Supporting Information

Toward Structure Prediction for Short Peptides Using the Improved SAAP Force Field Parameters[†]

Kenichi Dedachi,[‡] Taku Shimosato,[‡] Toshiya Minezaki,[§] and Michio Iwaoka[‡]

[‡]Department of Chemistry, School of Science, Tokai University, Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan

[§]Laboratory of General Education for Science and Technology, School of Science, Tokai University, Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan

Contents:

Details of the SAAP force field and the simulation algorithm.

Figure S1. Traces of the energies obtained for Met-enkephalin by the SAAP-MC (trajectory 1) (a) and AMBER-MD (trajectory 1) (b) simulations at 300 K in water.

Figure S2. Representative structures of Met-enkephalin obtained by the AMBER-MD simulation (trajectory 1) at 300 K in water.

Figure S3-1. Traces of the main-chain RMSD obtained for chignolin by the SAAP-MC (trajectories 1–10) simulations at 300 K in water.

Figure S3-2. Traces of the main-chain RMSD obtained for chignolin by the SAAP-MC (trajectories 11–20) simulations at 300 K in water.

Figure S4. The minor structures of chignolin with a representative form obtained by the SAAP-MC simulation at 300 K in water.

Figure S5. Representative structures of chignolin obtained by the AMBER-MD simulation at 300 K in water.

Table S1. Dihedral angles (degree) of each amino acid residue for representative structures A, A', B, and C of chignolin obtained by the SAAP-MC simulation at 300 K in water.

Table S2. The structures of chignolin with the smallest all-atom RMSD from the native structure obtained by the SAAP-MC simulation at 300 K in water.

Table S3. The structure of chignolin with the lowest total potential energy obtained by the SAAP-MC simulation at 300 K in water.

Table S4. Results of single-point *ab initio* calculation for representative structure A of chignolin in water.

Table S5. Results of single-point *ab initio* calculation for representative structure A' of chignolin in water.

Table S6. Results of single-point *ab initio* calculation for representative structure B of chignolin in water.

Table S7. Results of single-point *ab initio* calculation for representative structure C of chignolin in water.

Details of the SAAP force field and the simulation algorithm.

The details are described in the references [S1][S2]. In the SAAP force field, the potential energy functions of a polypeptide molecule are given by eq. (1).

$$E^{\text{total}} = E^{\text{SAAP}} + E^{\text{inter}} + E^{\text{others}} \approx E^{\text{SAAP}} + E^{\text{inter}} \qquad (1)$$

The total potential energy (E^{total}) is divided into single amino acid potentials of the individual amino acid units (E^{SAAP}) and the inter-amino acid interactions (E^{inter}). The correlation term between these terms (E^{others}) should be included to attain the exact potential energy function, but the E^{others} term is currently ignored as a constant because the contribution to the total potential energy would be small compared to the other two terms. The inter-amino acid interaction term (E^{inter}) is further divided into E^{ES} and E^{LJ} terms [eq. (2)], which correspond to electrostatic interactions and Lennard-Jones potentials (or van der Waals interactions), respectively, between the amino acid units. It should be noted that the E^{ES} and E^{LJ} terms do not include the atomic interactions within the amino acid units.

$$E^{\text{inter}} = E^{\text{ES}} + E^{\text{LJ}} \qquad (2)$$

Important features of the SAAP force field are two-fold. First, the structure of a polypeptide molecule is simply divided into the amino acid units. This is in distinct difference from previous all-atom force fields, where the molecule is divided into the individual atoms. Second, solvent effects are implicitly included in the SAAP force field parameters. The SAAP force field equips with different parameter sets in vacuo (D = 1.0), in ether (D = 4.335), and in water (D = 78.96) for the E^{SAAP} term, hence by using the SAAP force field the molecular simulation in water can be carried out within the same computation time to that in vacuo.

To develop the SAAP force field parameters, numerous ab initio calculations have to be performed for each amino acid unit (CHO-NH-CHR-CO-NH_2) in vacuo, in ether, and in water. In order to cover the whole conformational space, each dihedral angle, ϕ, ψ, χ_1, χ_2, etc., is divided with a resolution of 15 degree. For example, 13824 (= 24^3) structures are considered for valine, which has a three-dimensional conformation freedom (i.e., ϕ, ψ and χ_1). For the each structure, the SAAP force field parameters are defined with the relative energy in the conformational space, the Cartesian coordinates, the atomic charges, and the Lennard-Jones parameters. For the amino acids, except for Gly, Ala, Val and Pro, the side-chain separation approximation is applied in order to reduce the number of the structures to be considered. Although this approximation should cause a decrease in

the accuracy of the SAAP force field, it will take too long time to develop the parameters without the approximation.

In the SAAP simulation program, connection of the amino acid units (CHO-NH-CHR-CO-NH_2) is achieved as follows. A new amino acid unit is connected to the C-terminal of a pre-peptide so that the amide planes from the both sides are superimposed. Then, the potential energy of the new peptide is calculated as the sum of the potential energy of the pre-peptide, the SAAP energy of the connected amino acid unit (E^{SAAP}), and the electrostatic and Lennard-Jones potentials (E^{ES} and E^{LJ}, respectively) between the pre-peptide and the connected amino acid unit.

In each Monte Carlo step, one dihedral angle is selected randomly, and the value is changed at random within a given range of the angle (usually ±32 degree). When a Pro residue is present in the sequence, the "down" or "up" puckering conformer is randomly selected in each step. A new peptide structure is subsequently built up by using the new set of the dihedral angles, and the potential energy is calculated as described above. If the energy change (ΔE) meats the criterion defined by the Metropolis equation ($\xi \leq \exp(-\beta \Delta E)$; ξ = a randomly created number between 0 and 1, $\beta = 1/RT$, R = a gas constant, T = absolute temperature), the new structure is selected, otherwise it is abandoned.

References

[S1] M. Iwaoka, S. Tomoda, "The SAAP force field. A simple approach to a new all-atom protein force field by using single amino acid potential (SAAP) functions in various solvents," *J. Comput. Chem.*, vol. 24, no. 10, pp. 1192-1200, 2003.

[S2] M. Iwaoka, N. Kimura, D. Yosida, T. Minezaki, "The SAAP force field. Development of the single amino acid potentials for twenty proteinogenic amino acids and Monte Carlo molecular simulation for short peptides," *J. Comput. Chem.*, vol. 30, no. 13, pp. 2039-2055, 2009.

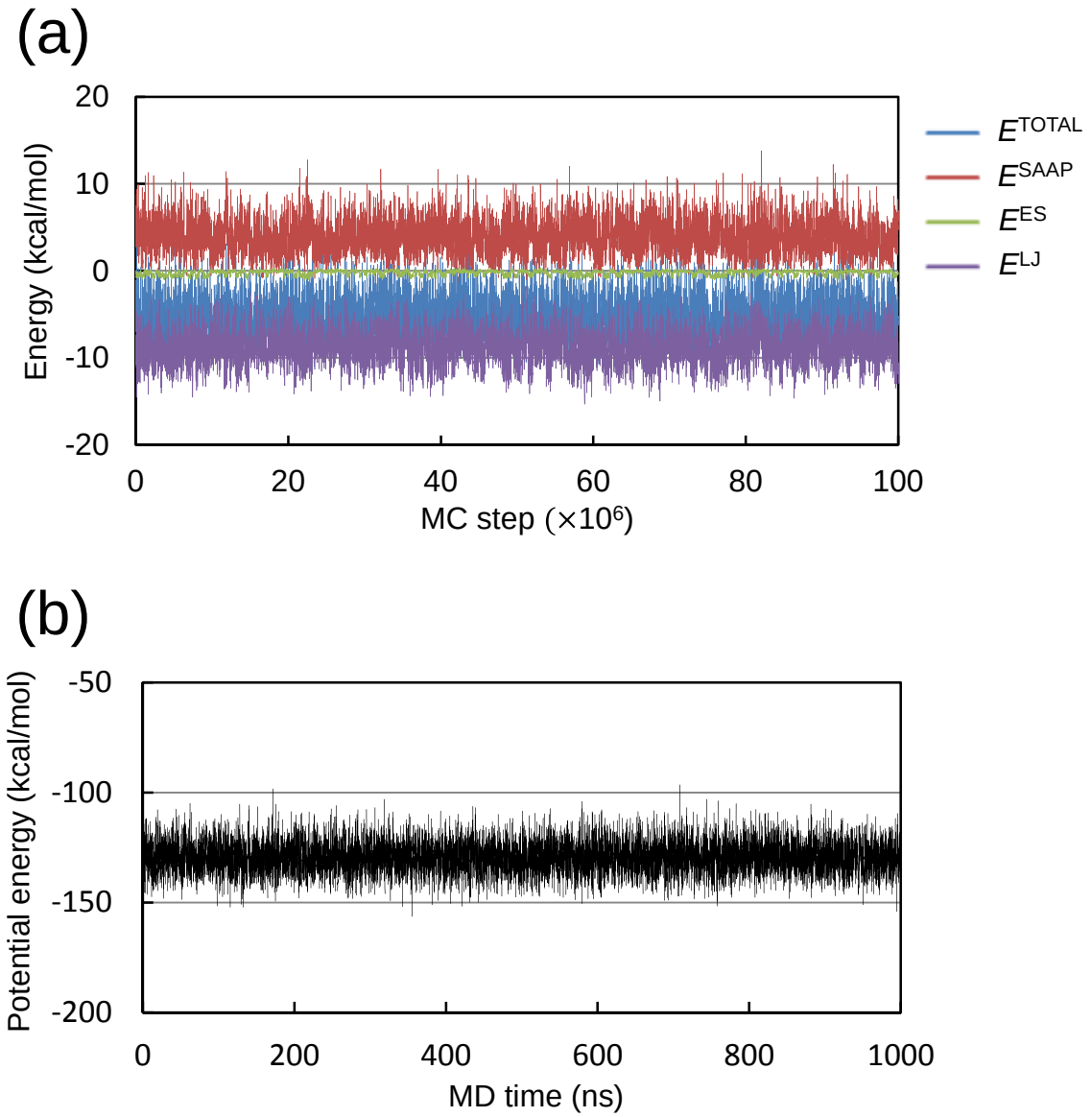

Figure S1. Traces of the energies obtained for Met-enkephalin by the SAAP-MC (trajectory 1) (a) and AMBER-MD (trajectory 1) (b) simulations at 300 K in water. The total energy (E^{TOTAL}) obtained by SAAP-MC are divided to E^{SAAP}, E^{ES}, and E^{LJ} terms.

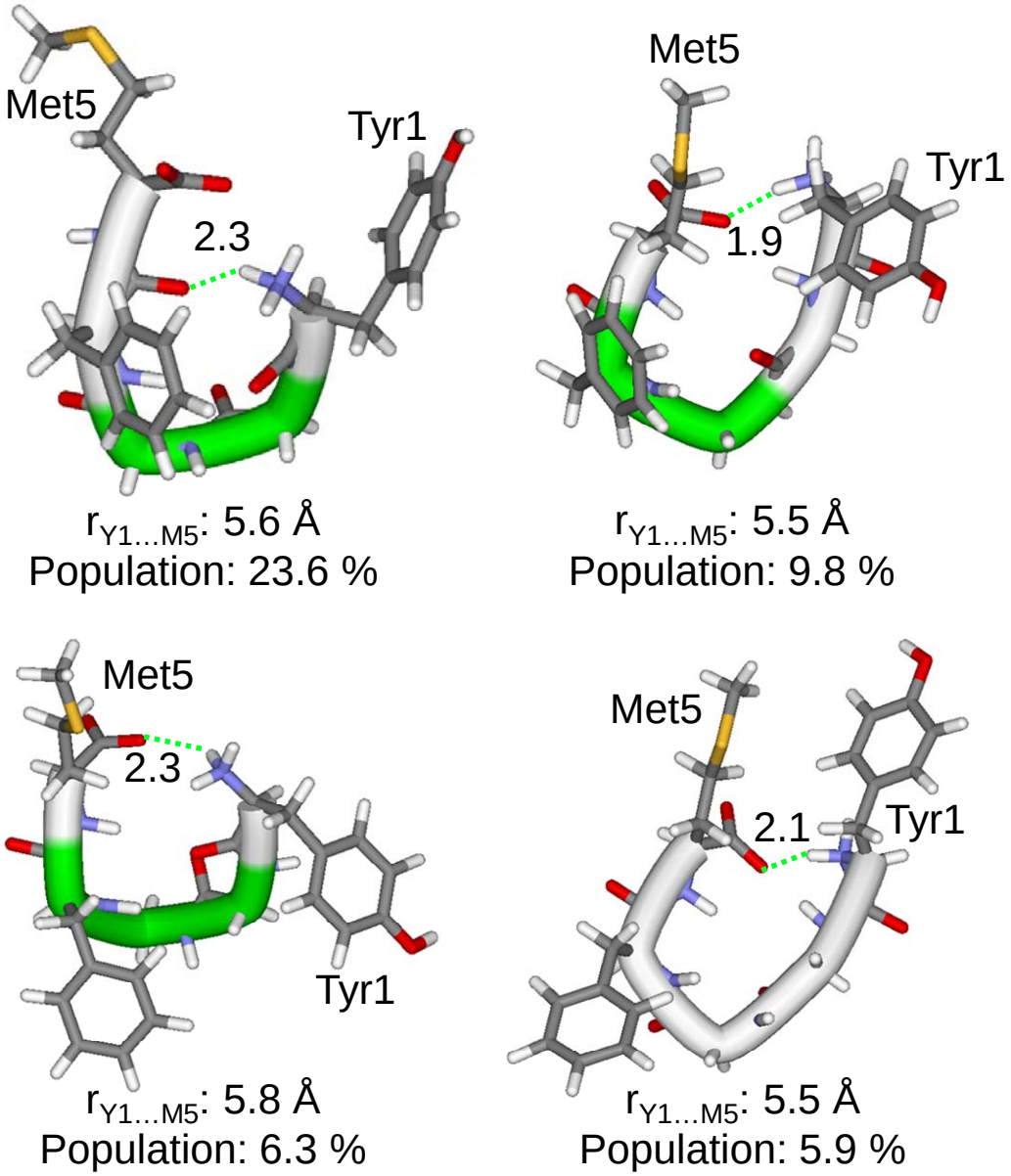

Figure S2. Representative structures of Met-enkephalin obtained by the AMBER-MD simulation (trajectory 1) at 300 K in water. The obtained 10,000 structures were analyzed by a structural clustering algorithm using the k-means method based on the main-chain RMSD. Distances of the hydrogen bonds are given in a unit of Å.

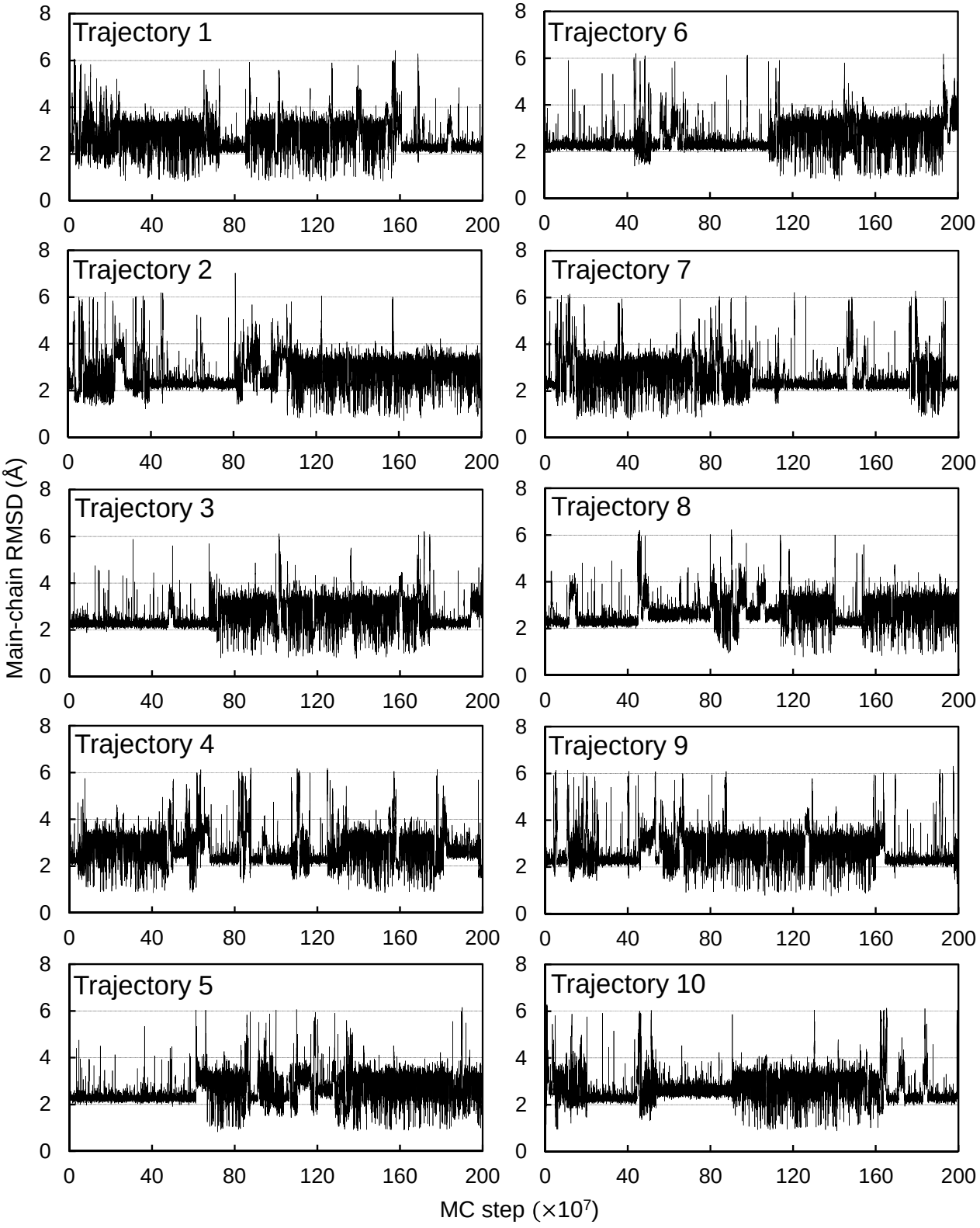

Figure S3-1. Traces of the main-chain RMSD obtained for chignolin by the SAAP-MC (trajectories 1–10) simulations at 300 K in water. The RMSD values were calculated with respect to the native structure.

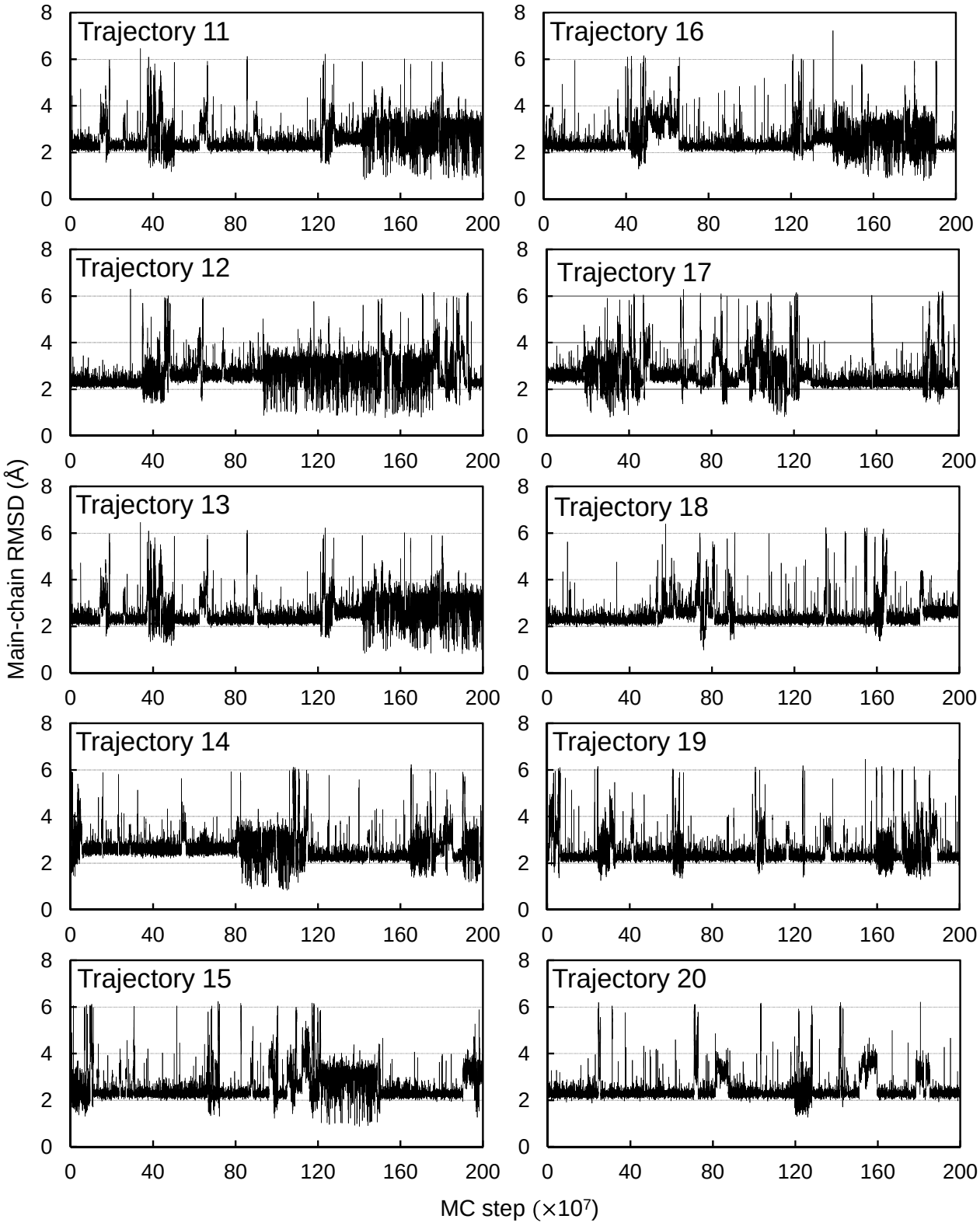

Figure S3-2. Traces of the main-chain RMSD obtained for chignolin by the SAAP-MC (trajectories 11–20) simulations at 300 K in water. The RMSD values were calculated with respect to the native structure.

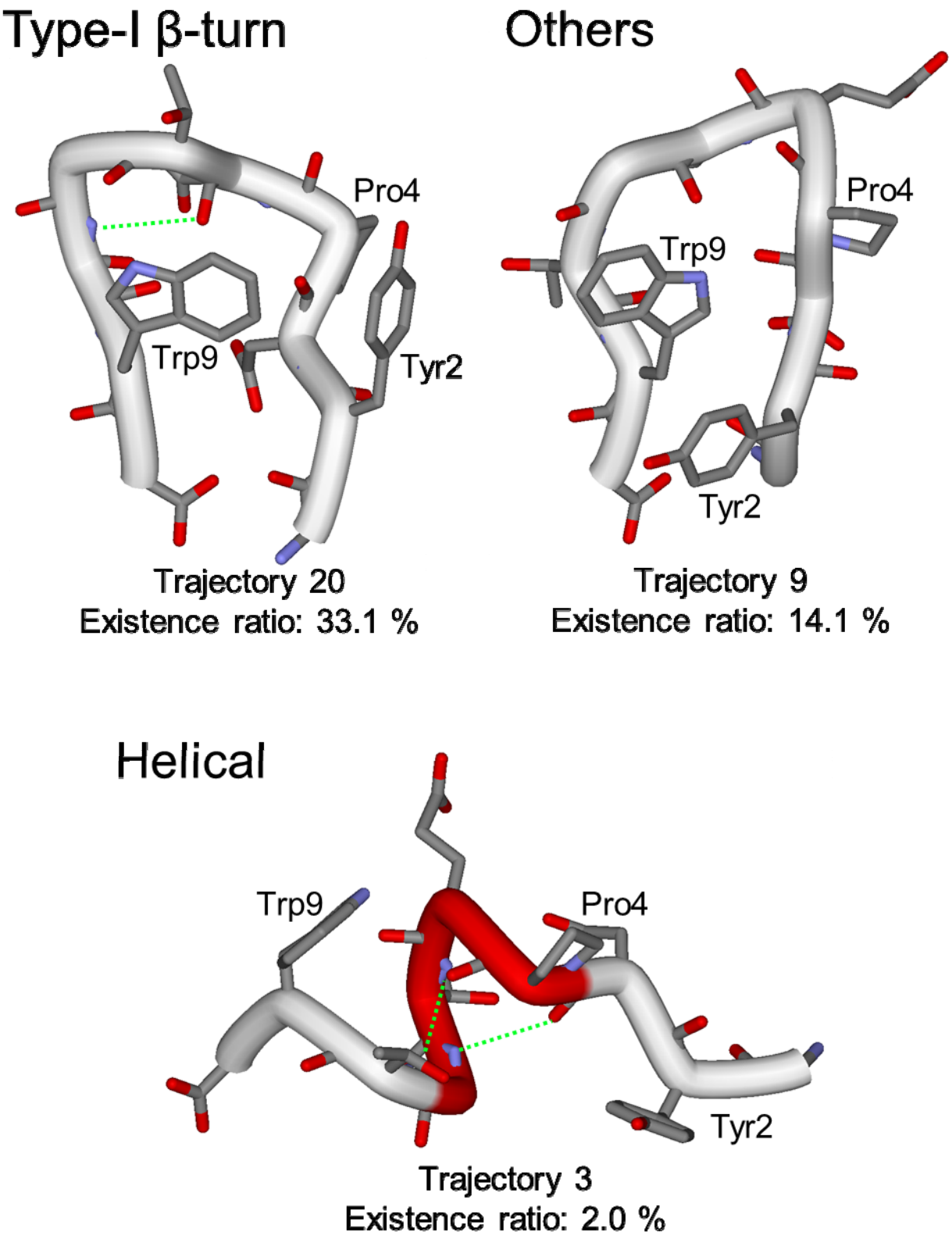

Figure S4. The minor structures of chignolin with a representative form obtained by the SAAP-MC simulation at 300 K in water. The hydrogen bonds are indicated by dotted lines.

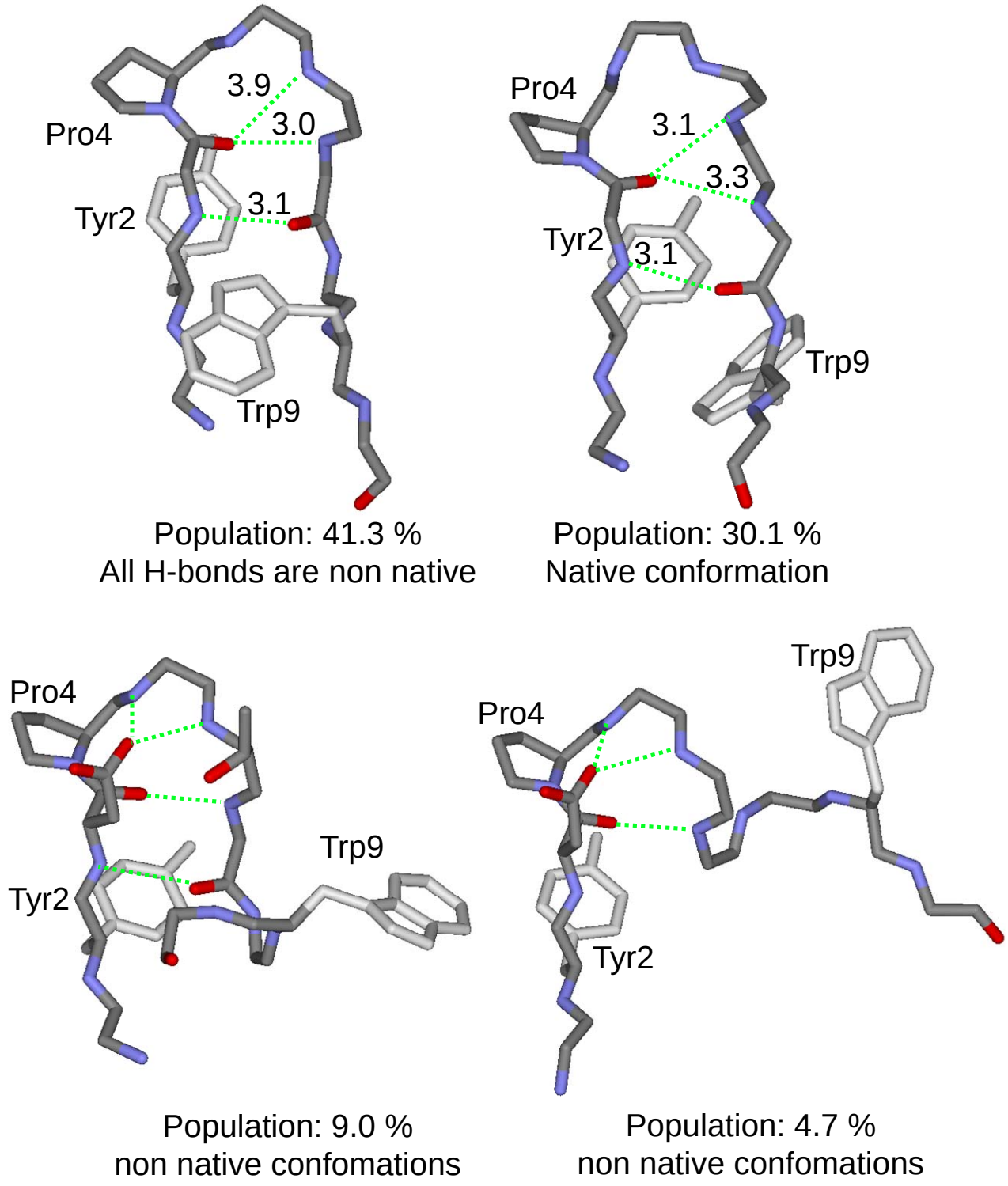

Figure S5. Representative structures of chignolin obtained by the AMBER-MD simulation at 300 K in water. The obtained 10,000 structures were analyzed by a structural clustering algorithm using the k-means method based on the main-chain RMSD values. The hydrogen bonds are indicated by dotted lines.

Table S1. Dihedral angles (degree) of each amino acid residue for representative structures A, A', B, and C of chignolin obtained by the SAAP-MC simulation at 300 K in water.

Structure A

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	-177.9	—	—	—
Tyr2	-103.3	116.7	-68.2	-87.3	-180.0
Asp3	-138.5	140.0	-99.3	76.7	—
Pro4	-73.1	-23.1	—	—	—
Glu5	-80.8	-25.5	28.3	178.7	128.6
Thr6	-82.9	-25.7	-56.3	-180.0	—
Gly7	97.0	18.5	—	—	—
Thr8	-114.4	139.2	82.2	-180.0	—
Trp9	-135.9	136.5	-53.1	-97.1	—
Gly10	165.6	-165.4	—	—	—

Structure A'

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	-169.1	—	—	—
Tyr2	-104.1	147.0	147.4	68.8	-180.0
Asp3	-129.1	141.3	-70.8	95.6	—
Pro4	-78.5	-24.9	—	—	—
Glu5	-92.3	-30.9	82.2	-177.7	80.8
Thr6	-79.6	-22.7	-51.2	-180.0	—
Gly7	51.5	40.3	—	—	—
Thr8	-120.4	125.7	-67.3	-180.0	—
Trp9	-99.3	148.1	-61.9	-111.8	—
Gly10	166.6	-150.8	—	—	—

Structure B

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	157.0	—	—	—
Tyr2	-112.7	141.0	163.6	94.9	-180.0
Asp3	-117.9	138.6	-79.9	101.4	—
Pro4	-62.2	-18.5	—	—	—
Glu5	-77.1	133.3	113.8	162.9	29.8
Thr6	-80.9	113.8	-63.2	-180.0	—
Gly7	131.7	-24.9	—	—	—
Thr8	-133.6	118.6	-61.4	-180.0	—
Trp9	-128.0	148.8	-69.8	-85.4	—
Gly10	168.5	-163.5	—	—	—

Structure C

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	-177.1	—	—	—
Tyr2	-109.8	143.9	170.7	-97.1	-180.0
Asp3	-115.0	149.3	-76.4	-89.5	—
Pro4	-65.4	-25.7	—	—	—
Glu5	-87.8	144.8	-118.7	-176.7	51.1
Thr6	-82.4	131.8	-72.6	-180.0	—
Gly7	106.2	-6.1	—	—	—
Thr8	-102.1	-27.3	43.8	-180.0	—
Trp9	-91.5	-21.3	59.3	-99.4	—
Gly10	141.0	14.2	—	—	—

Table S2. The structures of chignolin with the smallest all-atom RMSD from the native structure obtained by the SAAP-MC simulation at 300 K in water.

E^{TOTAL}: -21.43 kcal/mol.

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	169.1	—	—	—
Tyr2	-102.2	118.8	165.0	103.0	-15.3
Asp3	-99.6	129.6	-135.0	-100.1	—
Pro4	-83.5	-27.9	—	—	—
Glu5	-66.1	-28.3	76.9	-176.2	-101.6
Thr6	-95.6	-28.9	-164.7	52.6	—
Gly7	74.5	10.1	—	—	—
Thr8	-130.5	161.3	-169.4	59.6	—
Trp9	-54.8	134.9	-46.3	93.9	—
Gly10	153.1	6.0	—	—	—

Table S3. The structure of chignolin with the lowest total potential energy obtained by the SAAP-MC simulation at 300 K in water.

E^{TOTAL}: -33.36 kcal/mol.

	φ	ψ	χ_1	χ_2	χ_3
Gly1	—	-154.8	—	—	—
Tyr2	-119.7	138.8	173.6	84.5	173.6
Asp3	-114.6	150.8	-94.1	-64.0	—
Pro4	-75.3	-9.0	—	—	—
Glu5	-79.7	142.4	72.6	-178.8	87.4
Thr6	-79.0	124.8	-179.0	-70.3	—
Gly7	114.9	-22.8	—	—	—
Thr8	-134.4	128.2	-48.2	69.6	—
Trp9	-133.5	127.7	-20.3	103.6	—
Gly10	-172.2	173.5	—	—	—

Table S4. Results of single-point *ab initio* calculation for representative structure A of chignolin in water.

Calculation level: HF/IEFPCM/6-31+G(d,p)

SCF energy: -3800.42798 a.u.

No.	Atom	X	Y	Z	No.	Atom	X	Y	Z	No.	Atom	X	Y	Z
1	N	6.313	4.572	-0.307	47	C	-3.337	6.038	5.123	93	N	-1.986	1.010	-0.479
2	H	5.549	4.188	-0.847	48	H	-3.676	6.983	4.711	94	H	-1.761	1.959	-0.239
3	H	7.164	4.200	-0.658	49	H	-3.359	6.116	6.203	95	C	-1.057	-0.037	-0.100
4	H	6.333	5.554	-0.461	50	C	-4.192	4.883	4.598	96	H	-1.565	-0.970	-0.284
5	C	6.045	4.204	1.120	51	H	-5.248	5.115	4.552	97	C	0.299	-0.039	-0.930
6	H	6.036	5.100	1.721	52	H	-4.069	4.010	5.231	98	H	0.737	-1.019	-0.834
7	H	6.830	3.548	1.467	53	C	-3.597	4.572	3.208	99	C	0.178	0.245	-2.417
8	C	4.682	3.498	1.123	54	H	-3.615	3.508	2.999	100	H	1.162	0.197	-2.877
9	O	4.131	3.338	0.074	55	C	-4.408	5.220	2.090	101	H	-0.218	1.228	-2.593
10	N	4.218	3.116	2.312	56	O	-5.397	4.651	1.680	102	H	-0.446	-0.474	-2.919
11	H	4.759	3.400	3.111	57	N	-4.036	6.412	1.635	103	O	1.178	0.904	-0.379
12	C	2.901	2.543	2.531	58	H	-3.162	6.747	1.982	104	H	2.036	0.781	-0.749
13	H	2.504	2.344	1.548	59	C	-4.747	7.110	0.573	105	C	-0.847	0.086	1.413
14	C	2.884	1.182	3.338	60	H	-5.786	6.841	0.676	106	O	-0.731	1.168	1.949
15	H	1.866	0.942	3.594	61	C	-4.613	8.693	0.683	107	N	-0.797	-1.055	2.091
16	H	3.421	1.318	4.262	62	H	-5.460	9.074	1.242	108	H	-0.921	-1.904	1.568
17	C	3.467	-0.038	2.672	63	H	-4.656	9.137	-0.300	109	C	-0.586	-1.148	3.523
18	C	2.712	-0.845	1.858	64	C	-3.359	9.272	1.341	110	H	-0.120	-0.223	3.822
19	H	1.688	-0.604	1.691	65	H	-2.477	8.956	0.795	111	C	-1.911	-1.332	4.388
20	C	3.241	-1.957	1.244	66	H	-3.252	8.933	2.360	112	H	-1.651	-1.291	5.431
21	H	2.635	-2.558	0.606	67	C	-3.382	10.825	1.332	113	H	-2.322	-2.311	4.202
22	C	4.550	-2.300	1.454	68	O	-3.182	11.370	2.402	114	C	-3.025	-0.350	4.190
23	O	5.126	-3.366	0.863	69	O	-3.597	11.331	0.239	115	C	-3.285	0.727	4.930
24	H	4.489	-4.010	0.645	70	C	-4.323	6.559	-0.796	116	H	-2.760	1.093	5.771
25	C	5.324	-1.516	2.266	71	O	-5.093	6.630	-1.732	117	N	-4.391	1.383	4.472
26	H	6.337	-1.790	2.426	72	N	-3.125	5.995	-0.915	118	H	-4.791	2.183	4.879
27	C	4.781	-0.403	2.862	73	H	-2.604	5.919	-0.067	119	C	-4.887	0.712	3.413
28	H	5.396	0.190	3.498	74	C	-2.660	5.375	-2.148	120	C	-5.996	0.974	2.630
29	C	2.053	3.669	3.139	75	H	-3.135	5.911	-2.953	121	H	-6.620	1.820	2.810
30	O	2.366	4.208	4.181	76	C	-1.080	5.473	-2.325	122	C	-6.266	0.110	1.621
31	N	1.017	4.070	2.417	77	H	-0.635	4.921	-1.509	123	H	-7.115	0.283	0.999
32	H	0.841	3.580	1.557	78	C	-0.506	4.904	-3.607	124	C	-5.455	-0.997	1.386
33	C	0.143	5.170	2.775	79	H	0.571	5.025	-3.587	125	H	-5.696	-1.655	0.590
34	H	0.217	5.283	3.846	80	H	-0.882	5.430	-4.468	126	C	-4.362	-1.247	2.163
35	C	0.514	6.566	2.104	81	H	-0.727	3.856	-3.715	127	H	-3.750	-2.101	1.976
36	H	-0.389	7.070	1.802	82	O	-0.585	6.785	-2.279	128	C	-4.061	-0.384	3.197
37	H	1.113	6.389	1.225	83	H	-0.889	7.222	-1.513	129	C	0.430	-2.274	3.727
38	C	1.308	7.541	3.004	84	C	-3.194	3.939	-2.250	130	O	0.350	-3.315	3.107
39	O	2.495	7.277	3.133	85	O	-3.346	3.436	-3.344	131	N	1.395	-2.037	4.603
40	O	0.663	8.451	3.503	86	N	-3.496	3.288	-1.131	132	H	1.517	-1.123	4.986
41	C	-1.278	4.704	2.449	87	H	-3.404	3.791	-0.277	133	C	2.567	-2.862	4.819
42	O	-1.517	4.065	1.447	88	C	-4.132	1.994	-1.107	134	H	3.007	-3.153	3.871
43	N	-2.223	5.041	3.322	89	H	-4.692	1.846	-2.018	135	H	2.308	-3.774	5.346
44	C	-1.942	5.707	4.600	90	H	-4.827	1.957	-0.278	136	C	3.614	-2.066	5.651
45	H	-1.323	6.582	4.446	91	C	-3.207	0.798	-0.949	137	O	3.370	-0.869	5.800
46	H	-1.424	5.028	5.270	92	O	-3.631	-0.310	-1.199	138	O	4.548	-2.733	6.061

Table S5. Results of single-point *ab initio* calculation for representative structure A' of chignolin in water.

Calculation level: HF/IEFPCM/6-31+G(d,p)

SCF energy: -3800.43678 a.u.

No.	Atom	X	Y	Z	No.	Atom	X	Y	Z	No.	Atom	X	Y	Z
1	N	-1.069	-8.525	-5.290	47	C	-8.089	0.298	-4.319	93	N	-3.909	-3.593	2.157
2	H	-1.095	-8.594	-4.295	48	H	-9.114	0.059	-4.057	94	H	-4.469	-3.701	1.327
3	H	-0.353	-9.105	-5.635	49	H	-8.092	0.754	-5.301	95	C	-2.794	-4.495	2.366
4	H	-1.922	-8.893	-5.617	50	C	-7.457	1.200	-3.255	96	H	-2.224	-4.079	3.183
5	C	-0.892	-7.081	-5.630	51	H	-8.126	1.966	-2.887	97	C	-3.190	-5.980	2.762
6	H	-1.452	-6.855	-6.525	52	H	-6.576	1.688	-3.654	98	H	-3.906	-6.323	2.029
7	H	0.159	-6.897	-5.807	53	C	-7.015	0.221	-2.145	99	C	-2.071	-6.997	2.789
8	C	-1.377	-6.301	-4.400	54	H	-6.090	0.538	-1.682	100	H	-2.483	-7.963	3.053
9	O	-1.579	-6.915	-3.398	55	C	-8.040	0.153	-1.016	101	H	-1.332	-6.739	3.525
10	N	-1.526	-4.986	-4.550	56	O	-7.992	0.980	-0.129	102	H	-1.589	-7.086	1.832
11	H	-1.432	-4.639	-5.490	57	N	-8.996	-0.772	-1.076	103	O	-3.783	-6.090	4.025
12	C	-2.121	-4.119	-3.554	58	H	-8.909	-1.433	-1.824	104	H	-4.525	-5.528	4.085
13	H	-2.215	-4.707	-2.655	59	C	-10.057	-0.890	-0.084	105	C	-1.930	-4.392	1.104
14	C	-1.235	-2.836	-3.220	60	H	-10.115	0.069	0.403	106	O	-2.407	-4.528	-0.002
15	H	-0.619	-2.611	-4.076	61	C	-11.474	-1.230	-0.724	107	N	-0.651	-4.097	1.295
16	H	-0.571	-3.082	-2.403	62	H	-11.592	-0.681	-1.643	108	H	-0.334	-4.149	2.246
17	C	-1.933	-1.545	-2.847	63	H	-12.253	-0.903	-0.054	109	C	0.342	-4.082	0.239
18	C	-2.623	-1.394	-1.660	64	C	-11.795	-2.682	-1.064	110	H	-0.204	-4.070	-0.692
19	H	-2.649	-2.204	-0.968	65	H	-11.784	-3.290	-0.179	111	C	1.288	-2.803	0.282
20	C	-3.262	-0.224	-1.329	66	H	-11.068	-3.071	-1.756	112	H	2.028	-2.892	-0.500
21	H	-3.787	-0.120	-0.410	67	C	-13.195	-2.757	-1.723	113	H	1.818	-2.790	1.225
22	C	-3.225	0.846	-2.191	68	O	-13.226	-2.480	-2.902	114	C	0.667	-1.441	0.112
23	O	-3.858	1.978	-1.820	69	O	-14.113	-3.042	-0.985	115	C	0.840	-0.627	-0.935
24	H	-3.971	2.548	-2.554	70	C	-9.637	-1.877	1.013	116	H	1.368	-0.811	-1.840
25	C	-2.551	0.725	-3.374	71	O	-10.054	-1.722	2.146	117	N	0.222	0.579	-0.731
26	H	-2.510	1.550	-4.048	72	N	-8.830	-2.883	0.690	118	H	0.216	1.335	-1.363
27	C	-1.920	-0.465	-3.693	73	H	-8.476	-2.864	-0.243	119	C	-0.348	0.574	0.500
28	H	-1.402	-0.538	-4.621	74	C	-8.322	-3.834	1.671	120	C	-1.058	1.567	1.162
29	C	-3.537	-3.774	-4.032	75	H	-9.101	-3.957	2.407	121	H	-1.250	2.516	0.703
30	O	-3.795	-3.643	-5.209	76	C	-7.992	-5.252	1.029	122	C	-1.498	1.295	2.422
31	N	-4.445	-3.607	-3.084	77	H	-7.145	-5.115	0.373	123	H	-2.048	2.041	2.961
32	H	-4.136	-3.720	-2.135	78	C	-7.625	-6.362	1.989	124	C	-1.237	0.062	3.025
33	C	-5.817	-3.197	-3.319	79	H	-7.414	-7.259	1.419	125	H	-1.588	-0.116	4.017
34	H	-5.852	-2.822	-4.330	80	H	-8.438	-6.581	2.658	126	C	-0.542	-0.914	2.362
35	C	-6.887	-4.368	-3.186	81	H	-6.755	-6.118	2.571	127	H	-0.345	-1.854	2.838
36	H	-7.879	-3.943	-3.177	82	O	-9.042	-5.788	0.271	128	C	-0.089	-0.671	1.072
37	H	-6.731	-4.891	-2.257	83	H	-9.324	-5.174	-0.371	129	C	1.109	-5.410	0.306
38	C	-6.847	-5.418	-4.323	84	C	-7.140	-3.223	2.440	130	O	1.316	-5.972	1.361
39	O	-6.196	-6.421	-4.093	85	O	-6.875	-3.636	3.549	131	N	1.548	-5.884	-0.846
40	O	-7.462	-5.117	-5.335	86	N	-6.458	-2.230	1.881	132	H	1.330	-5.431	-1.709
41	C	-6.083	-2.022	-2.374	87	H	-6.707	-1.968	0.955	133	C	2.244	-7.136	-1.056
42	O	-5.644	-2.008	-1.244	88	C	-5.300	-1.609	2.482	134	H	1.750	-7.938	-0.523
43	N	-6.822	-1.032	-2.864	89	H	-5.583	-1.031	3.351	135	H	3.263	-7.080	-0.682
44	C	-7.225	-0.959	-4.275	90	H	-4.868	-0.934	1.755	136	C	2.276	-7.462	-2.576
45	H	-7.757	-1.854	-4.563	91	C	-4.199	-2.557	2.942	137	O	1.914	-6.550	-3.318
46	H	-6.351	-0.855	-4.906	92	O	-3.574	-2.307	3.948	138	O	2.661	-8.587	-2.847

Table S6. Results of single-point *ab initio* calculation for representative structure B of chignolin in water.

Calculation level: HF/IEFPCM/6-31+G(d,p)

SCF energy: -3800.39878 a.u.

No.	Atom	X	Y	Z	No.	Atom	X	Y	Z	No.	Atom	X	Y	Z
1	N	5.395	0.210	6.233	47	C	0.616	8.594	-0.375	93	N	-3.422	0.278	-1.006
2	H	4.625	-0.308	5.812	48	H	0.298	8.220	-1.345	94	H	-3.530	1.267	-0.846
3	H	5.226	0.220	7.228	49	H	1.355	9.368	-0.539	95	C	-2.090	-0.241	-1.263
4	H	6.264	-0.274	6.071	50	C	-0.587	9.088	0.433	96	H	-2.144	-1.300	-1.060
5	C	5.370	1.587	5.649	51	H	-1.305	9.644	-0.156	97	C	-1.550	-0.065	-2.743
6	H	5.906	1.563	4.709	52	H	-0.254	9.729	1.243	98	H	-1.616	0.981	-3.002
7	H	5.860	2.267	6.329	53	C	-1.194	7.802	1.037	99	C	-0.099	-0.485	-2.966
8	C	3.890	1.900	5.404	54	H	-1.569	7.981	2.037	100	H	0.158	-0.302	-4.002
9	O	3.142	0.973	5.305	55	C	-2.406	7.323	0.242	101	H	0.042	-1.542	-2.772
10	N	3.551	3.180	5.340	56	O	-3.471	7.877	0.430	102	H	0.594	0.065	-2.343
11	H	4.307	3.839	5.361	57	N	-2.263	6.340	-0.647	103	O	-2.329	-0.726	-3.710
12	C	2.223	3.663	5.014	58	H	-1.404	5.830	-0.573	104	H	-2.432	-1.627	-3.475
13	H	1.607	2.788	4.883	59	C	-3.397	5.709	-1.302	105	C	-1.205	0.386	-0.182
14	C	1.567	4.565	6.149	60	H	-4.114	6.496	-1.483	106	O	-1.123	1.592	-0.057
15	H	2.320	5.236	6.541	61	C	-3.046	5.041	-2.698	107	N	-0.611	-0.457	0.645
16	H	1.247	3.927	6.962	62	H	-2.042	5.301	-2.983	108	H	-0.708	-1.435	0.440
17	C	0.372	5.419	5.766	63	H	-3.695	5.445	-3.451	109	C	0.264	-0.068	1.734
18	C	-0.917	4.980	5.961	64	C	-3.143	3.521	-2.819	110	H	0.505	0.974	1.578
19	H	-1.082	4.014	6.386	65	H	-4.018	3.139	-2.329	111	C	-0.397	-0.218	3.176
20	C	-2.011	5.762	5.632	66	H	-2.301	3.049	-2.342	112	H	0.374	-0.106	3.921
21	H	-3.001	5.402	5.810	67	C	-3.195	3.022	-4.286	113	H	-0.795	-1.215	3.281
22	C	-1.826	7.010	5.088	68	O	-2.614	3.719	-5.107	114	C	-1.501	0.727	3.550
23	O	-2.858	7.829	4.773	69	O	-3.794	1.968	-4.454	115	C	-1.381	1.949	4.072
24	H	-3.630	7.335	4.596	70	C	-4.037	4.740	-0.296	116	H	-0.504	2.474	4.339
25	C	-0.547	7.478	4.897	71	O	-3.364	3.977	0.367	117	N	-2.606	2.525	4.257
26	H	-0.415	8.455	4.488	72	N	-5.362	4.803	-0.189	118	H	-2.770	3.403	4.669
27	C	0.531	6.689	5.240	73	H	-5.814	5.339	-0.907	119	C	-3.559	1.658	3.857
28	H	1.517	7.073	5.093	74	C	-6.154	3.824	0.541	120	C	-4.937	1.782	3.861
29	C	2.326	4.367	3.658	75	H	-5.526	3.492	1.355	121	H	-5.418	2.666	4.217
30	O	3.279	5.057	3.372	76	C	-7.488	4.390	1.175	122	C	-5.656	0.734	3.390
31	N	1.309	4.181	2.823	77	H	-8.060	4.836	0.373	123	H	-6.721	0.798	3.378
32	H	0.537	3.639	3.171	78	C	-8.392	3.384	1.858	124	C	-5.030	-0.420	2.924
33	C	1.165	4.841	1.542	79	H	-9.275	3.897	2.217	125	H	-5.624	-1.220	2.562
34	H	1.996	5.523	1.457	80	H	-7.904	2.932	2.703	126	C	-3.671	-0.535	2.925
35	C	1.204	3.860	0.290	81	H	-8.706	2.603	1.182	127	H	-3.202	-1.424	2.566
36	H	0.746	4.347	-0.558	82	O	-7.264	5.377	2.151	128	C	-2.910	0.516	3.398
37	H	0.639	2.971	0.515	83	H	-6.926	6.151	1.756	129	C	1.555	-0.875	1.569
38	C	2.616	3.411	-0.156	84	C	-6.333	2.619	-0.397	130	O	1.546	-1.999	1.109
39	O	2.952	2.292	0.193	85	O	-6.908	2.721	-1.460	131	N	2.661	-0.290	1.985
40	O	3.242	4.235	-0.805	86	N	-5.741	1.488	-0.016	132	H	2.659	0.666	2.277
41	C	-0.116	5.677	1.621	87	H	-5.391	1.461	0.916	133	C	4.008	-0.806	1.858
42	O	-1.121	5.247	2.147	88	C	-5.850	0.243	-0.737	134	H	4.189	-1.155	0.847
43	N	-0.057	6.890	1.081	89	H	-6.294	0.436	-1.703	135	H	4.162	-1.652	2.520
44	C	1.159	7.451	0.476	90	H	-6.484	-0.456	-0.210	136	C	5.031	0.313	2.210
45	H	1.681	6.698	-0.097	91	C	-4.525	-0.462	-0.984	137	O	4.541	1.411	2.484
46	H	1.820	7.820	1.253	92	O	-4.519	-1.655	-1.192	138	O	6.199	-0.042	2.171

Table S7. Results of single-point *ab initio* calculation for representative structure C of chignolin in water.

Calculation level: HF/IEFPCM/6-31+G(d,p)

SCF energy: -3800.43002 a.u.

No.	Atom	X	Y	Z	No.	Atom	X	Y	Z	No.	Atom	X	Y	Z
1	N	6.834	1.746	4.668	47	C	-0.243	9.375	-0.293	93	N	-2.117	0.629	-1.347
2	H	5.994	1.198	4.598	48	H	-0.671	8.933	-1.187	94	H	-2.404	1.365	-0.730
3	H	7.394	1.368	5.389	49	H	0.323	10.250	-0.589	95	C	-0.831	0.681	-2.028
4	H	7.334	1.637	3.822	50	C	-1.339	9.702	0.725	96	H	-0.851	-0.109	-2.759
5	C	6.401	3.156	4.922	51	H	-2.233	10.114	0.277	97	C	-0.577	2.067	-2.769
6	H	6.856	3.805	4.189	52	H	-0.972	10.417	1.452	98	H	0.487	2.135	-2.972
7	H	6.727	3.450	5.909	53	C	-1.606	8.361	1.444	99	C	-1.295	2.261	-4.097
8	C	4.868	3.152	4.818	54	H	-1.797	8.512	2.499	100	H	-1.007	3.219	-4.511
9	O	4.325	2.104	4.627	55	C	-2.855	7.667	0.906	101	H	-2.368	2.269	-3.959
10	N	4.266	4.324	4.952	56	O	-3.936	7.974	1.368	102	H	-1.044	1.490	-4.812
11	H	4.868	5.126	5.011	57	N	-2.730	6.793	-0.091	103	O	-0.948	3.175	-1.978
12	C	2.847	4.544	4.750	58	H	-1.794	6.478	-0.264	104	H	-0.208	3.499	-1.502
13	H	2.413	3.573	4.568	59	C	-3.831	5.983	-0.574	105	C	0.293	0.303	-1.053
14	C	2.113	5.192	6.004	60	H	-4.734	6.537	-0.360	106	O	1.315	-0.192	-1.487
15	H	2.678	6.047	6.335	61	C	-3.776	5.712	-2.141	107	N	0.118	0.528	0.250
16	H	2.110	4.480	6.811	62	H	-3.714	4.652	-2.329	108	H	-0.784	0.880	0.512
17	C	0.688	5.658	5.824	63	H	-2.871	6.162	-2.538	109	C	1.104	0.171	1.261
18	C	0.390	6.970	5.558	64	C	-4.932	6.220	-3.009	110	H	2.047	0.097	0.746
19	H	1.186	7.670	5.464	65	H	-5.024	7.294	-2.961	111	C	1.224	1.269	2.408
20	C	-0.905	7.409	5.419	66	H	-5.866	5.796	-2.665	112	H	1.553	2.201	1.963
21	H	-1.123	8.428	5.220	67	C	-4.757	5.800	-4.489	113	H	1.988	0.961	3.112
22	C	-1.943	6.529	5.541	68	O	-4.540	4.612	-4.664	114	C	-0.011	1.592	3.211
23	O	-3.203	6.989	5.381	69	O	-4.870	6.687	-5.314	115	C	-0.831	2.638	3.034
24	H	-3.794	6.550	5.951	70	C	-3.874	4.699	0.267	116	H	-0.780	3.408	2.300
25	C	-1.676	5.218	5.820	71	O	-2.861	4.163	0.661	117	N	-1.838	2.636	3.968
26	H	-2.479	4.529	5.919	72	N	-5.083	4.208	0.522	118	H	-2.543	3.323	4.053
27	C	-0.373	4.798	5.966	73	H	-5.824	4.626	-0.005	119	C	-1.666	1.573	4.796
28	H	-0.185	3.777	6.197	74	C	-5.297	2.896	1.107	120	C	-2.413	1.161	5.895
29	C	2.698	5.366	3.463	75	H	-4.453	2.717	1.757	121	H	-3.284	1.700	6.216
30	O	3.490	6.234	3.169	76	C	-6.616	2.783	1.979	122	C	-1.995	0.042	6.558
31	N	1.647	5.081	2.708	77	H	-7.419	3.233	1.407	123	H	-2.547	-0.301	7.410
32	H	1.005	4.404	3.078	78	C	-7.066	1.380	2.343	124	C	-0.857	-0.659	6.143
33	C	1.249	5.834	1.536	79	H	-7.988	1.448	2.910	125	H	-0.555	-1.529	6.687
34	H	1.966	6.635	1.427	80	H	-6.334	0.886	2.961	126	C	-0.127	-0.247	5.057
35	C	1.254	4.972	0.195	81	H	-7.250	0.775	1.469	127	H	0.745	-0.792	4.750
36	H	0.673	5.485	-0.560	82	O	-6.527	3.456	3.211	128	C	-0.530	0.887	4.359
37	H	0.794	4.010	0.387	83	H	-6.512	4.380	3.074	129	C	0.826	-1.241	1.796
38	C	2.664	4.707	-0.395	84	C	-5.215	1.866	-0.031	130	O	1.729	-1.873	2.306
39	O	3.072	5.555	-1.188	85	O	-5.805	2.029	-1.078	131	N	-0.385	-1.750	1.658
40	O	3.228	3.695	0.020	86	N	-4.441	0.802	0.198	132	H	-1.136	-1.153	1.384
41	C	-0.111	6.467	1.847	87	H	-4.039	0.720	1.105	133	C	-0.886	-2.951	2.303
42	O	-0.918	5.917	2.571	88	C	-4.298	-0.302	-0.719	134	H	-0.276	-3.187	3.164
43	N	-0.362	7.626	1.255	89	H	-5.086	-0.258	-1.457	135	H	-0.840	-3.799	1.625
44	C	0.620	8.351	0.438	90	H	-4.397	-1.236	-0.181	136	C	-2.364	-2.734	2.733
45	H	1.139	7.673	-0.225	91	C	-2.982	-0.374	-1.478	137	O	-2.776	-3.511	3.576
46	H	1.349	8.838	1.077	92	O	-2.767	-1.340	-2.179	138	O	-2.951	-1.835	2.131