POSSIM: Parameterizing complete second-order polarizable force field for proteins

Xinbi Li, Sergei Y. Ponomarev, Daniel L. Sigalovsky, John P. Cvitkovic and George A. Kaminski[*]

Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA 01609

AUTHOR EMAIL ADDRESS: gkaminski@wpi.edu

Supporting Information

This file contains values of potential energy parameters produced and used in the course of the reported work. In addition, details on the fitting done for specific side-chain residues are given.

[*] Author to whom correspondence should be addressed.

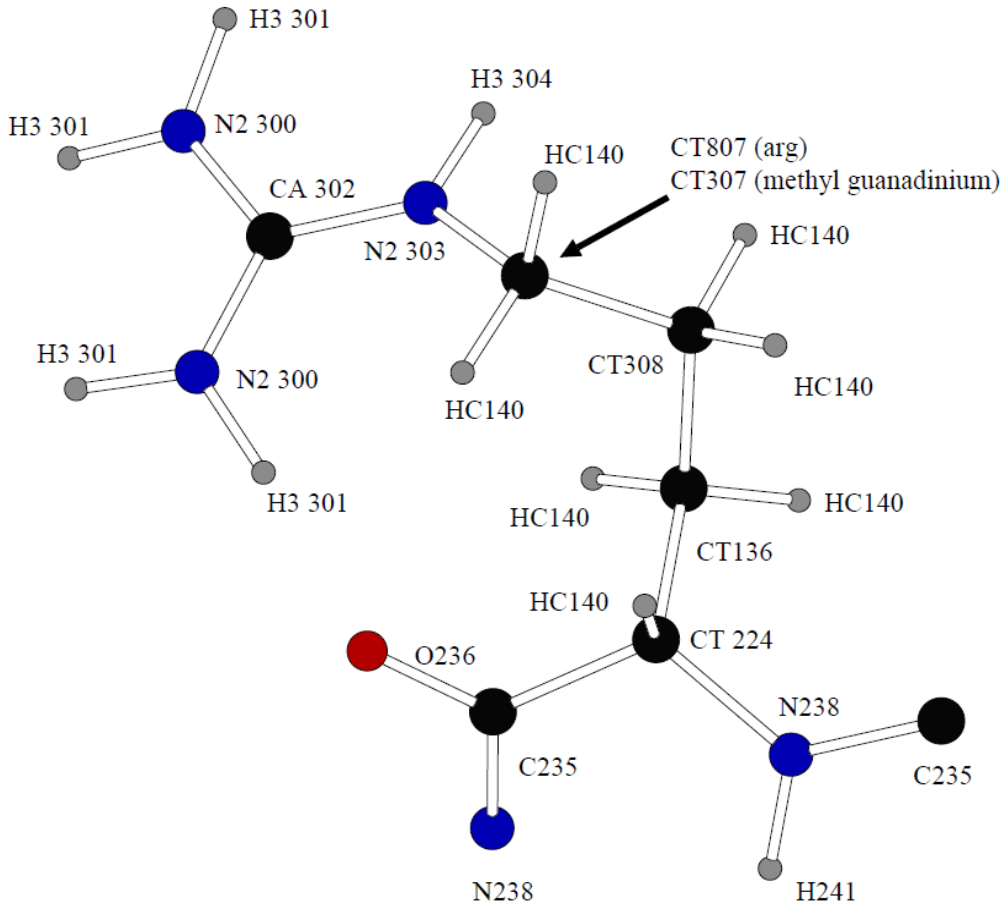

Figure S1. Methyl guanidinium/arginine side-chain atomtype assignments.

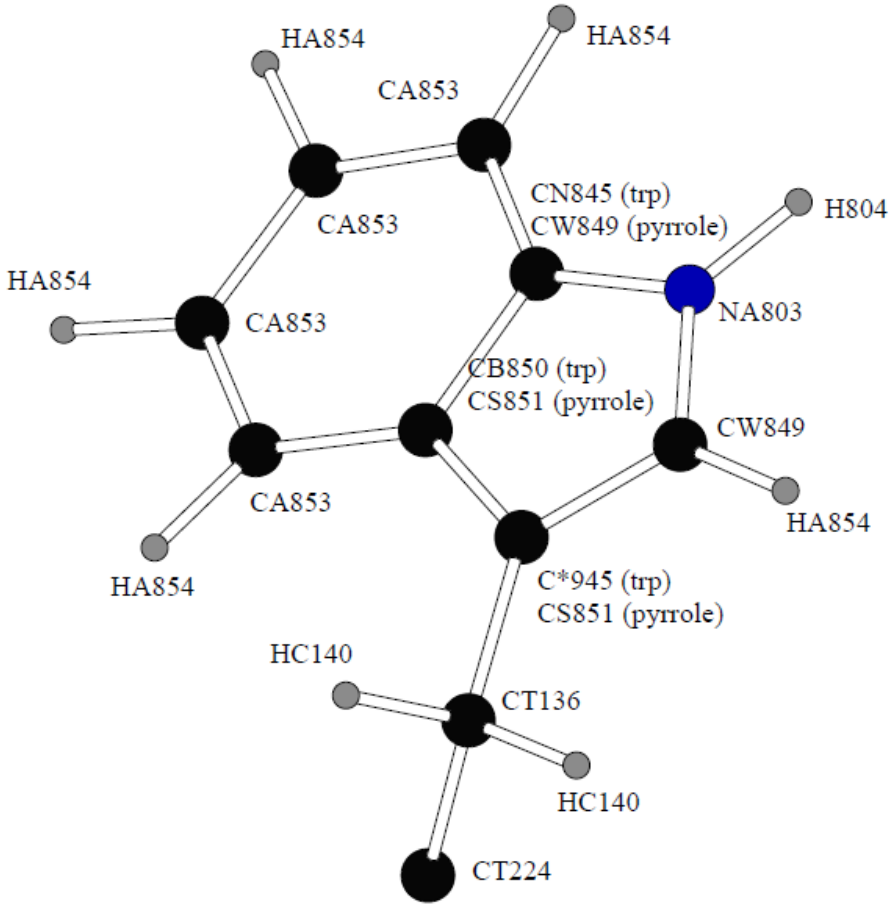

Figure S2. Pyrrole/tryptophan side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

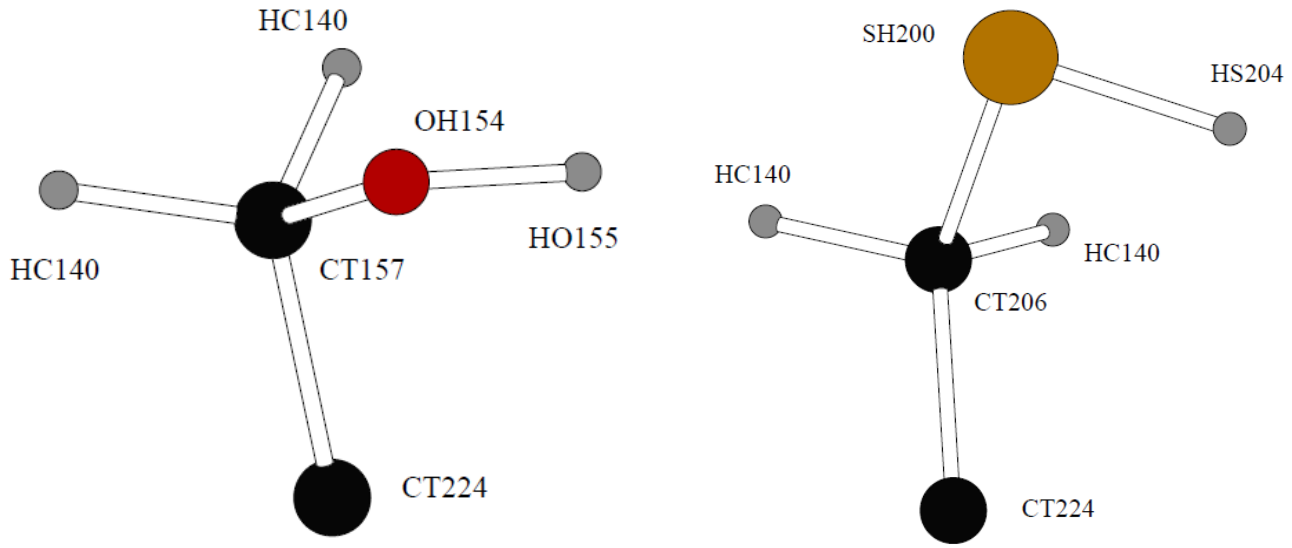

Figure S3. Serine (left) and cysteine (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

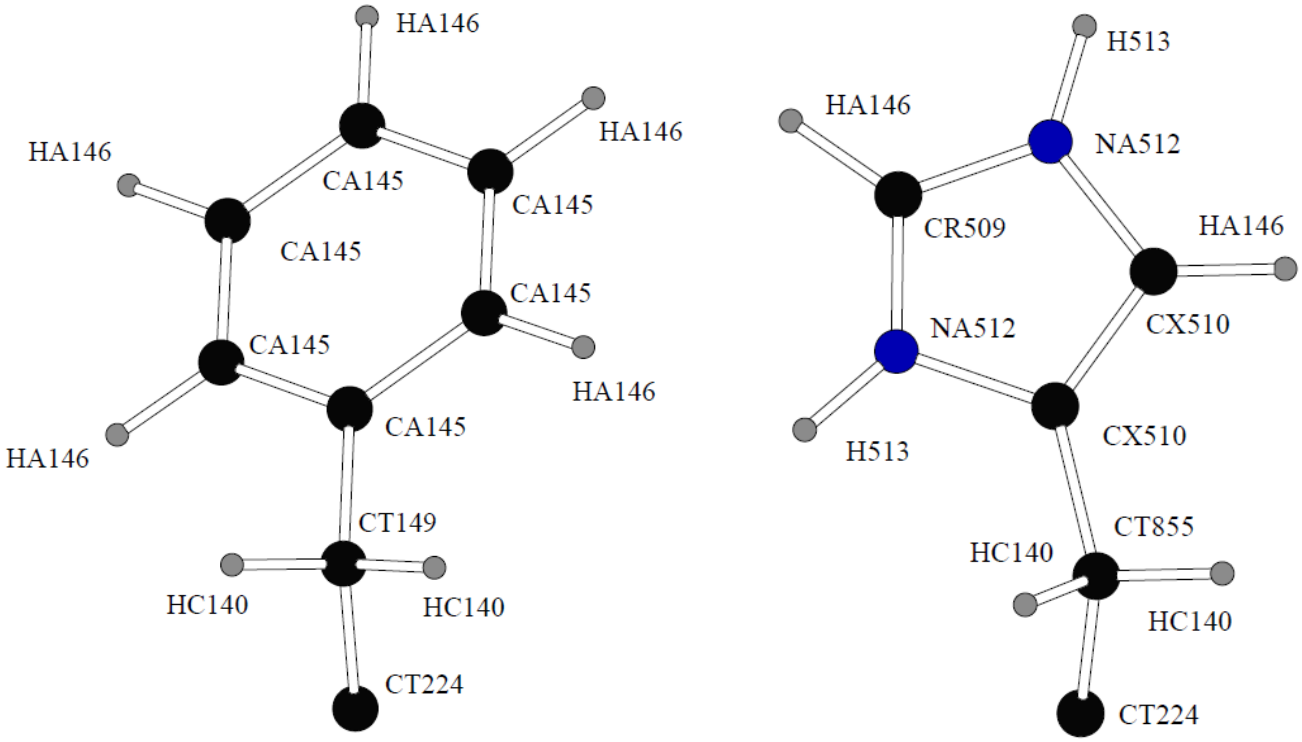

Figure S4. Phenylalanine (left) and protonated histidine (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

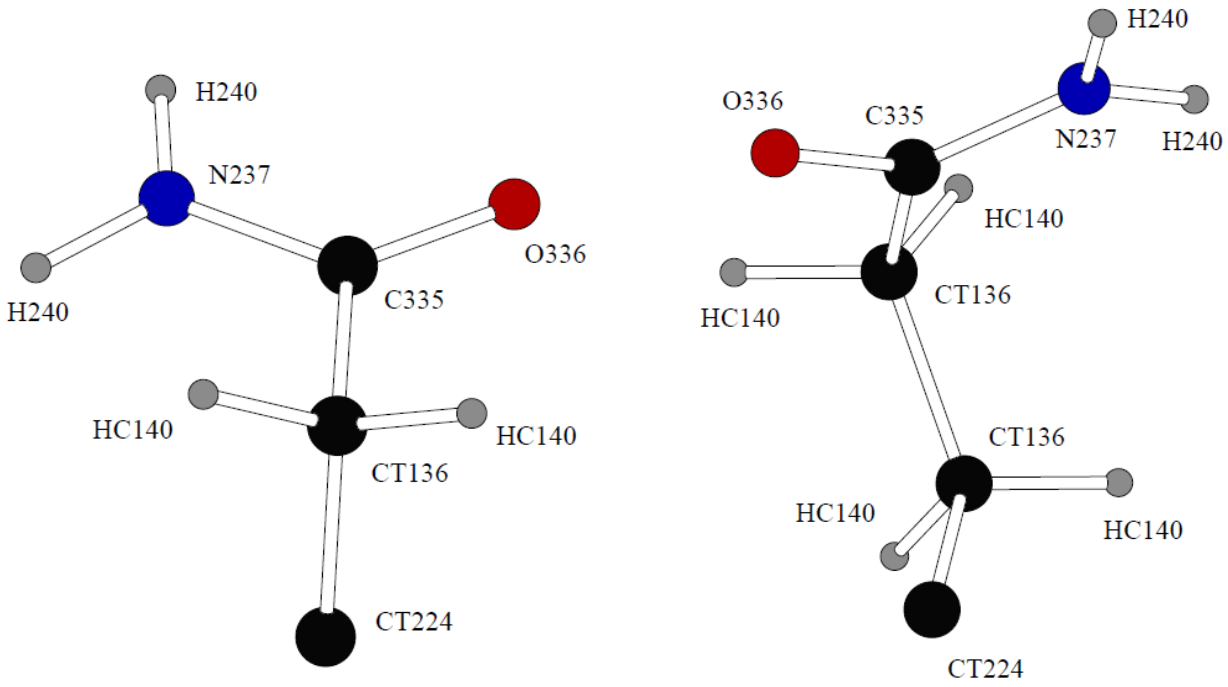

Figure S5. Asparagine (left) and glutamine (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

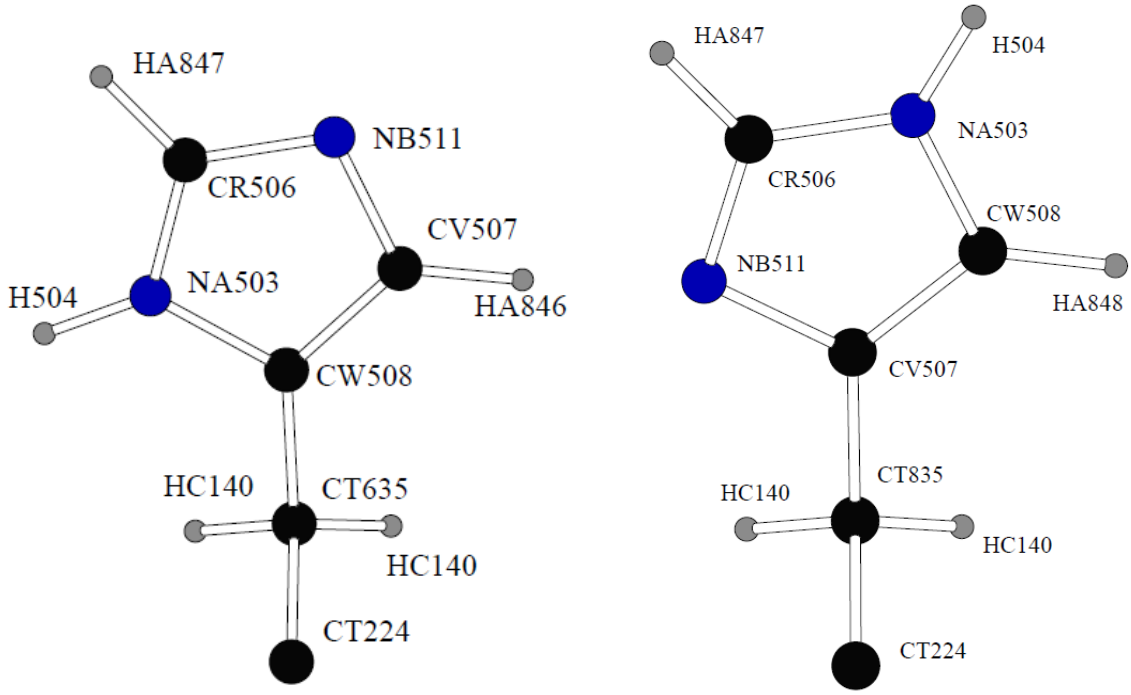

Figure S6. HID (left) and HIE (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

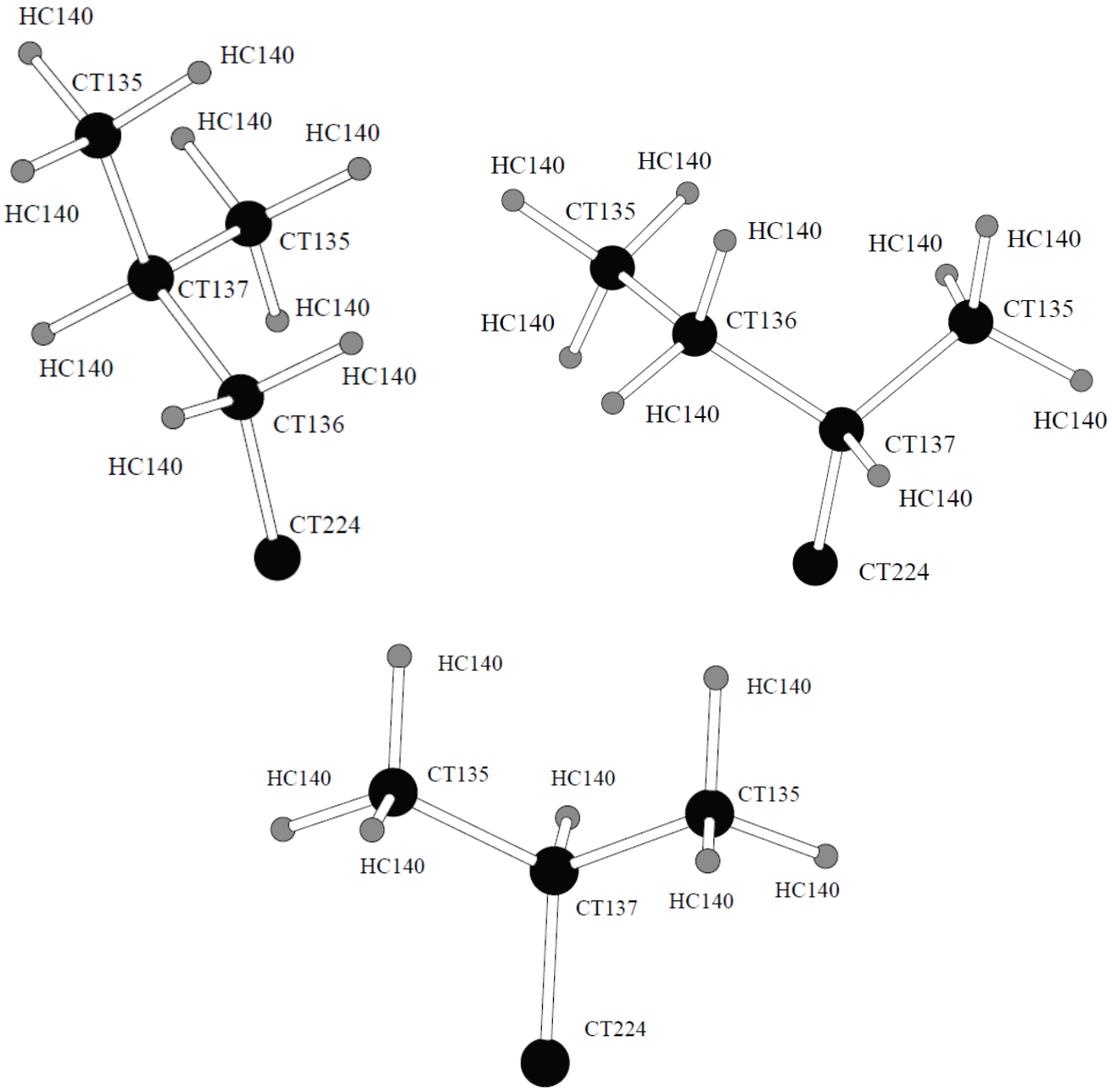

Figure S7. Leucine (top left) isoleucine (top right) and valine (bottom) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

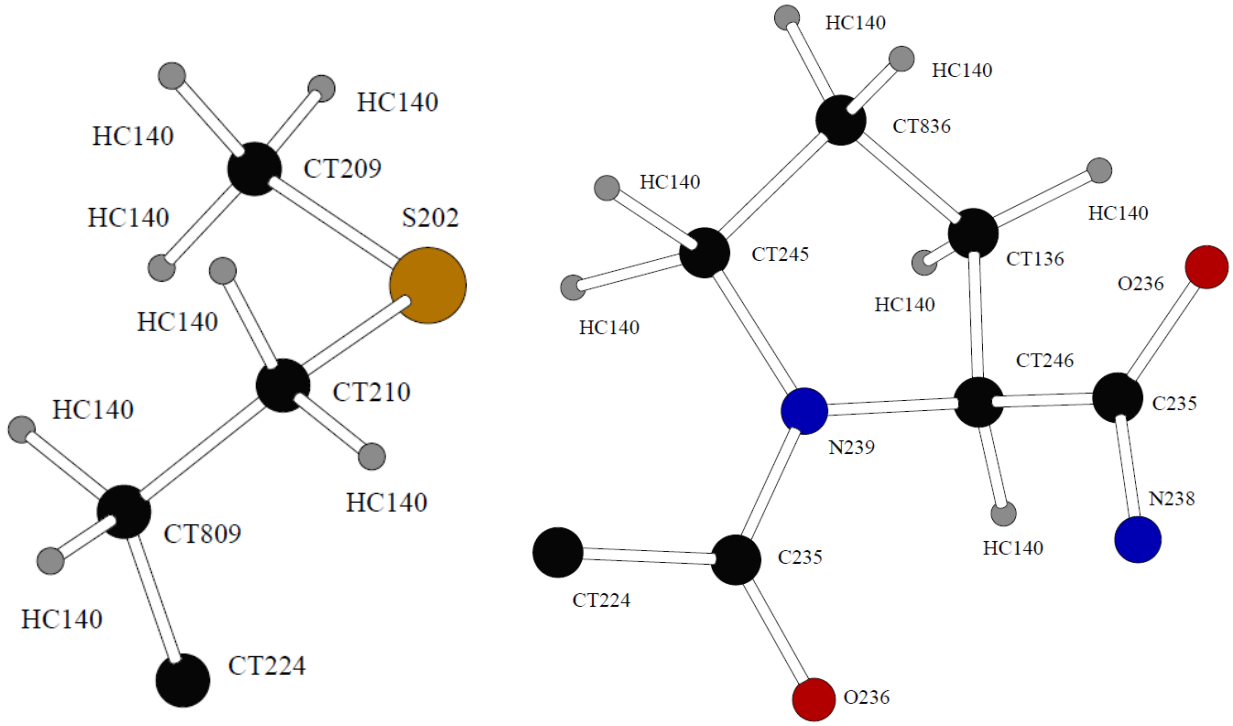

Figure S8. Methionine (left) and proline (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1 (except the amide nitrogen in proline as noted).

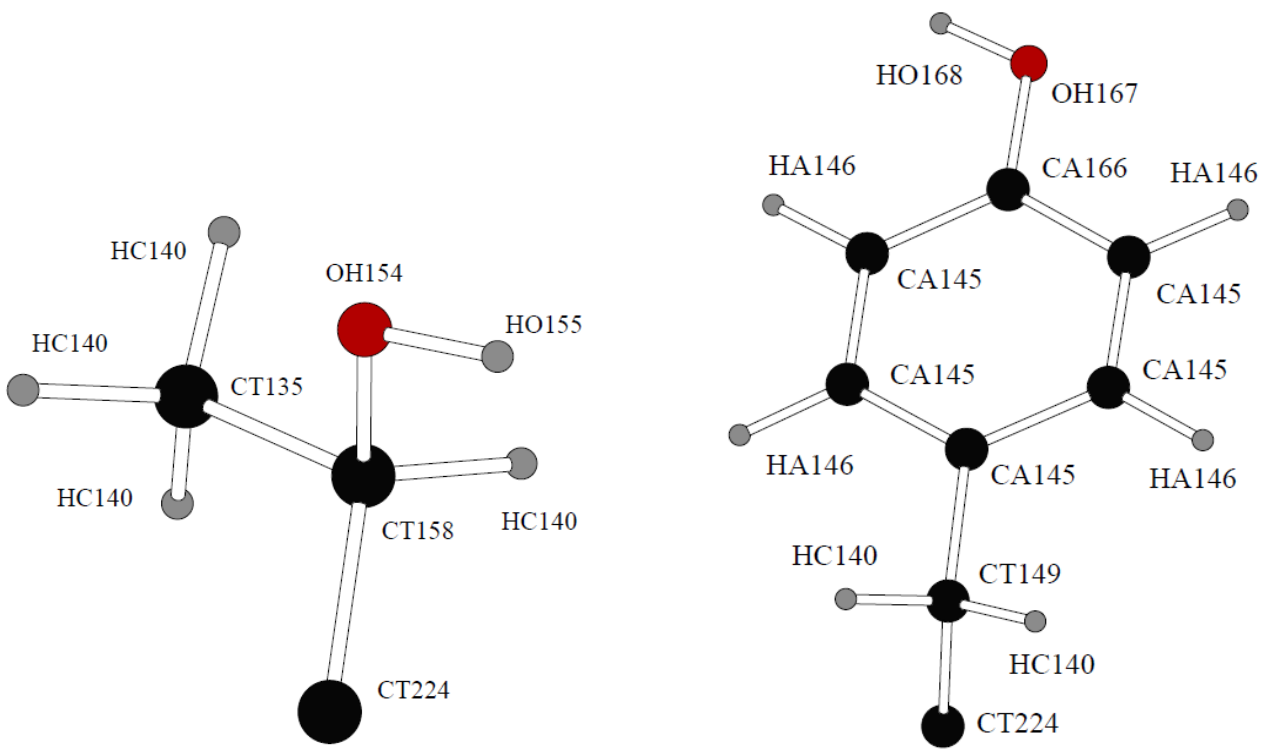

Figure S9. Threonine (left) and tyrosine (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

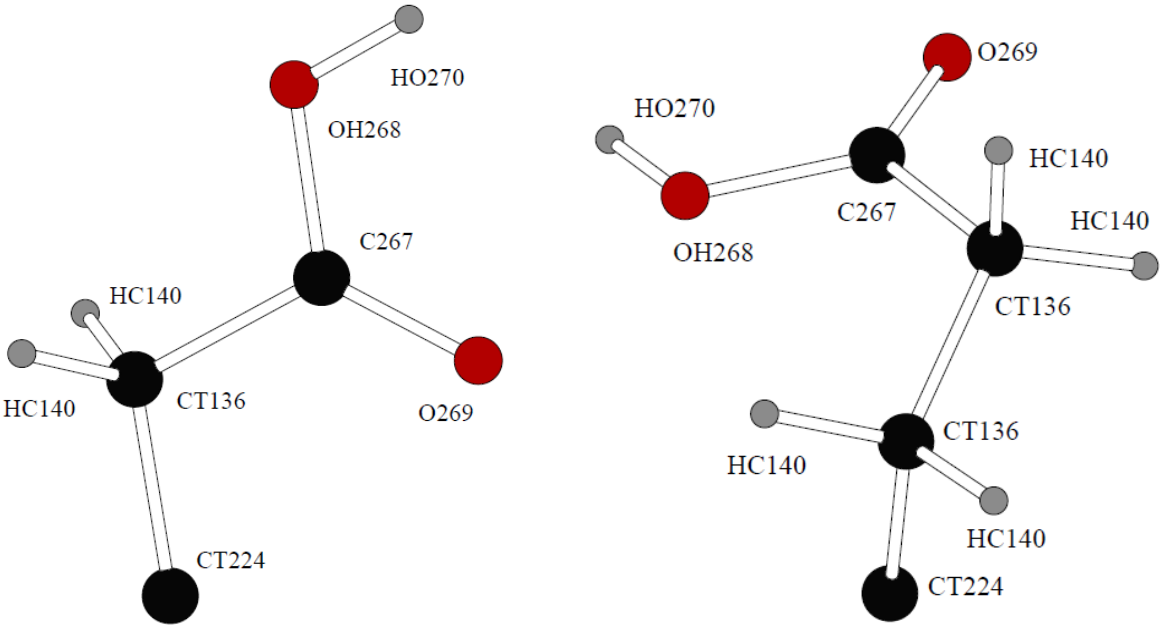

Figure S10. Protonated aspartic acid (left) and protonated glutamic acid (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

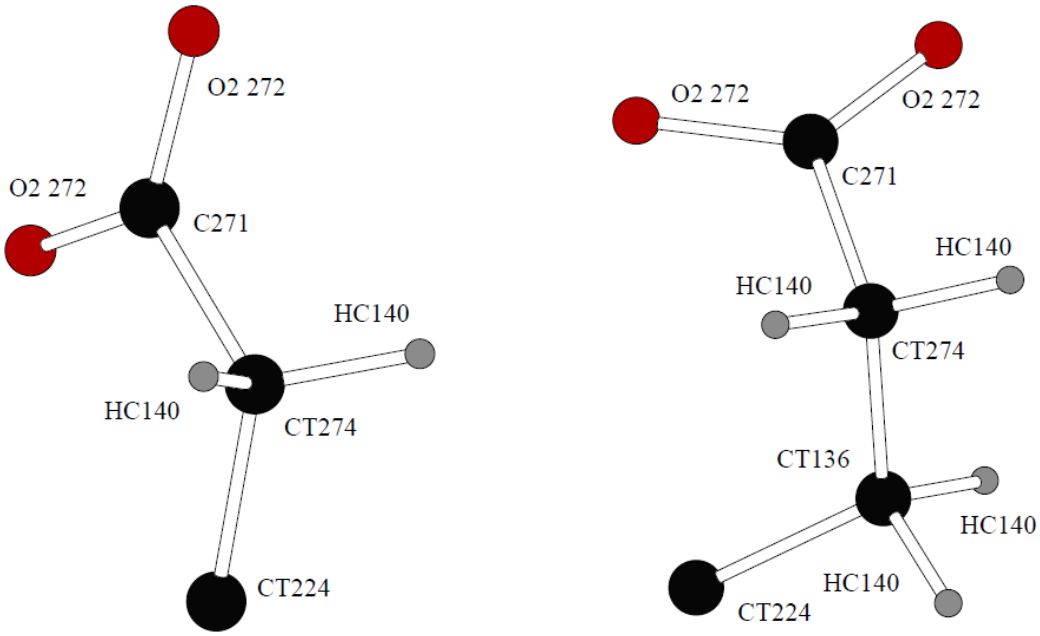

Figure S11. Aspartic acid (left) and Glutamic acid (right) side-chain atomtype assignments. The backbone atomtypes are as in Figure S1.

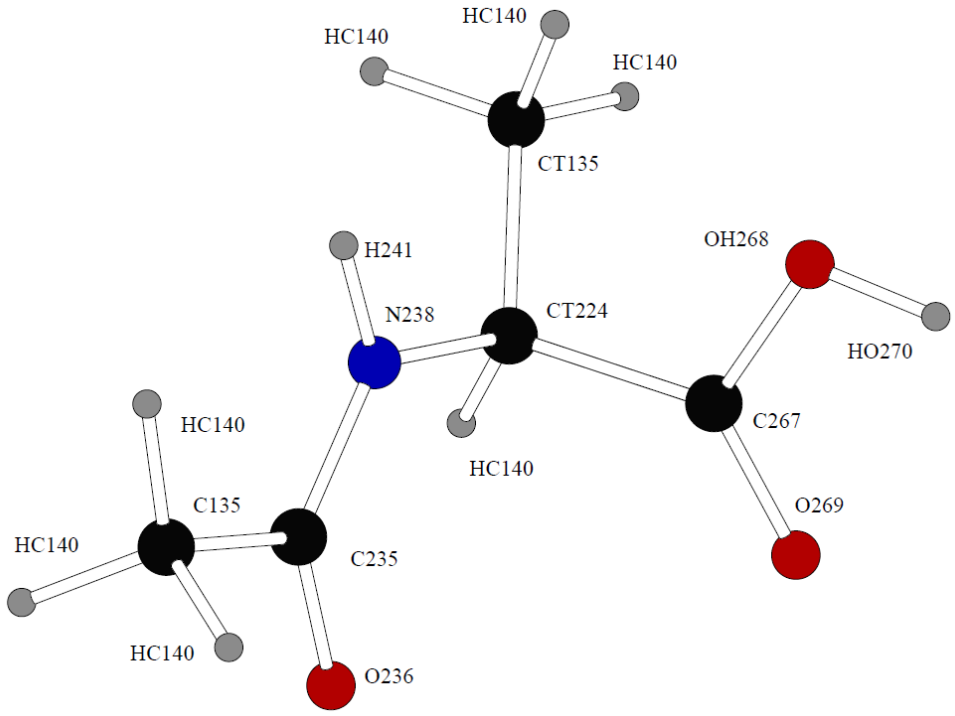

Figure S12. Systems used to fit parameters for protonated C-terminus.

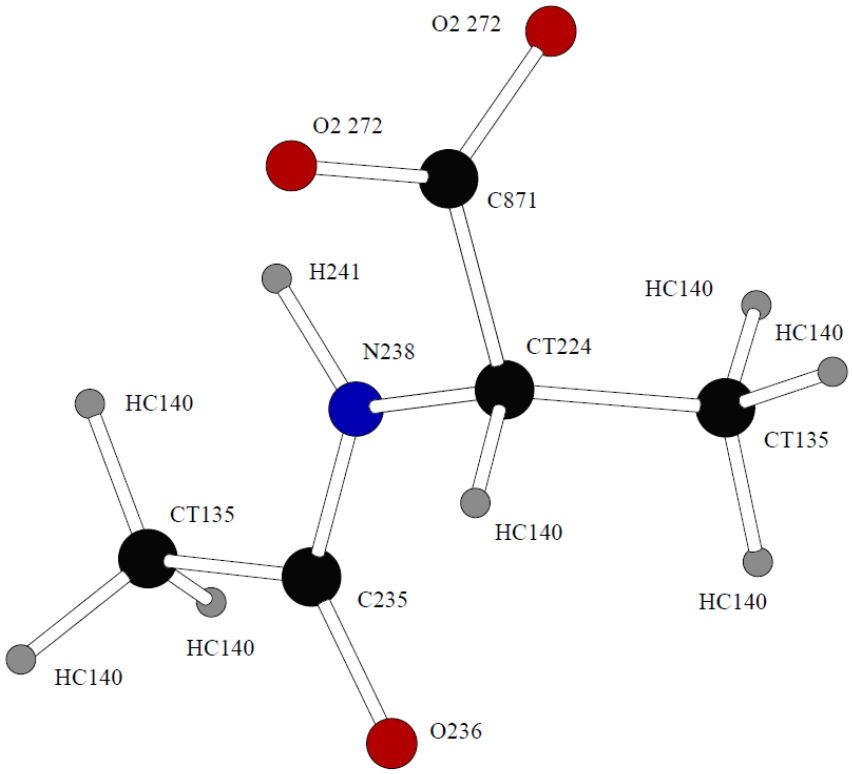

Figure S13. Systems used to fit parameters for deprotonated C-terminus.

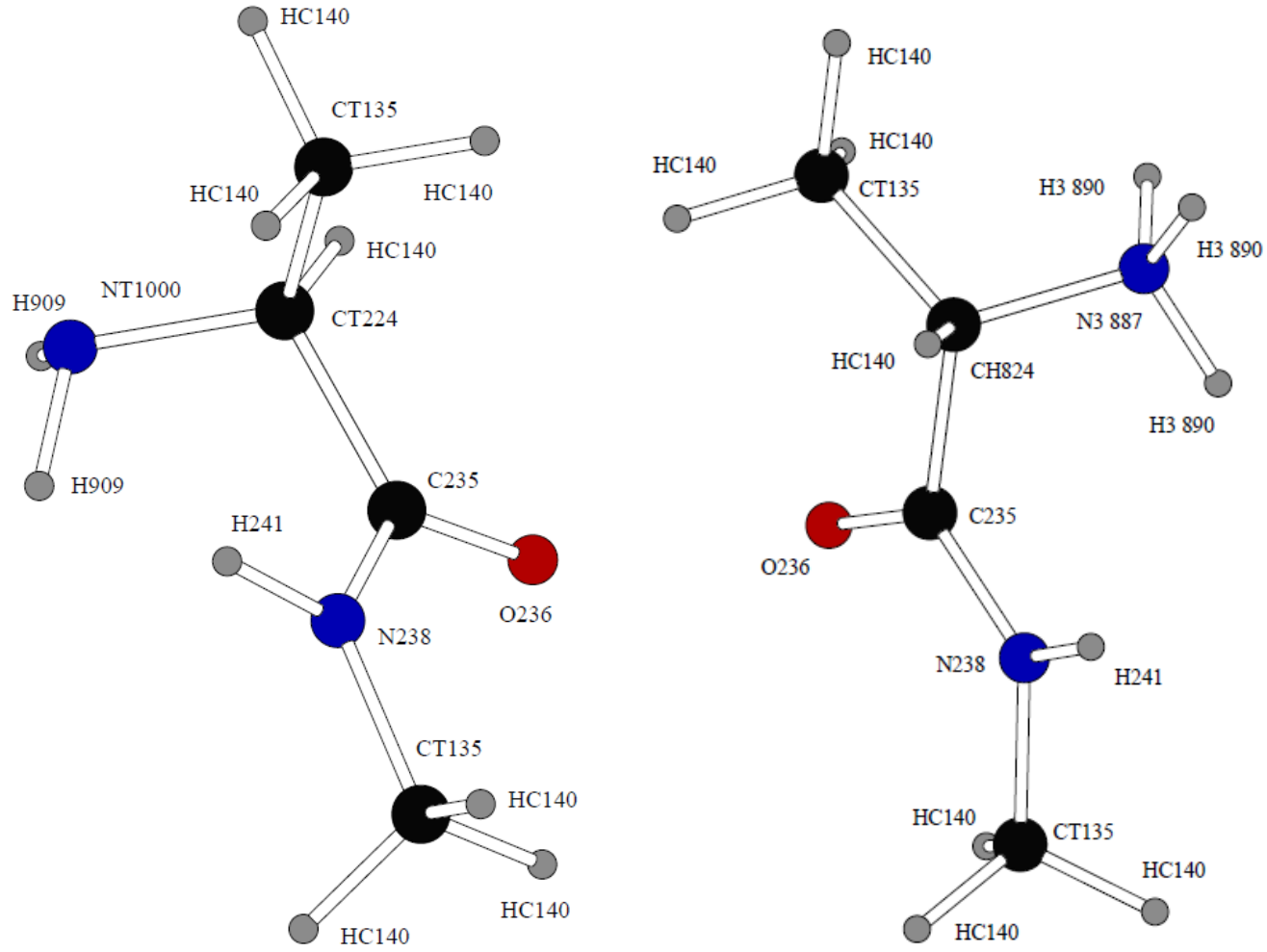

Figure S14. Systems used to fit parameters for deprotonated (left) and protonated (right) N-terminus.

Bond stretching and angle bending parameters taken from OPLS-AA (as implemented in BOSS version 4.8. For BOSS reference see: Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689-1700) without refitting.

All the atomtypes have R_{cut} threshold (as used in Equation 6) values equal to 0.8Å for charges and dipoles.

Table S1. Nonbonded Parameters in the POSSIM model. σ and ε are the Lennard-Jones constants; α^{-1} stands for the inverse polarizability.

Symbolic[a] and numeric atom types	**Description**	**Charge, electrons**	**σ,[b] Å**	**ε,[b] kcal/mol**	**α^{-1},[c] Å^{-3}**
General types (including some parameters derived previously)					
OW 111	O in POSSIM water	–0.702	3.270	0.175	1.300
HW 112	H in POSSIM water	0.351	0.0	0.0	3.300
CT 135	CH_3, alkanes	–0.180	3.500	0.066	0.5069
CT 136	CH_2, alkanes	–0.120	3.500	0.066	0.5069
CT 137	CH, alkanes	–0.06	3.500	0.066	0.5069
CT 138	CH_4, methane	–0.240	3.500	0.066	0.5069
HC 140	H on CT	0.060	2.500	0.030	–
CT 223	Gly C-α	0.048	3.500	0.066	0.5069
CT 224	peptide C-α	0.108	3.500	0.066	0.5069
C 235	C(=O) in amide, peptides	0.529	3.400	0.086	0.7797
O 236	C(=O) in amide, peptides	–0.558	3.220	0.152	0.8948
N 238	N in 2° amide, peptides	–0.378	3.350	0.170	0.6307
H 241	H(-N) in 2° amide, peptides	0.239	0.0	0.0	–
CT 242	H_3C(-N) in 2° N-Me amide	–0.012	3.500	0.066	0.5069
Methylammonium ($CH_3NH_3^+$) and Lys					
N3 287	RNH_3^+, Lys	–0.080	3.600	0.280	1.000
H3 290	RNH_3^+, Lys	0.360	0.0	0.00	–
CT 291	$CH_3NH_3^+$	–0.18	3.50	0.066	0.5069
CT 292	$RCH_2NH_3^+$, Lys C-ε	–0.12	3.50	0.066	0.5069
Benzene, phenol, Phe, Tyr, etc.					
CA 145	C in benzene, phenol (except C(-OH)), etc	–0.100	3.550	0.070	3.260
HA 146	H in benzene, phenol, etc.	+0.100	2.420	0.030	–
CT 149	RCH_2(-aryl), Phe and Tyr C-β	–0.020	3.500	0.066	0.5069
CA 166	C(-OH) in phenol, Tyr	–0.025	3.550	0.070	1.000

OH 167	C(-OH) in phenol, Tyr	–0.450	3.285	0.180	2.950
HO 168	-OH in phenol, Tyr	0.475	0.0	0.0	3.910
Thiols, sulfides (thioethers), disulfides, Cys, Cyx (disulfide Cys), Met					
SH 200	RSH, Cys	–0.266	3.700	0.450	0.5565
S 202	RSR', Met	–0.130	3.700	0.450	0.5565
S 203	R'SSR"$_3$, Cyx	–0.065	3.740	0.370	0.5565
HS 204	RSH, Cys	0.201	0.0	0.0	5.0828
CT 206	RCH$_2$SH, Cys C-β	–0.055	3.500	0.066	0.5069
CT 209	RSCH$_3$, RSSCH$_3$, Met C-ε	–0.115	3.500	0.066	0.5069
CT 210	RCH$_2$SR', Met C-γ	–0.055	3.500	0.066	0.5069
CT 217	CH$_3$SH	–0.115	3.500	0.066	0.5069
CT 809	Met C-β	–0.190	3.500	0.066	0.5069
Carboxylic acids, Asp, Glu, Ash, Glh (protonated Asp and Glu respectively)					
C 267	RCOOH, Ash C-γ, Glh C-δ	0.780	3.200	0.090	1.074
OH 268	-OH in RCOOH	–0.590	2.900	0.160	1.31
O 269	=O in RCOOH	–0.610	3.400	0.160	1.35
HO 270	-OH in RCOOH	0.420	0.0	0.0	–
C 271	RCOO$^-$, Asp C-γ, Glu C-δ	0.700	3.750	0.105	1.000
O2 272,	RCOO$^-$, Asp, Glu	–0.800	3.275	0.290	1.500
CT 273,	CH$_3$COO$^-$	–0.280	3.500	0.066	0.5069
CT 274	RCH$_2$COO$^-$, Asp C-β, Glu C-γ	–0.220	3.500	0.066	0.5069
Imidazole, Imidazolium, Hid, Hie, Hip (His with: H on N-δ, H on N-ε, H on both N resp.)					
NA 503	imidazole, Hid N-δ1, Hie N-ε2	–0.256	3.254	0.175	2.203
H 504	H(-NA) in imidazole, Hid, His	0.262	0.0	0.0	3.026
CR 506	imidazole, Hid and Hie C-ε1	0.287	3.550	0.070	1.8329
CV 507	imidazole, Hid C-δ2, Hie C-γ	0.188	3.550	0.070	1.8329
CW 508	imidazole, Hid C-γ, Hie C-δ2	–0.268	3.550	0.070	1.8329
CR 509	imidazolium, Hip C-ε1	0.385	3.550	0.070	3.260
CX 510	imidazolium, Hip C-γ C-δ2	0.215	3.550	0.070	3.260
NB 511	imidazole, Hid N-ε2, Hie N-δ1	–0.587	3.254	0.175	1.190
NA 512	imidazolium, Hip N-δ1 N-ε2	–0.5075	3.355	0.270	2.500
H 513	H(-NA) in imidazolium, Hip	0.450	0.0	0.0	–
CT 635	Hid C-β	0.067	3.500	0.066	0.5069

CT 835	Hie C-β	–0.001	3.500	0.066	0.5069
HA 146	H(-CR) and H(-CX) in imidazolium, Hip	0.100	2.420	0.030	–
HA 846	H(-CV) in imidazole, Hid	0.119	2.500	0.030	–
HA 847	H(-CR) in imidazole, Hid, Hie	0.068	2.500	0.030	–
HA 848	H(-CW) in imidazole, Hie	0.187	2.500	0.030	–
CT 855	Hip C-β	0.148	3.500	0.066	0.5069
Alcohols, Ser, Thr					
OH 154	ROH, Ser, Thr	–0.580	3.185	0.170	–
HO 155	ROH, Ser, Thr	0.350	0.0	0.0	1.68
CT 157	RCH$_2$OH, Ser C-β	0.110	3.500	0.066	0.5069
CT 158	RR'CHOH, Thr C-β	0.170	3.500	0.066	0.5069
CT 857	CH$_3$OH	0.050	3.500	0.066	0.5069
Acetamide, Asn, Gln					
N 237	RCONH$_2$, Asn, Gln	–0.501	3.250	0.170	0.400
H 240	RCONH$_2$, Asn, Gln	0.274	0.0	0.0	–
C 335	RCONH$_2$, Asn, Gln	0.449	3.400	0.086	0.9130
O 336	RCONH$_2$, Asn, Gln	–0.496	3.170	0.152	4.000
Methylguanidinium, Arg					
N2 300	[(NH$_2$)$_2$CNHCH$_3$]$^+$, Arg N-η	–0.920	3.420	0.170	1.400
H3 301	[(NH$_2$)$_2$CNHCH$_3$]$^+$, Arg	0.454	0.0	0.0	–
CA 302	[(NH$_2$)$_2$CNHCH$_3$]$^+$, Arg C-ζ	0.862	3.550	0.050	2.200
N2 303	[(NH$_2$)$_2$CNHCH$_3$]$^+$, Arg N-ε	–0.464	3.420	0.170	1.400
H3 304	[(NH$_2$)$_2$CNHCH$_3$]$^+$, Arg	0.390	0.0	0.0	–
CT 307	[(NH$_2$)$_2$CNHCH$_3$]$^+$	0.056	3.500	0.066	0.5069
CT 308	Arg C-γ	–0.120	3.500	0.066	0.5069
CT 807	Arg C-δ	0.116	3.500	0.066	0.5069
Pyrrole, Trp					
NA 803	pyrrole, Trp N-ε1	–0.387	3.750	0.120	0.700
H 804	H(-NA) in pyrole, Trp	0.387	0.0	0.0	–
CN 845	Trp C-ε2	0.0	3.550	0.070	3.26
CW 849	pyrrole, Trp C-δ1	–0.115	3.550	0.070	3.26
CB 850	Trp C-δ2	0.0	3.550	0.070	3.26

CS 851	pyrrole	–0.115	3.550	0.070	3.26
CA 853	pyrrole, Trp C-ε3 C-ζ2 C-ζ3 C-η2	–0.115	3.550	0.070	3.26
HA 854	H(-C_) in pyrrole, Trp	0.115	2.420	0.030	–
C* 945	Trp C-γ	0.0	3.550	0.070	3.26
Proline					
N 239	N in 3° amide, Pro	–0.027	3.350	0.170	0.6307
CT 245	RCONR'$\underline{C}H_2$, Pro C-δ	–0.212	3.500	0.066	0.5069
CT 246	RCONR'$\underline{C}$HR"R''', Pro C-α	–0.012	3.500	0.066	0.5069
CT 836	Pro C-γ	–0.020	3.500	0.066	0.5069
Peptide termini					
C 871	$\underline{C}OO^-$ peptide terminus	0.600	3.750	0.105	1.000
CT 824	C-α in NH_3^+ peptide terminus	0.108	3.500	0.066	0.5069
N3 887	$\underline{N}H_3^+$ peptide terminus	0.202	3.684	0.075	1.000
H3 890	$N\underline{H}_3^+$ peptide terminus	0.210	0.0	0.0	–
NT 900	$R\underline{N}H_2$	–0.772	3.3562	0.154	0.962
CT 903	$\underline{C}H_3NH_2$	0.095	3.500	0.066	0.5069
H 909	$RN\underline{H}_2$	0.2485	0.0	0.0	2.946
H 911	$C\underline{H}_3NH_2$	0.060	2.500	0.030	–
NT 1000	$\underline{N}H_2$ peptide terminus	–0.665	3.3562	0.154	0.926

[a]Symbolic types adapted from standard OPLS/BOSS types (see Figures S1-S14 and Table S1). Briefly: HC—H on CT; HA—H on aromatic C; H3—H on N3; H—H on various N; HO—H in OH alcohol; HS—H in SH thiol; CT—tetrahedral C; CA—aromatic C; C—carbonyl C; CB, CN, CR, CS, CV, CW, CX, C*—imidazole, imidazolium, His, pyrrole, Trp (see Figs. S2, S4, S6); N—amide N; NA, NB, N2—sp2 N; N3, NT—sp3 N, O—carbonyl O; OH—alcohol O; O2—carboxylate O, SH—thiol S; S—sulfide/disulfide S. [b]σ and ε are the Lennard-Jones constants (see Eq. 7). [c]α^{-1} stands for the inverse polarizability.

Table S2. Torsional parameters in the POSSIM model.

Torsion[a]	**Description**	**V_1, kcal/mol**	**V_2, kcal/mol**	**V_3, kcal/mol**
General types (including some parameters derived previously)				
HC–CT–CT–HC		0.0	0.0	0.3640
HC–CT–CT–CT		0.0	0.0	0.2100
CT–CT–CT–CT		0.980	–0.570	0.6400
HO–OH–CT–HC	methanol	0.0	0.0	0.3500
HC–CT–C–O	acids	0.0	0.0	0.0
CT–C–OH–HO	acids	1.244	6.048	0.0
O–C–OH–HO	acids	0.0	5.500	0.0
CT–CT–C–OH	acids	0.0	–2.140	0.0
HC–CT–CT–C	acids	0.0	0.0	0.185
CT–CT–CT–C	acids	0.223	0.706	0.0
HC–CT–C–O2	CH_3COO^-	0.0	0.0	0.0
HC–CT–N3–H3	RNH_3^+	0.000	0.000	0.249
HC–CT–CT–N3	RNH_3^+	0.000	0.000	0.210
CT–CT–N3–H3	RNH_3^+	0.000	0.000	0.355
HC–CT–SH–HS	CH_3SH	0.0	0.0	0.3916
HC–CT–S–CT	$RSCH_3$	0.0	0.0	0.515
CT–S–S–CT	CH_3SSCH_3, RSSR’	0.0	–6.850	1.711
HC–CT–S–S	$RSSCH_3$	0.0	0.0	0.366
Z–CA–X–Y[b]	improper torsion[b]	0.0	2.2	0.0
Z–N–X–Y[b]	improper torsion[b]	0.0	2.0	0.0
O–C–X–Y[b]	improper torsion[b]	0.0	21.0	0.0
Acetamide, NMA, peptides, etc.				
C–N–CT–HC	NMA	0.0	0.0	–0.2500
H–N–CT–HC	NMA	0.0	0.0	0.0
CT–C–N–CT	NMA	1.160	–1.733	0.0
O–C–N–CT	NMA	0.0	6.089	0.0
CT–C–N–H	acetamide, NMA	0.0	4.900	0.0
O–C–N–H	acetamide, NMA	0.0	4.900	0.0
C–N–CT–C	C–N–Cα–C, ϕ	2.000	–0.500	–3.772

N–CT–C–N	N–Cα–C–N, ψ	–2.837	3.942	–3.328
C–N–CT–CT	C–N–Cα–Cβ, ϕ'	–2.718	1.757	5.202
CT–CT–C–N	Cβ–Cα–C–N, ψ'	0.372	–0.915	3.321
HC–CT–C–N	NMA, peptides	0.0	0.0	–0.1365
N–CT–C–O	acetamide, NMA, peptides	0.0	0.0	0.0
CT–CT–C–O	peptides	0.0	0.0	0.0
HC–CT–C–O	acetamide, NMA, peptides	0.0	0.0	0.0
HC–CT–CT–N	peptides	0.0	0.0	0.348
HC–CT–CT–C	peptides	0.0	0.0	0.348
C–CT–N–H	peptides	0.0	0.0	0.0
CT–CT–N–H	peptides	0.0	0.0	0.0
HC–CT–C–N	acetamide	0.0	0.0	–0.140
N–C–C–O	NMA (improper)	0.0	21.00	0.0
C–N–C–H	NMA (improper)	0.0	2.000	0.0
Aromatics				
CA–CA–CA–CA	benzene	0.0	7.450	0.0
CA–CA–CA–HA	benzene	0.0	5.783	0.0
HA–CA–CA–HA	benzene	0.0	7.250	0.0
CA–CA–OH–HO	phenol	0.0	1.865	0.0
CA–CA–CA–OH	phenol	0.0	7.452	0.0
HA–CA–CA–OH	phenol	0.0	7.250	0.0
HC–CT–CA–CA	Phe, Tyr, Trp	0.0	0.0	0.0
CA–CA–CA–CT	Phe, Tyr, Trp	0.0	5.783	0.0
CT–CA–CA–CA	Phe, Tyr, Trp	0.0	7.250	0.0
HA–CW–CV–NA	imidazole	0.0	5.783	0.0
HA–CR–NA–CW	imidazole	0.0	5.783	0.0
CV–NB–CR–HA	imidazole	0.0	5.783	0.0
CR–NA–CW–HA	imidazole	0.0	5.783	0.0
NB–CV–CW–HA	imidazole	0.0	5.783	0.0
HA–CV–NB–CR	imidazole	0.0	5.783	0.0
CV–CW–NA–H	imidazole	0.0	5.783	0.0
H–NA–CR–NB	imidazole	0.0	5.783	0.0
HA–CW–CV–HA	imidazole	0.0	7.250	0.0

HA–CW–NA–H	imidazole	0.0	7.250	0.0
HA–CR–NA–H	imidazole	0.0	7.250	0.0
NA–CW–CV–NB	imidazole	0.0	7.450	0.0
CW–NA–CR–NB	imidazole	0.0	7.450	0.0
CW–CV–NB–CR	imidazole	0.0	7.450	0.0
CV–NB–CR–NA	imidazole	0.0	7.450	0.0
CV–CW–NA–CR	imidazole	0.0	7.450	0.0
CR-NB-CV-HA and CV-NB-CR-HA	imidazole	0.0	5.783	0.0
X–X–X–X[c]	imidazolium, Hid, Hie, Hip	0.0	7.450	0.0
H–X–X–X[c] (except H–N–C–C)	imidazolium, Hid, Hie, Hip	0.0	5.783	0.0
H–N–C–C	imidazolium	0.0	0.0	0.0
H–X–X–H[c]	imidazolium	0.0	7.250	0.0
Lysine				
N–CT–CT–CT	Lys χ_1	–3.862	–0.355	5.035
C–CT–CT–CT	Lys χ_1'	–6.000	–3.905	0.454
CT–CT–CT–N3	Lys χ_4	0.286	–2.595	–5.020
CT–CT–CT–CT		0.980	–0.570	0.640
Arginine and methylguanidinium				
HC–CT–N2–H3		0.0	0.0	0.0
HC–CT–N2–CA		0.0	0.0	0.045
CT–N2–CA–N2		0.0	4.000	0.0
H3–N2–CA–N2		0.0	2.400	0.0
N–CT–CT–CT	Arg χ_1	5.000	–2.500	4.500
C–CT–CT–CT	Arg χ_1'	2.300	1.650	–3.47
CT–CT–CT–CT	Arg χ_2	3.300	–1.430	4.840
CT–CT–CT–N2	Arg χ_3	–1.650	1.970	5.440
CT–CT–N2–CA	Arg χ_4	3.800	–4.050	1.290
Tryptophan and Pyrrole				
H–N–CA–CA		0.000	–0.500	0.000
CA–CT–CT–N	Trp χ_1	5.57441	–4.07773	2.91885

CA–CT–CT–C	Trp χ_1'	–1.78332	1.03626	5.500
CA–CA–CT–CT	Trp χ_2	–1.000	1.15243	–1.16324
Serine and Threonine				
HC–CT–CT–OH	Ser, Thr	0.0	0.0	0.468
N–CT–CT–OH	Ser χ_1	8.900	–5.375	6.322
C–CT–CT–OH	Ser χ_1'	–1.624	–2.672	–5.882
CT–CT–OH–HO	Ser χ_2	–0.740	–1.303	0.693
N–CT–CT–OH	Thr χ_1	5.000	–1.347	3.219
C–CT–CT–OH	Thr χ_1'	0.070	0.057	–4.721
CT–CT–OH–HO	Thr χ_2	–0.444	–1.317	1.098
N–CT–CT–CT	Thr	3.024	–0.969	3.497
C–CT–CT–CT	Thr	1.299	2.750	1.598
Phenylalanine and Tyrosine				
N–CT–CT–CA	Phe χ_1	0.003	0.639	0.580
C–CT–CT–CA	Phe χ_1'	–0.399	0.016	0.700
CT–CT–CA–CA	Phe χ_2	0.0	0.615	0.000
HC–CT–CT–CA	Phe, Tyr	–2.100	4.700	1.110
CA–CA–OH–HO	Tyr χ_6	0.0	1.865	0.0
CA–CA–CA–OH	Tyr χ_5	0.0	7.452	0.0
N–CT–CT–CA	Tyr χ_1	3.789	–2.784	2.555
C–CT–CT–CA	Tyr χ_2	–0.649	–1.232	3.073
CT–CT–CA–CA	Tyr χ_2	–3.882	2.369	5.000
Cysteine				
N–CT–CT–SH	Cys χ_1	1.286	1.243	–1.827
C–CT–CT–SH	Cys χ_1'	–1.671	0.098	3.455
CT–CT–SH–HS	Cys χ_2	–1.425	–0.164	0.537
HC–CT–CT–SH		0.0	0.0	0.0
HC–CT–SH–HS		0.0	0.0	0.392
Asparagine				
N–CT–CT–C	Asn χ_1	–0.832	–0.373	3.595
C–CT–CT–C	Asn χ_1'	–4.430	–1.053	–0.379

CT–CT–C–N	Asn χ_2	–1.392	0.118	–3.596
CT–CT–C–O	Asn χ_2'	–1.079	0.849	–3.251
HC–CT–CT–C		0.000	0.000	0.210
Glutamine				
N–CT–CT–CT	Gln χ_1	–1.830	3.988	1.397
C–CT–CT–CT	Gln χ_1'	–3.053	4.190	–1.378
CT–CT–CT–C	Gln χ_2	–1.588	2.497	–1.090
CT–CT–C–N	Gln χ_3	4.771	0.891	–0.241
CT–CT–C–O		0.0	0.250	0.0
Histidine (Hid and Hie)				
CT–CW–X–X[c]	Hid	0.0	5.783	0.0
CT–CW–NA–H	Hid	0.0	7.250	0.0
CT–CW–CV–HA	Hid	0.0	7.250	0.0
NB–CR–NA–H	Hid	0.0	7.250	0.0
HC–CT–CT–CW	Hid	0.0	0.0	0.0
HC–CT–CW–X[c]	Hid	0.0	0.0	0.0
HA–CR–NA–H	Hid	0.0	7.250	0.0
N–CT–CT–CW	Hid χ_1	3.172	–0.518	2.645
C–CT–CT–CW	Hid χ_1'	–0.597	–1.821	–2.950
CT–CT–CW–NA	Hid χ_2	–1.222	–1.022	–0.159
CT–CT–CW–CV	Hid χ_2'	1.780	–0.247	–3.516
CT–CV–X–X[c]	Hie	0.0	5.783	0.0
CT–CV–CW–HA	Hie	0.0	7.250	0.0
HA–CR–NA–H	Hie	0.0	7.250	0.0
HA–CW–NA–H	Hie	0.0	7.250	0.0
HC–CT–CT–CV	Hie	0.0	0.0	0.0
HC–CT–CV–X[c]	Hie	0.0	0.0	0.0
N–CT–CT–CV	Hie χ_1	0.641	–2.740	1.133
C–CT–CT–CV	Hie χ_1'	–1.202	–0.907	1.467
CT–CT–CV–NB	Hie χ_2	1.927	–0.416	0.884
CT–CT–CV–CW	Hie χ_2'	–2.506	–1.058	0.331
Protonated Histidine (Hip)				

CT–CX–X–X[c]	Hip	0.0	5.783	0.0
CT–CX–NA–H	Hip	0.0	7.250	0.0
CT–CX–CX–HA	Hip	0.0	7.250	0.0
HA–CX–NA–H	Hip	0.0	7.250	0.0
HA–CR–NA–CX	Hip	0.0	7.250	0.0
HA–CR–NA–H	Hip	0.0	7.250	0.0
HC–CT–CT–CX	Hip	0.0	0.0	0.0
HC–CT–CX–X[c]	Hip	0.0	0.0	0.0
CX–CX–NA–H	Hip	0.0	0.0	0.0
N–CT–CT–CX	Hip χ_1	–0.021	–2.290	3.251
C–CT–CT–CX	Hip χ_1'	–2.282	2.646	4.999
CT–CT–CX–NA	Hip χ_2	–2.148	–1.282	4.661
CT–CT–CX–CX	Hip χ_2'	4.090	1.786	–4.578
Leucine, Isoleucine, Valine				
N–CT–CT–CT	Leu χ_1	1.490	–0.083	–2.246
C–CT–CT–CT,	Leu χ_1'	–0.053	–0.252	4.216
CT–CT–CT–CT	Leu χ_2	1.450	–0.050	1.453
N–CT–CT–CT	Ile χ_1	1.699	–1.078	4.355
C–CT–CT–CT	Ile χ_1'	2.622	0.738	–1.807
CT–CT–CT–CT	Ile χ_2	–0.064	–0.185	0.292
N–CT–CT–CT	Val χ_1	3.198	–1.054	1.988
C–CT–CT–CT	Val χ_1'	1.857	2.177	0.821
Methionine				
HC–CT–S–CT	Met	0.0	0.0	0.647
N–CT–CT–CT	Met χ_1	5.000	–0.870	4.651
C–CT–CT–CT	Met χ_1'	2.500	–1.042	3.069
CT–CT–CT–S	Met χ_2	5.000	–5.000	0.000
CT–CT–S–CT	Met χ_3	–0.266	–0.461	–0.170
HC–CT–CT–S	Met	–0.147	–2.832	0.142
Proline				
CT–CT–CT–N		0.568	0.218	0.852
CT–CT–CT–C		–2.474	3.757	2.471

C–N–CT–HC		0.0	0.0	0.0
CT–CT–N–CT		0.980	–0.570	0.640
CT–N–CT–HC		0.0	0.0	0.0
C–CT–N–CT		4.914	5.000	–0.234
Aspartic and Glutamic Acids				
HC–CT–CT–C	Asp, Glu	0.0	0.0	0.210
N–CT–CT–C	Asp χ_1	–7.820	–7.830	7.550
C–CT–CT–C	Asp χ_1'	–6.330	3.210	5.610
CT–CT–C–O2	Asp χ_2	1.400	1.890	3.100
N–CT–CT–CT	Glu χ_1	–9.930	–1.010	–2.360
C–CT–CT–CT	Glu χ_1'	–3.990	–0.270	4.700
CT–CT–CT–C	Glu χ_2	–9.060	–9.940	9.930
CT–CT–C–O2	Glu χ_3	0.000	0.250	0.000
Protonated Aspartic (Ash) and Glutamic (Glh) Acids				
HC–CT–CT–C	Ash and Glh	0.0	0.0	0.210
HC–CT–C–O,	Ash and Glh	0.0	0.0	0.0
HC–CT–C–OH	Ash and Glh	0.0	0.0	0.0
O–C–OH–HO	Ash and Glh	0.0	5.500	0.0
CT–C–OH–HO	Ash and Glh	1.244	6.048	0.0
N–CT–CT–C	Ash χ_1	–5.000	–1.000	–1.500
C–CT–CT–C	Ash χ_1'	–0.704	1.100	1.500
CT–CT–C–O,	Ash χ_2	–5.000	2.700	–1.000
CT–CT–C–OH	Ash χ_2'	–1.000	1.500	0.594
N–CT–CT–CT	Glh χ_1	–0.319	4.308	–1.157
C–CT–CT–CT	Glh χ_1'	2.011	3.160	3.800
CT–CT–CT–C	Glh χ_2	3.500	1.252	0.000
CT–CT–C–O	Glh χ_3	–5.000	2.700	–1.000
CT–CT–CT–OH	Glh χ_3'	–1.000	1.5000	0.594
Peptide termini				
O–C–OH–HO	COOH terminus	0.0	5.500	0.0
CT–C–OH–HO	COOH terminus	1.244	6.048	0.0
CT–CT–C–O	COOH terminus	2.840	0.0	2.090

N–C–C–OH	COOH terminus	4.343	–0.698	–3.634
N–CT–C–O2	COO^- terminus	5.000	2.980	0.000
CT–CT–C–O2	COO^- terminus	0.000	0.370	1.000
HC–CT–N3–H3	NH2 terminus	0.000	0.000	0.249
HC–CT–CT–N3	NH2 terminus	0.000	0.000	0.210
CT–CT–N3–H3	NH2 terminus	0.151	1.648	0.920
N3–CT–CT–C	NH2 terminus	4.310	1.200	–3.070
HC–CT–NT–H	$NH3^+$ terminus	0.000	0.000	0.249
HC–CT–CT–NT	$NH3^+$ terminus	0.000	0.000	0.210
CT–CT–NT–H	$NH3^+$ terminus	0.792	3.914	–0.435
NT–CT–CT–C	$NH3^+$ terminus	4.310	1.200	–3.070

[a]Symbolic types adapted from standard OPLS/BOSS types (see Figures S1-S14 and Table S1). Briefly: HC—H on CT; HA—H on aromatic C; H3—H on N3; H—H on various N; HO—H in OH alcohol; HS—H in SH thiol; CT—tetrahedral C; CA—aromatic C; C—carbonyl C; CB, CN, CR, CS, CV, CW, CX, C*—imidazole, imidazolium, His, pyrrole, Trp (see Figs. S2, S4, S6); N—amide N; NA, NB, N2—sp2 N; N3, NT—sp3 N, O—carbonyl O; OH—alcohol O; O2—carboxylate O, SH—thiol S; S—sulfide/disulfide S. [b]X, Y, Z can be any atomtype. [c]X stands for either C or N. All the improper torsional parameters were adopted from OPLS–AA (as implemented in BOSS version 4.8 see: Jorgensen, W. L.; Tirado–Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689–1700) without any changes.

Details Regarding Side-Chain Fitting Results

Serine

For the serine amino acid, the backbone model was adopted without change from the alanine set,[11] and the side-chain parameters (with the exception of the torsional parameters noted below) were taken from methanol.[10] The side-chain torsional parameters related to χ_1 (χ_1 and χ_1') and χ_2 were fitted.

The results of comparing the final conformational analysis for serine dipeptide simulated with POSSIM compared to the quantum mechanical data are given in Table 2. The average energy RMSD is only 0.19 kcal/mol, compared with 0.34 kcal/mol in the previous version of the polarizable force field that we used (we will denote it as PFF)[7] and the same 0.34 kcal/mol in the refitted OPLS-AA.[7] The average error in the key dihedral angles (here and in the other cases, the key dihedrals are the backbone ϕ and ψ and side-chain χ or χ‘ torsions) as calculated with POSSIM is 6.3° vs. 8.1° and 4.9° in PFF and OPLS-AA, respectively. Overall, the POSSIM model performs well in simulating the serine residue.

Phenylalanine

The model for the phenylalanine residue was produced by merging the alanine backbone and benzene parameters from POSSIM with the torsional energy parameters for the χ_1 and χ_2 side-chain dihedrals fit in this work. The target and fitting quantum mechanical data were taken from previous work,[8] as was done for the majority of the other residues. The results of fitting the torsional parameters related to the side-chain dihedral angles are shown in Table 3. The energy RMS deviation was only 0.02 kcal/mol, same as for PFF and lower than the 0.15 kcal/mol for OPLS-AA,[7] although all the errors are sufficiently small. The average error in the key dihedrals for POSSIM was 8.9°, between the 9.5° and 7.5° values for PFF and OPLS-AA, respectively.

Cysteine

Alanine dipeptide and the CH_3SH molecule were used for the non-bonded parameters of this residue, as well as for the other parameters except for the torsions related to the χ_1 and χ_2 side-chain dihedrals. The results of fitting of these torsional parameters are shown in Table 4. It can be seen that the geometry is consistently close to the quantum mechanical target (with only one slight exception of

the ψ angle in the first conformation that differs by about 30°). The overall energy error is 0.25 kcal/mol, comparable with the 0.27 kcal/mol of PFF and somewhat better than the 0.35 kcal/mol error with OPLS-AA. The average angular deviation is 6.0°, not very different from PFF (4.8°) and OPLS-AA (5.8°) values and definitely in the acceptable range.

Asparagine and Glutamine

Asparagine and glutamine dipeptides were created by combining the alanine backbone and the acetamide parameters developed previously.[10] Results for the conformational studies and side-chain torsional fitting for these systems are given in Tables 5 and 6. For asparagine, the RMSD of the conformational energies was 0.14 kcal/mol, between the OPLS-AA value of 0.16 kcal/mol and the PFF result of 0.02 kcal.mol.[7] These deviations are well within the target accuracy. The average error in the key dihedrals (ϕ, ψ, χ_1 and χ_2) as compared to the quantum mechanical data was 8.6° with POSSIM, 8.7° with PFF, and 19.5° with OPLS-AA. POSSIM performs very well, with only one relatively large deviation of ca. 30° in ψ of the second conformer.

In case of glutamine, the average energy RMS error is 0.58 kcal/mol, compared to 0.92 kcal/mol for PFF and 0.96 for OPLS-AA.[7] This is a significant improvement, especially given that there are eleven conformers. The average angular deviation of 16.3° with POSSIM is comparable to the PFF and OPLS-AA average errors of 18.0° and 13.9°, respectively. The largest error is in the values for the χ_3 side-chain dihedral.

Histidine

We considered two electrostatically neutral histidine dipeptide forms, Hid (protonated nitrogen

atom in δ-position) and Hie (protonation at the ε-nitrogen). The quantum mechanical conformational data for the Hid form was not used in deriving parameters for the PFF force field[7] and we have produced them in the course of this project. The results of testing the conformational equilibrium for Hid are shown in Table 7. Average deviations in energy and the key dihedral angles are 0.94 kcal/mol and 15.6°. These values are generally consistent with the other residues given the number of conformers. We did not compare them with the PFF and OPLS-AA results as they were not reported in the previous work.

The torsional parameters related to the side-chain χ_1 and χ_2 dihedral angles in the Hie dipeptide were fitted to the same quantum mechanical set as used previously for the PFF and OPLS-AA force fields.[7] The results of the POSSIM calculations are presented in Table 8. We have managed to achieve an improvement of both the energy-related and angular results. The POSSIM RMS deviation of the conformational energy is only 0.68 kcal/mol (compared with the PFF result of 0.83 kcal/mol and the OPLS-AA result of 0.85 kcal/mol). The average error in the key conformational angles for POSSIM (8.7°) decreased by more than a factor of two compared to the average errors for PFF and OPLS-AA (18.2° and 18.7°, respectively).[7] It should be noted that conformers 4 and 5 are different in the quantum mechanical calculations, with the only geometrical difference worth mentioning being the ca. 26° shift in the value of the backbone ψ angle. However, all three force fields (POSSIM, PFF, and OPLS-AA) yield the same result of these two conformers converging to just one.

Overall, performance of the POSSIM polarizable force field for the neutral histidine residue is adequate and consistent with that for the other amino acids.

Leucine, Isoleucine, and Valine

These protein residues were constructed by combining aliphatic POSSIM parameters with the

backbone fitted to alanine. Results of leucine conformational fitting are given in Table 9. POSSIM performed reasonably well for all the conformers except for the very first one, with the backbone in the C5-conformation region. The quantum mechanical geometry is still reproduced by POSSIM very well. The average error in the key dihedrals is only 5.4°, compared with the 5.1° and 5.9° of the PFF and OPLS-AA results.[7] All these are in great agreement with the quantum mechanical data. The POSSIM conformational energy RMSD is 1.02 kcal/mol, noticeably greater than the PFF and OPLS-AA errors of ca. 0.35 kcal/mol. However, given that the larger RSMD is created mostly by the first conformer in the C5 backbone region, we believe that the overall performance of our leucine parameters is acceptable.

The isoleucine results are shown in Table 10. Here the performance of the parameters was uniformly good without any special conformational cases. The average RMSD in the conformational energies was 0.54 kcal/mol, and the average error in the key dihedral was 6.7°. This compares well with the PFF errors of 0.88 kcal/mol and 11.8° and the OPLS-AA deviations of 0.38 kcal/mol and 5.5°.[7]

Conformational data for the valine dipeptide is given in Table 11. The POSSIM force field produced conformational energies within 0.13 kcal/mol RMSD and the ϕ, ψ, and χ_1 angles within an average of 5.1° from the quantum mechanical results. The PFF deviations were 0.01 kcal/mol and 5.1°, and the OPLS-AA errors are 0.08–0.16 kcal/mol and 8.4 – 8.6°[7]. Thus, all the three force fields reproduce these conformational properties for valine adequately.

Methionine

Results of methionine dipeptide conformational energy optimization with POSSIM are shown in Table 12. The side-chain torsional parameters refitted in this case were those for χ_1, χ_1', χ_2, and χ_3, as well as torsional parameters for the H-C-C-S and C-C-S-H dihedrals. Average error in the conformational energies as obtained with POSSIM were 0.23 kcal/mol. This is much better than the

0.53 kcal/mol and 0.59 kcal/mol RMS deviations for the same system simulated with PFF and OPLS-AA, respectively.[7] The average error in the POSSIM key dihedrals (including the backbone ϕ and ψ) is 5.1° which is comparable to the average error in the key dihedrals as calculated with PFF (5.4°) and OPLS-AA (5.2°). Thus, the overall, the quality of the POSSIM parameters for this amino acid is very good.

Proline

Proline dipeptide represents a special case. Just like in the previous works,[7,8] we validated proline parameters by calculating energies with values of the N–C–C(O)–N angle constrained at its value corresponding to the energy minimum, as well as at positions where this angle deviated from the minimum by ±60° and 180°. The results are shown in Table 13. The average energy error of 0.74 kcal/mol is good, especially given that the maximum rotamer energy exceeds 12 kcal/mol. The corresponding errors with the PFF and OPLS-AA force fields were 1.27 and 1.54 kcal/mol.[7] However, it should be noted that we refitted torsional parameters for the N–C–C–C, C–C–C–C(O), and C(O)–C–N–C angles, and this refitting was not carried out in our previous projects involving torsional force field parameters fit for this residue.

Tryptophan

Results of torsional fitting for the tryptophan dipeptide are shown in Table 14. We have produced parameters for the C–C–C–N, C–C–C–C(O), and C–C–C–C dihedral angles (χ_1, χ_1', and χ_2, respectively). The conformational energy RMSD calculated previously with the OPLS-AA and PFF force fields were 0.50 and 0.49 kcal/mol, respectively, and the average deviations for the key dihedrals were 24.2° and 19.4°.[7] The average angular error obtained with the POSSIM simulations is 19.2°, but

the energy deviates from the quantum mechanical result by 0.75 kcal/mol. While this error is somewhat higher than that obtained for the previous version of the force field, its value as such is not outside of the range where conformational energy errors can be considered reasonable and acceptable. It appears that this system has a relatively flat potential energy surface with a number of shallow minima. This character of the energy landscape is reproduced by all the three force fields (OPLS-AA, PFF and POSSIM).

Threonine

For this system, the torsional parameters for χ_1, χ_1', χ_2, and related dihedrals were refitted. The results are shown in Table 15. The value of the average error in the key dihedrals was 6.9° with OPLS-AA and 8.9° with PFF, the previous version of the polarizable force field.[7] In parameterizing this residue with the POSSIM formalism, we have obtained a comparable error of 7.7°, as can be seen from the data in the table. The RMS deviation of the conformational energy was 0.76 kcal/mol as obtained with POSSIM. Application of the OPLS-AA and PFF force fields yielded errors of 0.87 and 0.75 kcal/mol, respectively. Overall, the quality of parameterization of this residue with the above methods appears to be adequate.

Tyrosine

For this residue, we refitted torsional parameters for the χ_1, χ_1', and χ_2 dihedral angles. The parameters related to χ_6 were adopted from torsional parameters for phenol[26] without change, and the remaining torsional coefficients which include atoms of the aromatic ring, were the same as in phenylalanine and benzene. The results of validating these parameters with the conformational calculations are presented in Table 16. The RMS deviation of the conformational energies is 0.27

kcal/mol. This value is the same as for PFF and is somewhat smaller than the OPLS-AA RMSD of 0.39 kcal/mol.[7] The average error in the key dihedral angles was found to be 13.7° with POSSIM and 8.9° and 8.1° with PFF and OPLS-AA, respectively.

Protonated Aspartic and Glutamic Acids (Ash and Glh, respectively)

In cases of these protonated carboxylic acid residues, we report the average errors in key dihedrals including the last angles (χ_3 for Ash and χ_4 for Glh) that contain the acid OH hydrogen, even though the values of these angles are not given in the tables describing the fitting results (Table 17 for the protonated aspartic acid and Table 18 for its glutamic acid counterpart). The inclusion of these values into the calculated averages is justified by the fact that the quantum mechanical deviations of their values from 0° and 180° are in some cases as big as 25.5°, thus reproducing their values is not simply a matter of having them approximately at the planar 0° or 180° values.

We have refitted the χ_1, χ_1', χ_2, and χ_2' torsional parameters for protonated Asp and χ_1, χ_1', χ_2, χ_3 and χ_3' ones for protonated Glu. The protonated acid group parameters were taken from previous work[26] without any changes.

As can be seen from Tables 17 and 18, the RMS deviations for the protonated aspartic acid and glutamic acid dipeptides as simulated with POSSIM were 0.26 kcal/mol and 0.92 kcal/mol, respectively. The latter error is somewhat on the larger side, but not unacceptable, given the absolute values of the conformational energies. The average errors in the key dihedrals for these two systems were 12.4° for Ash and 9.9° for Glh.

While the protonated forms of the Asp and Glu residues are not typical in proteins, they have to be parameterized for such applications as calculations of protein pKa shifts.

Aspartic and Glutamic Acids

As discussed in the Methods section, parameterization of charged residues was carried out with constrained geometry optimizations, thus only the POSSIM conformational energies (and not geometries) are compared to the quantum mechanical references.

Torsional parameters for the χ_1, χ_1,' and χ_2 dihedrals were refitted in the both cases. In addition, the H–C–C–C(O) torsional parameters were fitted for Asp. The resulting values are used in the both residues, supplemented with refitted parameters for C–C–C(O)–O in Glu.

Results of the torsional fitting are given in Tables 19 and 20. The average error for the aspartic acid dipeptide conformational energies is 0.71kcal/mol, and the error for glutamic acid is 1.48 kcal/mol. The OPLS-AA results varied between 0.16 kcal/mol and 1.95 kcal/mol for Asp (depending on the torsional parameter set) while the error was 1.53 kcal/mol for Glu.[7] The RMS deviations with PFF were 0.77 kcal/mol and 1.47 kcal/mol, for Asp and Glu respectively.[7] Overall, the performance of the POSSIM parameters is consistent of that of the previous generation polarizable PFF force field.

Protonated Histidine

This residue was parameterized by refitting the torsional parameters for the χ_1, χ_1', χ_2, and χ_2' dihedrals. The results are shown in Table 21. The average error in the conformational energies was less than 0.01 kcal/mol. This result was achieved without any torsional coefficients exceeding 5.0 kcal/mol in magnitude.

Arginine

The last charged side-chain residue we worked with in this project was arginine. The side-chain parameters were produced by fitting methylguanidine potential energy functions as described above. The torsional fitting for this amino acid was carried out by adjusting the Fourier coefficients for the χ_1, χ_1', χ_2, χ_3, and χ_4 dihedral angles. The results of this fitting are presented in Table 22.

The average error in the conformational energies was 1.05 kcal/mol; this falls between the PFF and OPLS-AA results of 0.79 and 1.15 kcal/mol, respectively.[7] It should be noted that the error in the POSSIM energies is defined almost entirely by minimum four, the highest energy and thus the least probable minima. Removing this minimum would reduce the average error to ca. 0.04 kcal/mol. Thus, we believe that the POSSIM parameters for this residue are adequate.