Supporting Information

Molecular Dynamics Trajectory Compression with a Coarse-Grained Model

Yi-Ming Cheng[1], Srinivasa Murthy Gopal[1], Sean M. Law,[1] Michael Feig[1,2*]

[1] *Department of Biochemistry and Molecular Biology*
[2] *Department of Chemistry*
Michigan State University, East Lansing MI, USA

*corresponding author: feig@msu.edu, (517) 432-7439

Table S1: Bond distances, angles, and dihedrals used in PRIMO protein reconstructions. The constructed and constrained atoms correspond to atoms A, B, C and D respectively as discussed in the original PRIMO paper[1]. All parameters are obtained from all atom explicit solvent simulations of dipeptides with the CHARMM force field[2, 3].

.

Residue	**Constructed Atom**	**Constrained Atoms/ CG Sites**	**Bond Length (Å)**	**Bond Angle** (degree)	**Dihedral** (degree)
Backbone	C^i	CO^i, N^{i+1}, $C_\alpha{}^i$	0.616	41.50	0.00
Arg	C_β	C_α, N, C	1.542	110.60	-120.95
Asn	C_γ	SC2, SC3, C_β	0.616	41.50	0.20
Cys	C_β	C_α, N, C	1.556	110.45	-122.35
Gln	C_β	C_α, N, C	1.559	110.75	-121.90
	C_δ	SC2, SC3, C_γ	0.616	40.15	1.50
Glu	C_β	C_α, N, C	1.553	111.20	-123.60
His	C_β	C_α, N, C	1.543	111.10	-122.5
	$C_{\varepsilon 1}$	SC3, SC2, C_γ	1.323	34.65	179.5
	$C_{\delta 2}$	SC3, SC2, C_γ	1.386	70.25	0.3
Leu	C_γ	C_β, SC2, SC3	1.543	21.65	-30.15
Lys	C_β	C_α, N, C	1.555	110.20	-122.95
Met	C_β	C_α, N, C	1.551	109.90	-123.80
Phe	C_β	C_α, N, C	1.559	110.95	-122.15
	$C_{\delta 1}$	SC2, SC3, C_γ	1.396	90.15	0.55
	$C_{\delta 2}$	SC3, SC2, C_γ	1.400	89.80	0.15
	C_ζ	SC3, $C_{\delta 2}$, C_γ	1.401	119.60	-0.05
Tyr	C_β	C_α, N, C	1.558	110.95	-122.95
	$C_{\delta 1}$	SC2, SC3, C_γ	1.407	90.40	0.10
	$C_{\delta 2}$	SC3, SC2, C_γ	1.396	90.15	-0.05
	C_ζ	SC3, $C_{\delta 2}$, C_γ	1.400	119.65	1.00
Pro	C_β	C_α, N, C	1.540	103.20	-122.05
	C_δ *(non N-term)*	N^i, $C_\alpha{}^i$, C^{i-1}	1.464	111.60	178.70
	C_δ *(N-term)*	N, C_α, C_β	1.460	111.60	12.20
Ser	C_β	C_α, N, C	1.544	112.60	-120.80
Thr	C_β	C_α, SC1, SC2	1.523	20.25	44.2
Trp	C_β	C_α, N, C	1.557	109.75	-123.50
	$C_{\delta 2}$	SC3, SC4, SC2	1.403	59.75	-0.25
	$C_{\varepsilon 2}$	SC2, SC4, SC3	1.357	25.25	0.10
	$C_{\delta 1}$	SC2, $C_{\varepsilon 2}$, $C_{\delta 2}$	1.377	109.95	0.15
	$C_{\zeta 3}$	SC3, $C_{\delta 2}$, $C_{\varepsilon 2}$	1.394	118.30	2.15
	$C_{\eta 2}$	SC4, $C_{\varepsilon 2}$, $C_{\delta 2}$	1.396	117.40	1.95
Val	C_β	C_α, SC1, SC2	1.530	22.05	-30.65

Table S2: Bond distances, angles, and dihedrals used in hydrogen atom reconstructions. The constructed and constrained atoms correspond to atoms A, B, C and D respectively as discussed in the original PRIMO paper[1]. All parameters are obtained from all atom explicit solvent simulations of dipeptides with the CHARMM force field[2, 3].

Residue	**Constructed Atom**	**Constrained Atoms/ CG Sites**	**Bond Length (Å)**	**Bond angle** (degree)	**Dihedral** (degree)
backbone	H_N	N, C_α, C^{i-1}	0.997	116.0	-178.6
	H_α	C_α, N, C	1.080	109.0	117.0
N-terminus	H_{T1}	N, CA, C	0.997	109.5	52.5
	H_{T2}	N, CA, C	0.997	109.5	-64.9
	H_{T3}	N, CA, C	0.997	109.5	174.7
Acetylated N-terminus	H_{y1}	$C_{\alpha y}$, C_y, O_y	1.111	110.1	180.0
	H_{y2}	$C_{\alpha y}$, H_{y1}, C_y	1.111	108.4	120.0
	H_{y3}	$C_{\alpha y}$, H_{y1}, H_{y2}	1.111	108.4	115.5
N-methylamide C-terminus	H_{NT}	N_T, $C_{\alpha T}$, C	0.997	116.0	-178.6
	H_{T1}	$C_{\alpha T}$, N, HN_T	1.111	110.1	180.0
	H_{T2}	$C_{\alpha T}$, H_{T1}, N	1.111	108.4	120.0
	H_{T3}	$C_{\alpha T}$, H_{T1}, H_{T2}	1.111	108.4	115.5
Gly	$H_{\alpha 1}$	C_α, C, N	1.080	109.0	117.0
	$H_{\alpha 2}$	C_α, N, C	1.080	109.0	117.0
Ala	$H_{\beta 1}$	C_β, C_α, N	1.111	111.3	60.9
	$H_{\beta 2}$	C_β, $H_{\beta 1}$, C_α	1.111	107.3	120.6
	$H_{\beta 3}$	C_β, $H_{\beta 1}$, $H_{\beta 2}$	1.111	107.9	115.5
Arg	$H_{\beta 1}$	C_β, C_γ, C_α	1.111	108.1	123.0
	$H_{\beta 2}$	C_β, C_α, C_γ	1.111	108.9	121.1
	$H_{\gamma 1}$	C_γ, C_δ, C_β	1.111	107.1	121.8
	$H_{\gamma 2}$	C_γ, C_β, C_δ	1.111	108.3	119.4
	$H_{\delta 1}$	C_δ, N_ε, C_γ	1.111	110.8	119.0
	$H_{\delta 2}$	C_δ, C_γ, N_ε	1.111	111.3	118.7
	H_ε	N_ε, C_δ, C_ζ	0.997	116.0	-178.6
	$H_{\eta 11}$	NH_1, C_ζ, N_ε	1.000	116.9	2.6
	$H_{\eta 12}$	NH_1, C_ζ, $H_{\eta 11}$	1.000	118.3	-176.4
	$H_{\eta 21}$	NH_2, C_ζ, N_ε	1.000	119.2	-0.3
	$H_{\eta 22}$	NH_2, $H_{\eta 21}$, C_ζ	1.000	119.3	-178.2
Asn	$H_{\beta 1}$	C_β, C_γ, C_α	1.111	107.5	119.9
	$H_{\beta 2}$	C_β, C_α, C_γ	1.111	109.8	120.2
	$H_{\delta 21}$	$N_{\delta 2}$, C_γ, $O_{\delta 1}$	1.000	117.6	-0.7
	$H_{\delta 22}$	$N_{\delta 2}$, $H_{\delta 21}$, C_γ	1.000	121.0	-178.1
Asp	$H_{\beta 1}$	C_β, C_γ, C_α	1.111	108.3	121.0
	$H_{\beta 2}$	C_β, C_α, C_γ	1.111	108.6	120.1
Cys	$H_{\beta 1}$	C_β, S_γ, C_α	1.111	109.3	122.1
	$H_{\beta 2}$	C_β, C_α, S_γ	1.111	109.5	121.7
	$H_{\gamma 1}$	S_γ, C_β, C_α	1.330	96.2	180.0
Gln	$H_{\beta 1}$	C_β, C_γ, C_α	1.111	107.2	121.5

	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	108.0	123.2
	$H_{\gamma 1}$	C_{γ}, C_{δ}, C_{β}	1.111	108.5	121.9
	$H_{\gamma 2}$	C_{γ}, C_{β}, C_{δ}	1.111	109.8	120.0
	$H_{\varepsilon 22}$	$N_{\varepsilon 2}$, C_{δ}, $O_{\varepsilon 1}$	1.000	118.4	-177.9
	$H_{\varepsilon 21}$	$N_{\varepsilon 2}$, $H_{\varepsilon 22}$, C_{δ}	1.000	120.9	170.4
Glu	$H_{\beta 1}$	C_{β}, C_{γ}, C_{α}	1.111	107.9	120.8
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	109.2	121.6
	$H_{\gamma 1}$	C_{γ}, C_{δ}, C_{β}	1.111	108.4	122.4
	$H_{\gamma 2}$	C_{γ}, C_{β}, C_{δ}	1.111	110.5	120.0
His	$H_{\beta 1}$	C_{β}, C_{γ}, C_{α}	1.111	108.6	122.7
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	107.8	119.8
	$H_{\delta 2}$	$C_{\delta 2}$, $N_{\varepsilon 2}$, C_{γ}	1.080	120.7	-177.9
	$H_{\varepsilon 1}$	$C_{\varepsilon 1}$, $N_{\delta 1}$, $N_{\varepsilon 2}$	1.080	123.5	-179.6
Ile	H_{β}	C_{β}, $C_{\gamma 1}$, $C_{\gamma 2}$	1.111	106.5	116.8
	$H_{\gamma 21}$	$C_{\gamma 2}$, C_{β}, C_{α}	1.111	111.2	64.5
	$H_{\gamma 22}$	$C_{\gamma 2}$, $H_{\gamma 21}$, C_{β}	1.111	106.6	121.2
	$H_{\gamma 23}$	$C_{\gamma 2}$, $H_{\gamma 21}$, $H_{\gamma 22}$	1.111	106.3	115.8
	$H_{\delta 1}$	C_{δ}, $C_{\gamma 1}$, C_{β}	1.111	110.4	59.6
	$H_{\delta 2}$	C_{δ}, $H_{\delta 1}$, $C_{\gamma 1}$	1.111	107.3	120.0
	$H_{\delta 3}$	C_{δ}, $H_{\delta 1}$, $H_{\delta 2}$	1.111	108.2	116.6
	$H_{\gamma 11}$	$C_{\gamma 1}$, C_{δ}, C_{β}	1.111	108.1	123.2
	$H_{\gamma 12}$	$C_{\gamma 1}$, C_{β}, C_{δ}	1.111	109.4	121.8
Leu	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	107.5	121.1
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	107.1	122.6
	H_{γ}	C_{γ} , $C_{\delta 1}$, $C_{\delta 2}$	1.111	107.3	-116.4
	$H_{\delta 11}$	$C_{\delta 1}$, C_{γ} , C_{β}	1.111	110.6	59.1
	$H_{\delta 12}$	$C_{\delta 1}$, $H_{\delta 11}$, C_{γ}	1.111	108.8	120.8
	$H_{\delta 13}$	$C_{\delta 1}$, $H_{\delta 11}$, $H_{\delta 12}$	1.111	108.6	118.3
	$H_{\delta 21}$	$C_{\delta 2}$, C_{γ} , C_{β}	1.111	110.0	63.2
	$H_{\delta 22}$	$C_{\delta 2}$, $H_{\delta 21}$, C_{γ}	1.111	109.2	120.8
	$H_{\delta 23}$	$C_{\delta 2}$, $H_{\delta 21}$, $H_{\delta 22}$	1.111	107.9	117.6
Lys	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	107.4	121.5
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	109.2	120.5
	$H_{\gamma 1}$	C_{γ} , C_{δ}, C_{β}	1.111	109.7	120.2
	$H_{\gamma 2}$	C_{γ} , C_{β}, C_{δ}	1.111	108.6	121.6
	$H_{\delta 1}$	C_{δ}, C_{ε}, C_{γ}	1.111	108.3	121.2
	$H_{\delta 2}$	C_{δ}, C_{γ} , C_{ε}	1.111	109.1	121.4
	$H_{\varepsilon 1}$	C_{ε}, N_{ζ}, C_{δ}	1.111	107.2	120.6
	$H_{\varepsilon 2}$	C_{ε}, C_{δ}, N_{ζ}	1.111	110.6	119.2
	$H_{\zeta 1}$	N_{ζ}, C_{ε}, C_{δ}	1.111	110.2	62.3
	$H_{\zeta 2}$	N_{ζ}, $H_{\zeta 1}$, C_{ε}	1.111	108.5	121.0
	$H_{\zeta 3}$	N_{ζ}, $H_{\zeta 1}$, $H_{\zeta 2}$	1.111	108.7	117.3
Met	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	109.1	120.9
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	108.4	120.8
	$H_{\gamma 1}$	C_{γ} , S_{δ}, C_{β}	1.111	108.6	122.2
	$H_{\gamma 2}$	C_{γ} , C_{β}, S_{δ}	1.111	107.5	121.7

	$H_{\varepsilon 1}$	C_{ε}, S_{δ}, C_{γ}	1.111	111.3	64.1
	$H_{\varepsilon 2}$	C_{ε}, $H_{\varepsilon 1}$, S_{δ}	1.111	107.6	121.4
	$H_{\varepsilon 3}$	C_{ε}, $H_{\varepsilon 1}$, $H_{\varepsilon 2}$	1.111	107.0	115.4
Phe	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	108.1	121.1
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	108.6	119.5
	$H_{\delta 1}$	$C_{\delta 1}$, C_{γ} , $C_{\varepsilon 1}$	1.080	119.7	179.4
	$H_{\delta 2}$	$C_{\delta 2}$, C_{γ} , $C_{\varepsilon 2}$	1.080	120.2	179.1
	$H_{\varepsilon 1}$	$C_{\varepsilon 1}$, $C_{\delta 1}$, C_{ζ}	1.080	119.3	179.5
	$H_{\varepsilon 2}$	$C_{\varepsilon 2}$, $C_{\delta 2}$, C_{ζ}	1.080	119.7	179.1
	H_{ζ}	C_{ζ}, $C_{\varepsilon 1}$, $C_{\varepsilon 2}$	1.080	120.5	179.5
Pro	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	111.8	119.5
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	111.3	118.7
	$H_{\gamma 1}$	C_{γ} , C_{β}, C_{δ}	1.111	112.7	119.9
	$H_{\gamma 2}$	C_{γ} , C_{δ}, C_{β}	1.111	111.2	118.8
	$H_{\delta 1}$	C_{δ}, C_{γ}, N	1.111	109.9	117.8
	$H_{\delta 2}$	C_{δ}, N, C_{γ}	1.111	111.5	119.3
Ser	$H_{\beta 1}$	C_{β}, C_{α}, O_{γ}	1.111	109.7	-119.5
	$H_{\beta 2}$	C_{β}, O_{γ}, C_{α}	1.111	107.7	-121.5
	$H_{\gamma 1}$	O_{γ}, C_{β}, C_{α}	0.960	106.1	175.7
Tyr	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	107.7	121.6
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	110.3	119.6
	$H_{\delta 1}$	$C_{\delta 1}$, C_{γ} , $C_{\varepsilon 1}$	1.080	119.3	-178.6
	$H_{\delta 2}$	$C_{\delta 2}$, C_{γ} , $C_{\varepsilon 2}$	1.080	120.1	-176.9
	$H_{\varepsilon 1}$	$C_{\varepsilon 1}$, $C_{\delta 1}$, C_{ζ}	1.080	119.1	179.3
	$H_{\varepsilon 2}$	$C_{\varepsilon 2}$, $C_{\delta 2}$, C_{ζ}	1.080	119.5	-179.6
	H_{η}	O_{η}, C_{ζ}, $C_{\varepsilon 2}$	0.960	106.7	-178.7
Thr	H_{β}	C_{β}, C_{α}, $O_{\gamma 1}$	1.111	110.0	121.2
	$H_{\gamma 1}$	$O_{\gamma 1}$, C_{β}, C_{α}	0.960	104.5	70.0
	$H_{\gamma 21}$	$C_{\gamma 2}$, C_{β}, C_{α}	1.111	110.6	59.3
	$H_{\gamma 22}$	$C_{\gamma 2}$, $H_{\gamma 21}$, C_{β}	1.111	108.4	122.2
	$H_{\gamma 23}$	$C_{\gamma 2}$, $H_{\gamma 21}$, $H_{\gamma 22}$	1.111	107.3	115.5
Trp	$H_{\beta 1}$	C_{β}, C_{γ} , C_{α}	1.111	108.1	121.5
	$H_{\beta 2}$	C_{β}, C_{α}, C_{γ}	1.111	107.8	121.1
	$H_{\delta 1}$	$C_{\delta 1}$, $N_{\varepsilon 1}$, C_{γ}	1.080	124.8	-178.4
	$H_{\zeta 2}$	$C_{\zeta 2}$, CH_2, $C_{\varepsilon 2}$	1.080	121.0	-179.3
	$H_{\eta 2}$	CH_2, $C_{\zeta 3}$, $C_{\zeta 2}$	1.080	120.3	-179.8
	$H_{\zeta 3}$	$C_{\zeta 3}$, $C_{\varepsilon 3}$, CH_2	1.080	119.5	-179.1
	$H_{\varepsilon 3}$	$C_{\varepsilon 3}$, $C_{\delta 2}$, $C_{\zeta 3}$	1.080	121.0	179.1
Val	H_{β}	C_{β}, $C_{\gamma 1}$, $C_{\gamma 2}$	1.111	107.8	-118.0
	$H_{\gamma 11}$	$C_{\gamma 1}$, C_{β}, C_{α}	1.111	111.6	57.3
	$H_{\gamma 12}$	$C_{\gamma 1}$, $H_{\gamma 11}$, C_{β}	1.111	107.8	122.1
	$H_{\gamma 13}$	$C_{\gamma 1}$, $H_{\gamma 11}$, $H_{\gamma 12}$	1.111	108.2	116.8
	$H_{\gamma 21}$	$C_{\gamma 2}$, C_{β}, C_{α}	1.111	110.8	58.5
	$H_{\gamma 22}$	$C_{\gamma 2}$, $H_{\gamma 21}$, C_{β}	1.111	108.8	119.5
	$H_{\gamma 23}$	$C_{\gamma 2}$, $H_{\gamma 21}$, $H_{\gamma 22}$	1.111	108.8	116.5

Table S3: Test set used for hydrogen reconstruction accuracy analysis with numbers of hydrogen atoms in different categories following Li et al.[4]. 6 structures were omitted from the original set of structures by Li et al. because of missing residues or atoms that complicate all-atom reconstruction from PRIMO.

	Categories of hydrogen atoms					
X-ray	sp3H1	sp3H2	sp3H3	sp2H1	sp2H2	spH1
1ab1	99	92	81	13	14	17
1g66	416	432	318	83	48	60
1gci	567	524	513	70	96	58
1m40	590	666	549	77	104	42
1muw	839	962	687	202	170	40
1ssx	370	332	324	55	76	40
1ucs	143	166	177	4	18	7
1vyr	787	898	678	138	160	53
1xvo	471	440	387	75	70	66
1yk4	103	156	69	24	6	10
2h5c	370	332	324	55	76	40
3pyp	262	324	222	75	30	18
Neutron diffraction						
1c57	521	534	453	119	62	57
1cq2	334	434	339	78	28	14
1xqn	521	534	453	119	62	57
1lzn	281	310	201	65	78	28
1ntp	461	500	393	81	48	65
2gve	839	988	660	207	182	37
1vcx	104	156	75	30	2	9
1io5	281	310	201	65	78	28
2mb5	334	434	339	78	28	14
5rsa	268	310	192	47	50	39

Table S4: Bond distances, angles, and dihedrals used in alternate PRIMO protein reconstructions as in Table S1.

Residue	**Constructed Atom**	**Constrained Atoms/ CG Sites**	**Bond Length (Å)**	**Bond Angle** (degree)	**Dihedral** (degree)
Arg	C_γ	C_δ, C_β, SC1	1.53	33.10	0.0
Asn	$O_{\delta 1}$	C_γ, C_β, $N_{\delta 2}$	1.24	121.40	179.45
	$N_{\delta 2}$	C_γ, $O_{\delta 1}$, C_β	1.35	121.65	179.35
Asp	$O_{\delta 1}$	C_γ, C_β, SC2	1.26	119.2	0.0
	$O_{\delta 2}$	C_γ, C_β, $O_{\delta 1}$	1.26	118.7	180.0
Cys	S_γ	C_β, C_α, SC1	1.83	112.95	0.0
Gln	C_γ	C_β, C_α, SC1	1.55	114.2	0.0
	C_δ	C_γ, C_β, (SC2+SC3)/2	1.53	111.9	0.0
	$O_{\varepsilon 1}$	C_δ, C_γ, (C_δ–SC2+SC3)	1.232	121.7	180.0
	$N_{\varepsilon 2}$	C_δ, C_γ, $O_{\varepsilon 1}$	1.35	117.2	180.0
Glu	C_γ	C_β, C_α, SC1	1.55	114.35	0.0
	C_δ	C_γ, C_β, (SC2+SC3)/2	1.515	111.9	0.0
	$O_{\varepsilon 1}$	C_δ, C_γ, (C_δ–SC2+SC3+ 0.61*(C_δ - C_γ)$_N$)^a	1.26	119.1	180.0
	$O_{\varepsilon 2}$	C_δ, C_γ, $O_{\varepsilon 1}$	1.26	118.95	180.0
His	C_γ	C_β, $N_{\delta\,1}$, SC1	1.51	26.75	0.0
Ile	C_δ	$C_{\gamma 1}$, C_β, SC2	1.534	113.7	0.0
	$C_{\gamma 2}$	C_β, C_α, SC1	1.55	108.45	0.0
Leu	$C_{\delta\,1}$	C_γ, C_β, SC2	1.54	111.45	0.0
	$C_{\delta\,2}$	C_γ, C_β, SC3	1.54	111.10	0.0
Lys	C_γ	C_δ, C_β, SC1	1.54	33.90	0.0
Met	C_γ	S_δ, C_β, SC1	1.82	30.10	0.0
Phe	C_γ	C_β, SC2, SC3	1.51	18.5	0.0
	$C_{\delta\,1}$	C_γ, C_β, SC2	1.41	120.3	0.0
	$C_{\delta\,2}$	C_γ, C_β, SC3	1.41	120.2	0.0
	$C_{\varepsilon\,1}$	$C_{\delta 1}$, C_γ, C_β	1.40	120.7	180.0
	$C_{\varepsilon\,2}$	$C_{\delta 2}$, C_γ, C_β	1.40	120.95	180.0
Tyr	C_γ	C_β, SC2, SC3	1.51	18.5	0.0
	$C_{\delta\,1}$	C_γ, C_β, SC2	1.41	120.45	0.0
	$C_{\delta\,2}$	C_γ, C_β, SC3	1.41	120.45	0.0
	$C_{\varepsilon\,1}$	$C_{\delta 1}$, C_γ, C_β	1.40	120.35	180.0
	$C_{\varepsilon\,2}$	$C_{\delta 2}$, C_γ, C_β	1.40	120.45	180.0
	C_ζ	$C_{\varepsilon 1}$, $C_{\delta 2}$, C_γ	1.40	119.65	1.0
	O_η	C_ζ, $C_{\varepsilon 2}$, $C_{\varepsilon 2}$	1.41	119.85	180.0
Pro	C_γ	C_β, C_δ, SC1	1.53	38.75	0.0
Ser	O_γ	C_β, C_α, SC1	1.427	112.15	0.0

Thr	$O_{\gamma 1}$	C_{β}, C_{α}, SC1	1.42	111.7	0.0
	$C_{\gamma 2}$	C_{β}, C_{α}, SC2	1.54	110.45	0.0
Trp	C_{γ}	SC1, $C_{\delta 1}$, $C_{\delta 2}$	0.755	35.75	-0.3
Val	$C_{\gamma 1}$	C_{β}, C_{α}, SC1	1.54	110.55	0.0
	$C_{\gamma 2}$	C_{β}, C_{α}, SC2	1.54	110.8	0.0

a. The subscripted N denotes the corresponding unit vector.

Table S5: Reconstruction accuracy (RMSD) for heavy atoms with alternative reconstruction protocol before and after minimization with CHARMM (see caption of Table 5 for details). Standard deviations are given in parentheses.

	Alt.	Alt/Min
ALL	0.099(0.031)	0.101(0.030)
Backbone	0.022(0.021)	0.034(0.019)
Side-Chain	0.138(0.044)	0.139(0.044)
ARG	0.042(0.026)	0.046(0.026)
ASN	0.038(0.017)	0.044(0.022)
ASP	0.025(0.015)	0.030(0.020)
CYS	0.128(0.083)	0.188(0.109)
GLN	0.212(0.143)	0.210(0.140)
GLU	0.293(0.120)	0.288(0.117)
HIS	0.063(0.035)	0.055(0.030)
ILE	0.196(0.186)	0.191(0.182)
LEU	0.043(0.017)	0.044(0.020)
LYS	0.050(0.031)	0.031(0.029)
MET	0.061(0.032)	0.058(0.037)
PHE	0.105(0.067)	0.100(0.064)
PRO	0.096(0.047)	0.093(0.048)
SER	0.144(0.072)	0.144(0.074)
THR	0.051(0.022)	0.049(0.027)
TRP	0.046(0.016)	0.042(0.015)
TYR	0.102(0.070)	0.100(0.071)
VAL	0.036(0.027)	0.038(0.027)

Table S6: Reconstruction accuracy (RMSD) for hydrogen atoms with alternative reconstruction protocol before and after minimization with CHARMM (see caption of Table 5 for details). Standard deviations are given in parentheses.

	Alt	Alt/Min
All/Average	0.201 *(0.218)*	0.199 *(0.217)*
sp3H3	0.253 *(0.212)*	0.250 *(0.210)*
sp3H2	0.195 *(0.160)*	0.193 *(0.158)*
sp3H1	0.116 *(0.078)*	0.113 *(0.078)*
sp2H2	0.212 *(0.223)*	0.212 *(0.223)*
sp2H1	0.155 *(0.070)*	0.150 *(0.068)*
spH1	1.044 *(0.597)*	1.045 *(0.598)*

Figure S1. Correlation between C_α-SC2 and C_α-C_β bond lengths. The blue points represent averages where data is present and extrapolated values based on the trends at either side of the data range.

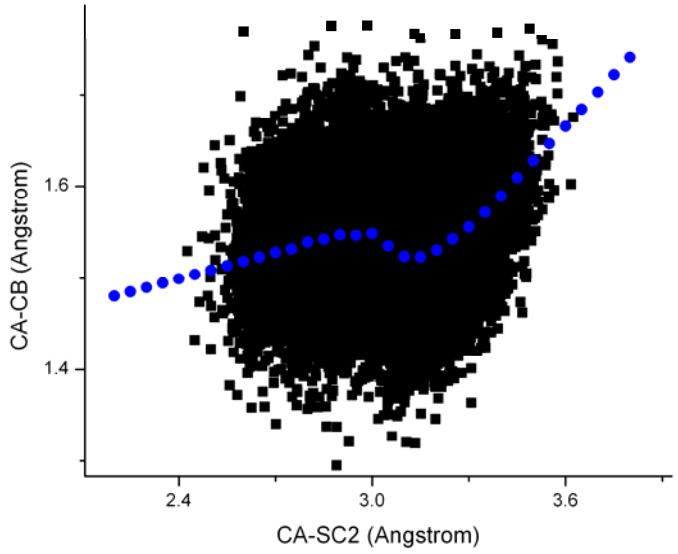


Figure S2. Correlation between C_α-SC2 and C_α-$C_{\gamma 1}$ bond lengths as in Fig. S1.

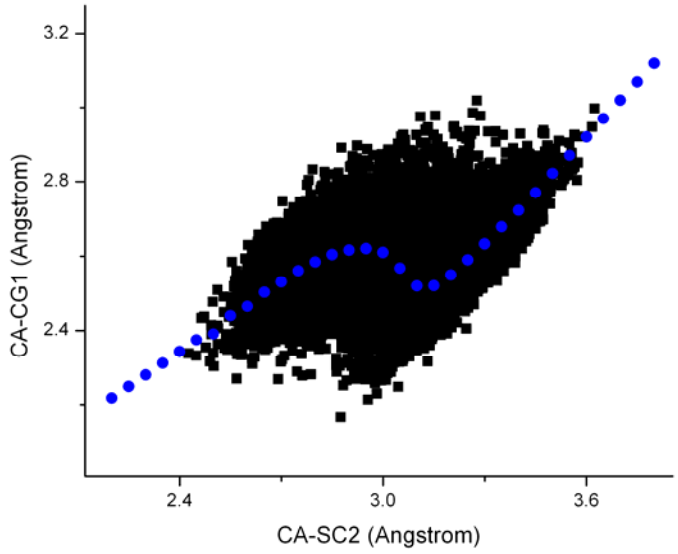


Figure S3. Correlation between C_α-SC2 bond lengths and C_α-C_β-$C_{\gamma 1}$ bond angles as in Figure S1.

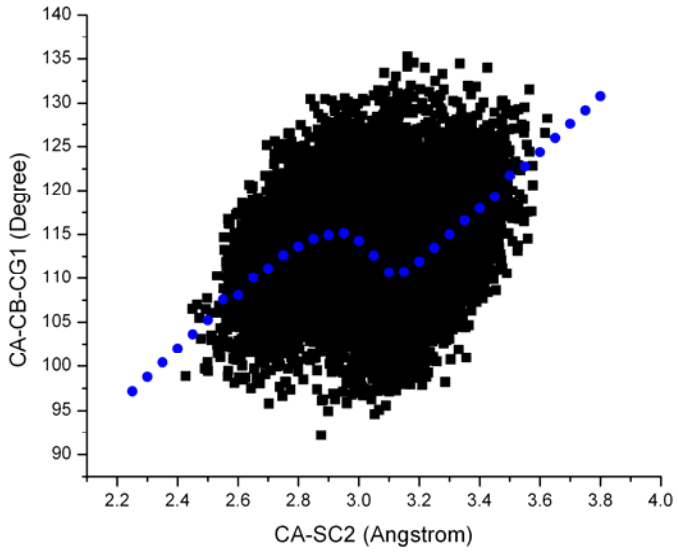


Figure S4. Correlation between C_α-SC2 and C_β-SC2 bond lengths as in Figure S1.

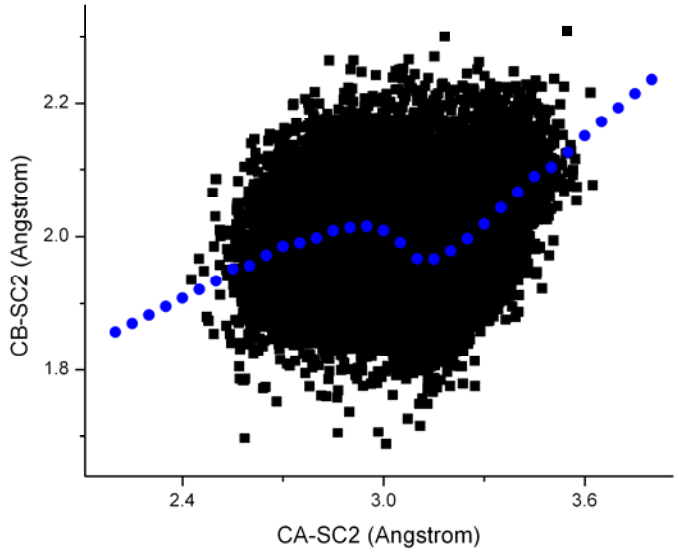


Figure S5. Correlation between SC1-SC2 and SC1-$C_{\gamma 1}$ bond lengths as in Figure S1.

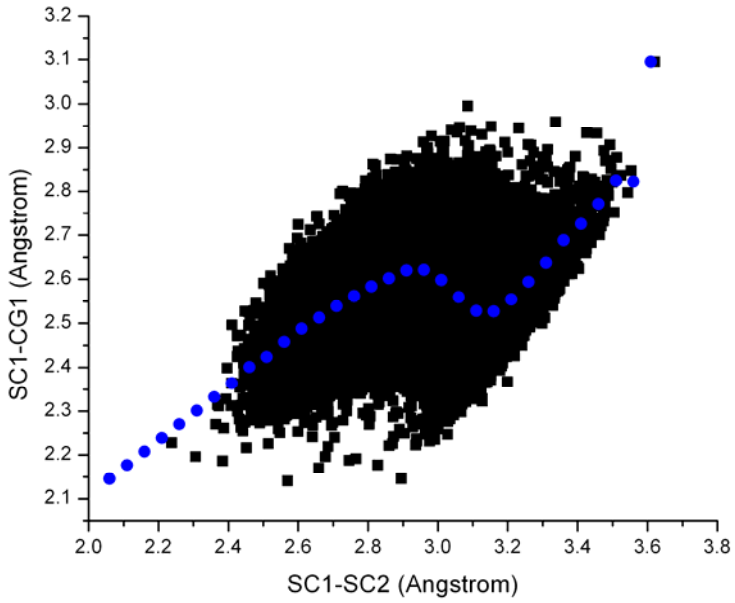
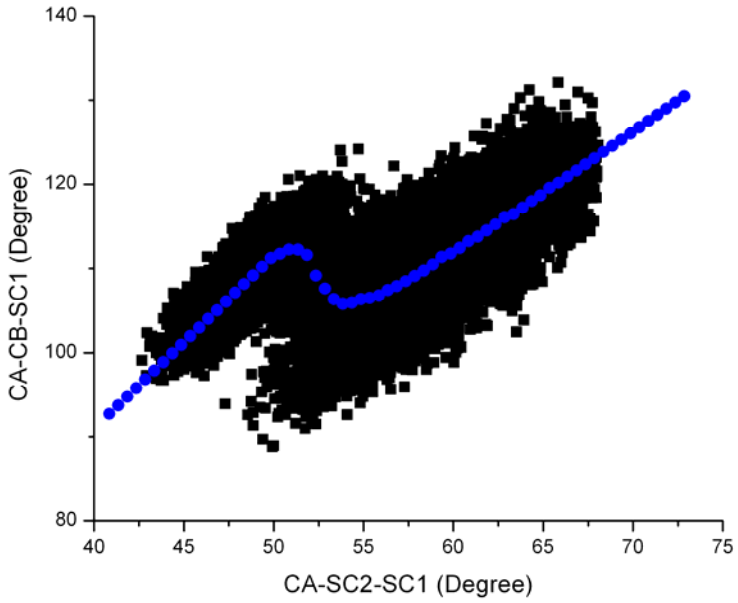


Figure S6. Correlation between C_α-SC2-SC1 and C_α-C_β-SC1 bond angles as in Figure S1.



References

[1] Gopal, S.M., Mukherjee, S., Cheng, Y.M., Feig, M. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. Proteins. 2010, 78, 1266-81.
[2] MacKerell, A.D., Bashford, D., Bellott, Dunbrack, R.L., Evanseck, J.D., Field, M.J., et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. The Journal of Physical Chemistry B. 1998, 102, 3586-616.
[3] MacKerell, A.D., Feig, M., Brooks, C.L. Improved Treatment of the Protein Backbone in Empirical Force Fields. Journal of the American Chemical Society. 2004, 126, 698-9.
[4] Li, Y., Roy, A., Zhang, Y. HAAD: A quick algorithm for accurate prediction of hydrogen atoms in protein structures. PLoS One. 2009, 4, e6701.