SUPPORTING INFORMATION

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

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

[1]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA

[2]Departments of Biological Sciences, Chemistry, Bioengineering, and Computer Science and Engineering, Lehigh University, PA, USA

[3]Leadership Computing Facility, Argonne National Laboratory, 9700 Cass Ave, Argonne, IL, USA

[4]Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

[5]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

[†]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)	carbohydate 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

DIHE + CMAP	16078.083	16078.234	16078.083	NA
IMPR	230.055	229.848	230.054	NA
BONDED TOTAL	332908.686	332941.030	328723.853	NA
VDW	144567.429	144569.180	144619.359	NA
ELEC	-1373004.180	-1372962.236	-1368357.441	NA
NON BONDED TOTAL	-1228436.751	-1228393.056	-1223738.082	NA
TOTAL	-895528.064	-895452.026	-895014.229	NA
% Change in TOTAL	0.000%	0.008%	0.057%	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
	Hyaluronate lyase			
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

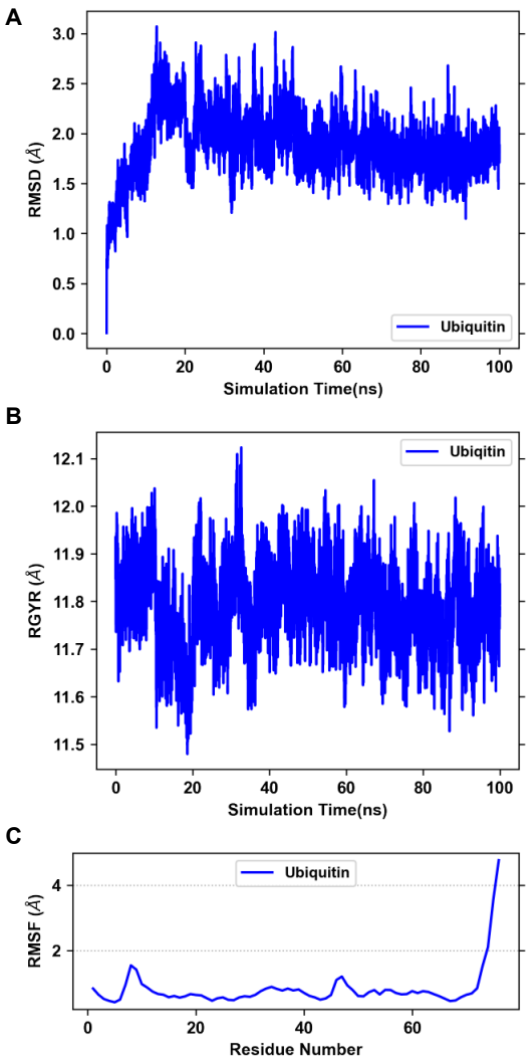

Figure S1. (A) RMSD of ubiquitin backbone atoms with respect to the initial structure. (B) RGYR of all protein atoms. (C) RMSF of backbone atoms of each residue across 100 ns trajectory.

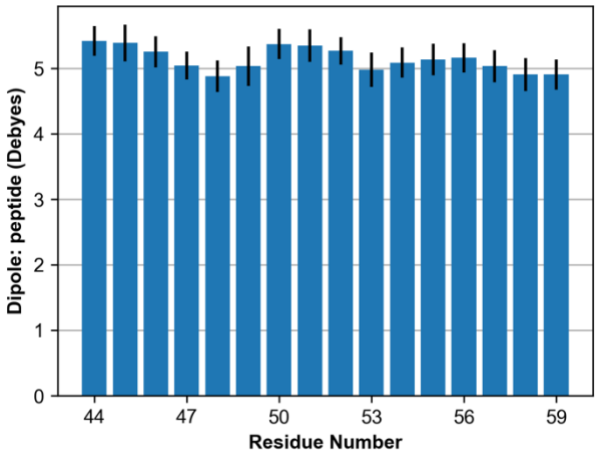

Figure S2. Total dipole moments of the peptide bonds of Tat RBD calculated as an average over 200 ns trajectory sampled every 1 ns. The error bars represent one standard deviation.

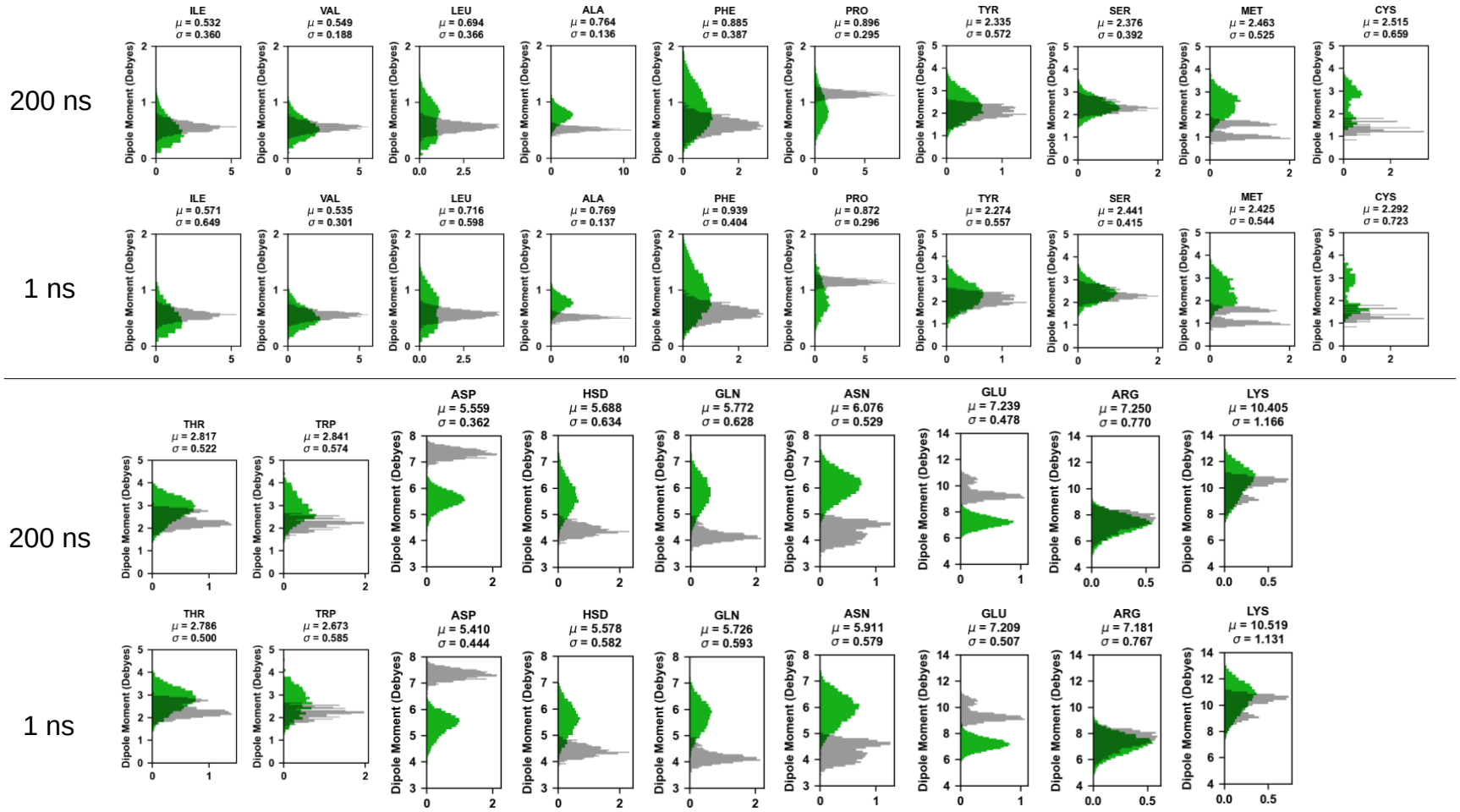

Figure S3. Comparison of dipole moment distributions of amino acid side chains calculated from 200 ns versus 1ns trajectories simulated with the Drude FF.

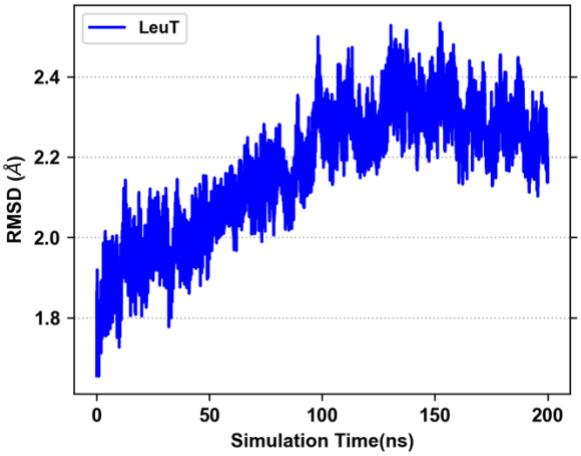

Figure S4. RMSD of LeuT non-hydrogen atoms with respect to the initial structure.

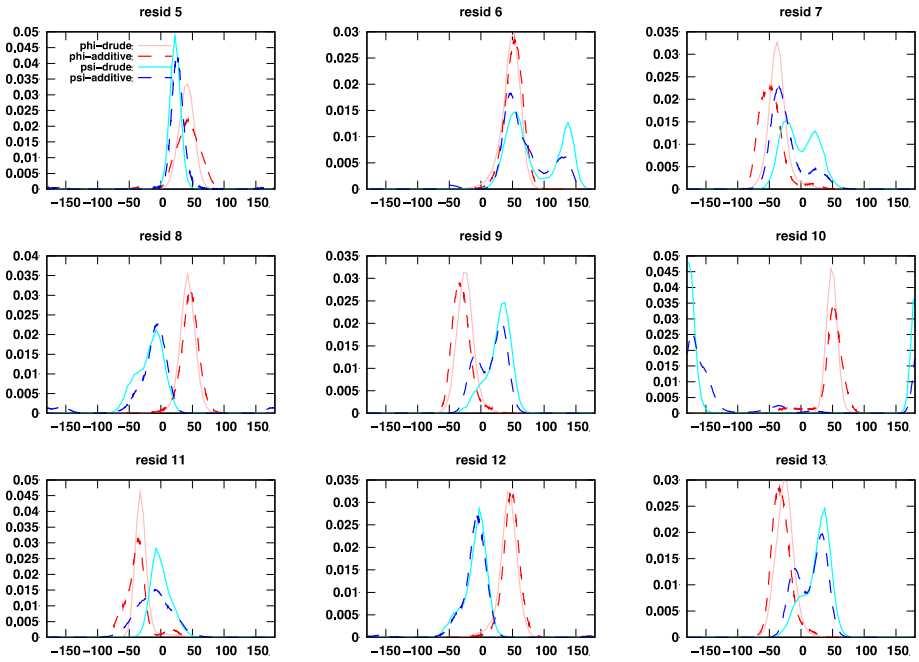

Figure S5. Comparison of glycosidic torsion angle distributions of *M. catarrhalis* serotype C oligosaccharide.

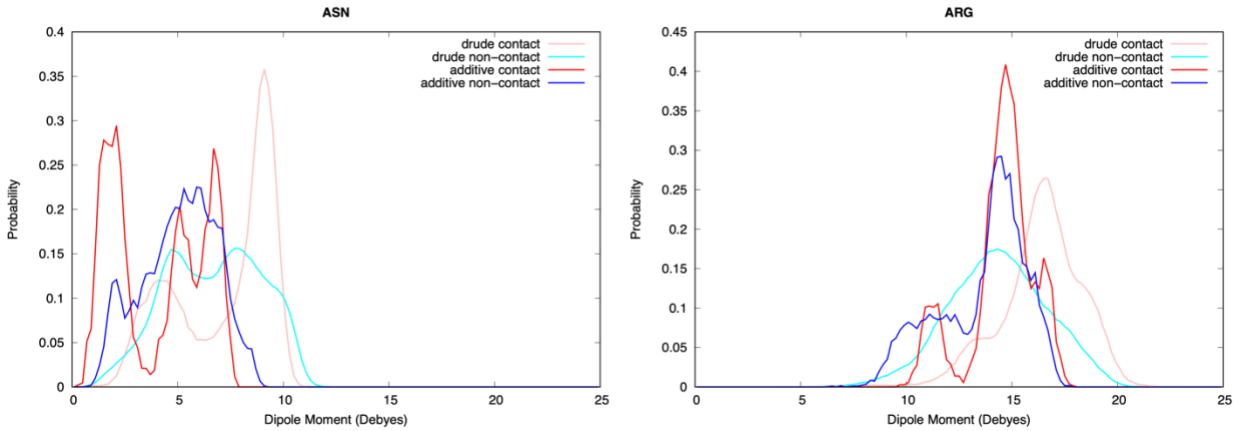

Figure S6. Comparison of dipole moments between the hyaluronan-contact and non-contact protein residues.