Supporting Information for

Optimization and Validation of the Classical Drude Polarizable Protein Force Field

Fang-Yu Lin,[a] Jing Huang,[a,b] Poonam Pandey,[a] Chetan Rupakheti,[c] Jing Li,[c] Benoît Roux,[c], and Alexander D. MacKerell Jr [a], [‡]

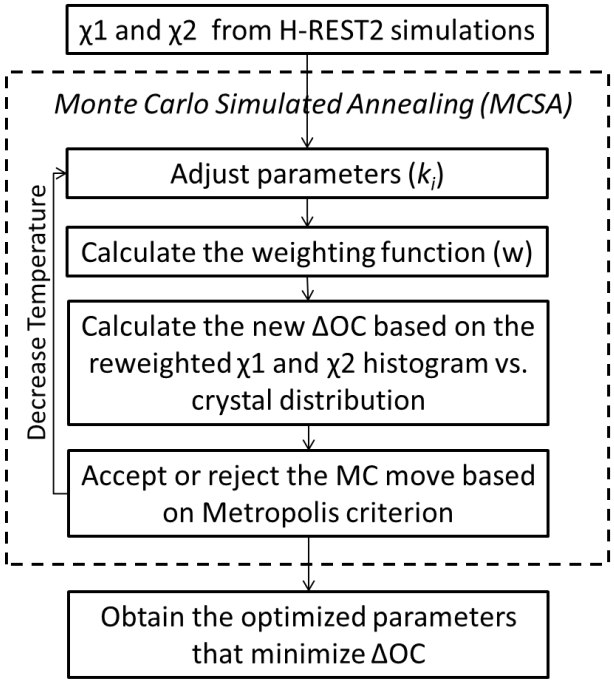
[a]Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn Street, Baltimore, MD 21201, USA.
[b]Westlake University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang, China
[c]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, 60637, USA.
Corresponding author: [‡]alex@outerbanks.umaryland.edu

Table S1. Ratio of MM/QM molecular polarizabilities of each dipeptide that contains N-acetylated and N'-methylamidated termini from five MCSA runs respectively. In each run, MCSA fitted parameters that optimized based on different QM target scale (0.70, 085, and 1.00) were used to calculate the MM/QM ratio.

Residue	Target Scale	MM/QM				
		Run1	Run2	Run3	Run4	Run5
ASN	1.00	0.92	0.91	0.91	0.89	0.91
	0.85	**0.85**	**0.85**	**0.85**	**0.85**	**0.85**
	0.70	0.83	0.82	0.83	0.83	0.83
ARG	1.00	0.85	0.85	0.85	0.85	0.86
	0.85	**0.84**	**0.83**	**0.83**	**0.84**	**0.84**
	0.70	0.76	0.76	0.76	0.76	0.76
ASP	1.00	0.77	0.76	0.77	0.76	0.77
	0.85	**0.77**	**0.76**	**0.76**	**0.76**	**0.77**
	0.70	0.71	0.71	0.70	0.70	0.70
CYS	1.00	0.81	0.81	0.80	0.79	0.80
	0.85	**0.81**	**0.81**	**0.80**	**0.79**	**0.80**
	0.70	0.73	0.72	0.73	0.73	0.73
GLN	1.00	0.89	0.88	0.88	0.92	0.89
	0.85	**0.85**	**0.85**	**0.85**	**0.84**	**0.84**
	0.70	0.77	0.76	0.78	0.78	0.76
GLU	1.00	0.77	0.77	0.77	0.75	0.77
	0.85	**0.76**	**0.75**	**0.75**	**0.75**	**0.77**
	0.70	0.70	0.71	0.71	0.71	0.71
HSD	1.00	0.89	0.88	0.88	0.87	0.88
	0.85	**0.84**	**0.85**	**0.84**	**0.83**	**0.84**
	0.70	0.80	0.79	0.80	0.79	0.80
HSE	1.00	0.90	0.89	0.89	0.88	0.89
	0.85	**0.85**	**0.84**	**0.84**	**0.84**	**0.84**
	0.70	0.80	0.80	0.80	0.80	0.80
HSP	1.00	0.93	0.92	0.94	0.92	0.94
	0.85	**0.86**	**0.85**	**0.86**	**0.85**	**0.86**
	0.70	0.86	0.85	0.86	0.85	0.86
ILE	1.00	0.94	0.93	0.93	0.92	0.93
	0.85	**0.85**	**0.85**	**0.85**	**0.85**	**0.85**
	0.70	0.85	0.84	0.85	0.85	0.85
LEU	1.00	0.89	0.90	0.90	0.95	0.89
	0.85	**0.85**	**0.84**	**0.84**	**0.84**	**0.83**
	0.70	0.80	0.79	0.78	0.78	0.79
LYS	1.00	0.87	0.87	0.90	0.88	0.91
	0.85	**0.85**	**0.84**	**0.84**	**0.84**	**0.84**
	0.70	0.75	0.75	0.73	0.77	0.74
MET	1.00	0.85	0.84	0.84	0.83	0.84
	0.85	**0.82**	**0.81**	**0.81**	**0.81**	**0.81**
	0.70	0.75	0.75	0.75	0.75	0.75
THR	1.00	0.86	0.86	0.85	0.86	0.86
	0.85	**0.84**	**0.84**	**0.84**	**0.84**	**0.84**
	0.70	0.78	0.78	0.78	0.78	0.78
TYR	1.00	0.86	0.85	0.85	0.84	0.85
	0.85	**0.82**	**0.82**	**0.82**	**0.82**	**0.82**
	0.70	0.76	0.76	0.76	0.76	0.76
TRP	1.00	0.91	0.83	0.90	0.89	0.90
	0.85	**0.84**	**0.84**	**0.84**	**0.84**	**0.84**
	0.70	0.81	0.81	0.81	0.81	0.81

PHE	1.00	0.91	0.89	0.89	0.89	0.89
	0.85	**0.83**	**0.83**	**0.83**	**0.84**	**0.85**
	0.70	0.80	0.80	0.80	0.81	0.80
SER	1.00	0.87	0.87	0.86	0.86	0.87
	0.85	**0.84**	**0.85**	**0.85**	**0.85**	**0.85**
	0.70	0.78	0.78	0.78	0.79	0.78
VAL	1.00	0.92	0.92	0.91	0.92	0.91
	0.85	**0.87**	**0.86**	**0.87**	**0.84**	**0.87**
	0.70	0.84	0.83	0.84	0.83	0.84

Figure S1. Flow diagram of the χ1 and χ2 optimization using the reweighting method.



$$w = e^{-(E'-E_0)/kT}$$

Energy (E) of the involved dihedral:

$$E = \sum_{m=1}\sum_{i=1} k_i \cos\theta_m$$

Figure S2. Secondary structural traces of the proteins used in the present study. PDB identifiers are given in a) to s).

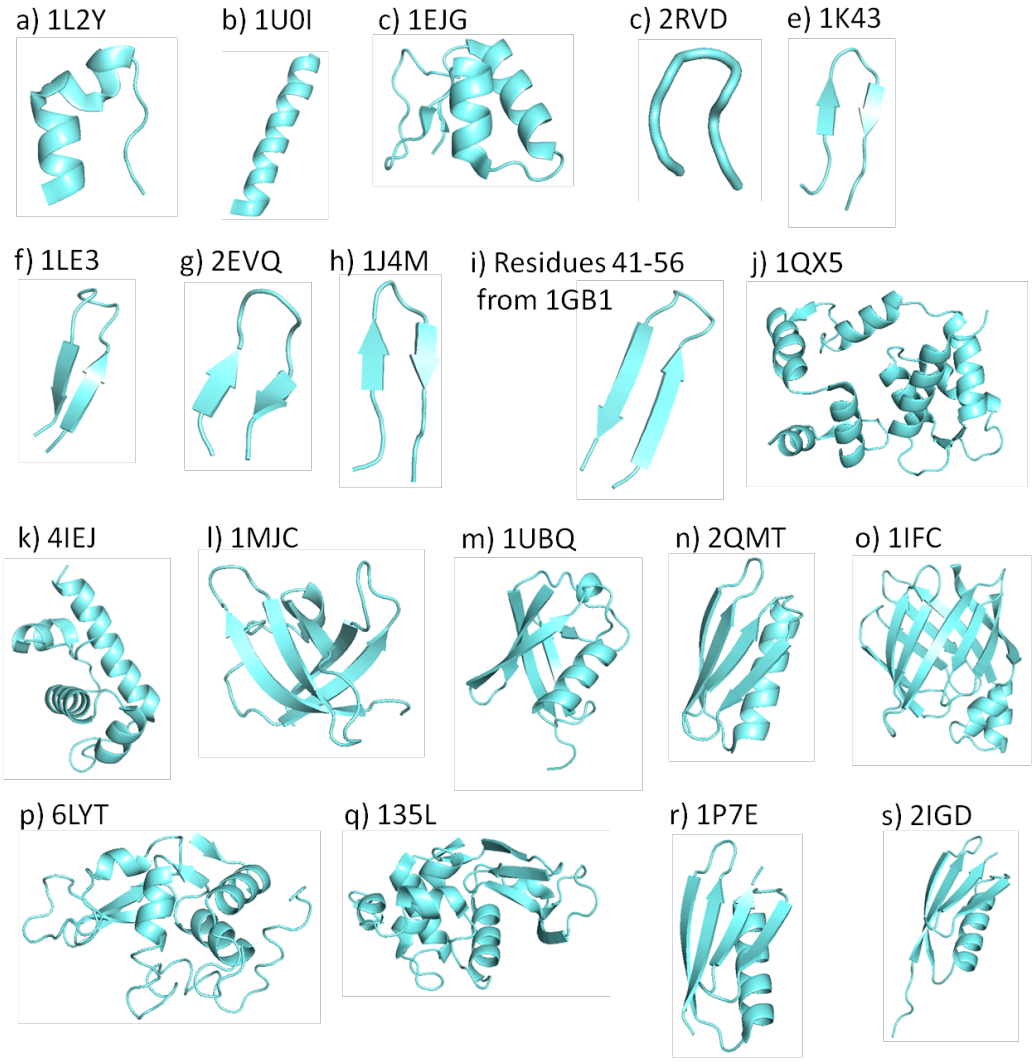
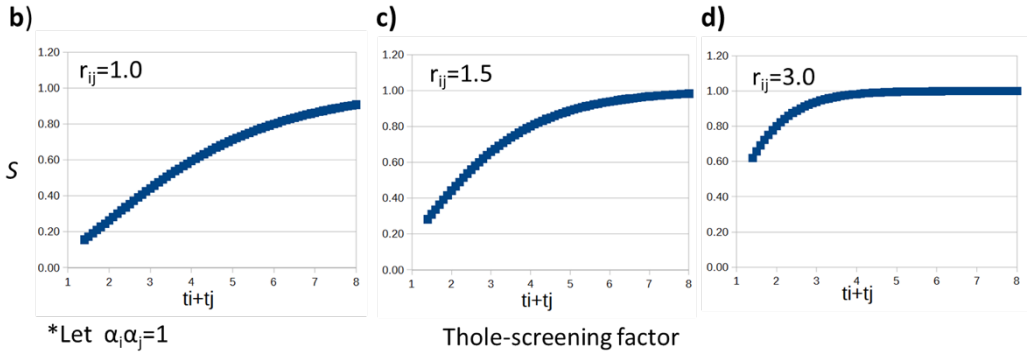


Figure S3. a) Damping of the Coulomb interaction by the Thole screening function S_{ij}. Charges on atoms i and j are q_i and q_j, respectively, separated by a distance r_{ij}. b)-d) Thole screening function S_{ij} as a function of (t_i+t_j) with r_{ij} set to 1.0, 1.5, and 3.0 respectively, where α is atomic polarizability, t is Thole parameter, and $\alpha_i\alpha_j$ is assumed to be equal to 1.

a)

$$U(r_{ij}) = \frac{\text{const. } q_i * q_j}{r_{ij}} * S_{ij}$$

$$S_{ij}(r_{ij}) = 1 - \left[\left(1 + \frac{(t_i + t_j)r_{ij}}{2(\alpha_i\alpha_j)^{1/6}}\right)\right] \exp\left[\frac{-(t_i + t_j)r_{ij}}{(\alpha_i\alpha_j)^{1/6}}\right]$$



*Let $\alpha_i\alpha_j$=1

Thole-screening factor

Figure S4. Analysis of nuclei-Drude particle distance distributions (Å) in Ala-Lys-Ala tripeptide aqueous solution system. a) Illustration of the Ala-Lys-Ala tripeptide in solution showing the Cβ/Cγ/Cδ carbons (atom types CB, CG and CD for carbons). b-c) The nuclei-Drude particle distance distributions between DCB...CB (in blue), DCG...CG (in red) and DCD…CD (in green) with the b) Drude-2013 and c) Drude-2019 FFs, where DCB, DCG and DCD are Drude particles of the CB, CG, and CD carbons, respectively.

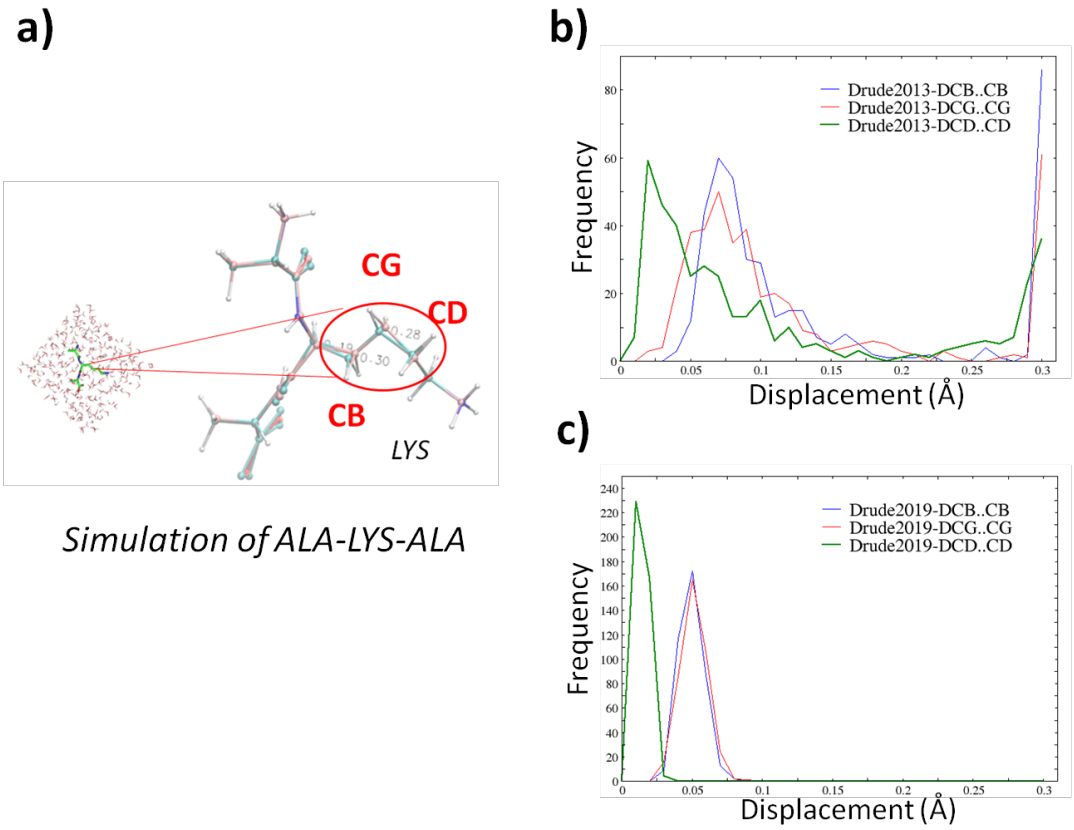


Table S2. Average RMS differences and RMS fluctuations of the RMSD with respect to the crystal or NMR structures of the backbone non-hydrogen atoms (Å).

PDB	C36m		Drude-2013		Drude-2019	
	AVG	RMSF	AVG	RMSF	AVG	RMSF
Number of amino acids < 50 aa						
1L2Y	1.66	0.40	4.98	1.12	1.51	0.51
1U0I	6.48	2.18	5.69	1.69	1.99	0.48
1EJG	4.01	0.15	1.60	0.32	1.37	0.13
2RVD	1.10	0.09	4.19	1.05	1.13	0.08
1K43	2.10	0.47	1.75	0.63	2.06	0.16
1LE3	0.99	0.36	5.35	1.36	1.76	0.47
2EVQ	4.30	3.23	4.43	1.76	1.07	0.29
1J4M	1.94	0.36	1.52	0.33	1.97	0.19
GB1 hairpin	2.38	0.42	7.78	2.49	7.19	2.32
AVERAGE	2.77	0.85	4.14	1.19	2.23	0.51
Number of amino acids > 50 aa						
1QX5	8.45	2.33	7.88	0.61	7.22	0.58
4IEJ	1.09	0.17	3.67	0.89	2.55	0.34
1MJC	7.78	10.10	2.54	0.72	1.89	0.45
1UBQ	2.23	0.47	3.16	0.55	2.40	0.37
2QMT	0.96	0.25	2.25	0.82	1.46	0.26
1IFC	1.49	0.28	2.99	0.25	2.45	0.55
6LYT	1.38	0.27	3.68	0.97	2.05	0.44
135L	1.73	0.33	2.95	0.26	2.84	0.51
1P7E	0.77	0.18	2.21	0.41	2.02	0.21
2IGD	10.04	7.95	5.71	1.95	1.41	0.23
AVERAGE	3.59	2.23	3.70	0.74	2.63	0.39

Table S3. Average RMS differences and RMS fluctuations of the RMSD with respect to the crystal or NMR structures of the side chain non-hydrogen atoms (Å)

PDB	C36m		Drude-2013		Drude-2019	
	AVG	RMSF	AVG	RMSF	AVG	RMSF
Number of amino acids < 50 aa						
1L2Y	2.98	0.52	7.87	1.30	2.82	0.50
1U0I	8.95	2.02	7.72	1.37	4.18	0.70
1EJG	3.84	0.10	2.85	0.41	2.70	0.22
2RVD	2.61	0.50	6.67	0.95	3.67	0.24
1K43	5.05	0.88	3.97	0.96	5.57	0.31
1LE3	1.89	0.41	8.25	2.05	2.22	0.44
2EVQ	6.28	3.68	7.16	1.62	2.53	0.72
1J4M	3.99	0.62	3.54	0.71	5.13	0.31
GB1 hairpin	4.00	0.59	9.78	2.33	9.23	2.41
AVERAGE	4.40	1.04	6.42	1.30	4.23	0.65
Number of amino acids > 50 aa						
1QX5	9.39	2.14	9.44	0.67	8.61	0.58
4IEJ	2.16	0.16	5.43	0.99	3.86	0.40
1MJC	8.81	9.39	3.60	0.81	2.92	0.39
1UBQ	3.04	0.29	3.72	0.45	3.21	0.34
2QMT	1.89	0.29	3.63	0.95	2.82	0.31
1IFC	2.38	0.26	4.17	0.29	3.34	0.49
6LYT	2.48	0.26	5.65	1.42	3.40	0.57
135L	2.83	0.43	4.15	0.30	4.40	0.71
1P7E	1.73	0.19	3.76	0.52	3.37	0.20
2IGD	11.58	8.31	7.40	2.25	2.51	0.23
AVERAGE	4.63	2.17	5.09	0.86	3.84	0.42

Figure S5. The hydrogen bond distances in the hairpin of the GB3 domain (PDB:2IGD) in a), using b) the original Drude-2013 protein FF (Drude-2013), c) Drude-2013 with optimized backbone parameters only (Drude-2013-opt-backbone), and d) final optimized Drude-2019 protein FF (Drude-2019).

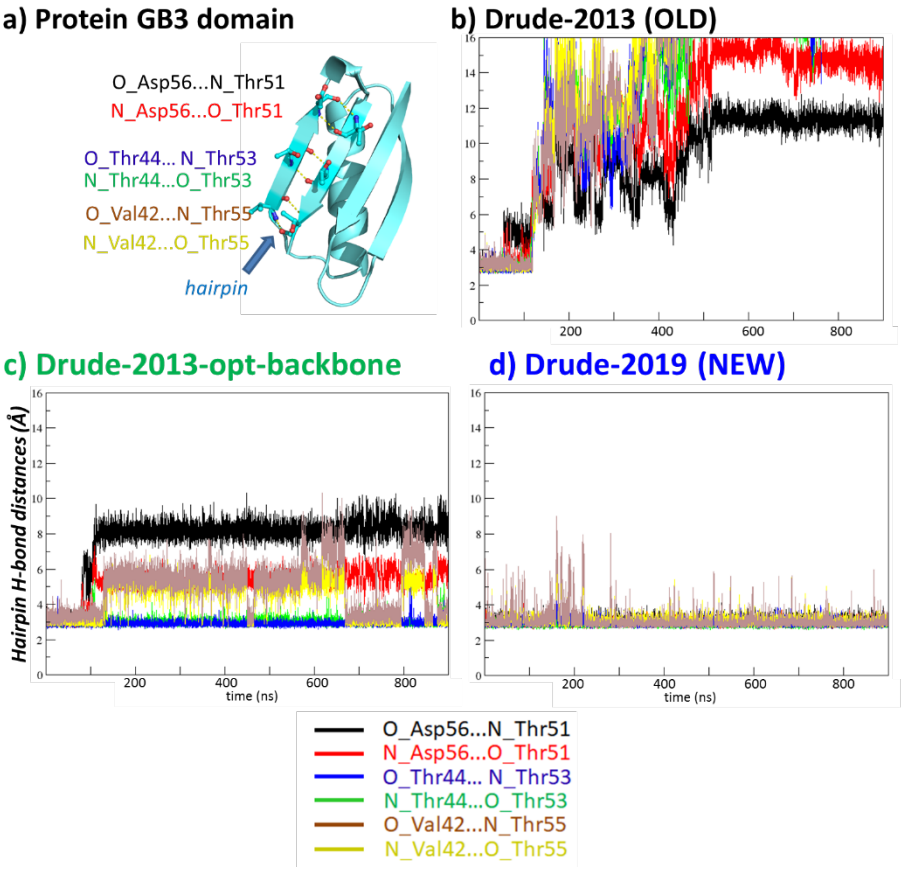


Figure S6. Overall ϕ, ψ distributions for the a) C36m, b) Drude-2013 and c) optimized Drude-2019 models. Simulated data included results from PDB: 1QX5, 4IEJ, 1MJC, 1UBQ, 2QMT, 1IFC, 6LYT, 135L, 1P7E, and 2IGD. Simulation results are presented as inverted Boltzmann distributions which are free energies (FE, kcal/mol) obtained from Boltzmann weighting, FE = kTln(P), of the population, P, in 1˚x 1˚ bins summed over all the proteins where k and T are the Boltzmann constant and temperature, 298˚, respectively.

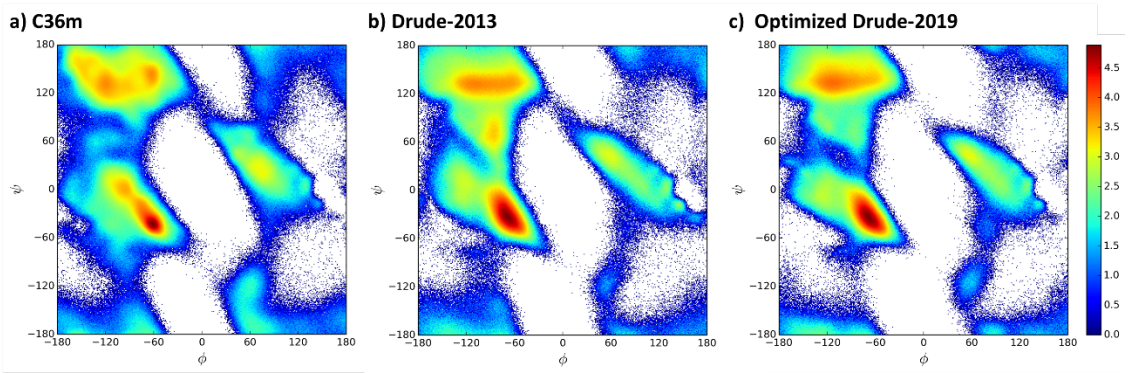


Figure S7. Comparison between experimental and calculated $^{h3}J_{NC}$ couplings in a) ubiquitin (PDB: 1UBQ), b) Protein GB1 domain (PDB: 2QMT), c) Cold-shock protein A (PDB: 1MJC), d) Apocalmodulin (PDB: 1QX5), and e) Intestinal fatty acid binding protein (PDB:1IFC). Red, black, and blue circles denote the values obtained from additive FF (in red), Drude-2013 FF (in black), and optimized Drude-2019 protein FF (in blue).

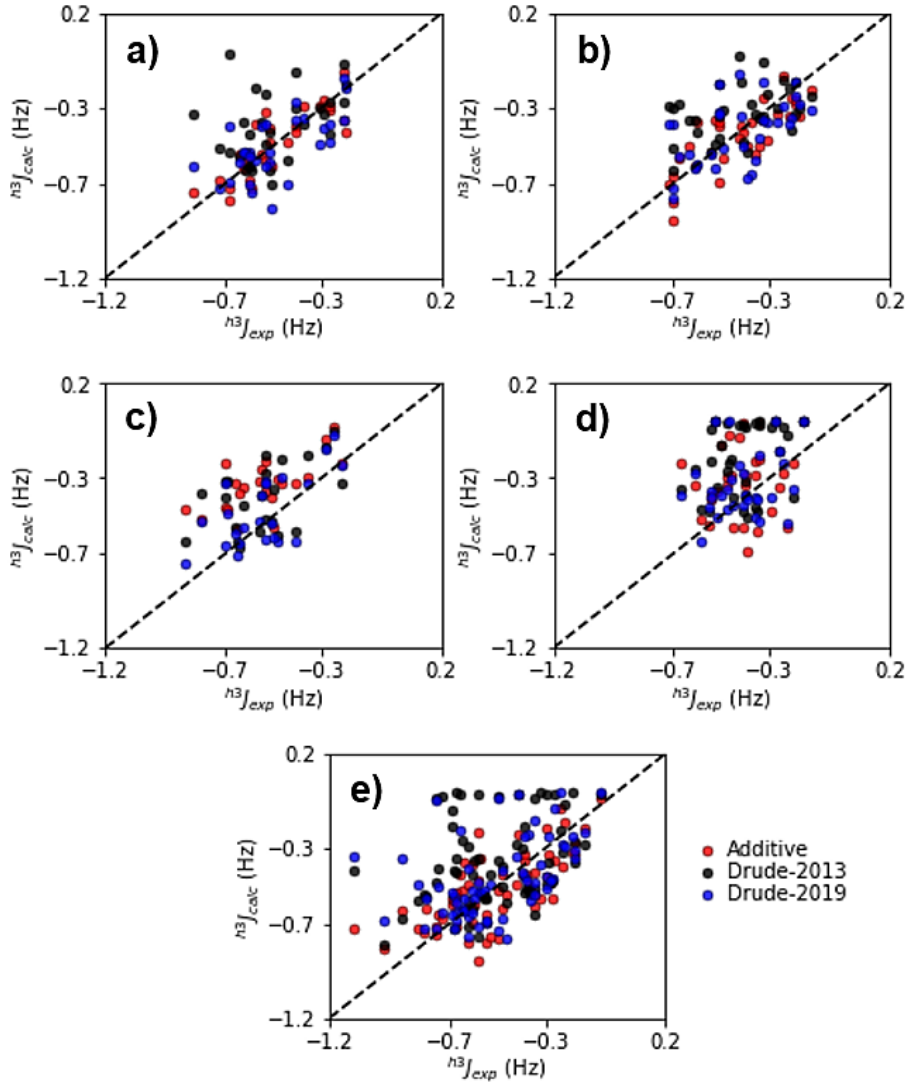


Figure S8. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in ubiquitin (PDB: 1UBQ) from 1μs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B} = [\mathbf{8\pi^2/3}] \times \mathbf{RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

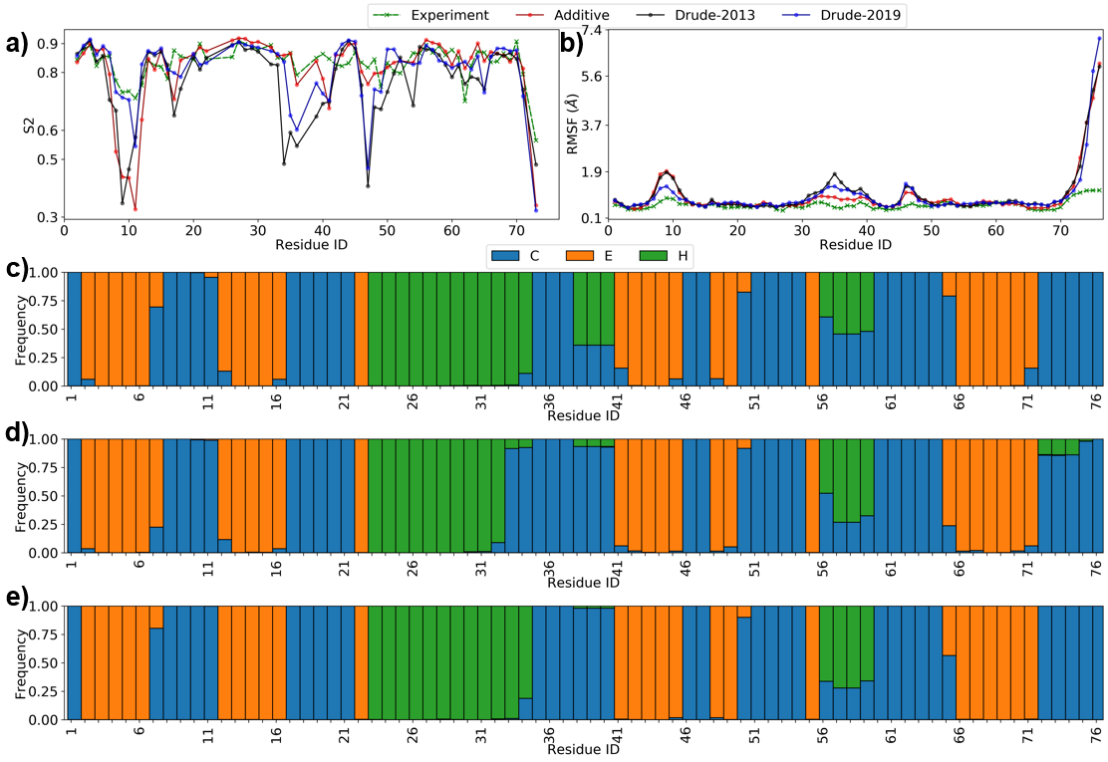


Figure S9. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in Protein GB1 domain (PDB: 2QMT) from 1 μs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B} = [\mathbf{8\pi^2/3}] \times \mathbf{RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

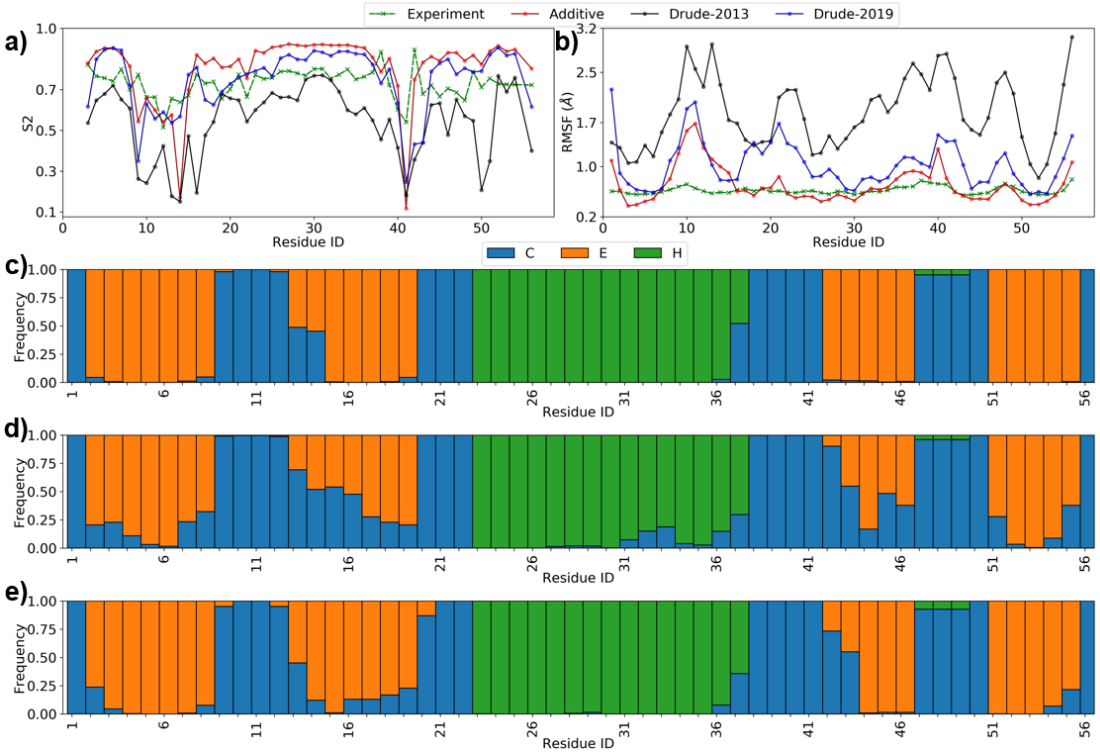


Figure S10. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in Cold-shock protein A (PDB: 1MJC) from 1 μs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B = [8\pi^2/3] \times RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

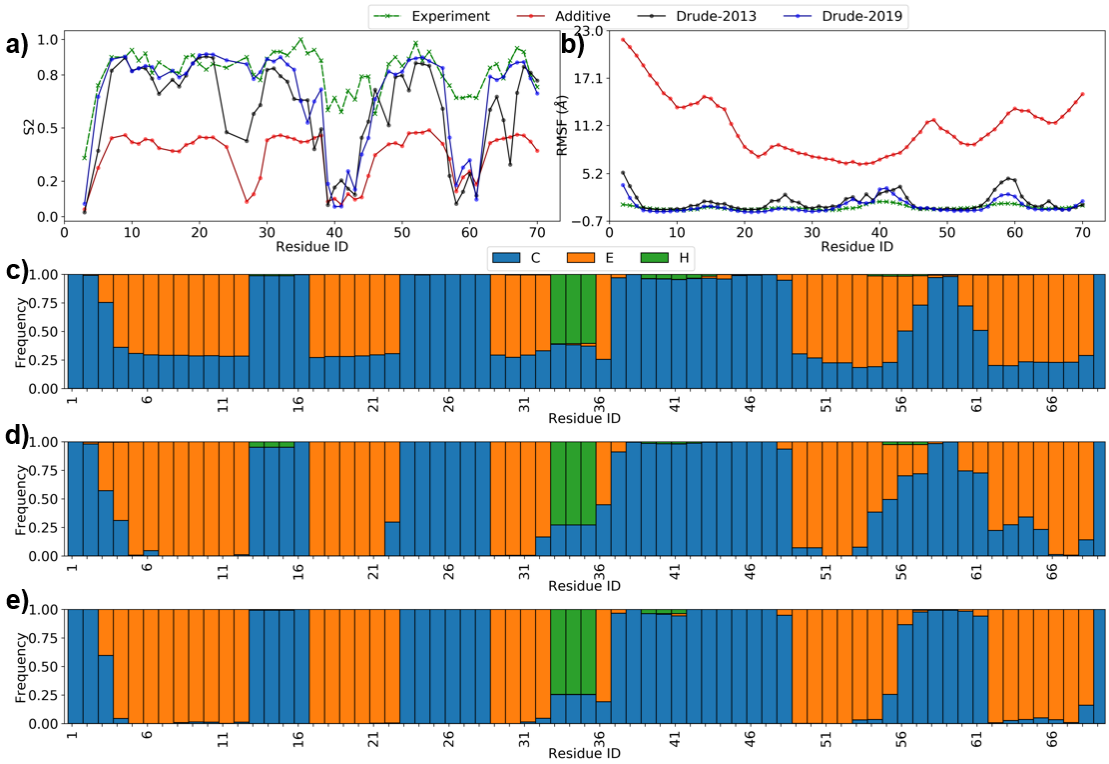


Figure S11. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in Apocalmodulin (PDB: 1QX5) from 1 μs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B} = [\mathbf{8\pi^2/3}] \times \mathbf{RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

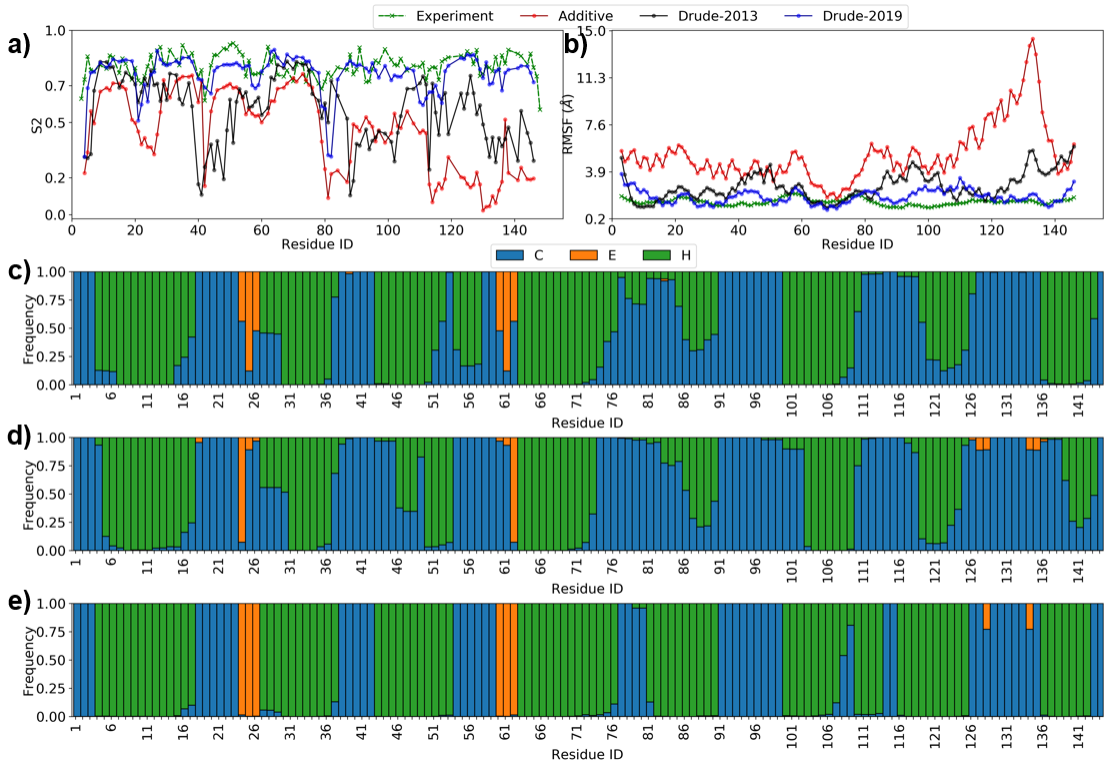


Figure S12. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in Intestinal fatty acid binding protein (PDB: 1IFC) system from 1 μs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B} = [\mathbf{8\pi^2/3}] \times \mathbf{RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

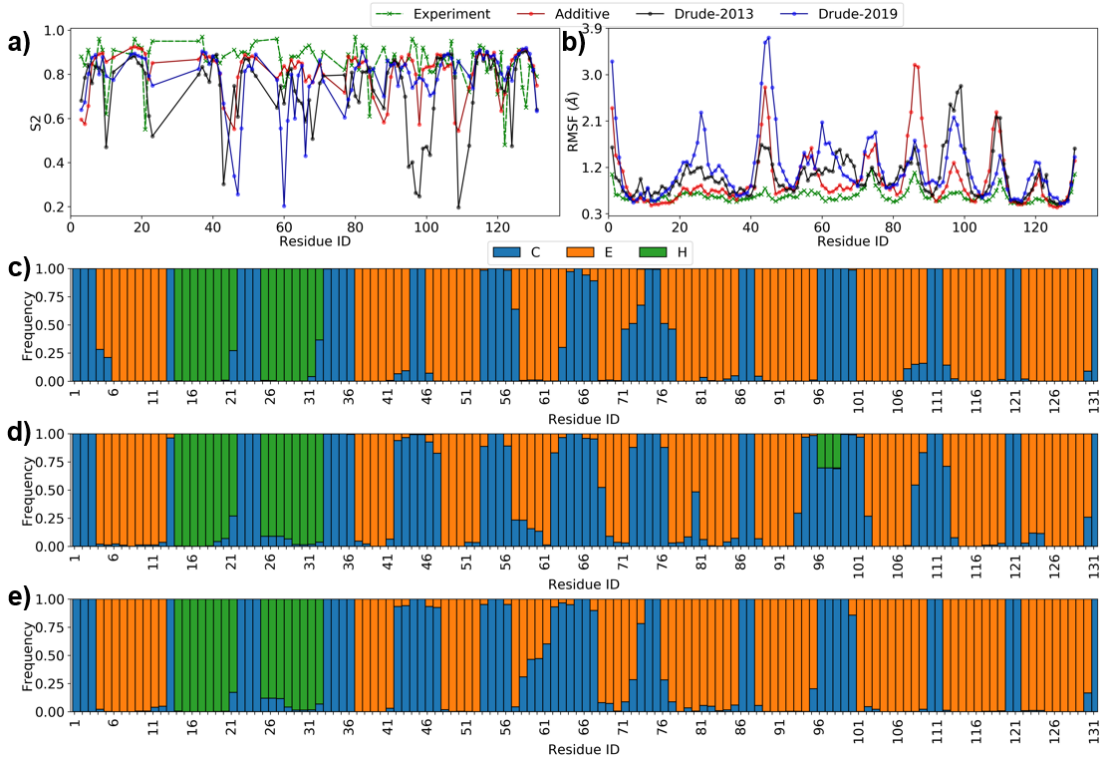


Figure S13. Analysis of NMR backbone order parameters, S2, RMSF, and secondary structure based on DSSP per residue in Hen lysozme (PDB: 6LYT) system from 1 µs MD simulations. a) Comparison of backbone order parameter S2 and b) RMSF, between experimental values (green), C36m (red), Drude-2013 (black), Drude-2019 (blue), respectively. The experimental RMSF were derived from the B-factors (B), according to the formula $\mathbf{B} = [\mathbf{8\pi^2/3}] \times \mathbf{RMSF^2}$. Fractional population analysis of the secondary structures [helix (H), sheet(E), and loops or random coil (C)] from MD simulation trajectories using c) C36m, d) Drude-2013, and e) Drude-2019 FFs.

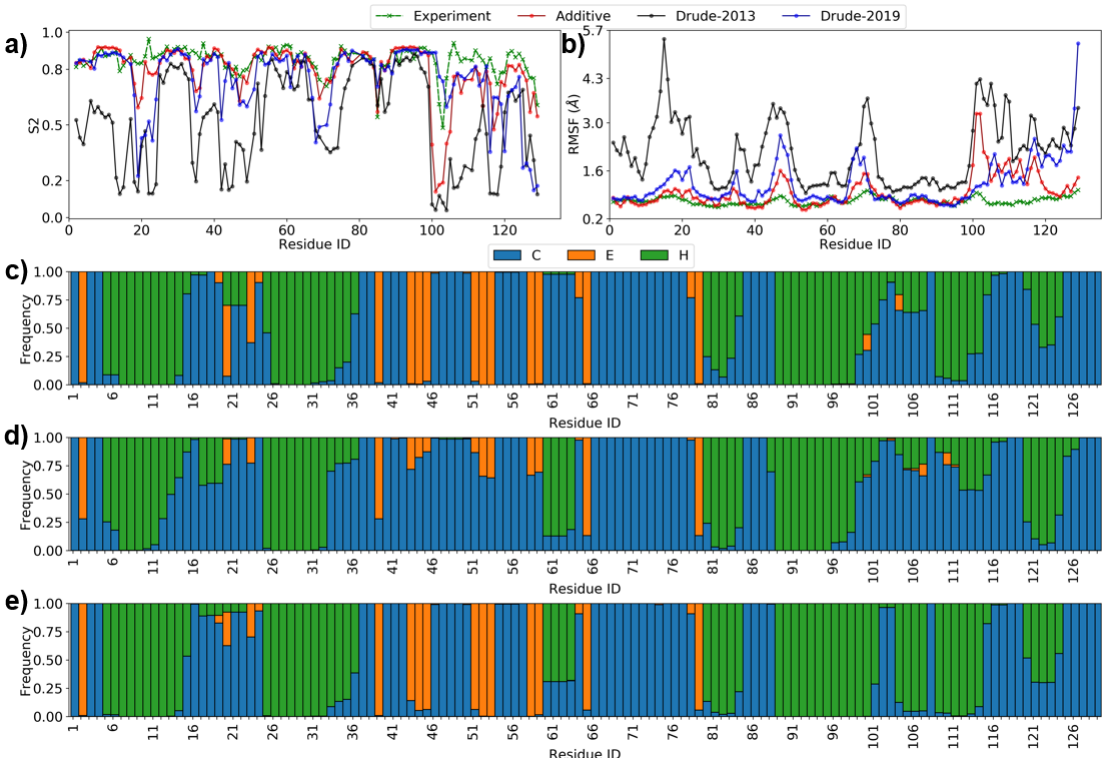


Figure S14. Hydrogen-bond, structural, dipole and DSSP based secondary structure analysis of a simulation of the HP7 peptide (PDB ID: 2EVQ) at 370 K. (a) Pictorial representation of NMR structure of HP7 with hydrogen-bond interactions between peptide bond carbonyl oxygen (O) and nitrogen hydrogen (HN) of residues Thr2, Asn4, Lys9, and Thr11. (b) Heatmap showing DSSP-defined secondary structures versus time. (c) RMSD with respect to the NMR structure for Cα atoms versus time. (d) Upper panel: Distance between Thr2:O-Thr11:HN (red) and Thr4:HN-Thr11:O atoms versus time; middle and lower panels: Dipole moment components ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$) and total dipole moment ($\boldsymbol{\mu}_R$) of peptide backbone (C,O,N,H,Cα,Ha) for Thr4 and Thr11, respectively, versus time. (e) Upper panel: Distance of Asn4:O-Lys9:HN (red) and Asn4:HN-Lys9:O versus time; middle and lower panels: Dipole moment components ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$) and total dipole moment ($\boldsymbol{\mu}_R$) of peptide backbone for Asn4 and Lys9, respectively, versus time. The DSSP assignment codes are H: Alpha helix, B: Residue in isolated beta-bridge, E: Extended strand, participates in beta ladder, G: 3/10 helix, I: pi helix), T: hydrogen bonded turn and S: bend, while L indicates loops or irregular/random-coil elements.

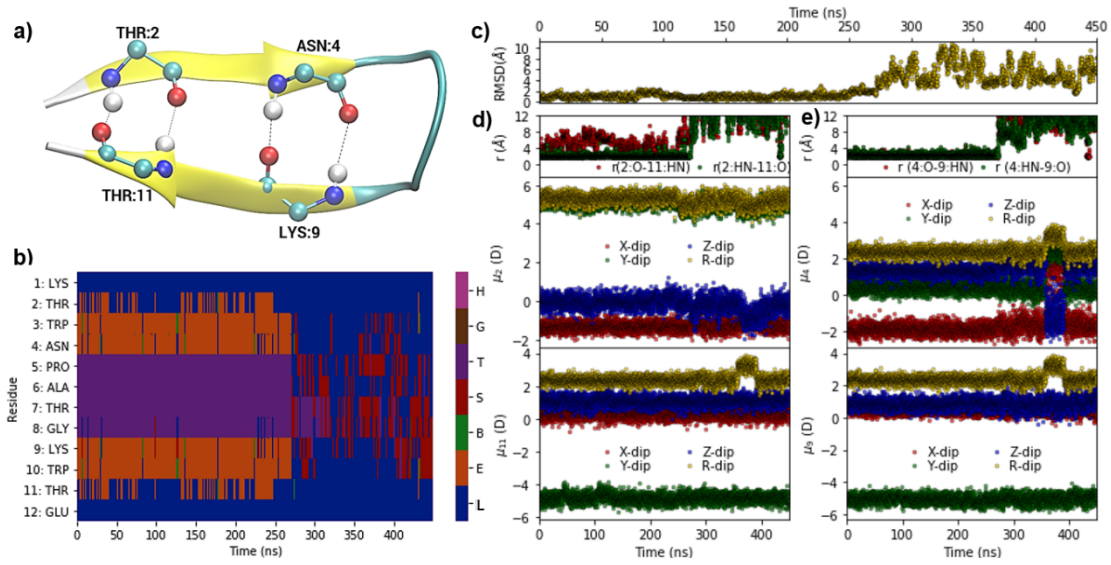


Figure S15. Probability distribution of dipole moment components and total values ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$, and $\boldsymbol{\mu}_R$) corresponding to the sampled secondary structures for (a) Thr2, (b) Trp3, (c) Asn4, (d) Lys9 from the HP7 simulation. The DSSP assignment codes are H: Alpha helix, B: Residue in isolated beta-bridge, E: Extended strand, participates in beta ladder, G: 3/10 helix, I: pi helix), T: hydrogen bonded turn and S: bend, R: random coil, and L indicates loops or irregular/random-coil elements.

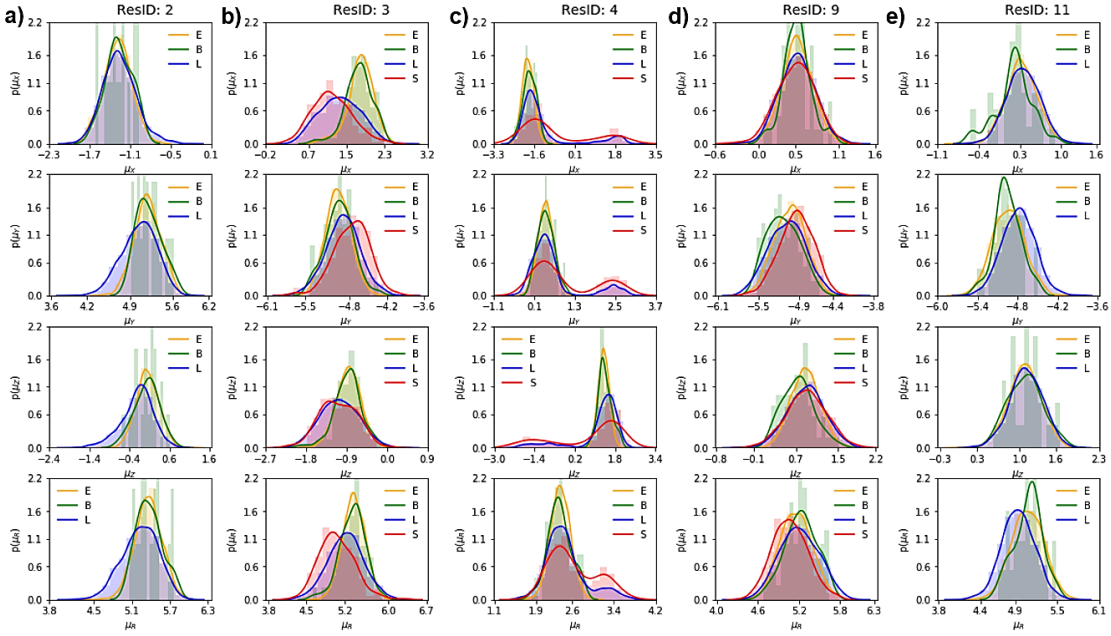


Figure S16. Hydrogen-bond, structural, dipole and DSSP based secondary structure analysis of a simulation of the Trp-cage peptide (PDB ID: 1L2Y) at 400 K. (a) Pictorial representation of NMR structure of Trp cage, with hydrogen-bond interactions between peptide bond carbonyl oxygen (O) of Ile4 or Gln5 and nitrogen hydrogen (HN) of Lys8 or Asp9. (b) Heatmap showing DSSP-defined secondary structures versus time. (c) RMSD (for helical region) with respect to the NMR structure for Cα atoms versus time. (d) Upper panel: Distance between Ile4:O-Lys8:HN versus time; middle and lower panels: Dipole moment components ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$) and total dipole moment ($\boldsymbol{\mu}_R$) of peptide backbone (C,O,N,H,Cα,Ha) for Ile4, respectively, versus time. (e) Upper panel: Distance of Gln5:O-Asn9:NH versus time; lower panel: Dipole moment components ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$) and total dipole moment ($\boldsymbol{\mu}_R$) of peptide backbone for Gln5, respectively, versus time. The DSSP assignment codes are H: Alpha helix, B: Residue in isolated beta-bridge, E: Extended strand, participates in beta ladder, G: 3/10 helix, I: pi helix, T: hydrogen bonded turn and S: bend, while L indicates loops or irregular/random-coil elements.

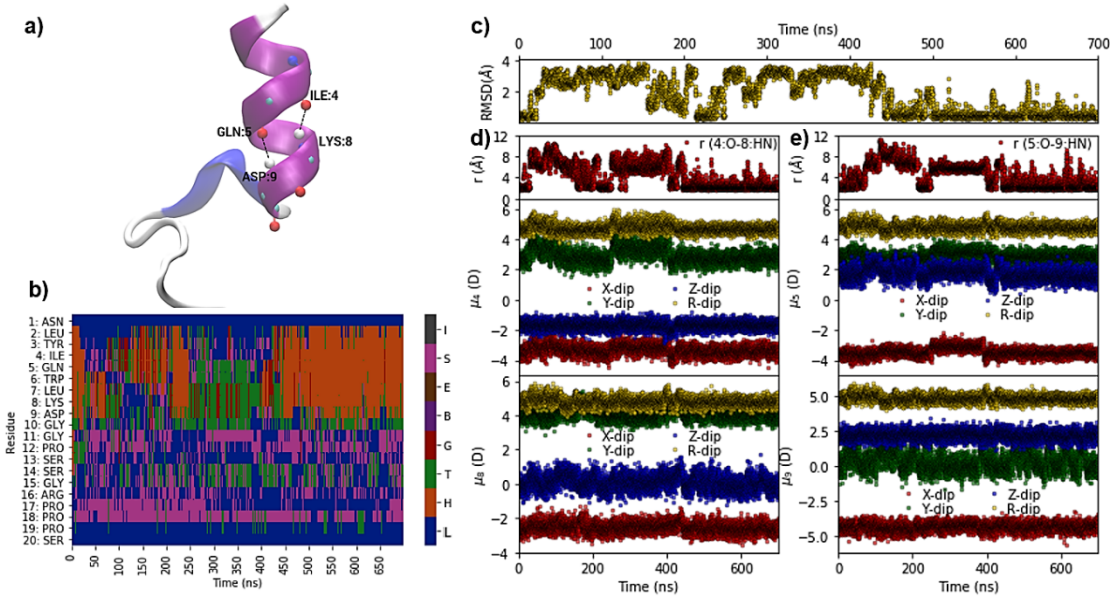


Figure S17. Probability distribution of dipole moment components and total values ($\boldsymbol{\mu}_x$, $\boldsymbol{\mu}_y$, $\boldsymbol{\mu}_z$, and $\boldsymbol{\mu}_R$) corresponding to the sampled secondary structures for (a) Ile4 and (b) Gln5, (c) Lys8, (d) Asp9 from the Trp cage simulation. The DSSP assignment codes are H: Alpha helix, B: Residue in isolated beta-bridge, E: Extended strand, participates in beta ladder, G: 3/10 helix, I: pi helix), T: hydrogen bonded turn and S: bend, and L indicates loops or irregular/random-coil elements.

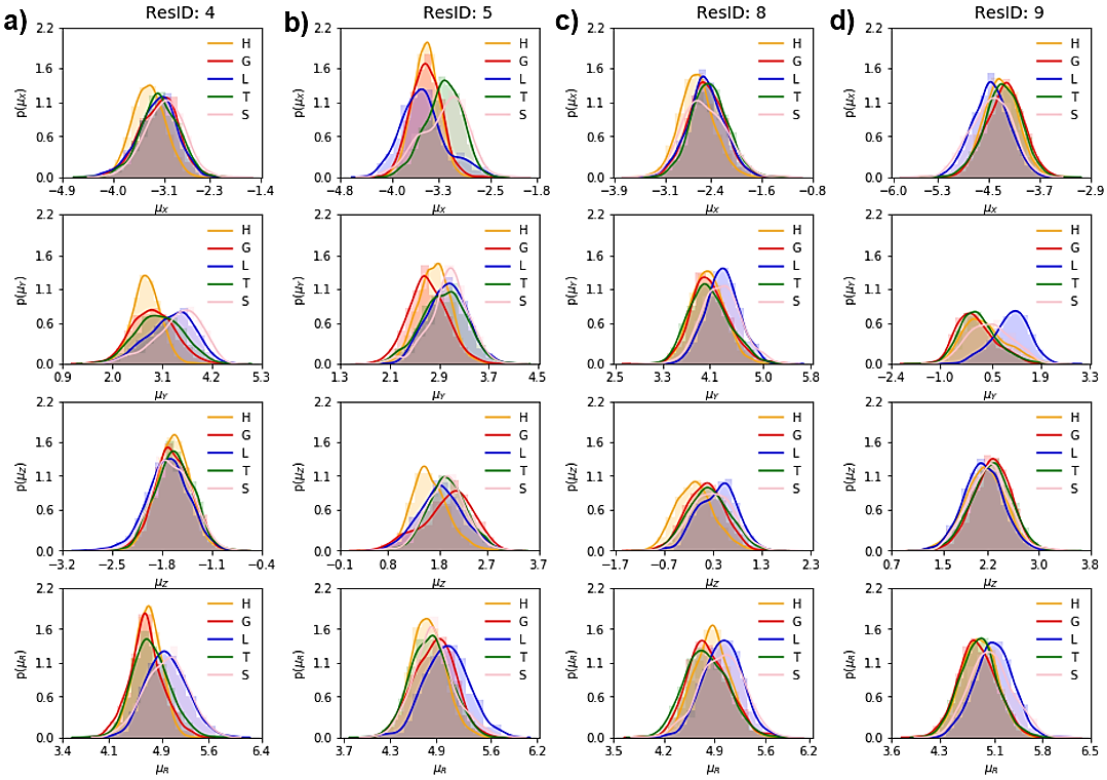


Figure S18. RMSD analysis based on Cα atoms for a) Trp-cage (PDB ID: 1L2Y), b) IAAL-K3 (PDB ID: 1U0I), c) Crambin (PDB ID: 1EJG), d) Cln025 (PDB ID: 2RVD), e) MBH12 (PDB ID: 1K43), f) Tryptophan Zipper 4 (PDB ID: 1LE3), g) HP7 (PDB ID: 2EVQ), h) 14-residue peptide (MBH12, PDB ID: 1J4M), and i) GB1 hairpin (PDB ID: 1GB1), using additive FF (in red), Drude-2013 FF (in black), and optimized Drude-2019 protein FF (in blue).

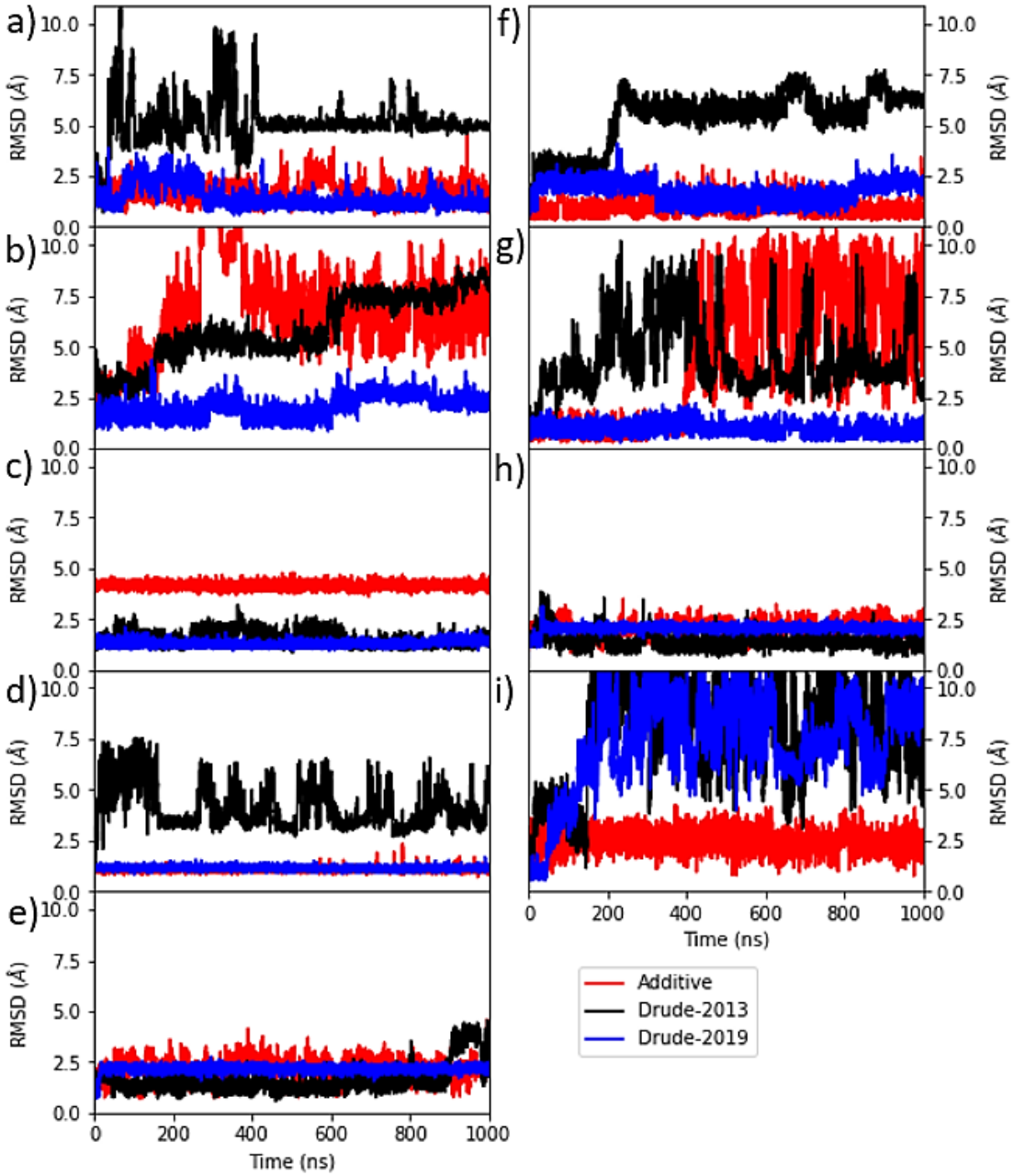


Figure S19. RMSD analysis based on Cα atoms for a) Apocalmodulin (PDB ID: 1QX5), b) DMAP1 (PDB ID: 4IEJ), c) Cold-shock protein A (PDB ID: 1MJC), d) Ubiquitin (PDB ID: 1UBQ), e) Protein GB1 domain (PDB ID: 2QMT), f) Intestinal fatty acid binding protein (PDB ID: 1IFC), g) Hen lysozyme (PDB ID: 6LYT), h) Lysozyme (PDB ID: 135L), i) Protein GB3 domain (PDB ID: 1P7E), and j) Protein GB3 domain (PDB ID: 2IGD), using additive FF (in red), Drude-2013 FF (in black), and optimized Drude-2019 protein FF (in blue).