperatures and friction coefficients for both real atoms and Drude particles is used with the same values as for CHARMM and NAMD. Differently for OpenMM, the pressure is maintained using a Monte Carlo barostat set to 1 atm with a coupling frequency of 100 steps. The timestep is set to 1 fs and coordinates are saved every 10 ps. PME is used for the electrostatic interactions and the LJ interactions are switched over 10 to 12 Å with a long-range isotropic correction. The Ewald error tolerance is set to 0.0001. The SHAKE algorithm is implemented in OpenMM to constrain all covalent bonds that involve hydrogen atoms, and the Drude hardwall is set to 0.2 Å.

**GROMACS:** The GROMACS program<sup>104</sup> allows for both GPU and CPU simulations of the Drude polarizable systems.<sup>85</sup> Parallelization on CPU hardware is currently supported via OpenMP parallelization. The integrator implemented for extended Lagrangian dynamics in Drude systems is the same two-step velocity-Verlet algorithm as implemented in CHARMM. Currently, only NVT

simulations are supported in GROMACS for the Drude FF using a dual Nosé-Hoover thermostat. Real atoms are coupled to a thermostat at a target temperature with a coupling constant of 0.1 ps and Drude particles are coupled to a 1 K thermostat with a coupling constant of 0.005 ps. The timestep is set to 1 fs and coordinates are saved every 10 ps. The Drude hard wall constraint is set to a maximum value of 0.2 Å. Bonds involving hydrogen atoms are constrained via LINCS.<sup>105</sup> The short-range LJ potential is switched to zero over 10–12 Å. Electrostatic forces are computed with the smooth PME algorithm<sup>94</sup> with a real-space cutoff of 12 Å.

### Dipole Moment Calculations

Dipole moments of selected moieties were calculated from the trajectories and presented as probability distributions. Both the CHARMM additive and Drude FFs define the topologies such that the molecules can be divided into the groups of atoms with integer charges. While this feature not only facilitates the modular approach to building new chemical entities, it also provides an opportunity to analyze dipole moments of specific moieties in a highly specific way. The dipole moment is calculated from the charges of all the particles in the moiety and their spatial relationship. In the case of the Drude FF, this includes all Drude particles and lone pairs, as well as atoms. Presented in **Table S4** is the list of atom definitions for integer charge groups. When the sum of the charges in the moiety equals zero, the dipole moment is orientation independent. However, if the charge does not equal zero, the dipole moment is position dependent. In such cases, the structures from each simulation snapshot were aligned with the crystal structure based on the backbone heavy atoms and the coordinates of the selected particles were translated to define their center of geometry as the origin prior to the dipole moment calculation. This was performed with all groups of integer -1 or +1 (i.e., acidic and basic amino acids), as well as with PRO residues in the Drude FF whose total charge is 0.178.

### Simulation Systems

**Table 1** summarizes the diverse biomolecular systems used in this study to test *Drude Prepper* and its simulation protocols, and the respective system images are shown in **Figure 1**. For each system, we performed pre-equilibration for 1 ns, equilibration for 100 ps, and production for 100 ns, 200 ns, or 5 μs using different MD engines as described.

**Table 1.** Simulation system descriptions.

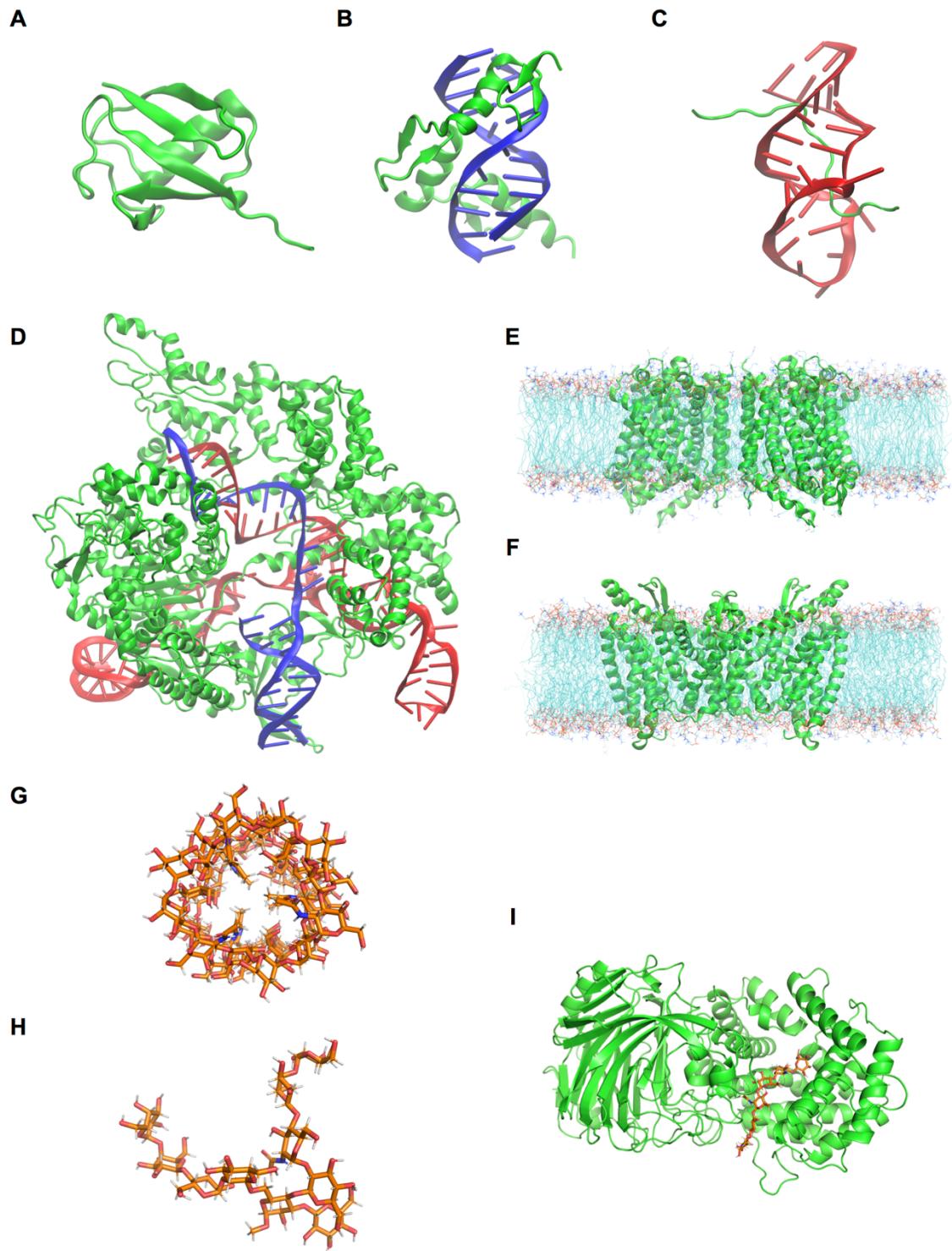
System name	Molecular type	# particles <sup>†</sup>	Box size (Å <sup>3</sup> )	Pre-equilibration / Equilibration / Production
Ubiquitin	Protein	41,694	64 × 64 × 64	NAMD / GROMACS / GROMACS
TF-DNA	Protein + DNA	51,026	66 × 66 × 66	CHARMM / NAMD / NAMD
HIV-1 TAR-TAT RBD	Protein + RNA	50,444	69 × 69 × 69	CHARMM / OpenMM / OpenMM
Cas9-sgRNA-DNA	Protein + DNA + RNA	426,781	140 × 140 × 140	OpenMM / OpenMM / OpenMM
LeuT in a POPC bilayer	Protein + Lipid	411,236	155 × 155 × 114	OpenMM / OpenMM / OpenMM
MalT-maltose complex	Protein + Lipid + Carbohydrate	382,991	150.5 × 150.5 × 112.0	OpenMM / OpenMM / OpenMM
<i>E. coli</i> O176-antigen OS	Carbohydrate	198,373	108 × 108 × 108	OpenMM / OpenMM / OpenMM
<i>M. catarrhalis</i> OS	Carbohydrate	12,468	43 × 43 × 43	OpenMM / OpenMM / OpenMM

HA lyase–HA substrate	Protein + Carbohydrate	208,974	$110 \times 110 \times 110$	OpenMM / OpenMM / OpenMM
-----------------------	------------------------	---------	-----------------------------	--------------------------

<sup>†</sup>Including Drude particles and virtual lone pairs.

**Ubiquitin:** Human erythrocytic ubiquitin coordinates were obtained from the 1.8 Å resolution crystal structure (PDB ID 1UBQ).<sup>106</sup> The protein was solvated in a cubic box with 7,921 water molecules. For this system, the protein carries no net charge and simulations were performed without any neutralizing counterions or added salt. The initial box size was  $64 \times 64 \times 64 \text{ \AA}^3$  (**Figure 1A**).

**Transcription factor-DNA complex:** A simulation system of a zinc finger transcription factor (TF) binding with a DNA fragment was setup using the 1.6 Å resolution X-ray structure (PDB ID 1AAY).<sup>107</sup> The double strand DNA contains 11 base pairs, and the TF-DNA complex was solvated in a cubic water box containing 9,493 water molecules and 150 mM KCl (including neutralization cations, 28 K<sup>+</sup> and 25 Cl<sup>-</sup> ions). The initial box size was  $66 \text{ \AA} \times 66 \text{ \AA} \times 66 \text{ \AA}^3$  (**Figure 1B**).



**Figure 1.** Molecular visualizations of the simulation systems in this study: (A) Ubiquitin, (B) TF-DNA, (C) HIV-1 TAR-TAT RBD, (D) Cas9-sgRNA-DNA. (E) LeuT in a bilayer, (F) MalT-maltose complex in a bilayer, (G) *E. coli* O176 O-antigen oligosaccharide, (H) *M. catarrhalis* serotype C oligosaccharide, and (I) HA lyase – HA substrate. Coloring scheme: green cartoon for protein, red cartoon for RNA, blue cartoon for DNA, cyan lines for lipids, and orange sticks for carbohydrates.

**HIV-1 TAR with Tat RNA binding domain (RBD):** Interaction of human immunodeficiency virus (HIV) trans-activation-response (TAR) RNA element with the trans-activator protein (Tat) is critical towards viral mRNA production.<sup>108</sup> A novel conformational rearrangement in the TAR RNA binding site allows Tat RNA binding domain to recognize it.<sup>109</sup> System construction was initiated from model 1 of the solution NMR structure (PDB ID 6MCE).<sup>110</sup> The RNA-peptide complex was solvated in a cubic box of  $69 \times 69 \times 69 \text{ \AA}^3$  with TIP3P water and 0.15 M KCl buffer (including neutralization cations, 48 K<sup>+</sup> and 27 Cl<sup>-</sup> ions) (**Figure 1C**).

**Cas9-sgRNA-DNA:** The simulation setup for Cas9-sgRNA-DNA complex started with a cryo-EM structure (PDB ID 6O0Z) at 3.3 Å resolution.<sup>111</sup> The complex was solvated in a cubic box of  $140 \times 140 \times 140 \text{ \AA}^3$ . Overall negative charge on the complex was neutralized by adding K<sup>+</sup> ions to the system along with 150 mM KCl buffer (324 K<sup>+</sup> and 214 Cl<sup>-</sup> ions) (**Figure 1D**).

**LeuT in a POPC bilayer:** Leucine transporter LeuT is a 12 transmembrane-helix protein that utilizes established electrochemical potential of Na<sup>+</sup> ions to facilitate the transport of neutral amino acids. The LeuT protein (PDB ID 3TT1)<sup>112</sup> at a resolution of 3.1 Å was embedded in a palmitoyloleoyl-phosphatidylcholine (POPC) bilayer with 290 POPC at the top leaflet and 297 POPC at the bottom leaflet. 53,950 water molecules were added along with 150 mM NaCl (including neutralization ions, 146 Na<sup>+</sup> and 154 Cl<sup>-</sup> ions), reaching an initial system size of  $155 \times 155 \times 114 \text{ \AA}^3$  (**Figure 1E**).

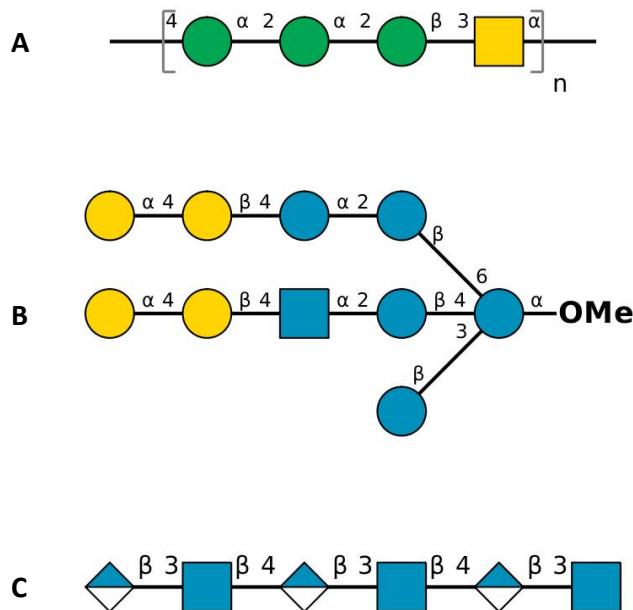
**MalT-maltose complex in a mixed bilayer of DOPC and DOPE:** MalT, a member of the enzyme IIC family, is a membrane-embedded sugar transport protein that is part of the phosphoenolpyruvate-dependent phosphotransferases.<sup>113-115</sup> The initial simulation system was built using PDB ID 5IWS, 2.55 Å resolution, for MalT from *Bacillus cereus*.<sup>116</sup> To properly recognize the maltose substrate, the PDB structure was downloaded from RCSB not OPM.<sup>117</sup> Since PDB:5IWS contains only MalT monomer and a maltose substrate, a dimeric structure was prepared using the “Generation of Biologically Functional Unit” option in *PDB Reader & Manipulator*.<sup>118</sup> In addition, since the protein structure from RCSB is not properly oriented to a membrane whose normal is along the Z-axis and center is at Z = 0, the PPM option of the OPM server<sup>117</sup> was used to orient the MalT dimer. Then, the structure was embedded in a 3:1 mixed bilayer of dioleoyl-phosphatidylethanolamine (DOPE) and dioleoyl-phosphatidylcholine (DOPC): 213 DOPE and 71 DOPC at the top leaflet and 228 DOPE and 76 DOPC at the bottom leaflet. 49,355 TIP3P water molecules and 0.15 M KCl (including neutralization ions, 133 K<sup>+</sup> and 153 Cl<sup>-</sup>) were added to the system, and the initial system size was  $150.5 \times 150.5 \times 112.0 \text{ \AA}^3$  (**Figure 1F**).

**Escherichia coli O176 O-antigen oligosaccharide:** O-antigen polysaccharides constitutes part of the lipopolysaccharide molecule present in the outer membrane of Gram-negative bacteria,<sup>119</sup> which have diverse chemical compositions even within a single species.<sup>120</sup> In a previous study, Patel et al. studied the conformational preferences of an *Escherichia coli* O176 O-antigen oligosaccharide at the atomic level using NMR spectroscopy and MD simulations.<sup>121</sup> An initial structure of ten repeating units of O176 O-antigen structures (**Figure 2A**) was built by *Glycan Modeler*<sup>24</sup> and immersed at the center of a cubic box ( $108 \times 108 \times 108 \text{ \AA}^3$ ) including 150 mM KCl (112 K<sup>+</sup> and 112 Cl<sup>-</sup>) (**Figure 1G**).

**Moraxella catarrhalis serotype C oligosaccharide:** *Moraxella catarrhalis* is a pathogenic Gram-negative bacterium that causes otitis media and sinusitis in children.<sup>122</sup> In a previous study, all-atom MD simulations were conducted to understand the conformational properties of three major serotypes A, B, and C that are responsible for approximately 95% of the clinical isolates.<sup>123</sup> In this work, we built a serotype C oligosaccharide (**Figure 2B**). Lipid A and two Kdo residues in serotype C were replaced by O-methylation. An initial structure of the oligosaccharide was built by *Glycan*

*Modeler* and immersed at the center of a cubic box of TIP3P water molecules ( $43 \times 43 \times 43 \text{ \AA}^3$ ) with 150 mM KCl (6  $\text{K}^+$  and 6  $\text{Cl}^-$ ) (**Figure 1H**).

**Hyaluronate lyase in complex with hyaluronan substrate:** Hyaluronate lyase enzymes degrade hyaluronan, which is a linear glycosaminoglycan consisting of a polymer with *N*-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) repeating units linked via alternating  $\beta$ 1-4 and  $\beta$ 1-3 glycosidic bonds.<sup>124</sup> It is present predominantly as a main polysaccharide component of the host connective tissues including bacterial toxins.<sup>125</sup> An initial complex structure of hyaluronate lyase enzyme with hexasaccharide hyaluronan substrate (**Figure 2C**) was set up using PDB ID 1LOH<sup>125</sup> with a resolution of 2.0 Å. The terminal GlcA ring at the non-reducing end of the substrate was distorted in the crystal structure,<sup>125</sup> so its coordinates were refined using *Glycan Modeler*. This initial complex was placed into the center of a cubic box ( $110 \times 110 \times 110 \text{ \AA}^3$ ), solvated with 42,412 TIP3P water molecules and 150 mM KCl (124  $\text{K}^+$  and 107  $\text{Cl}^-$  ions including neutralization ions) (**Figure 1I**).



**Figure 2.** Symbol representation of carbohydrate sequences of (A) *E. coli* O176 O-antigen repeating unit ( $n = 10$ ), (B) *M. catarrhalis* serotype C oligosaccharide, and (C) hyaluronan hexasaccharide in PDB ID 1LOH: blue circle for D-glucose (Glc), green circle for D-mannose (Man), yellow circle for D-galactose (Gal), blue square for *N*-acetyl-D-glucosamine (GlcNAc), yellow square for *N*-acetyl-D-galactosamine (GalNAc), and half-filled blue diamond for D-glucuronic acid (GlcA).

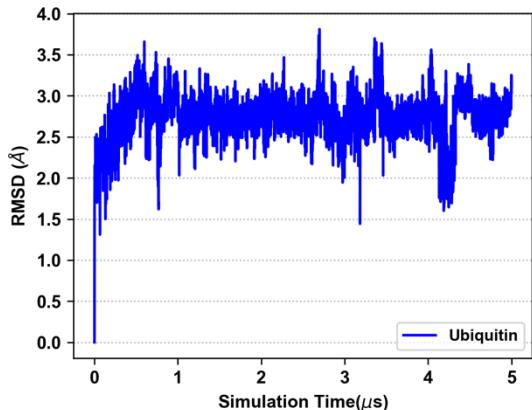
## RESULTS AND DISCUSSION

The unique functionality of CHARMM-GUI is its versatility to allow users with the option to choose widely used simulation packages for their simulations. With Drude MD simulations now supported by CHARMM, NAMD, OpenMM, and GROMACS, we have been able to transfer this feature into *Drude Prepper* as well. To demonstrate it, MD simulations of various systems in **Table 1** were carried out with different MD simulation software. While the pre-equilibration with the additive FF

can be performed by using any of the simulation packages, it is recommended to perform the equilibration and production MD simulations with the Drude FF using the same simulation package. In this study, we have performed 1 ns pre-equilibration for each system from which the last coordinates along with the additive PSF file were then used as input data for *Drude Prepper*. To evaluate the implementation of the Drude model in NAMD, OpenMM, and GROMACS, as well as the conversion of the CHARMM PSF to the data structures required by those programs by *Drude Prepper* and the consistency of the respective inputs, we performed a single-point energy component analysis on the minimized system coordinates (**Table S5**). The total energy differences for NAMD, OpenMM, and GROMACS compared to CHARMM were less than  $\pm 0.1\%$  except the carbohydrate-containing systems, which were less than  $\pm 0.3\%$ .

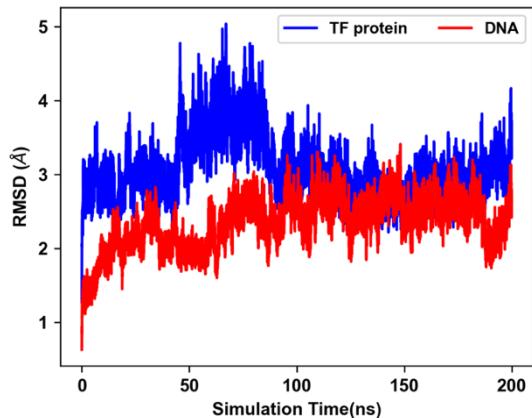
Using the input files generated by *Drude Prepper*, each system was equilibrated for 100 ps followed by 100 or 200 ns production MD simulation, or 5  $\mu$ s performed in the case of ubiquitin. To evaluate the stability of the systems, root mean square deviations (RMSD) of backbone atoms with respect to the initial structure, root mean square fluctuation (RMSF) of each residue, and radius of gyration (RGYR) of the macromolecules were calculated. The production trajectories were aligned to the backbone atoms of the respective protein, nucleic acid, or carbohydrates segments for these calculations. In addition, given the variation of the electronic structure of the molecules associated with the explicit treatment of polarization in the Drude FF, variations in dipole moments for selected moieties were calculated. The calculation of dipole moments, which can have significant impact on the dynamics and interactions,<sup>126, 127</sup> was facilitated by the presence of neutral or integer charge group in the Drude FF. For example, the charges of the protein backbone C, O, N, H, C $\alpha$ , and H $\alpha$  atoms sum to zero as do charges of the neutral protein sidechains. Definitions of the atoms, Drude particles, and lone pairs included in the different moieties used for the dipole moment calculations are in **Table S4**. In the case of charged groups, the individual moieties were RMSD aligned based on the non-hydrogen atoms to a reference frame based on the structure used to initiate the MD simulations

**Ubiquitin:** The solvated structure of ubiquitin remained stable over 100 ns GROMACS simulation with no systematic structural deviation, as demonstrated by stable backbone RMSD (**Figure S1A**) and RGYR (**Figure S1B**). After an initial increase to  $\sim 3\text{ \AA}$ , the RMSD relaxed back to  $\sim 2\text{ \AA}$  for the remainder of the 100 ns simulation. The RMSF analysis revealed that the C-terminal residues were very flexible during the simulation, accounting for most of the structural deviation (**Figure S1C**). We also ran the equilibration and production for the same system using OpenMM to determine if any variations in the structure would occur in an extended long-time MD. The production simulation was carried out for 5  $\mu$ s and the trajectory was saved every 1 ns. The protein stayed stable over the entire simulation with a backbone RMSD of  $\sim 3\text{ \AA}$  (**Figure 3**).



**Figure 3.** Backbone RMSD of ubiquitin from the 5  $\mu$ s MD simulation performed using OpenMM.

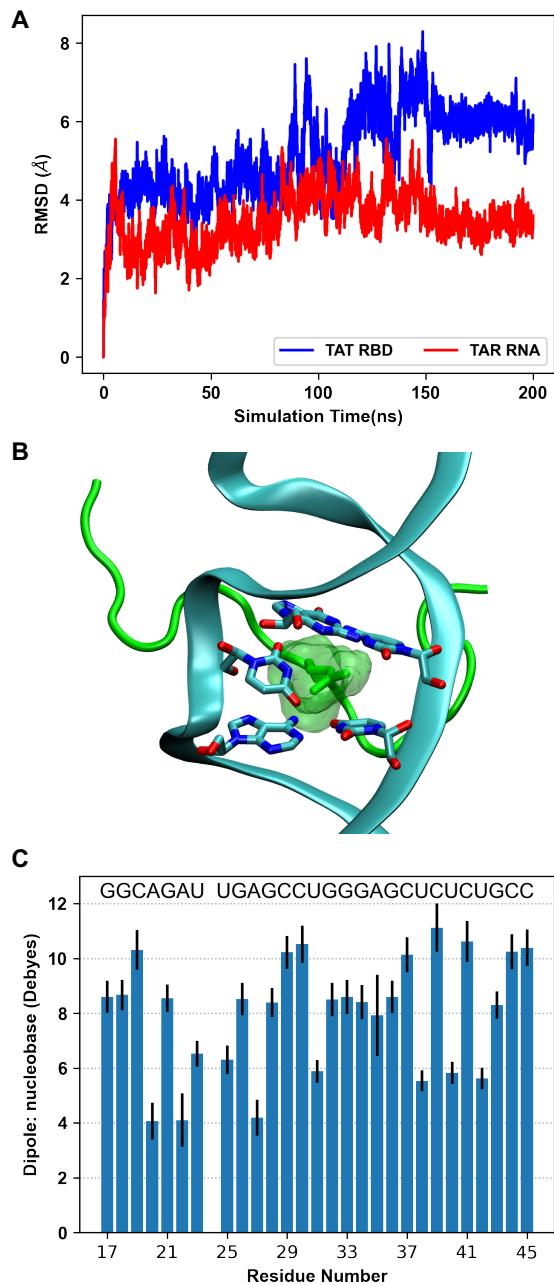
**Transcription factor-DNA complex:** TF proteins modulate the transcription of genetic information from DNA to mRNA by binding specifically to DNA sequences. A model TF-DNA system was simulated with the Zif268 zinc finger TF bound to a 11 base-pair double strand DNA. The TF-DNA interactions were well maintained throughout the 200 ns MD simulation. The heavy atom RMSD of TF protein or DNA was computed by aligning the DNA structure to its initial structure. As illustrated in **Figure 4**, the DNA structures are very stable with RMSD fluctuating between 2 Å and 3 Å. The protein RMSD increases to ~4.5 Å after 50 ns MD simulation and then decrease to ~3 Å and remains stable out to 200 ns, indicating conformational dynamics of the TF protein.



**Figure 4.** RMSD of TF-protein and DNA heavy atoms after alignment of the DNA structure with respect to the initial structure.

**HIV-1 TAR with Tat RBD:** The MD simulation was started with an NMR structure (PDB ID 6MCE) of HIV-1 TAR RNA-Tat RBD peptide complex that contains an arginine sandwich motifs.<sup>110</sup> We find that this interaction was short lived during MD simulation as Arg49 and Arg52 lost the initial interactions with the sandwiching nucleobases. The peptide RMSD is higher than the TAR RNA with most deviation coming from exposed N and C termini of the peptide (**Figure 5A**). However,

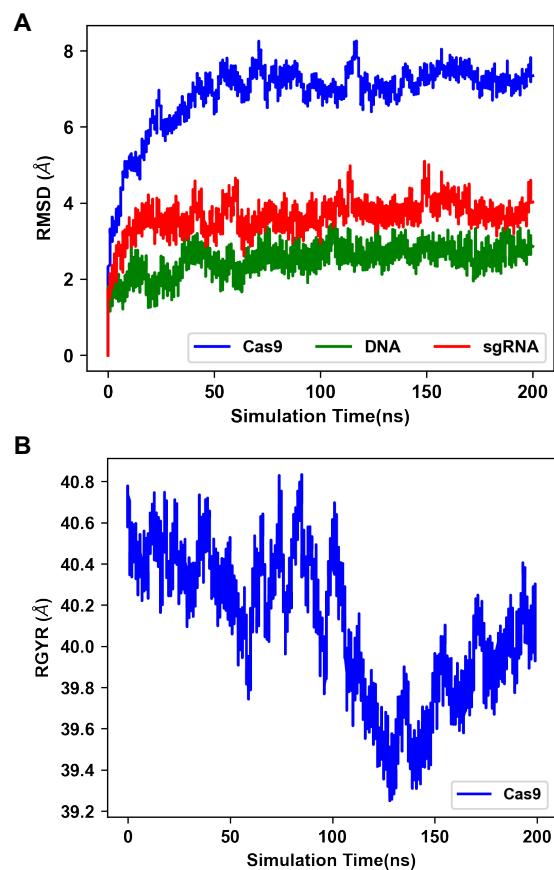
the TAR RNA over the simulation adapts a conformation with Arg52 sandwiched between two base-pairs (**Figure 5B**), while Arg49 interacts with two flipped bases. Such interactions stabilize the RMSD during last 50 ns. We find significant standard deviations (up to  $\pm 1.5$  D) in dipole moments of some nucleobases (**Figure 5C**). Peptide bond dipole moments also show slight variations in the mean values across the residues (**Figure S2**).



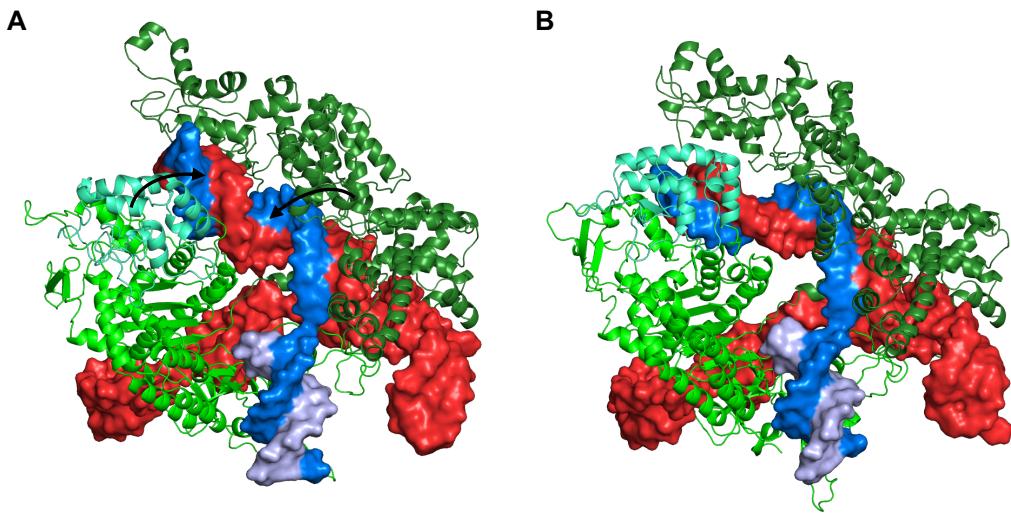
**Figure 5.** (A) RMSD of HIV-1 TAR RNA and Tat RBD with respect to their initial structure. (B) Snapshot of HIV-1 TAR-TAT RBD complex at 200 ns, showing Arg52 sandwiched between two base-pairs. Protein is shown as green cartoon and RNA is shown as cyan cartoon. (C) Dipole

moments of nucleobases of HIV-1 TAR RNA calculated as an average over the 200 ns sampled every 1 ns. The error bars represent one standard deviation.

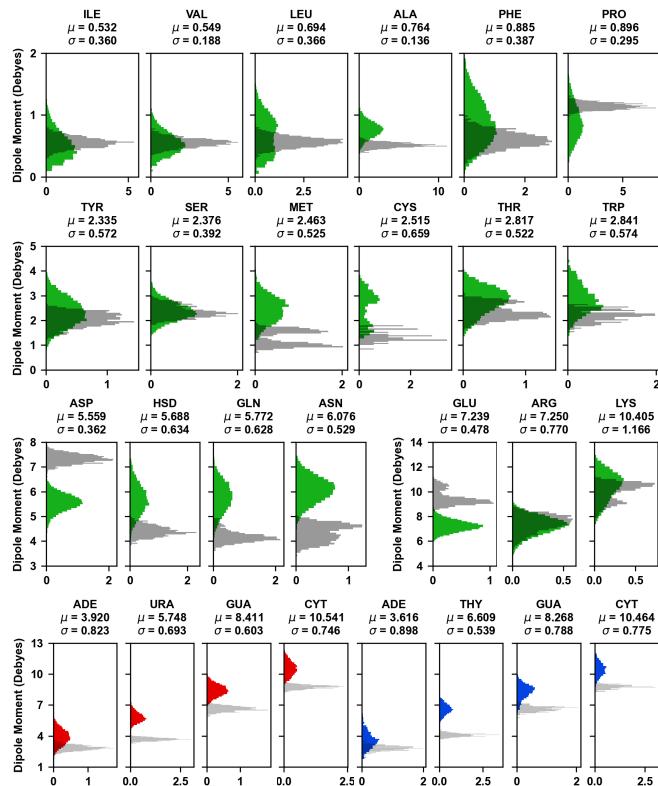
**Cas9-sgRNA-DNA:** This multimolecular complex is the largest of the selected systems in terms of number of atoms including Drude particles and lone pairs (**Table 1**). The RMSD of Cas9 protein, DNA, and RNA components shows that the complex adapts a stable conformation after 50 ns (**Figure 6A**). The major conformational change comes from the recognition lobe (**Figure 7**), which was also observed in previous MD simulations of the complex.<sup>23</sup> The protein RGYR time series shows significant reduction as the recognition lobe and HNH domain tightly wrap around the portion showing interactions between target DNA strand and 5' end of sgRNA (**Figure 6B**). We calculated the distributions of dipole moments of amino acid sidechains and nucleobases (**Figure 8**). Compared to the additive FF, the Drude FF appears to provide broader fluctuations in the dipole moment to respond to the interactions. We find that there are major differences in the mean dipole moments of amino acids such as ASP, GLU, ASN, and GLN. In case of nucleobases, we find that dipole moments are consistently higher with the Drude FF. The creation of sets of atoms whose total charge is an integer in the Drude and CHARMM FFs is critical for the dipole moment comparisons. Such groups include peptide bonds, amino acid sidechains, nucleobases, heads/tails of lipids, etc. The selection definitions for atoms, Drude particles and lone pairs in each group is in **Table S4**.



**Figure 6.** (A) RMSD of different components of Cas9 complex following alignment of each component to its initial structure. (B) RGYR of Cas9 protein.



**Figure 7.** Conformations of Cas9-sgRNA-DNA complex at (A) 0 ns and (B) 200 ns. The protein is shown as green cartoon with the recognition lobe in forest green and the HNH domain in cyan. The polynucleotides are shown as surface with sgRNA in red, target DNA strand in marine blue, and non-target DNA strand in light blue. The arrows in (A) show the movement of the recognition lobe and HNH domain during the simulation that results in a conformation in (B).

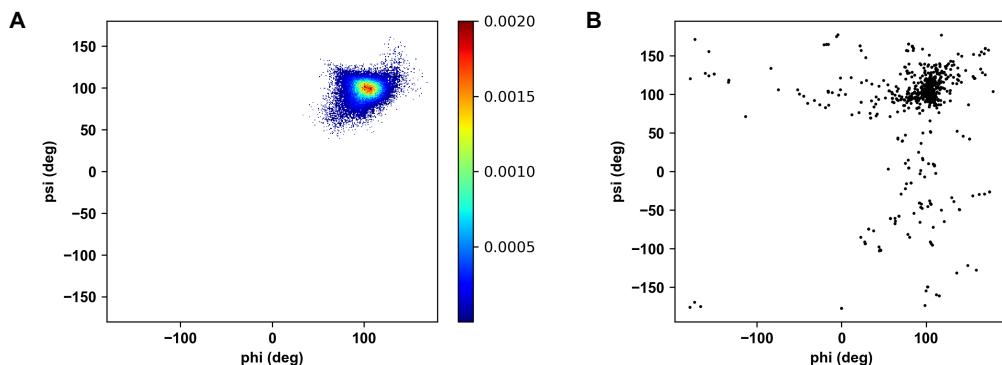


**Figure 8.** Distributions of dipole moments of amino acid sidechains and nucleobases in the Cas9 complex. The colored histograms correspond to the Drude FF (Cas9 – green; sgRNA – red; DNA

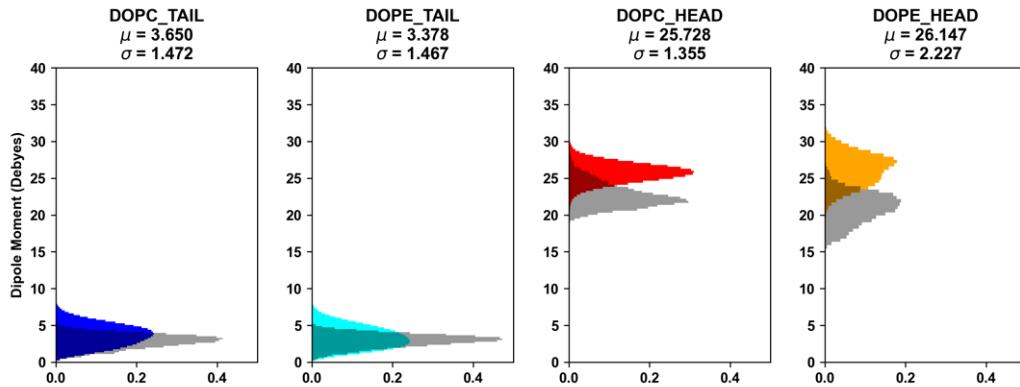
– blue) calculated based on 200 ns simulation, while the transparent gray histograms represent the additive FF calculated based on the 1 ns pre-equilibration trajectory. The histograms are normalized and the mean and standard deviation from Drude simulations is shown above each subplot. Comparison of Drude FF dipole moment distributions at 1 ns and 200 ns is presented in **Figure S3**.

**LeuT in a POPC bilayer:** The availability of the high-quality Drude FF parameters for both protein and lipids allows the study of membrane proteins with charge polarization. A LeuT homodimer in an outward-open conformation was embedded in a POPC lipid bilayer and subjected to a 200 ns MD simulations using OpenMM. The overall protein structure is stable during the MD trajectory. The slowly increasing RMSD mainly reflected the conformational changes of extracellular loops (**Figure S4**).

**MalT-maltose complex in a mixed bilayer of DOPC and DOPE:** The Drude FF has also been tested for a protein-carbohydrate complex in a membrane. The MalT homodimer with a maltose in each binding site was embedded in a mixed bilayer of DOPC:DOPE=1:3 (**Figure 1F**). The distribution of glycosidic torsion angles of maltose (**Figure 9A**) shows good agreement with that obtained from GFDB (Glycan Fragment Database)<sup>128</sup> (**Figure 9B**) that is composed of a glycosidic torsion angle distributions of selected glycan sequences in the PDB. The dipole moments of head and tail groups of DOPC and DOPE lipids were calculated (**Figure 10**). There is no significant difference in the dipole moments of tail groups between DOPC and DOPE lipids, which is expected as they have identical acyl tails. The phosphatidyl ethanolamine (PE) head group shows a peak at higher dipole moments compared to phosphatidyl choline (PC), indicating that the PE head group is slightly more polar than the PC head group. The calculated dipole moments of lipid tail groups from the Drude simulations show broader distribution compared to the additive FF. In addition, Drude lipid head groups show higher dipole moments than the additive FF lipids.

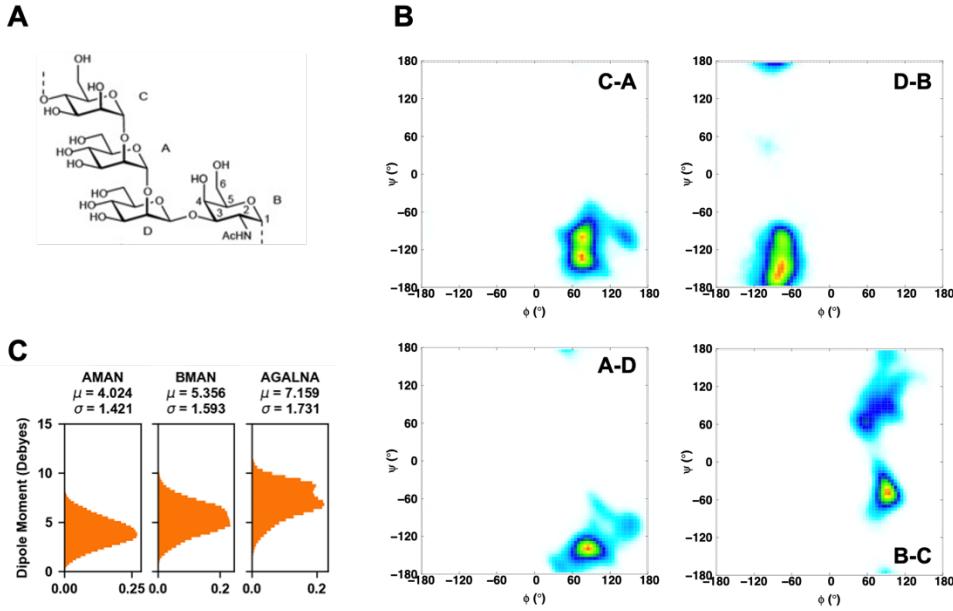


**Figure 9.** Distributions of glycosidic torsion angles of maltose in (A) MalT-maltose complex simulation and from (B) all PDB maltose structures.



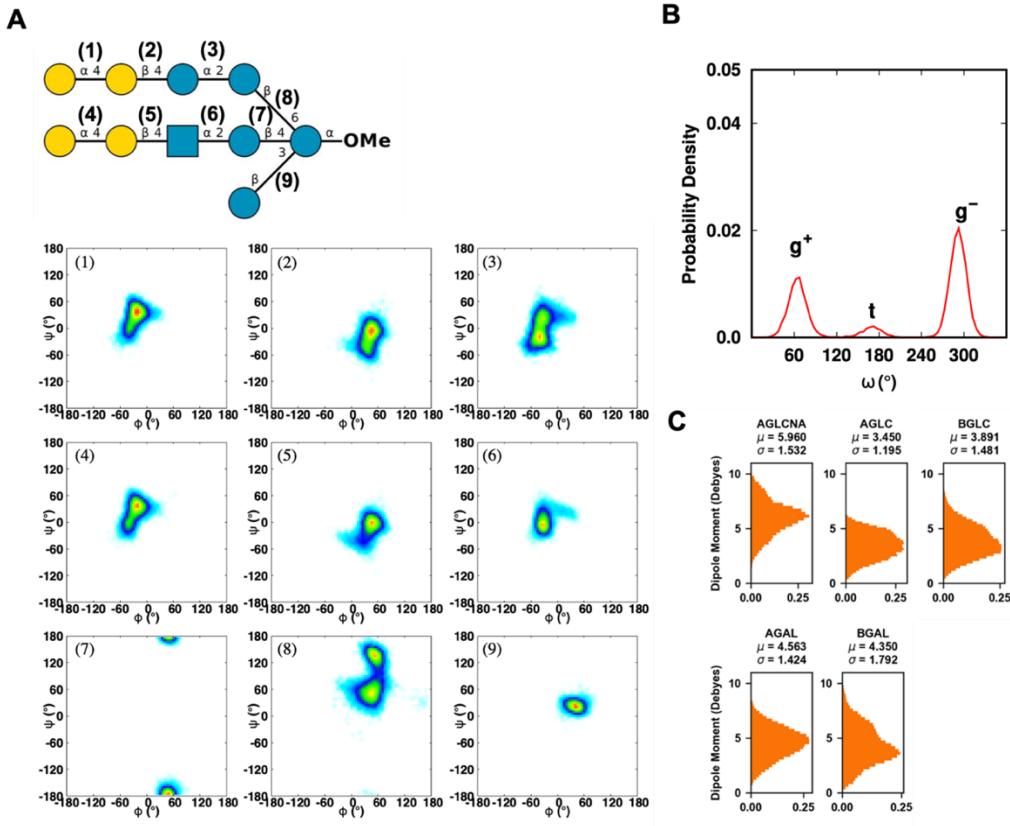
**Figure 10.** Distributions of dipole moments of membrane lipid head and tail groups over 200 ns simulations. The colored histograms correspond to the Drude FF (DOPC tail – blue; DOPE tail – cyan; DOPC head – red; DOPE head - orange), while the transparent gray histograms represent the additive FF calculated based on the 1 ns pre-equilibration trajectory. The histograms are normalized and the mean and standard deviation from Drude simulations is shown above each subplot.

***Escherichia coli* O176 O-antigen oligosaccharide:** We also ran a 200 ns simulation of the 10 repeating unit of O176 O-antigen oligosaccharide in solution with the Drude FF to calculate the  $\phi/\psi$  glycosidic torsions, where  $\phi = \text{O}5'-\text{C}1'-\text{On}-\text{Cn}$ ,  $\psi = \text{C}1'-\text{On}-\text{Cn}-\text{C}$  ( $n-1$ ), and  $n$  is the linkage position. The O176 O-antigen sequence is  $[\alpha\text{-D-Man}(1\rightarrow 2)\alpha\text{-D-Man}(1\rightarrow 2)\beta\text{-D-Man}(1\rightarrow 3)\alpha\text{-D-GlcNAc}(1\rightarrow 4)]_{10}$  (Figure 2A). The distribution of each glycosidic linkage is presented as a contour plot in Figure 11B. Except the  $\psi$  angle distribution of  $\alpha\text{-D-GlcNAc}(1\rightarrow 4)\alpha\text{-D-Man}$  (i.e., B-C in Figure 11A), the glycosidic torsion angles from the Drude simulation agree well with a previous study.<sup>121</sup> Figure 11C shows the dipole moment distribution of each carbohydrate type, revealing large variations of dipole moments with different average values even for  $\alpha\text{Man}$  and  $\beta\text{Man}$ . Note that for the carbohydrates, the charge on the monosaccharide units in the additive FF does not sum to be integer, so it is not possible to calculate well-defined dipole moments for a meaningful comparison.



**Figure 11.** (A) Schematic structure of the repeating units of the O-antigen polysaccharides from *E. coli* O176. (B) Glycosidic torsion angles  $\phi$  and  $\psi$  distributions.  $\phi = \text{O}5'\text{-C}1'\text{-On-C}n$ ,  $\psi = \text{C}1'\text{-On-C}n\text{-C}$  ( $n$  is the linkage position). The probability range is from white, 0, to blue, 0.1, to green, 0.3, to yellow, 0.7, and to red, 1. (C) Distributions of dipole moments of sugar monomers in the Drude FF.

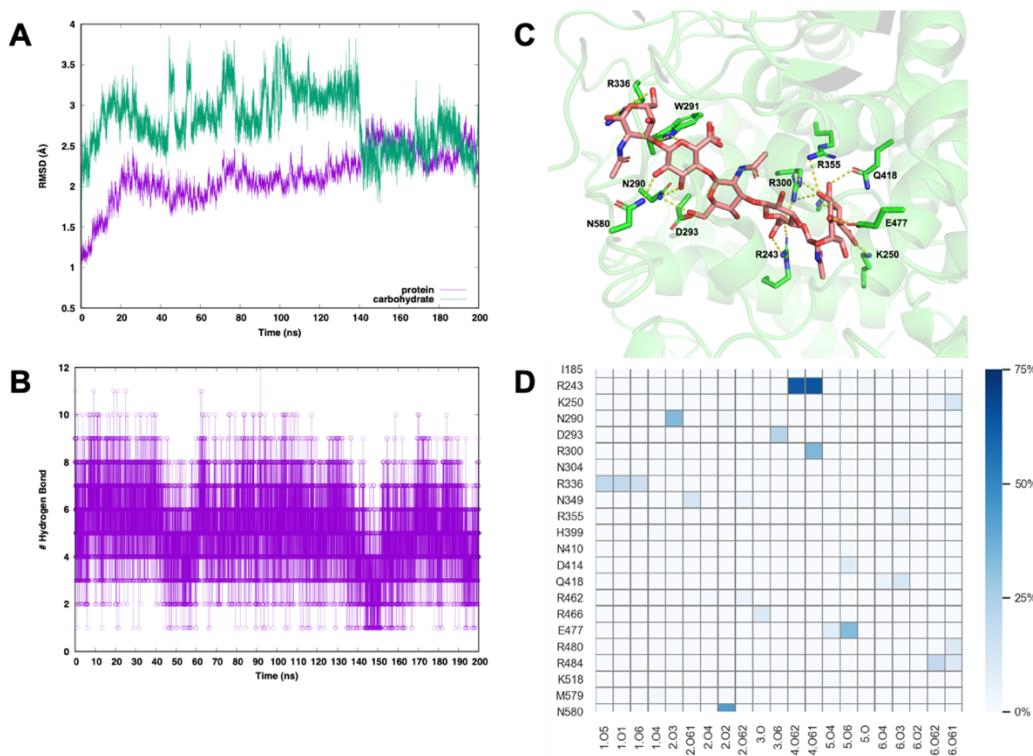
***Moraxella catarrhalis* serotype C oligosaccharide:** Previously, we used the CHARMM additive FF to perform 1  $\mu$ s simulation of symmetric bilayer and solution systems of three serotypes A, B, and C lipooligosaccharides of *M. catarrhalis*. Here, the Drude FF has been tested with a serotype C oligosaccharide solution system. The distributions of  $\phi/\psi$  glycosidic torsion angles over 200 ns are presented in **Figure 12A**, and these 2D distributions are consistent with the one in our previous study (**Figure S5**). In the case of  $\omega$  torsion angle of  $\beta$ 1-6 linkage at the first root residue, *gauche-gauche* (g-) conformations are dominant, almost identical to our previous study, but *trans-gauche* (t) conformations have a slightly bigger population. The ratio between rotamer conformations g-:g+:t from the Drude simulation is 58.4:34.7:6.9, which is consistent with the experimental result in that methyl  $\alpha$ -D-glucopyranose shows a rotamer conformational ratio of 57:38:5.<sup>129</sup> **Figure 12C** shows the dipole moment distribution of each carbohydrate type, again revealing large variations of dipole moments with different average values for different carbohydrate types.



**Figure 12.** (A) Two-dimensional distributions of  $\phi/\psi$  glycosidic torsion angles of serotype C oligosaccharide of *M. catarrhalis*. The glycosidic torsion angle definitions are defined in **Figure 11**. The probability range is from white, 0, to blue, 0.1, to green, 0.3, to yellow, 0.7, and to red, 1. (B) The torsion angle  $\omega$  is defined as O6-C6-C5-O5, where  $g^+$  denotes *gauche-trans*,  $g^-$  denotes *gauche-gauche*, and  $t$  denotes *trans-gauche*. (C) Distributions of dipole moments of sugar monomers.

**Hyaluronate lyase in complex with hyaluronan substrate:** *Streptococcus pneumoniae* hyaluronate lyase enzyme in complex with a hexasaccharide hyaluronan substrate (PDB ID: 1LOH) was simulated for 200 ns with the Drude FF. The overall complex structure was stable during the simulation (**Figure 13A**). The hexasaccharide substrate remains stably bound in the active site of lyase enzyme. To understand the stability of the substrate binding to the active site, hydrogen bond analysis was carried out. A time-series of the intermolecular hydrogen bonds between the lyase enzyme and the hexasaccharide is shown in **Figure 13B**. At least more than 3 hydrogen bonds exist in ~80% of the trajectory. Especially, R243 and R300 form these hydrogen bonds with the forth sugar (O61 of GlcA) in ~64% and ~35%, N580 with the second sugar (O2 of GlcA) in ~44%, N290 with the second sugar (O3 of GlcA) in ~34%, E477 with the fifth sugar (O6 of GlcNAc) in ~33%, and D293 with the third sugar (O6 of GlcNAc) in ~23% (**Figure 13C,D**). This result is well matched with the previous site-directed mutation study showing that R243 and N580 are involved in substrate binding, whereas N349, H399, and Y408 are involved in the catalytic process. The lower number of hydrogen bonds in the later residues are understandable because this enzyme is the Y408F mutant (inactive) form. Clearly, the hyaluronan substrate is stably bound due to the charge complementarity between substrate negative charges and the multiple basic residues lining the cleft surface. Such electrostatic interactions in the binding cleft are well

characterized in the Drude simulation, as the dipole moment distributions of amino acids around the substrate within 4.5 Å (i.e., substrate-contact residues) are wider than the non-contact protein residues (**Figure S6**).



**Figure 13.** (A) Backbone RMSD of hyaluronate lyase (purple) and hyaluronan substrate (green). (B) The number of hydrogen bonds between a hyaluronan substrate and its enzyme (C) A snapshot of complex structure at 30 ns (D) Hydrogen bonding frequencies between a hyaluronan substrate and its enzyme. The y-axis labels the interacting amino acid residues, and the x-axis labels the interacting atoms of the substrate. The color bar represents the frequency of hydrogen bonding observed in the simulation trajectory.

## CONCLUSIONS

In this work, we present *Drude Prepper*, a new utility in CHARMM-GUI, a web-based tool that will facilitate the use of the Drude FF by the scientific community for studies of a multitude of biomolecular systems. The Drude FF for proteins was first released in 2013 and only a few research groups have been using it as compared to the number of studies using the additive FF. CHARMM-GUI, by providing a simplified platform to diverse users, has played an important role in facilitating the use of the CHARMM additive FF, and we anticipate that the *Drude Prepper* will give the same access to the CHARMM Drude FF. Here, we have demonstrated the best practices to use *Drude Prepper* by applying it to various systems including proteins, DNA, RNA, lipids, carbohydrates, atomic ions, and their combinations. *Drude Prepper* takes the advantage of using the same residue nomenclature as in the CHARMM additive FF to allow for the additive PSF to be readily converted to its Drude counterpart. This avoids the major effort that would be required to build each simulation system from scratch. The inputs generated by *Drude Prepper* are highly consistent across the four simulation packages (CHARMM, NAMD, OpenMM, and GROMACS), and the energies calculated with the 4 simulation programs with the generated data structures

and inputs are in agreement. The inputs for GROMACS are currently available for protein only systems, and there are ongoing efforts to extend the inputs to other classes of molecules. While the analysis of simulations performed in this study demonstrate the applicability of the Drude FF, they also highlight the ability of the Drude FF to provide potentially important information based on dipole moment variations occurring in the simulations.

**Acknowledgements:** Financial support from the NIH GM133754 (to JAL), GM131710 (to ADM Jr.) and GM138472 (to WI). Computational support from the University of Maryland Computer-Aided Drug Design Center and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

**Conflict of Interest:** ADM Jr. is co-founder and CSO of SilcsBio LLC.

## REFERENCES

1. Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.-t.; Rémy, H.-W.; Stark, H.; Wiesmann, C., Cryo-EM in drug discovery: achievements, limitations and prospects. *Nature Reviews Drug Discovery* **2018**, *17* (7), 471-492.
2. Penfold, J.; Wagner, N. J., Editorial overview: Recent applications of x-ray and neutron scattering techniques in colloid and interfacial science, characterised by increasing diversity and complexity. *Current Opinion in Colloid & Interface Science* **2019**, *42*.
3. Loura, L. M.; Prieto, M., FRET in Membrane Biophysics: An Overview. *Front Physiol* **2011**, *2*, 82.
4. Sotomayor, M.; Schulten, K., Single-molecule experiments in vitro and in silico. *Science* **2007**, *316* (5828), 1144-8.
5. van der Kamp, M. W.; Mulholland, A. J., Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52* (16), 2708-28.
6. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., Machine learning for molecular and materials science. *Nature* **2018**, *559* (7715), 547-555.
7. Glaser, J.; Nguyen, T. D.; Anderson, J. A.; Lui, P.; Spiga, F.; Millan, J. A.; Morse, D. C.; Glotzer, S. C., Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **2015**, *192*, 97-107.
8. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C.; Anton, A Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91-97.
9. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435-447.
10. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F., The GROMOS Biomolecular Simulation Program Package. *Journal of Physical Chemistry A* **1999**, *103*, 3596-3607.
11. Brooks, B. R.; Brooks, C. L., 3rd; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, *30* (10), 1545-614.
12. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26* (16), 1668-1688.

13. Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A., Implementation of Accelerated Molecular Dynamics in NAMD. *Computational science & discovery* **2011**, *4* (1).
14. Jung, J.; Mori, T.; Kobayashi, C.; Matsunaga, Y.; Yoda, T.; Feig, M.; Sugita, Y., GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *WIREs Computational Molecular Science* **2015**, *5* (4), 310-323.
15. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W., Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018**, *14* (10), 5273-5289.
16. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659.
17. Dequidt, A.; Devémy, J.; Pádua, A. A. H., Thermalized Drude Oscillators with the LAMMPS Molecular Dynamics Simulator. *Journal of Chemical Information and Modeling* **2016**, *56* (1), 260-268.
18. Bussi, G.; Tribello, G. A., Analyzing and Biasing Simulations with PLUMED. *Methods Mol Biol* **2019**, *2022*, 529-578.
19. Hospital, A.; Andrio, P.; Fenollosa, C.; Cicin-Sain, D.; Orozco, M.; Gelpí, J. L., MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics* **2012**, *28* (9), 1278-1279.
20. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859-1865.
21. Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Roux, B.; Im, W., CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues. *Advances in protein chemistry and structural biology* **2014**, *96*, 235-65.
22. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., 3rd; MacKerell, A. D., Jr.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **2016**, *12* (1), 405-13.
23. Lee, J.; Hitzenberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W., CHARMM-GUI supports the Amber force fields. *J Chem Phys* **2020**, *153* (3), 035103.
24. Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W., CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, *29* (4), 320-331.

25. Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S. J.; Patel, D. S.; Beaven, A. H.; Lee, K. I.; Rui, H.; Park, S.; Lee, H. S.; Roux, B.; MacKerell, A. D., Jr.; Klauda, J. B.; Qi, Y.; Im, W., CHARMM-GUI 10 years for biomolecular modeling and simulation. *J Comput Chem* **2017**, *38* (15), 1114-1124.
26. Choi, Y. K.; Park, S.-J.; Park, S.; Kim, S.; Kern, N. R.; Lee, J.; Im, W., CHARMM-GUI Polymer Builder for Modeling and Simulation of Synthetic Polymers. *Journal of Chemical Theory and Computation* **2021**.
27. Foley, B. L.; Tessier, M. B.; Woods, R. J., Carbohydrate force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2* (4), 652-697.
28. Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A., ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J Chem Theory Comput* **2014**, *10* (10), 4515-4534.
29. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T., Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J Chem Theory Comput* **2016**, *12* (8), 3926-47.
30. Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: a refined force field for DNA simulations. *Nat Methods* **2016**, *13* (1), 55-8.
31. Aytenfisu, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H., Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J Chem Theory Comput* **2017**, *13* (2), 900-915.
32. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W., 2nd; MacKerell, A. D., Jr., CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J. Chem. Theory Comp.* **2011**, *7* (10), 3162-3180.
33. Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D., Jr., Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J Comput Chem* **2011**, *32* (9), 1929-1943.
34. Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr., Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comp.* **2012**, *8* (1), 348-362.
35. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell Jr., A. D., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory and Comp.* **2012**, *8*, 3257-3273.
36. Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D., Jr.; Pastor, R. W.; J.B., K., CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature. *Biophysical Journal* **2014**, *107*, 134-145.

37. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71–73.
38. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* **2016**, *12* (1), 281-96.
39. Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L., Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *Journal of Chemical Theory and Computation* **2015**, *11* (7), 3499-3509.
40. Dodd, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L., LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research* **2017**, *45* (W1), W331-W336.
41. Soares, T. A.; Hunenberger, P. H.; Kastenholz, M. A.; Kraeutler, V.; Lenz, T.; Lins, R.; Oostenbrink, C.; van Gunsteren, W., An improved nucleic acid parameter set for the GROMOS force field. *J. Comp. Chem.* **2005**, *26*, 725-737.
42. Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F., Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34*, 273-284.
43. Reif, M. M.; Winger, M.; Oostenbrink, C., Testing of the GROMOS Force-Field Parameter Set 54A8: Structural Properties of Electrolyte Solutions, Lipid Bilayers, and Proteins. *Journal of Chemical Theory and Computation* **2013**, *9* (2), 1247-1264.
44. Halgren, T. A.; Damm, W., Polarizable force fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236-242.
45. Rick, S. W.; Stuart, S. J., Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. *Rev. Comp. Chem.* **2002**, *18*, 89-146.
46. Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J., Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.
47. Zhu, X.; Lopes, P. E.; MacKerell, A. D., Jr., Recent Developments and Applications of the CHARMM force fields. *Wiley Interdiscip Rev Comput Mol Sci* **2012**, *2* (1), 167-185.
48. Levine, Z. A.; Shea, J. E., Simulations of disordered proteins and systems with conformational heterogeneity. *Curr Opin Struct Biol* **2017**, *43*, 95-103.
49. Huang, J.; MacKerell, A. D., Jr., Force field development and simulations of intrinsically disordered proteins. *Curr Opin Struct Biol* **2018**, *48*, 40-48.
50. Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J., Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of BPTI in Liquid Water. *J. Chem. Theory Comp.* **2005**, *1*, 169-180.

51. Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B., Polarizable Simulations with Second order Interaction Model - force field and software for fast polarizable calculations: Parameters for small model systems and free energy calculations. *J Chem Theory Comput* **2009**, *5* (11), 2935-2943.
52. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J Chem Theory Comput.* **2013**, *9*, 4046-4064.
53. Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P., AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation* **2018**, *14*, 2084–2108.
54. Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III, CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem* **2004**, *25*, 1504-1514.
55. Zhong, Y.; Patel, S., Binding structures of tri-N-acetyl-beta-glucosamine in hen egg white lysozyme using molecular dynamics with a polarizable force field. *J Comput Chem* **2013**, *34* (3), 163-74.
56. Oostenbrink, C.; van Gunsteren, W. F., Free energies of ligand binding for structurally diverse compounds. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (19), 6750-6754.
57. Patel, S.; Davis, J. E.; Bauer, B. A., Exploring ion permeation energetics in gramicidin A using polarizable charge equilibration force fields. *J. Am. Chem. Soc.* **2009**, *131* (39), 13890-1.
58. Soniat, M.; Hartman, L.; Rick, S. W., Charge Transfer Models of Zinc and Magnesium in Water. *J Chem Theory Comput* **2015**, *11* (4), 1658-67.
59. Kunz, A. P.; van Gunsteren, W. F., Development of a nonlinear classical polarization model for liquid water and aqueous solutions: COS/D. *J. Phys. Chem. A* **2009**, *113* (43), 11570-11579.
60. Rick, S. W., A polarizable, charge transfer model of water using the drude oscillator. *J Comput Chem* **2016**, *37* (22), 2060-6.
61. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., Jr., An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem Rev* **2016**, *116* (9), 4983-5013.
62. Ren, P.; Wu, C.; Ponder, J. W., Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. *J Chem Theory Comput* **2011**, *7* (10), 3143-3161.
63. Zhang, C.; Bell, D.; Harger, M.; Ren, P., Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *J Chem Theory Comput* **2017**, *13* (2), 666-678.
64. Savelyev, A.; MacKerell, A. D., Jr., All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J Comput Chem* **2014**, *35*, 1219-1239.
65. Savelyev, A.; MacKerell, A. D., Jr., Balancing the interactions of ions, water and DNA in the Drude polarizable force field. *J Phys Chem B* **2014**, *118*, 6742-675.

66. Chowdhary, J.; Harder, E.; Lopes, P. E.; Huang, L.; MacKerell, A. D., Jr.; Roux, B., A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. *J. Phys. Chem. B* **2013**, *117*, 9142–9160.
67. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr., Polarizable Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comp.* **2013**, *9*, 5430–5449.
68. He, X.; Lopes, P. E. M.; MacKerell, A. D., Jr., Polarizable Empirical Force Field for Acyclic Poly-Alcohols Based on the Classical Drude Oscillator. *Biopolymers* **2013**, *99*, 724-738.
69. Patel, D. S.; He, X.; MacKerell, A. D., Jr., Polarizable empirical force field for hexopyranose monosaccharides based on the classical Drude oscillator. *J Phys Chem B* **2015**, *119* (3), 637-652.
70. Aytenfisu, A. H.; Yang, M.; MacKerell, A. D., Jr., CHARMM Drude Polarizable Force Field for Glycosidic Linkages Involving Pyranoses and Furanoses. *J Chem Theory Comput* **2018**, ASAP Article.
71. Yang, M.; Aytenfisu, A. H.; MacKerell, A. D., Jr., Proper Balance of Solvent-Solute and Solute-Solute Interactions in the Treatment of the Diffusion of Glucose using the Drude Polarizable Force Field. *Carbohydrate Research* **2018**, *457*, 41-50.
72. Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B., Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J Chem Theory Comput* **2010**, *6* (3), 774-786.
73. Luo, Y.; Jiang, W.; Yu, H.; MacKerell, A. D., Jr.; Roux, B., Simulation study of ion pairing in concentrated aqueous salt solutions with a polarizable force field. *Faraday Discussions* **2013**, *160*, 135 - 149.
74. Lin, F.-Y.; MacKerell, A. D., Polarizable Empirical Force Field for Halogen-Containing Compounds Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2018**, *14* (2), 1083-1098.
75. Heid, E.; Fleck, M.; Chatterjee, P.; Schröder, C.; MacKerell, A. D., Toward Prediction of Electrostatic Parameters for Force Fields That Explicitly Treat Electronic Polarization. *Journal of Chemical Theory and Computation* **2019**, *15* (4), 2460-2469.
76. Harder, E.; Anisimov, V. M.; Whitfield, T.; MacKerell, A. D., Jr.; Roux, B., Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. *J. Phys. Chem. B* **2008**, *112*, 3509-3521.
77. Baker, C. M.; Lopes, P. E.; Zhu, X.; Roux, B.; MacKerell, A. D., Jr., Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. *J Chem Theory Comput* **2010**, *6* (4), 1181-1198.
78. Lemkul, J. A.; MacKerell, A. D., Jr., Balancing the Interactions of Mg<sup>2+</sup> in Aqueous Solution and with Nucleic Acid Moieties For a Polarizable Force Field Based on the Classical Drude Oscillator Model. *J Phys Chem B* **2016**, *120* (44), 11436-11448.

79. Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell, A. D., Jr., Molecular Dynamics Simulations of Explicitly Solvated Drude Polarizable Systems on GPUs: Implementation, Validation, and Benchmark. *J. Comp. Chem.* **2018**, *39*, 1682–1689.
80. Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; McCammon, J. A., Accelerated Molecular Dynamics Simulations with the AMOEBA Polarizable Force Field on Graphics Processing Units. *J. Chem. Theory Comput.* **2013**, *9* (11), 4684-4691.
81. Lagardere, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P., Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science* **2018**, *9* (4), 956-972.
82. Lamoureux, G.; Roux, B., Modelling Induced Polarizability with Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 5185-5197.
83. Jiang, W.; Hardy, D. J.; Phillips, J. C.; Mackerell, A. D., Jr.; Schulten, K.; Roux, B., High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2* (2), 87-92.
84. Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell, A. D., Jr., Molecular dynamics simulations using the drude polarizable force field on GPUs with OpenMM: Implementation, validation, and benchmarks. *J. Comput. Chem.* **2018**, *39* (21), 1682-1689.
85. Lemkul, J. A.; Roux, B.; van der Spoel, D.; MacKerell, A. D., Jr., Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model. *J. Comput. Chem.* **2015**, *36* (19), 1473-9.
86. Jo, S.; Song, K. C.; Desaire, H.; MacKerell Jr, A. D.; Im, W., Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *Journal of Computational Chemistry* **2011**, *32* (14), 3135-3141.
87. Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W., Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinformatics* **2017**, *33* (19), 3051-3057.
88. Pandey, P.; Aytenfisu, A. H.; MacKerell, A. D.; Mallajosyula, S. S., Drude Polarizable Force Field Parametrization of Carboxylate and N-Acetyl Amine Carbohydrate Derivatives. *Journal of Chemical Theory and Computation* **2019**, *15* (9), 4982-5000.
89. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell Jr, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2010**, *31* (4), 671-690.
90. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926-935.
91. Durell, S. R.; Brooks, B. R.; Bennaim, A., Solvent-induced forces between 2 hydrophilic groups. *J. Phys. Chem.* **1994**, *98* (8), 2198-2202.

92. Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical Physics Letters* **2006**, *418* (1-3), 245-249.
93. Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D., Six-site polarizable model of water based on the classical Drude oscillator. *The Journal of Chemical Physics* **2013**, *138* (3), 034508.
94. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577-8593.
95. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327-341.
96. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4* (2), 187-217.
97. Steinbach, P. J.; Brooks, B. R., New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry* **1994**, *15* (7), 667-683.
98. Allen, M. P.; Tildesley, D. J., *Computer simulations of liquids*. Clarendon Press: Oxford, England, 1987.
99. Leonard, A. N.; Simmonett, A. C.; Pickard, F. C.; Huang, J.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W., Comparison of Additive and Polarizable Models with Explicit Treatment of Long-Range Lennard-Jones Interactions Using Alkane Simulations. *Journal of Chemical Theory and Computation* **2018**, *14* (2), 948-958.
100. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81* (1), 511-519.
101. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31* (3), 1695-1697.
102. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994**, *101* (5), 4177-4189.
103. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995**, *103* (11), 4613-4621.
104. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
105. Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (1), 116-122.
106. Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J., Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology* **1987**, *194* (3), 531-544.

107. Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O., Zif268 protein&#x2013;DNA complex refined at 1.6&#xe5;: a model system for understanding zinc finger&#x2013;DNA interactions. *Structure* **1996**, *4* (10), 1171-1180.
108. Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Heaphy, S.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A.; Valerio, R., Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86* (18), 6925-6929.
109. Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **1992**, *257* (5066), 76.
110. Pham, V. V.; Salguero, C.; Khan, S. N.; Meagher, J. L.; Brown, W. C.; Humbert, N.; de Rocquigny, H.; Smith, J. L.; D'Souza, V. M., HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nature Communications* **2018**, *9* (1), 4266.
111. Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonović, M.; Subramaniam, S., Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nature Structural & Molecular Biology* **2019**, *26* (8), 679-685.
112. Krishnamurthy, H.; Gouaux, E., X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **2012**, *481* (7382), 469-474.
113. Cao, Y.; Jin, X.; Levin, E. J.; Huang, H.; Zong, Y.; Quick, M.; Weng, J.; Pan, Y.; Love, J.; Punta, M.; Rost, B.; Hendrickson, W. A.; Javitch, J. A.; Rajashankar, K. R.; Zhou, M., Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* **2011**, *473* (7345), 50-54.
114. Ren, Z.; Lee, J.; Moosa, M. M.; Nian, Y.; Hu, L.; Xu, Z.; McCoy, J. G.; Ferreon, A. C. M.; Im, W.; Zhou, M., Structure of an EIIC sugar transporter trapped in an inward-facing conformation. *Proceedings of the National Academy of Sciences* **2018**, *115* (23), 5962.
115. Lee, J.; Ren, Z.; Zhou, M.; Im, W., Molecular Simulation and Biochemical Studies Support an Elevator-type Transport Mechanism in&#xa0;EIIC. *Biophysical Journal* **2017**, *112* (11), 2249-2252.
116. McCoy, Jason G.; Ren, Z.; Stanevich, V.; Lee, J.; Mitra, S.; Levin, Elena J.; Poget, S.; Quick, M.; Im, W.; Zhou, M., The Structure of a Sugar Transporter of the Glucose EIIC Superfamily Provides Insight into the Elevator Mechanism of Membrane Transport. *Structure* **2016**, *24* (6), 956-964.
117. Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L., OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **2011**, *40* (D1), D370-D376.
118. Jo, S.; Cheng, X.; Islam, S. M.; Huang, L.; Rui, H.; Zhu, A.; Lee, H. S.; Qi, Y.; Han, W.; Vanommeslaeghe, K.; MacKerell, A. D.; Roux, B.; Im, W., Chapter Eight - CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues. In *Advances in protein chemistry and structural biology*, Karabencheva-Christova, T., Ed. Academic Press: 2014; Vol. 96, pp 235-265.

119. Im, W.; Khalid, S., Molecular Simulations of Gram-Negative Bacterial Membranes Come of Age. *Annual Review of Physical Chemistry* **2020**, 71 (1), 171-188.
120. Iguchi, A.; Iyoda, S.; Kikuchi, T.; Ogura, Y.; Katsura, K.; Ohnishi, M.; Hayashi, T.; Thomson, N. R., A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Research* **2014**, 22 (1), 101-107.
121. Patel, D. S.; Blasco, P.; Widmalm, G.; Im, W., *Escherichia coli* O176 LPS structure and dynamics: A NMR spectroscopy and MD simulation study. *Current Research in Structural Biology* **2020**, 2, 79-88.
122. Hays, J. P., *Moraxella catarrhalis*: A mini review. *Journal of Pediatric Infectious Diseases* **2009**, 4, 211-220.
123. Gao, Y.; Lee, J.; Widmalm, G.; Im, W., Preferred conformations of lipooligosaccharides and oligosaccharides of *Moraxella catarrhalis*. *Glycobiology* **2019**, 30 (2), 86-94.
124. Li, S.; Kelly, S. J.; Lamani, E.; Ferraroni, M.; Jedrzejas, M. J., Structural basis of hyaluronan degradation by *Streptococcus pneumoniae* hyaluronate lyase. *The EMBO Journal* **2000**, 19 (6), 1228-1240.
125. Jedrzejas, M. J.; Mello, L. V.; de Groot, B. L.; Li, S., Mechanism of Hyaluronan Degradation by *Streptococcus pneumoniae* Hyaluronate Lyase: STRUCTURES OF COMPLEXES WITH THE SUBSTRATE. *Journal of Biological Chemistry* **2002**, 277 (31), 28287-28297.
126. Lemkul, J. A.; Huang, J.; MacKerell, A. D., Induced Dipole–Dipole Interactions Influence the Unfolding Pathways of Wild-Type and Mutant Amyloid  $\beta$ -Peptides. *The Journal of Physical Chemistry B* **2015**, 119 (51), 15574-15582.
127. Lemkul, J. A., Same fold, different properties: polarizable molecular dynamics simulations of telomeric and TERRA G-quadruplexes. *Nucleic Acids Research* **2020**, 48 (2), 561-575.
128. Jo, S.; Im, W., Glycan fragment database: a database of PDB-based glycan 3D structures. *Nucleic Acids Research* **2013**, 41 (D1), D470-D474.
129. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences* **2001**, 98 (19), 10541.

## SUPPORTING INFORMATION

### CHARMM-GUI Drude Prepper for Molecular Dynamics Simulation Using the Classical Drude Polarizable Force Field

Abhishek A. Kognole<sup>1†</sup>, Jumin Lee<sup>2†</sup>, Sang-Jun Park<sup>2†</sup>, Sunhwan Jo<sup>3†</sup>, Payal Chatterjee<sup>1</sup>, Justin A. Lemkul<sup>4</sup>, Jing Huang<sup>5</sup>, Alexander D. MacKerell, Jr.<sup>1\*</sup>, and Wonpil Im<sup>2\*</sup>

<sup>1</sup>Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA

<sup>2</sup>Departments of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University, PA, USA

<sup>3</sup>Leadership Computing Facility, Argonne National Laboratory, 9700 Cass Ave, Argonne, IL, USA

<sup>4</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

<sup>5</sup>Key Laboratory of Structural Biology of Zhejiang Province, Institute of Biology, Westlake Institute for Advanced Study, School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou, Zhejiang, 310024, China

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: alex@outerbanks.umaryland.edu and wonpil@lehigh.edu

**Table S1.** Summary of current CHARMM Drude FF topology and parameter files.

Toppar file name	Types of molecules
master_protein_2019g	Water, ions, amino acids,
nucleic_acid_2017c	Nucleobases, nucleosides, methylated nucleobases
carbohydrate_2019b	Polyalcohols, hexapyranoses, furanoses, Carboxylate and N-Acetyl Amine Carbohydrate Derivatives
lipid_2017c	Saturated and Unsaturated Zwitterionic Lipids
model_2019g	Alkanes, alcohols, amides, aromatics, heterocycles, halogen-containing compounds, molecular ions
d_aminoacids_2019g	D-amino acids

**Table S2.** Descriptions of input and stream files in *Drude Prepper*.

Filename	Description
uploaded_files (directory)	Directory that stores user-uploaded additive FF PSF and PDB/CRD files
toppar (directory)	Directory with the C36 topology and parameter files
toppar.str	CHARMM stream file that reads the C36 top/par files
toppar_drude (directory)	Directory with the Drude Polarizable FF topology and parameter files
toppar_drude.str	CHARMM stream file that read the Drude Polarizable FF top/par files
step1_reader.inp	CHARMM input to read, parse, and learn the topology of input system and prepare separate coordinate files and informative files for step2_drude.inp
step1_reader_glycan.str	CHARMM stream file that contain carbohydrate sequence (if exists)
step2_drude.inp	CHARMM input to read output from step1_reader and build the system with the Drude polarizable FF and setup parameters for PBCs.
step3_input.inp	CHARMM input to perform short minimization and generate restraint files for user-selected simulation package
step4_equilibration.inp	CHARMM input to perform equilibration
step5_production.inp	CHARMM input to perform production
crystal_image.str	Crystal transformation file for PBCs
checkfft.py	Python script to calculate the number of grid points for PME FFT

**Table S3.** Definitions of backbone and sidechain heavy atoms for the harmonic restraints during minimization, equilibration and analysis purposes.

Type of system	Atoms for backbone (BB)	Atoms of sidechain (SC)
Protein	C,O,N,CA,	protein and not (BB or hydrogen)
DNA/RNA	O5',C5',C4',C3',O3',P,O1P,O2P	DNA/RNA and not (BB or hydrogen)
Carbohydrates	C+, O5 (bonded to C1), O6 (bonded to C2)	carbohydrate and not (BB or hydrogen)
Lipids	-	

**Table S4.** CHARMM atom selection definitions for dipole moment calculations.

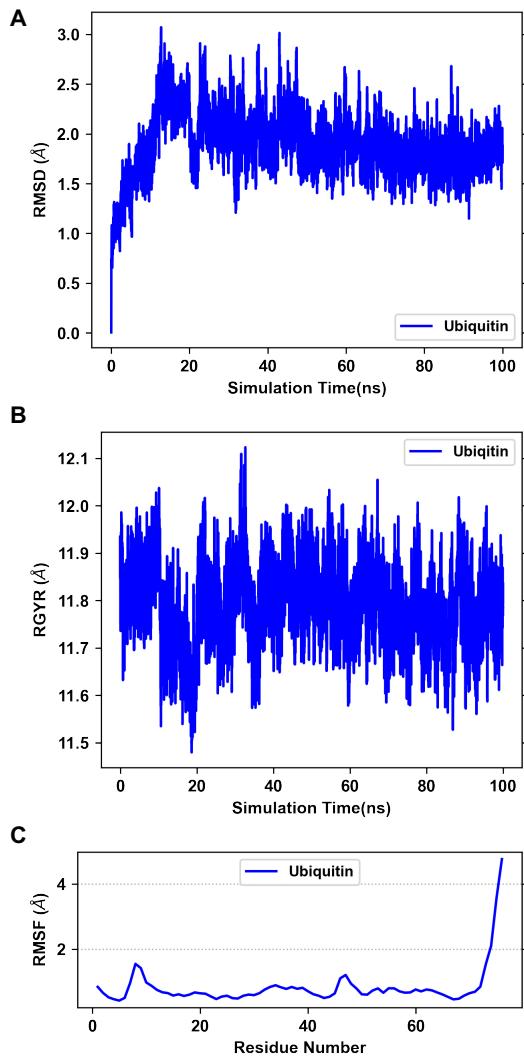
Atom group of interest	Atom selections to parse	Example command to calculate
peptide bond (between res1 and res2)	define pb_res1 sele type C .or. type DC .or. type O .or. type DO .or. type LPOA .or. type LPOB end  define pb_res2 sele type N .or. type DN .or. type HN .or. type CA .or. type DCA .or. type HA* end	coor dipo sele segid PROA .and. ((resid @res1 .and. pb_res1) .or. (resid @res2 .and. pb_res2)) end
amino acid side chain	define not_sc sele (type C .or. type DC .or. type O .or. type DO .or. type LPOA .or. type LPOB) .or. (type N .or. type DN .or. type HN .or. type CA .or. type DCA .or. type HA*) .or. - (resname GLY) .or. (type HT* .or. type OT* .or. type DOT* .or. type LPT*) end	coor dipo sele segid PROA .and. resid @res1 .and. .not. not_sc end
a nucleobase (A/T/U/G/C)	define not_base sele (type P .or. type DP .or. type O1P .or. type DO1P .or. type O2P .or. type DO2P .or. type O5' .or. type DO5' .or. type C5' .or. type DC5' .or. type H5' .or. type H5" .or. type C4' .or. type DC4' .or. type H4' .or. type O4' .or. type DO4' .or. type LPRA .or. type LPRB .or. type LPX .or. type C1' .or. type DC1' .or. type H1' .or. type C2' .or. type DC2' .or. type H2" .or. type O2' .or. type DO2' .or. type H2' .or. type LPHA .or. type LPHB .or. type C3' .or. type DC3' .or. type H3' .or. type O3' .or. type DO3' .or. type H3T .or. type H5T .or. type LP5A .or. type LP5B .or. type LP3A .or. type LP3B) end	coor dipo sele segid RNAA .and. resid @res1 .and. .not. not_base end
carbohydrate/oligosaccharide	define SEL1 sele .bonded. resid @resid .and. .not. resid @resid .and. type C* end  define SEL2 sele .bonded. SEL1 .and. resid @resid end  define SEL3 sele .bonded. SEL2 .and. type LPDB* end  define SEL4 sele .bonded. resid @resid .and. .not. resid @resid .and. type O* end  define SEL5 sele .bonded. SEL4 .and. ( type DO* .or. type LPDB* ) end	coor dipo sele segid CARB .and. ( resid @resid .or. SEL4 .or. SEL5) .and. .not. SEL3 show end
lipid molecule	define HEAD sele .bygroup. type C1* end define TAIL sele .not. HEAD end	coor dipo sele segid MEMB .and. resid @resid .and. HEAD end coor dipo sele segid MEMB .and. resid @resid .and. TAIL end
model compound	define not_model sele none end	coor dipo sele segid HETA .and. resname @residue .and. resid @res1 end

**Table S5.** Energy component analysis.

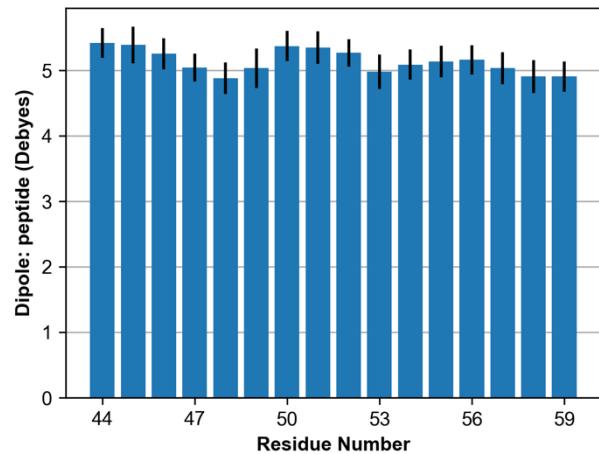
	Ubiquitin			
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	30413.983	30415.756	30307.062	30415.590
ANGL + UREY	1192.620	1192.266	1192.619	1192.260
DIHE + CMAP	705.546	705.567	705.546	705.570
IMPR	11.874	11.875	11.874	11.876
BONDED TOTAL	32324.023	32325.464	32217.101	32325.300
VDW	16195.158	16196.134	16198.157	16198.920
ELEC	-137385.924	-137382.775	-137275.575	-137271.160
NON BONDED TOTAL	-121190.766	-121186.641	-121077.418	-121072.240
TOTAL	-88866.743	-88861.170	-88860.317	-88746.940
% Change in TOTAL	0.000%	0.006%	0.007%	0.135%
	TF-DNA			
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	961.672	961.104	962.561	NA
ANGL + UREY	1924.009	1924.036	1924.284	NA
DIHE + CMAP	1240.353	1240.385	1240.392	NA
IMPR	11.113	11.112	11.112	NA
BONDED TOTAL	4137.147	4136.636	4138.348	NA
VDW	35289.733	35289.643	35294.310	NA
ELEC	-129139.215	-129135.434	-129114.402	NA
NON BONDED TOTAL	-93849.482	-93845.791	-93820.092	NA
TOTAL	-89712.334	-89709.155	-89681.744	NA
% Change in TOTAL	0.000%	0.004%	0.034%	NA
	HIV-1 TAR-TAT RBD			
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	42643.804	42640.577	39865.454	NA
ANGL + UREY	859.036	859.109	859.036	NA
DIHE + CMAP	1062.892	1062.850	1062.892	NA
IMPR	2.851	2.864	2.851	NA
BONDED TOTAL	44568.582	44565.400	41790.232	NA
VDW	24744.850	24744.986	24747.752	NA
ELEC	-191385.703	-191369.068	-188511.145	NA
NON BONDED TOTAL	-166640.853	-166624.082	-163763.393	NA
TOTAL	-122072.271	-122058.682	-121973.161	NA
% Change in TOTAL	0.000%	0.011%	0.081%	NA
	Cas9-sgRNA-DNA			
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	291070.682	291102.548	286885.826	NA
ANGL + UREY	25529.866	25530.400	25529.890	NA

LeuT in POPC				
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	220414.221	220432.129	217963.404	NA
ANGL + UREY	47379.398	47378.064	47379.394	NA
DIHE + CMAP	103606.202	103606.444	103606.204	NA
IMPR	2128.175	2128.432	2127.197	NA
BONDED TOTAL	373527.996	373545.068	371076.199	NA
VDW	148639.003	148697.431	148690.569	NA
ELEC	-1100236.360	-1100153.591	-1097714.948	NA
NON BONDED TOTAL	-951597.357	-951456.160	-949024.379	NA
TOTAL	-578069.359	-577911.092	-577948.179	NA
% Change in TOTAL	0.000%	0.027%	0.021%	NA
Escherichia coli O176 O-antigen oligosaccharide				
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	155588.274	155607.751	155583.492	NA
ANGL + UREY	311.454	311.529	311.453	NA
DIHE + CMAP	1554.262	1554.277	1554.262	NA
IMPR	0.512	0.512	0.512	NA
BONDED TOTAL	157454.502	157474.069	157449.719	NA
VDW	112792.473	112793.192	112804.988	NA
ELEC	-737584.588	-736253.941	-736237.896	NA
NON BONDED TOTAL	-624792.115	-623460.749	-623432.908	NA
TOTAL	-467337.613	-465986.681	-465983.190	NA
% Change in TOTAL	0.000%	0.289%	0.290%	NA
Moraxella catarrhalis serotype C oligosaccharide				
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	10055.599	10055.914	10054.493	NA
ANGL + UREY	53.564	53.554	53.564	NA
DIHE + CMAP	384.478	384.490	384.478	NA
IMPR	0.368	0.361	0.368	NA

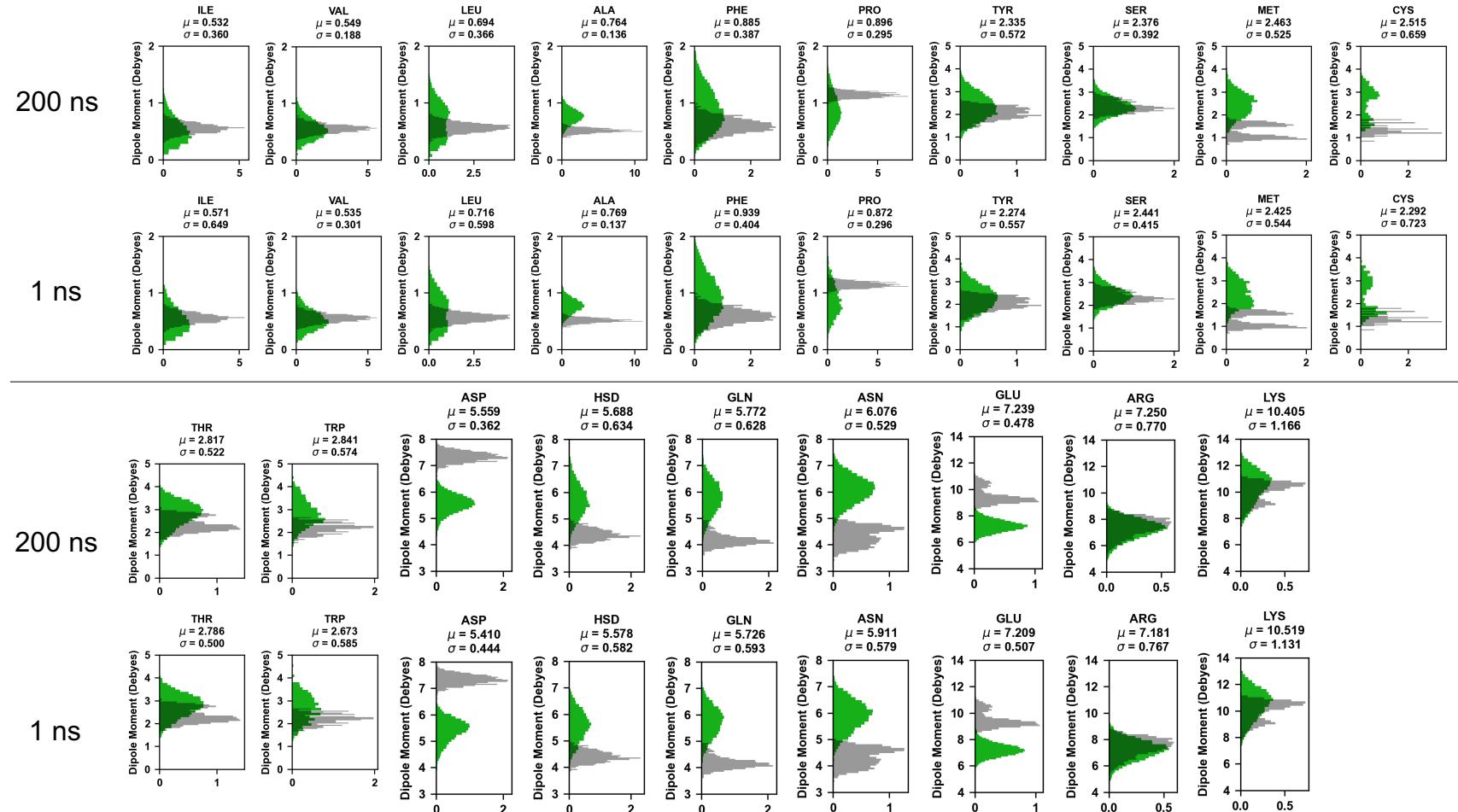
BONDED TOTAL	10494.009	10494.319	10492.903	NA
VDW	7151.726	7151.804	7152.527	NA
ELEC	-47452.734	-47367.389	-47367.211	NA
NON BONDED TOTAL	-40301.007	-40215.585	-40214.684	NA
TOTAL	-29806.998	-29721.267	-29721.781	NA
% Change in TOTAL	0.000%	0.288%	0.286%	NA
	<b>Hyaluronate lyase</b>			
ENERGY COMPONENT	CHARMM c44b2	NAMD 2.14	OpenMM 7.5.1	GROMACS
BOND	149940.563	149954.990	149078.280	NA
ANGL + UREY	11898.806	11897.481	11898.802	NA
DIHE + CMAP	6889.268	6889.242	6889.268	NA
IMPR	178.000	177.977	178.000	NA
BONDED TOTAL	168906.637	168919.690	168044.351	NA
VDW	94478.099	94477.996	94520.446	NA
ELEC	-705000.320	-703661.081	-702794.657	NA
NON BONDED TOTAL	-610522.221	-609183.084	-608274.211	NA
TOTAL	-441615.584	-440263.395	-440229.860	NA
% Change in TOTAL	0.000%	0.306%	0.314%	NA



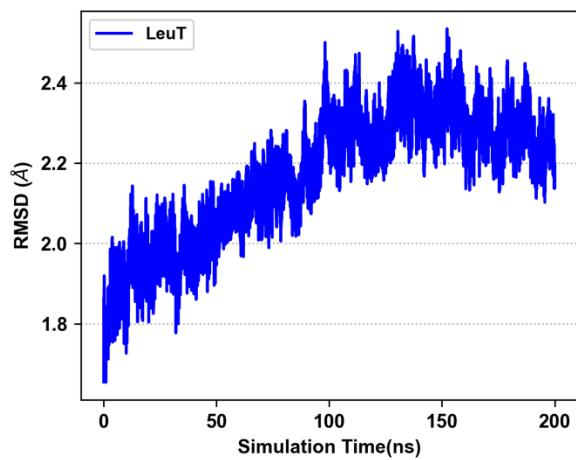
**Figure S1.** (A) RMSD of ubiquitin backbone atoms with respect to the initial structure. (B) RGYR of all protein atoms. (C) RMSF of backbone atoms of each residue across 100 ns trajectory.



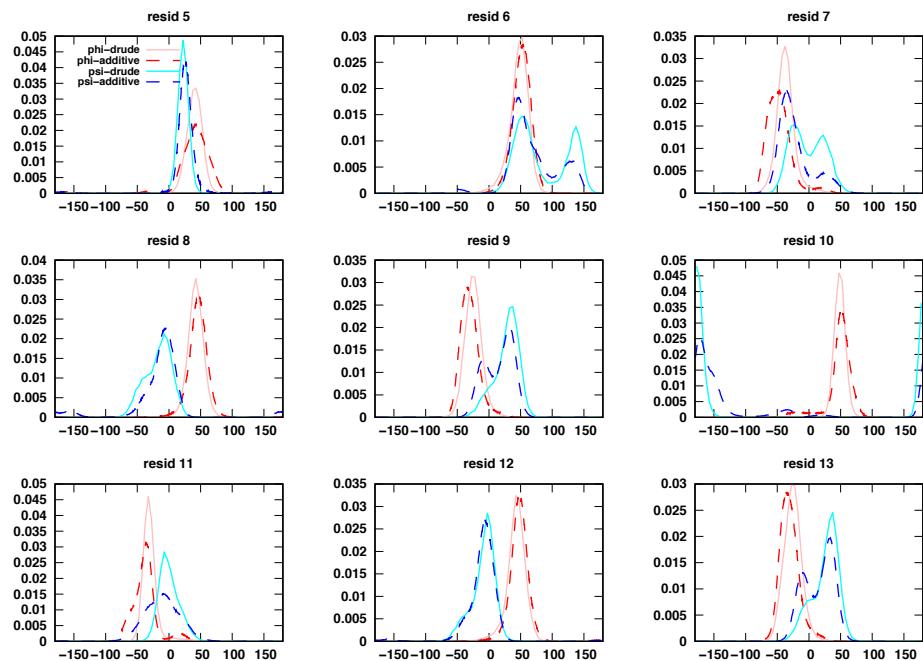
**Figure S2.** Total dipole moments of the peptide bonds of Tat RBD calculated as an average over 200 ns trajectory sampled every 1 ns. The error bars represent one standard deviation.



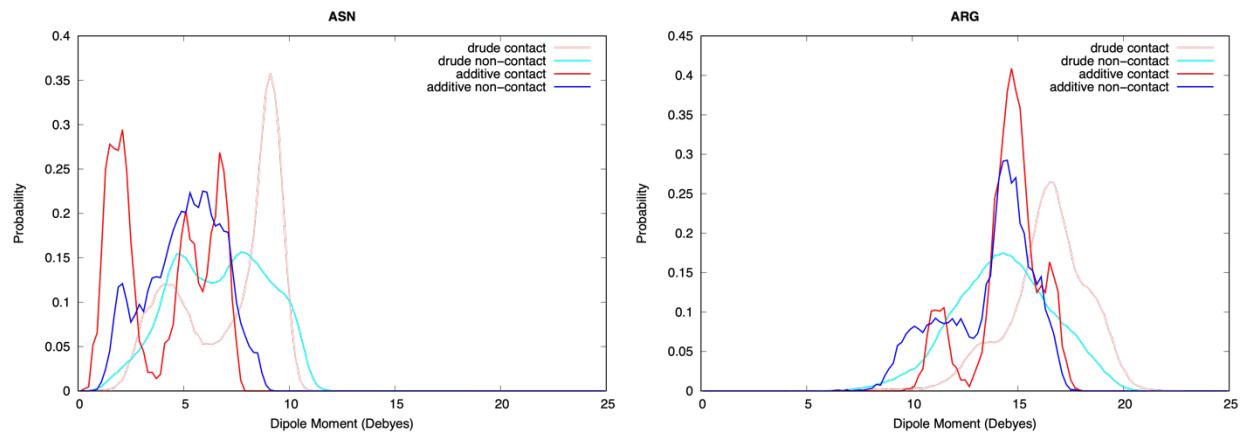
**Figure S3.** Comparison of dipole moment distributions of amino acid side chains calculated from 200 ns versus 1ns trajectories simulated with the Drude FF.



**Figure S4.** RMSD of LeuT non-hydrogen atoms with respect to the initial structure.



**Figure S5.** Comparison of glycosidic torsion angle distributions of *M. catarrhalis* serotype C oligosaccharide.



**Figure S6.** Comparison of dipole moments between the hyaluronan-contact and non-contact protein residues.