Accurate Structure Prediction and Conformational Analysis of Cyclic Peptides with Residue-Specific Force Fields

Hao Geng,[†] Fan Jiang,[*,†] and Yun-Dong Wu[*,†,‡]

[†]School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, 518055, China

[‡]College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China

*E-mail: jiangfan@pku.edu.cn (F.J.) and wuyd@pkusz.edu.cn (Y.-D.W.)

Supporting Information

1. Residue-Specific Force Fields (RSFFs)

The development of a force field involves the balance between accuracy, complexity, and transferability. For the simulation of biomolecules such as protein, rather high accuracy in the description of the conformational energetics is needed. As demonstrated by Best et al.,[1] a modification of backbone ψ potential of about 1 kJ/mol can result in dramatic change of the α-helix content of a poly-Ala-based peptide. Besides, commonly-used force fields also have problems in describing the different intrinsic conformational propensities of different amino acid (AA) residues.[2] On the other hand, for simulation speed and the ease of implementation and parameterization, it would be very useful to increase the force field accuracy without increasing its complexity (i.e. without introducing non-classical potential functions such as electronic polarization).

Therefore, we choose to sacrifice the transferability by using different force field parameters for different AA residues, thanks to the very limited chemical space of these building blocks of peptides/proteins. Also, we are interested in the behaviors of peptides/proteins in aqueous solution, which is quite different from those in vacuum. The parameterization of RSFFs is based on the conformational free energy surfaces of short peptides in aq. solution, not the conformational energies of isolated peptides. Due to the lack of high-resolution free energy surfaces from direct experiments, we use the statistics of many protein crystal structures instead. Specifically, we constructed the so-called 'coil library' from the residues outside the secondary structure regions (helices, sheets, turns), to approximate the intrinsic conformational preferences of AA residues without the influence of inter-residue backbone hydrogen-bonding.

The overall potential energy function of a RSFF is within the framework of classical force field:

$$V_{\text{total}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + V_{\text{local-LJ}} + V_{\text{LJ}} + V_{\text{local-Coulomb}} + V_{\text{Coulomb}} \quad (1)$$

All the bond stretching (V_{bond}), bond-angle bending (V_{angle}), van der Waals interactions (Lennard-Jones potential V_{LJ}), and electrostatic interactions (Coulomb potential, V_{Coulomb}) are adopted from the OPLS-AA/L force field (for RSFF1) and AMBER-99SB force field (for RSFF2). On the other hand, the dihedral angle potential (V_{torsion}) for all the rotable bonds were reparameterized using the statistical free energy surfaces (potential of mean force, PMF) for protein coil library as reference, such that the backbone ϕ, ψ and side-chain χ PMF obtained for dipeptide simulations agree excellently with the reference data. In both RSFF1 & RSFF2, the 1-4 L-J interactions ($V_{\text{local-LJ}}$) are scaled by a factor of 0.50, whereas the 1-4 Coulomb interactions ($V_{\text{local-Coulomb}}$) are not scaled down to achieve

more balanced intra-residue electrostatic interactions. Unlike commonly-used protein force fields, special parameters were used for some 1-5 and 1-6 van der Waals interactions (included in the $V_{\mathrm{local-LJ}}$) to optimized the coupling between neighboring torsions.

Unlike most protein force fields, RSFFs were parameterized based on free energies instead of potential energies. However, the parameters of a dihedral-angle θ can be efficiently optimized by adding corrections equal to the difference between the coil-library PMF and the simulated PMF:

$$V_{\mathrm{new}}(\theta) - V_{\mathrm{old}}(\theta) = G_{\mathrm{coil}}(\theta) - G_{\mathrm{MD,old}}(\theta) \qquad (2)$$

A ϕ, ψ free energy decomposition approach was invented to decompose the difference between coil-library and simulated 2D Ramachandran plots into separate ϕ and ψ corrections. The parameters for specially treated 1-5/1-6 L-J interactions were modified manually. The detailed procedure of parameterization can be found in our previous works.

The RSFF1 force field has been parameterized with the TIP4P-Ew water model, and the RSFF2 has been parameterized with the TIP3P water model. Thus, they are usually used with these two water models, respectively.

Simulations of all dipeptides (blocked amino acids, Ac-X-Nme) using RSFF1 & RSFF2 give $^{3}J_{\mathrm{HNa}}$-coupling constants in excellent agreement with NMR measurements, much better than some recent force fields. Also, an independent benchmark study using 256 two-residue peptides (Ac-X-Y-NH_2) has shown that RSFF2 can improve the modeling of sequence-specific conformational behavior of peptides.

A brief description of the usage of the RSFF1 force field in *Gromacs*:

Prepare the system with built-in OPLS-AA/L force field and tip4p water, and then run

```
python g_mod_top_RSFF1.py x.top x.top
```

to modify the topology file and add all new parameters into it.

Also, make sure that the RSFF1.itp & tip4pEw.itp files are in the directory of the topology file.

A brief description of the usage of the RSFF2 force field in *Gromacs*:

Prepare the system with built-in Amber-99SB force field and tip3p water, and then run

```
python g_mod_top_RSFF2.py x.top x.top
```

to modify the topology file and add all new parameters into it.

Also, make sure that the RSFF2.itp & tip3p.itp are in the directory of the topology file.

All the necessary files for running RSFF1/RSFF2 simulations in the *Gromacs* have been uploaded as Supporting material of this paper. You can easily find all force field parameters in the Python program g_mod_top_RSFF1.py and g_mod_top_RSFF2.py.

2. Simulation Details

All simulations were carried out using the *Gromacs* 4.5.4.[3] As same as our previous works, TIP4P-Ew[4] and TIP3P[5] water models were used with RSFF1 and RSFF2 force fields, respectively. The electrostatics were treated using the particle-mesh Ewald (PME)[6] with a real-space cutoff of 0.9 nm and van der Waals interactions were also cutoff at 0.9 nm with long-range dispersion correction for energy and pressure. A velocity rescaling thermostat[7] with τ_T = 0.2 ps and a Berendsen barostat[8] with τ_P = 0.5 ps were used to maintain constant temperature and constant pressure (for NPT simulations). All bonds involving hydrogen were constrained using LINCS,[9] and a time step of 2 fs was used for cyclic peptide simulations and 3 fs for protein simulations.

Cyclic peptides. Each cyclic peptide was initially constructed using the *HyperChem* software. Each was solvated with 508-1034 water molecules depending on the size. After energy minimization and a 300 K NPT MD simulation to get the box volume, a 600 K (except for RSFF2 simulations of peptides 4, 5, 14, 20, 500K were used to avoid *cis-trans* isomerization of peptide bonds) NVT MD simulation of 30 ns was performed to obtain the initial structures for REMD. For each REMD simulation, 24 to 32 replicas were used with temperature range from 300 K to 600 K. However, for RSFF2 simulations of peptides 4, 5, 14, 20, *T* range of 300-500K is used to avoid *cis-trans* isomerization of peptide bond. The intermediate temperatures were chosen following a recent study[10] to obtain uniform exchange rate, and exchanges were attempted between neighboring replicas every 1.0 ps. The structures were saved every 1.0 ps. For all pentapeptides and hexapeptides, each replica was simulated for 100 ns, and first 10 ns was discarded in all analyses. For larger cyclic peptides, 200 ns per replica was simulated, and first 20 ns was discarded.

Linear peptides. Each peptide has the sequence as same as the corresponding cyclic peptide (as in Table 1 and Table S1). The N-terminal and C-terminal of each linear peptide was blocked by acetyl (Ace-) and N-methylamino (-Nme) groups, respectively. Their initial structures (α-helix) were constructed using the *HyperChem* software. Each peptide was solvated with 534-1238 water molecules depending on the size. After energy minimization and a 300 K NPT MD simulation to get the box volume, a 500 K NVT MD simulation of 30 ns was performed to obtain the initial structures for REMD. For each REMD simulation, 24 to 32 replicas were used with temperature range from 300 K to 500 K. Exchanges were attempted between neighboring replicas every 1.5 ps. The structures were saved every 0.6 ps. For all penta- and hexa- peptides, each replica was simulated for 90 ns, and first 9

ns was discarded in analyses. For larger peptides, 180 ns per replica was simulated, and first 18 ns was discarded. We use a slightly shorter simulation time for each linear peptide because it usually has lower energy barriers for conformational transitions compared with the cyclic one.

Table S1. Simulation settings for the 20 cyclic peptides and their linear counterparts.

			cyclic peptides			linear peptides	
	sequence	size	run time (ns)	# water molecules OPLS-AA/L & RSFF1	# water molecules ff99SB-ildn & RSFF2	run time (ns)	# water molecules RSFF2
1	GPGaP	5	100 x 24	512	543	90 x 24	543
2	fPGaP	5	100 x 24	508	539	90 x 24	534
3	GPSaP	5	100 x 24	509	541	90 x 24	541
4	GPfGA	5	100 x 24	510	541	90 x 24	534
5	GPfGV	5	100 x 24	508	541	90 x 24	534
6	FlGfLG	6	100 x 24	588	599	90 x 24	639
7	GHGAYG	6	100 x 24	602	587	90 x 24	643
8	GPLTLF	6	100 x 24	591	585	90 x 24	639
9	VSAPGF	6	100 x 24	599	590	90 x 24	644
10	pFTFWF	6	100 x 24	583	577	90 x 24	625
11	GTFLYV	6	100 x 24	587	586	90 x 24	634
12	GSPSWLV	7	200 x 28	720	706	180 x 28	863
13	GYGPLIL	7	200 x 28	717	705	180 x 28	863
14	AAYPPIGV	8	200 x 28	714	705	180 x 28	867
15	aGPfaGPf	8	200 x 28	722	706	180 x 28	865
16	aGPFaGPF	8	200 x 28	721	704	180 x 28	863
17	AGPFAGPF	8	200 x 28	717	706	180 x 28	862
18	PPAGLATF	8	200 x 28	718	701	180 x 28	858
19	$(VPG)_4$	12	200 x 32	1034	925	180 x 32	1231
20	$(APGVGV)_2$	12	200 x 32	1031	927	180 x 32	1238

There are some unnatural D-amino acids (indicated by lower-case letters) in these peptides. However, all dihedral-angle potential functions used in our force field modifications are even functions, and all distances and angles are not affected by enantiomeric inversion. Thus, RSFF1 & RSFF2 are expected to be equally applicable to both L- and D- enantiomers of an amino acid.

Proteins. The four small model proteins simulated in this work are: the third Ig-binding domain of streptococcal protein G (GB3), the bovine pancreatic trypsin inhibitor (BPTI), the human ubiquitin (Ubq), and the hen egg-white lysozyme (HEWL). Their Protein Data Bank (PDB) structures (PDB ID:

1P7E, 5PTI, 1UBQ, and 6LYT, respectively) were used as initial structures. Around 2500, 2500, 2600, 4000 water molecules were added for the simulations of 1P7E (56 residues), 5PTI (58 residues), 1UBQ (76 residues), and 6LYT (129 residues), respectively. Each system is neutralized by adding counter ions. After energy minimization, a 2 ns NPT simulation was carried out with 1000 kJ/nm^2 position restraints on all main-chain atoms and the system is gradually heated up to 300 K during the first 1 ns. Each production run is an unrestrained 0.8 μs NPT simulation at 300 K, and the first 0.2 μs is discarded in the calculation of conformational entropy.

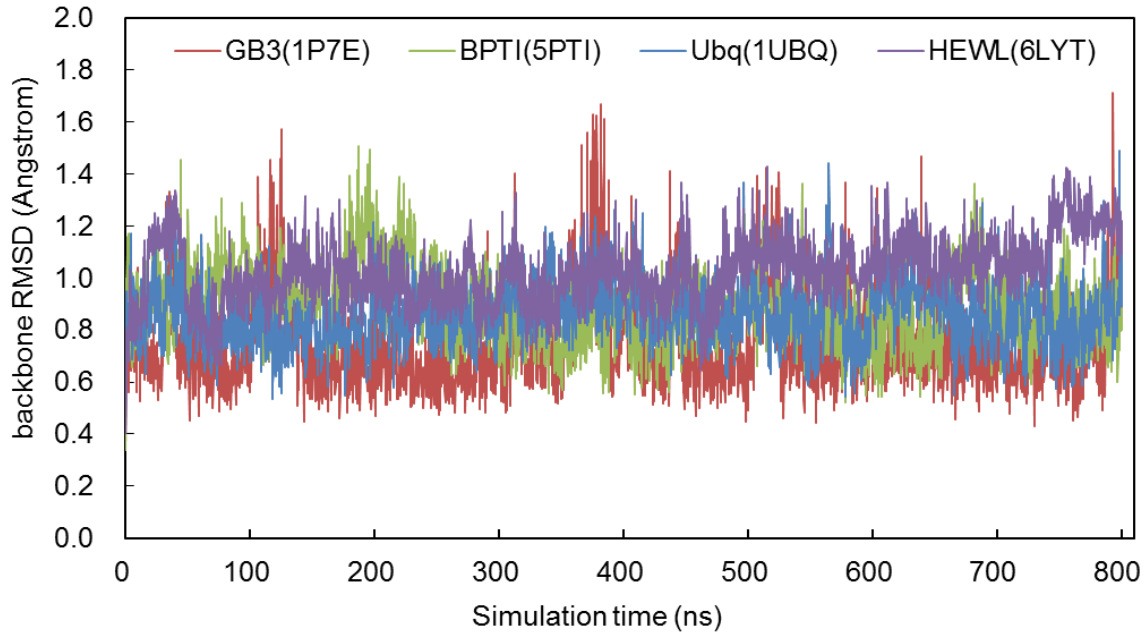


Figure S1. The time evolution of the backbone RMSD (in Å) to corresponding experimental structure, from the MD simulation of each protein using RSFF2 force field. For the RMSD calculations of the GB3, BPTI, Ubq simulations, a few flexible C-terminal residues (residues 56, 56-58, 72-76, respectively) were excluded. For HEWL, the flexible loop of residues 68-72 was excluded in the RMSD calculation.

3. Clustering Analysis

Each clustering analysis was performed using 30000 frames sampled evenly from a simulation trajectory at 300K, following the Rodriguez-Laio[11] algorithm. We define distance d_{ij} between two date points (ϕ_i, ψ_i) and (ϕ_j, ψ_j) on the Ramachandran map of a residue as:

$$d_{ij} = \sqrt{(\phi_i - \phi_j)^2 + (\psi_i - \psi_j)^2}$$

The local density ρ_i around point *i* is defined as

$$\rho_i = \sum_j \chi(d_{ij} - d_c)$$

Where $\chi(x) = 1$ if $x < 0$ and $\chi(x) = 0$ otherwise, and d_c is a cutoff distance, which is set to 5 degrees here. Basically, ρ_i is equal to the number of points that within a radius of d_c around point *i*. Then, we compute the minimum distance between point *i* and any other point with higher density:

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij})$$

The all cluster centers (density maxima) were defined as points with $\delta_i > 60$ degrees. After the cluster centers have been found, each observed (ϕ, ψ) point is assigned to the same cluster as its nearest neighbor of higher density. Residues in the same (ϕ, ψ) cluster belong to the same single-residue-level conformation.

For an entire structure of a cyclic peptide, its conformation can be encoded as a string of the conformational states of all residues along the sequence. Structures with the same conformational code belong to the same conformation. The relative free energy of each conformation was calculated based on:

$$\Delta G(\text{confomation}_i) = -RT \frac{p(\text{confomation}_i)}{p(\text{confomation}_0)}$$

Here, $p(\text{conformation}_i)$ is the probability of a certain conformation and $p(\text{conformation}_0)$ is the probability of the most populated conformation.

In a few cases, a small amount (<10%) of structures with *cis*- peptide bond(s) can be found in the REMD trajectory. These structures were excluded in the population analysis and free energy calculation.

The representative structure of each CP conformation was chosen as the structure with maximal ρ_{sum} (The summation of local densities of all residues in one structure, $\rho_{sum} = \sum_{i=1}^{n\_res} \rho_i$), among all structures belong to this conformation. The ϕ, ψ of each residue in the representative structure was

usually quite close to the corresponding cluster center in the Ramachandran plot.

4. Calculation of Local ϕ, ψ Free energies (ΔG_{local})

Previous studies indicate that the statistical ϕ, ψ distribution from a protein coil library (residues outside secondary structure regions) can give a very good approximation of the intrinsic conformational preference of a single amino acid residue.[12,13,14,15,16] Based on this, the intrinsic propensity of a given residue to adopt a certain conformation can be quantified by free energy:[17]

$$\Delta G_{local}(\text{conformer}) = -RT \ln p(\text{conformer})$$

Here, p(conformer) is the intrinsic probability of adopting a certain conformer. As same as our previous works,[18] a 2D Gaussian kernel was used to estimate the (unnormalized) probability density of any given ϕ, ψ for a given residue type:

$$p(\phi,\psi) = \sum_i w_i \exp[-\frac{\Delta^2(\phi,\psi,\phi_i,\psi_i)}{2\sigma^2}]$$

Here i count for all residues of the given type in a coil library, and w_i is the $1/m$ weighting for m identical chains in one PDB structure. Δ^2 is the squared distance between the given conformation (ϕ, ψ) and the i^{th} observed conformation (ϕ_i, ψ_i) from the coil library, considering the periodicity of the dihedral angles:

$$\Delta^2(\phi,\psi,\phi_i,\psi_i) = \min(|\phi_i-\phi|, 360°-|\phi_i-\phi|)^2 + \min(|\psi_i-\psi|, 360°-|\psi_i-\psi|)^2$$

Here, our Coil-3[16] coil library is used, which has also been used to obtain reference data for RSFF1 and RSFF2 developments.

In all calculations, smoothing parameter σ is set to 20 °, such that the results from different simulations can be directly compared. For the cyclic peptides and globular proteins, their experimental structures were used to calculate each residue's local free energy. For the linear peptides corresponding to the 20 CPs, there is no experimental structure. Thus, for each peptide, 100 structures evenly sampled in its RSFF2 simulation (300 K replica) were used, and the average value was reported. The standard error of the mean was estimated to be in range 0.1 ~ 0.3 kJ/mol.

Some peptides contain D-amino acid residues, which are not in the protein coil library. For each of them, the inverse (-ϕ, -ψ) was used in the calculation of its local conformational free energy.

5. Calculation of Backbone Conformational Entropies

The conformational entropy was calculated based on the Boltzmann-Shannon entropy:

$$S = -R\sum_{i=1}^{N} p_i \ln p_i$$

where p_i is the probability distribution in a conformational space divided into N equal sized bins. The detailed calculation follows the method used in a previous work, with a bin size of 10 °.[13] Briefly, we calculated the conformational entropy of a single residue j ($S^{(1)}_j$) based on the its 2-dimensional (ϕ_j, ψ_j) distribution. To include the influence of neighboring residue, we also calculated the conformational entropy of residue pairs: $S^{(2)}_{j-1,j}$ and $S^{(2)}_{j,j+1}$ based on the 4-dimensional distributions (ϕ_j, ψ_j, ϕ_j, ψ_j) and (ϕ_j, ψ_j, ϕ_j, ψ_j), respectively. Finally, the reported residue-wise conformational entropy for j^{th} residue is:

$$S_j = S^{(1)}_j + [S^{(2)}_{j-1,j} + S^{(2)}_{j,j+1}]/2$$

Any residue in a cyclic peptide has its preceding (j-1) and successive (j+1) residues. However, for the 20 linear peptides and the four proteins, the backbone entropies of each N-terminal residue and each C-terminal residue were not calculated, because they do not have preceding and successive residues, respectively. Thus, for linear peptides or proteins in Figure 4, there can be less points for $-T\Delta S_{BB}$ data compared with those for ΔG_{local} data, since all residues were calculated for the later.

6. Treatment of Symmetry for CPs with Repeated Sequences

Some CPs have repeated sequences, such as aGPFaGPF and $(VPG)_4$. In the conformational clustering analysis of these CPs, equivalent conformational sequences are considered as identical CP conformation. As an example: for aGPFaGPF, 11222222 and 22221122 belong to one conformation. Another example: for $(VPG)_4$, 111112113112 and 112113112111 belong to one conformation.

For these cyclic peptides, the RMSD of a certain structure with respect to corresponding crystal structure was calculated as the lowest one in all possible structure alignments: two alignments for each of aGPFaGPF, aGPFaGPF, aGPFaGPF, $(APGVGV)_2$, and four alignments for $(VPG)_4$.

In the calculation of ΔS_{BB} (Figure 4), for CPs with repeated sequences, only one residue among residues in equivalent positions has been considered, because they have nearly the same ϕ, ψ distributions from the simulation. On the other hand, ΔG_{local} for cyclic peptides were calculated based on their crystal structures. In the crystal structures, some equivalent residues have very similar ϕ, ψ angles, and only one residue was chosen for analysis. However, the ΔG_{local} of residues in equivalent positions but with quite different observed conformations were all calculated in Figure 4.

7. References

(1) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix−coil transition of polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004-9015.

(2) Meral, D.; Toal, S.; Schweitzer-Stenner, R.; Urbanc, B. Water-centered interpretation of intrinsic pPII propensities of amino acid residues: in vitro-driven molecular dynamics study. J. Phys. Chem. B **2015**, *119*, 13237-13251.

(3) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. *Bioinformatics* **2013**, *29*, 1−10.

(4) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.

(5) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926

(6) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(7) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(8) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.

(9) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(10) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2025-2027.

(11) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science* **2014**, *344*, 1492–1496.

(12) Serrano L. Comparison between the φ distribution of the amino acids in the protein

database and NMR data indicates that amino acids have various φ propensities in the random coil conformation. *J. Mol. Biol.* **1995**, *254*, 322-333.

(13) Fiebig K M, Schwalbe H, Buck M, et al. Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem.* **1996**, *100*, 2661-2666.

(14) Smith L J, Bolin K A, Schwalbe H, et al. Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **1996**, *255*, 494-506.

(15) Avbelj F, Grdadolnik S G, Grdadolnik J, et al. Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Natl. Acad. Sci.* **2006**, *103*, 1272-1277.

(16) Jiang F, Han W, Wu Y D. The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3413-3428.

(17) Zhou, C.-Y.; Jiang, F.; Wu, Y.-D. Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER ff99SB. *J. Phys. Chem. B* **2014**, *119*, 1035–1047.

(18) Jiang F, Wu Y D. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J. Am. Chem. Soc.* **2014**, *136*, 9536-9539.

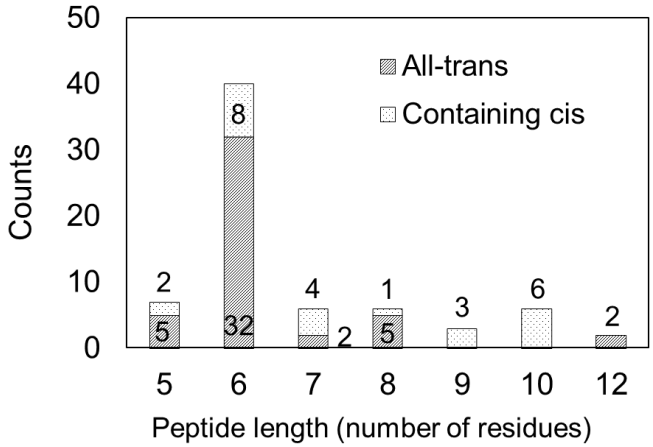


Figure S2. The length distribution of small cyclic peptides retrieved from CSD. The numbers of peptides with all-*trans* peptide bonds and ones with *cis* peptide bonds were both given.

Table S2. Sequences and β-turn patterns of 32 all-*trans* hexapeptides.

	CCDC code	sequence[a]
1	CAHWEN	FlGfLG[b]
2	BAMLIK	GHGAYG
3	VAWTAQ	PLTLFG
4	YEXJIV	VSAPGF
5	GAJDUQ	pFTFWF
6	ZUKRAY	VGTFLV
7	AAGAGG10	AAGAGG
8	AAGGAG10	AAGGAG
9	BIHXUL10	PVFFAG
10	BIXPAZ10	FlGLFG
11	CAMVES	pPAAAA
12	CGDLLL10	lLGlLG
13	CGLEGL	LGGLGG
14	CGLPGL	PGGPGG
15	CLPGDH	LFGlfG
16	CYBGPP	PfGPfG
17	CYCAPP	PfAPfA
18	CYDGPA	PaGPaG
19	CYHEXG	GGGGGG
20	DICWET	GSGGSG
21	DUYTIA	PFFpFF
22	GAJFAY	pFTFWF
23	GGAAGG	GGGaaG
24	GUFXOU	PFFpAA
25	JEHMAK	PGYGPL
26	PAPRVA	fPVfPV
27	PHLEGL10	FlGFLG
28	VEGROO	LFGFLG
29	VUGMEP	pFASFF
30	BIHTUH	FPAFPA[c]
31	XOPZEH	FfFfFf[d]
32	BUYXOI	PPpPPp[d]

[a]Lowercase letter indicates d-amino acid residues. [b]The residues in black form β-turns with the two resides in red as the opposite residues. [c]dimer with intermolecular H-bonds in crystal structure. [d]formed entirely by hydrophobic residues.

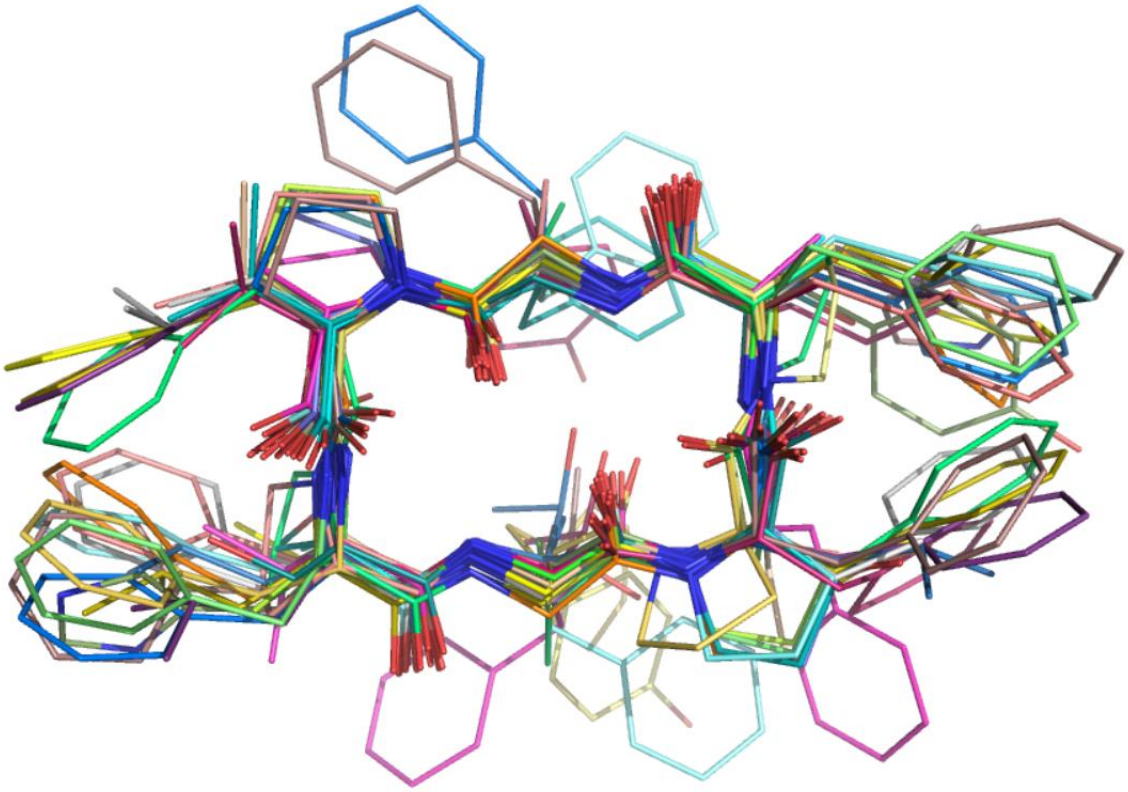


Figure S3. Superposition of crystal structures of all-*trans* hexapeptides in two-β-turns structures.

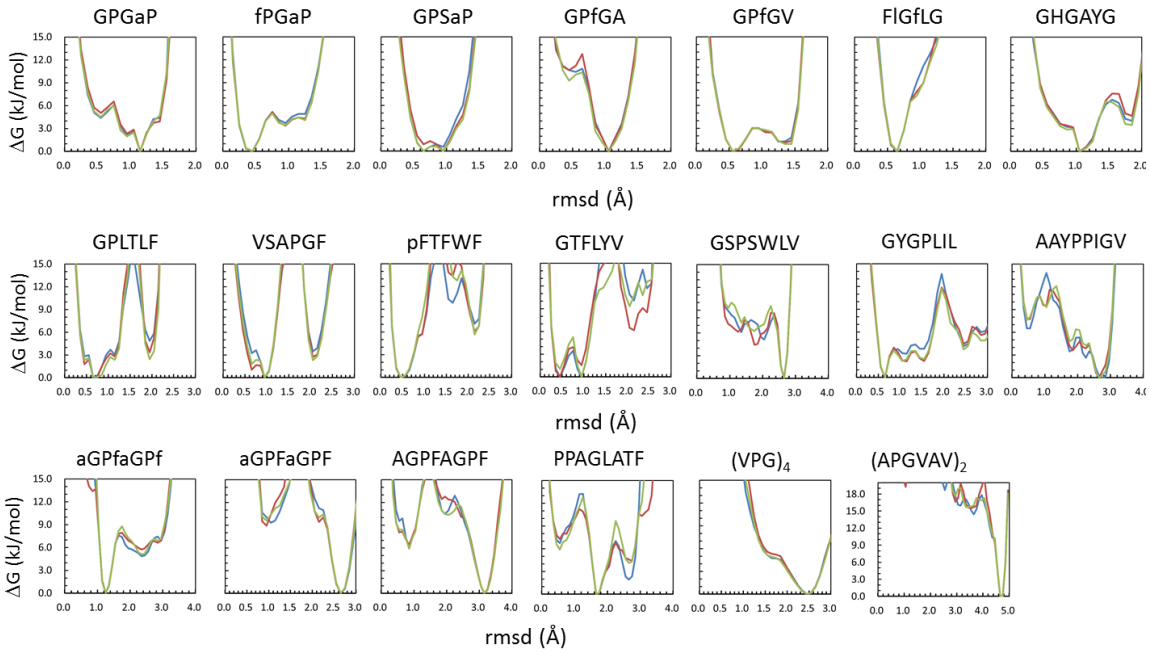


Figure S4. The convergence of the OPLS-AA/L simulations of the 20 CPs. In each plot, the three free energy curves of the RMSD (to crystal structure) are from three equal-length non-overlapping time windows. For penta- and hexa- peptides: blue for 10 ns – 40 ns, red for 40 ns – 70 ns, and green for 70ns – 100ns. For larger peptides: blue for 20 ns – 80 ns, red for 80 ns – 140 ns, and green for 140ns – 200ns. The following plots of this type use the same drawings.

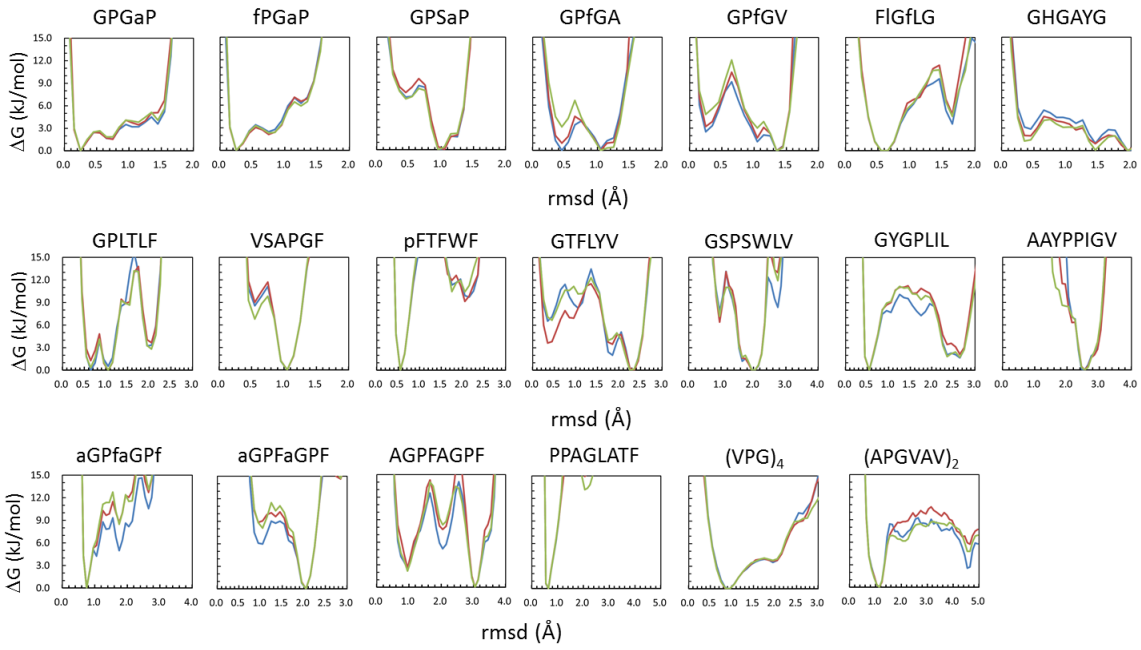


Figure S5. The convergence of the Amber99sb-ildn simulations of the 20 CPs, by comparing free energy curves (blue, red, green) from three equal-length non-overlapping time windows.

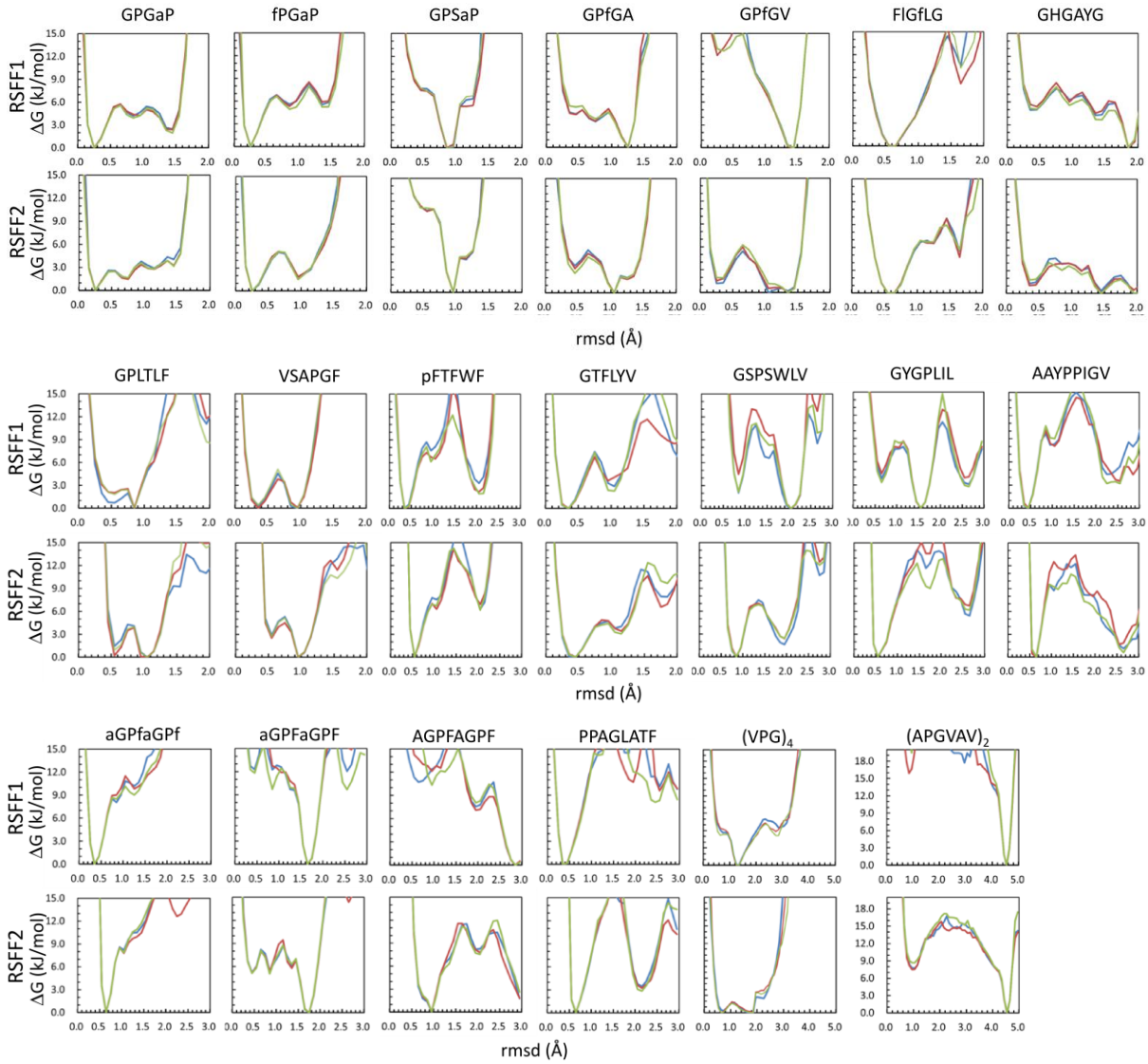


Figure S6. The convergence of the RSFF1 simulations and the RSFF2 simulations of the 20 CPs, by comparing free energy curves (blue, red, green) from three equal-length non-overlapping time windows.

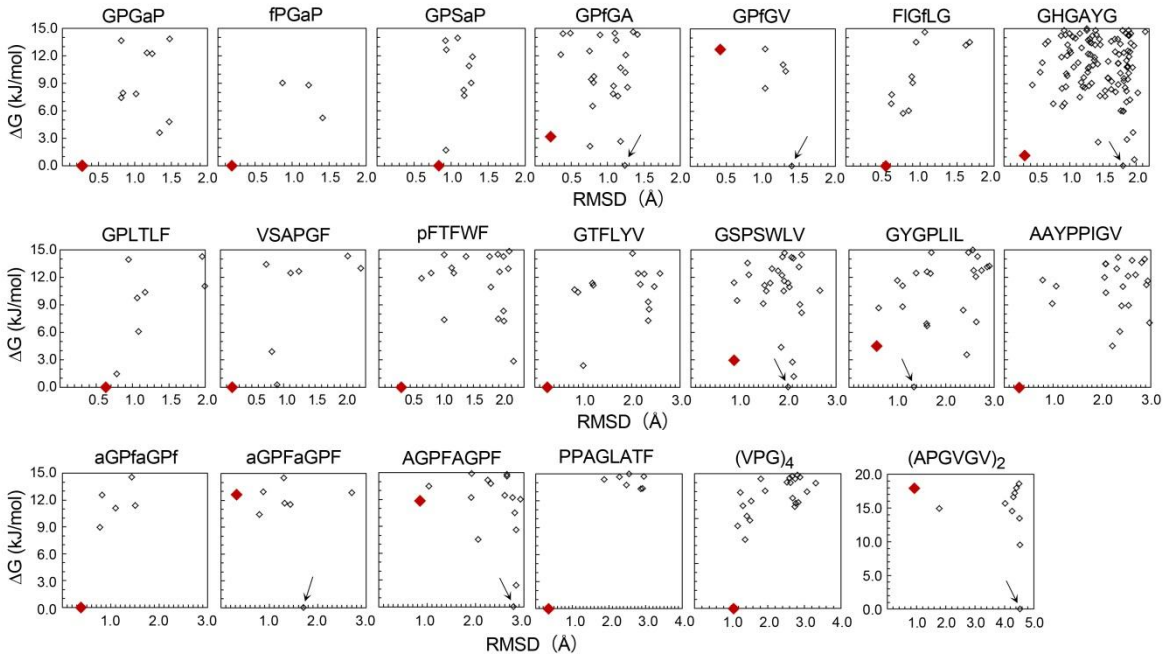


Figure S7. The relative free energies (ΔG) of conformations within 15kJ/mol (20kJ/mol for $(APGVGV)_2$) plotted against their RMSD (mainchain and Cβ atoms) to the crystal structure, for each cyclic peptide simulated using RSFF1 force field. In each plot, the red diamond indicates the crystal-like conformation, and the arrow marks the conformation with lowest free energy.

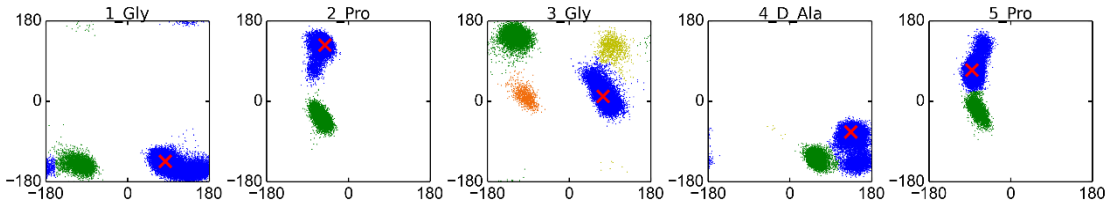
Table S3. The simulation results of 20 cyclic peptides using RSFF2 force field with TIP3P and TIP4P-Ew water models.

	sequence[a]	size	RSFF2 TIP3P			RSFF2 TIP4P-Ew		
			ΔG[b]	rmsd[c]	rmsd[d]	ΔG[b]	rmsd[c]	rmsd[d]
1	GPGaP	5	0.0	0.2	0.2	0.0	0.2	0.2
2	fPGaP	5	0.0	0.2	0.2	0.0	0.2	0.2
3	GPSaP	5	0.0	0.9	0.9	2.1	0.8	1.1
4	GPfGA	5	1.2	0.4	1.0	0.0	0.3	0.3
5	GPfGV	5	1.0	0.3	1.4	5.0	0.2	1.4
6	FlGfLG	6	0.0	0.5	0.5	0.0	0.5	0.5
7	GHGAYG	6	0.0	0.3	0.3	0.0	0.4	0.4
8	GPLTLF	6	1.2	0.5	1.1	0.0	0.5	0.5
9	VSAPGF	6	1.0	0.6	1.0	3.2	0.6	0.9
10	pFTFWF	6	0.0	0.5	0.5	0.0	0.5	0.5
11	GTFLYV	6	0.0	0.3	0.3	0.0	0.4	0.4
12	GSPSWLV	7	0.0	0.8	0.8	0.0	0.8	0.8
13	GYGPLIL	7	0.0	0.5	0.5	0.0	0.4	0.4
14	AAYPPIGV	8	0.0	0.6	0.6	0.0	0.7	0.7
15	aGPfaGPf	8	0.0	0.6	0.6	0.0	0.6	0.6
16	aGPFaGPF	8	5.3	0.3	1.8	10.1	0.5	1.7
17	AGPFAGPF	8	1.8	0.8	0.9	7.8	0.8	2.9
18	PPAGLATF	8	0.0	0.7	0.7	0.0	0.6	0.6
19	$(VPG)_4$	12	0.0	0.5	0.5	2.4	0.7	1.8
20	$(APGVGV)_2$	12	6.6	1.0	4.6	15.3	1.0	4.6

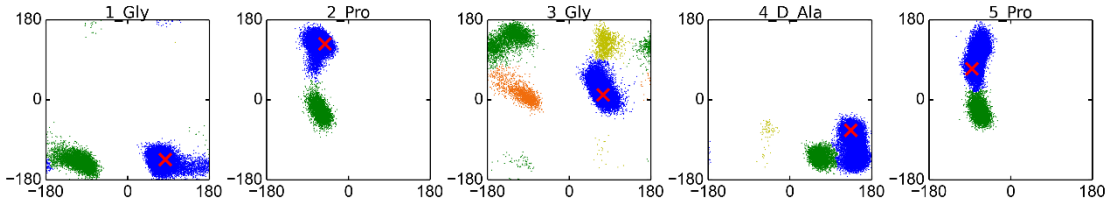
[a,b,c,d]The meanings of these columns are the same as Table 1.

Cyclic peptide 1, sequnce: GPGaP

RSFF1



RSFF2



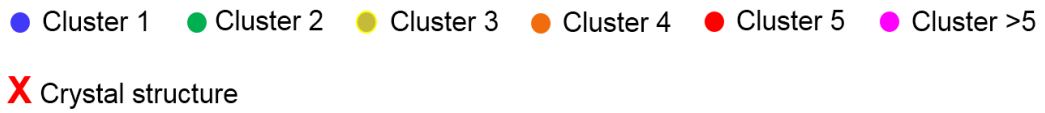
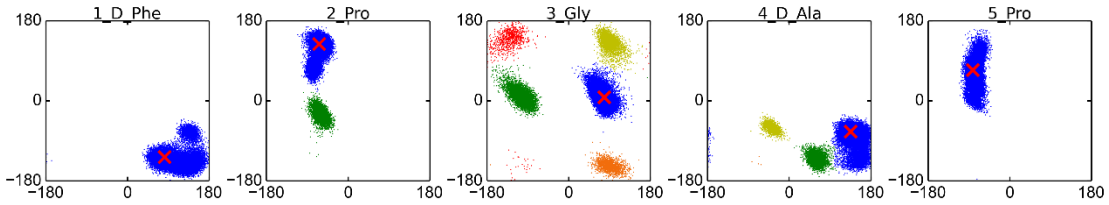


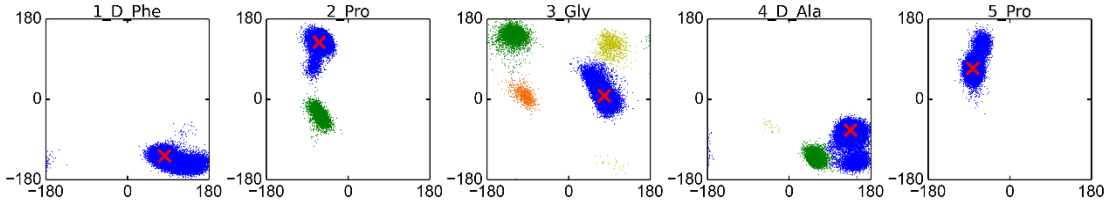
Figure S8. Clustering analysis of the ϕ (x axis)/ψ (y axis) plot of each residue based on the Rodriguez-Laio algorithm. Points are colored according to the cluster to which they are assigned. For each residue, the each cluster is defined as one conformation and labeled by numbers in the order of decreasing population. Red cross show the position of crystal structure. The lowercase in sequences show D amino acids.

Cyclic peptide 2, sequnce: fPGaP

RSFF1

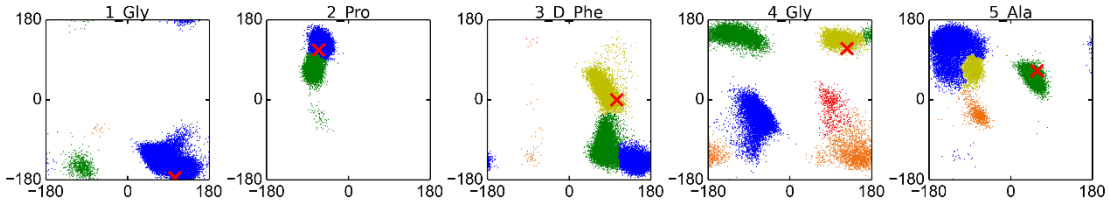


RSFF2



peptide 3, sequnce: GPfGA

RSFF1



RSFF2

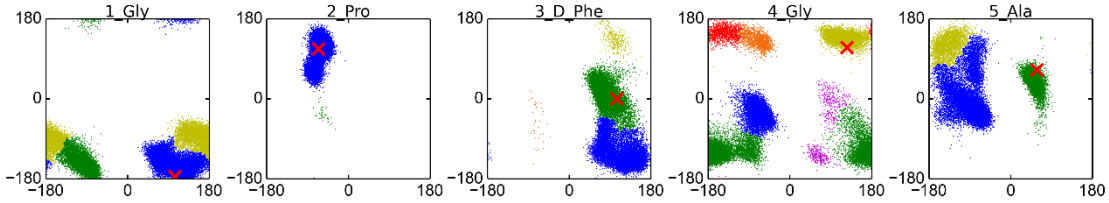
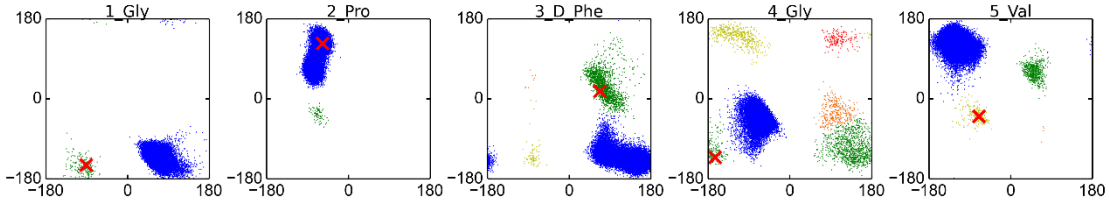


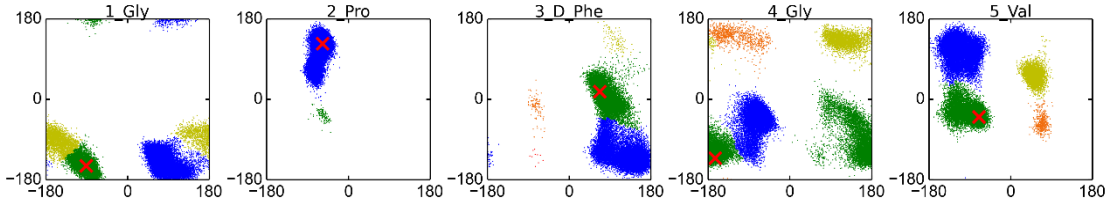
Figure S8. (*Continued*).

peptide 4, sequnce: GPfGV

RSFF1

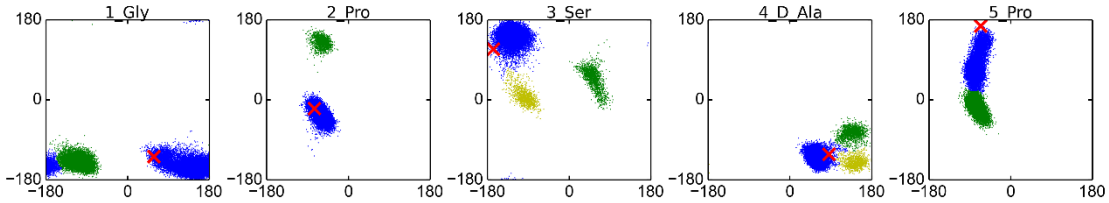


RSFF2



peptide 5, sequnce: GPSaP

RSFF1



RSFF2

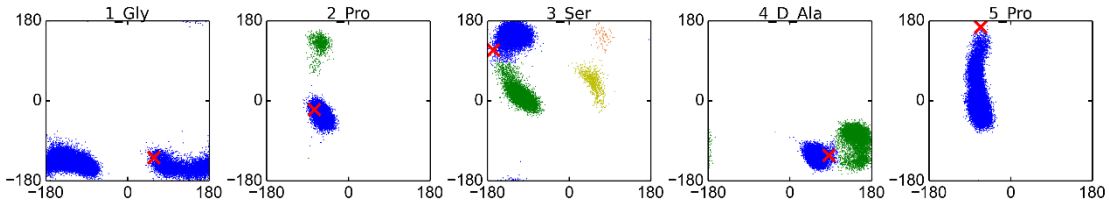
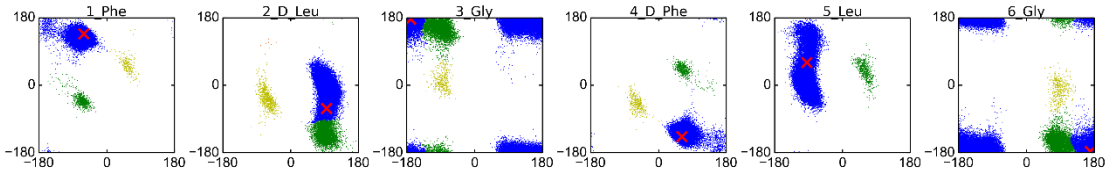


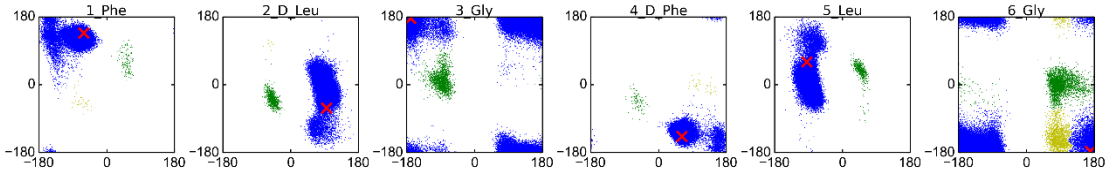
Figure S8. (*Continued*).

peptide 6, sequnce: FlGfLG

RSFF1

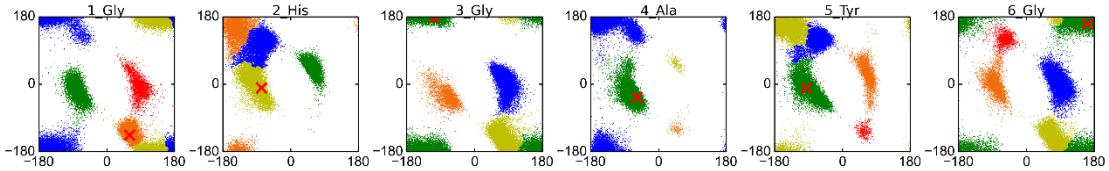


RSFF2



peptide 7, sequnce: GHGAYG

RSFF1



RSFF2

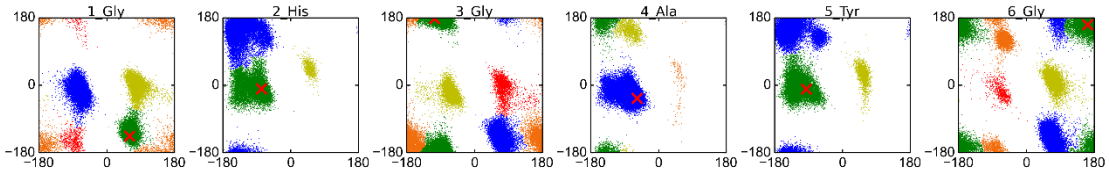
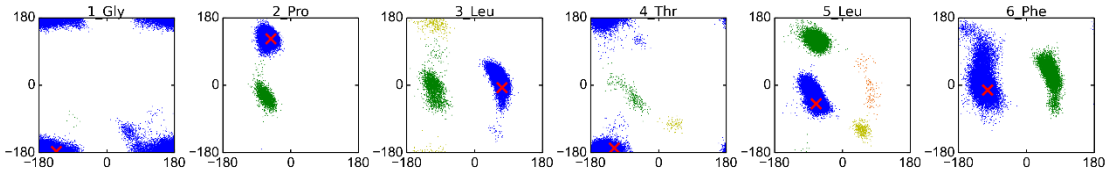


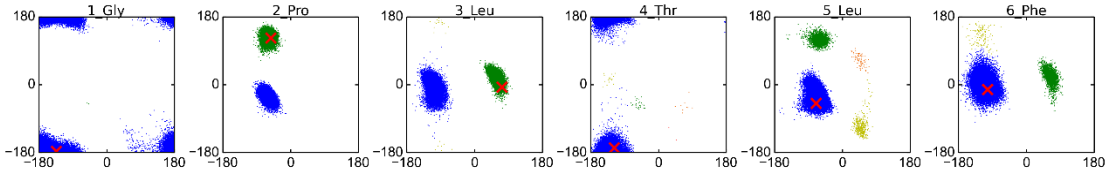
Figure S8. (*Continued*).

peptide 8, sequnce: GPLTLF

RSFF1

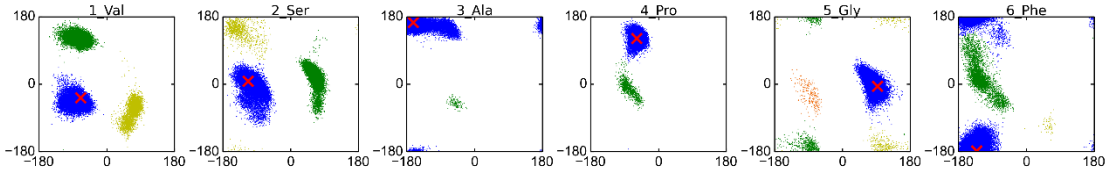


RSFF2



peptide 9, sequnce: VSAPGF

RSFF1



RSFF2

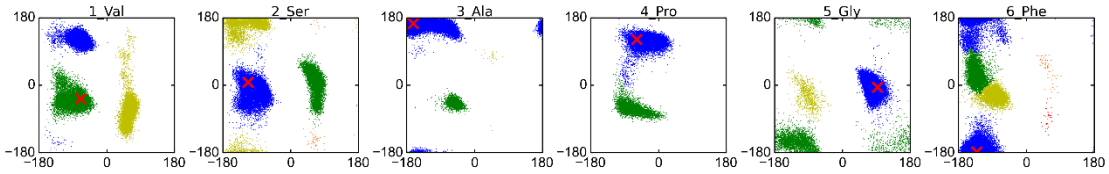
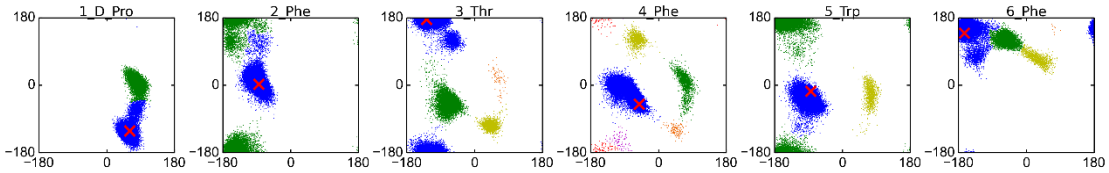


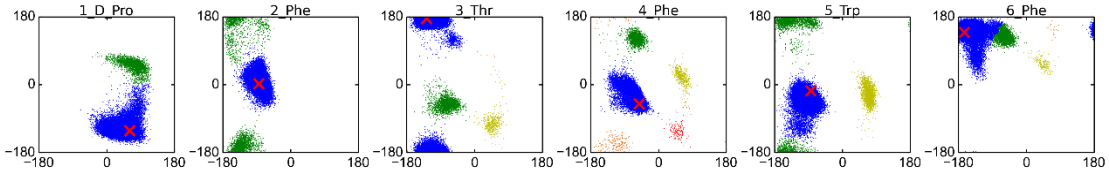
Figure S8. (*Continued*).

peptide 10, sequnce: pFTFWF

RSFF1

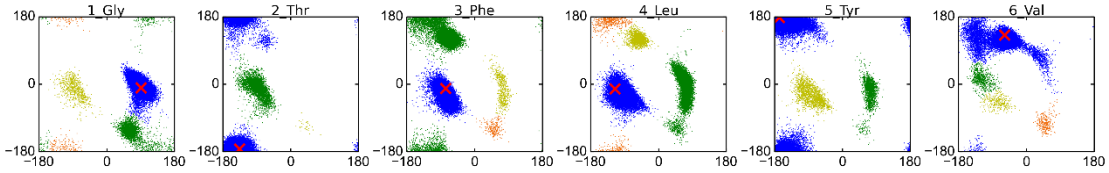


RSFF2



peptide 11, sequnce: GTFLYV

RSFF1



RSFF2

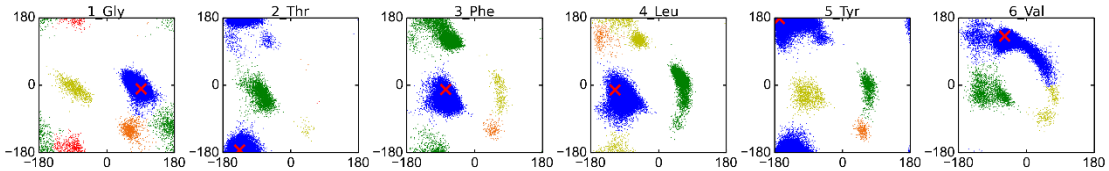
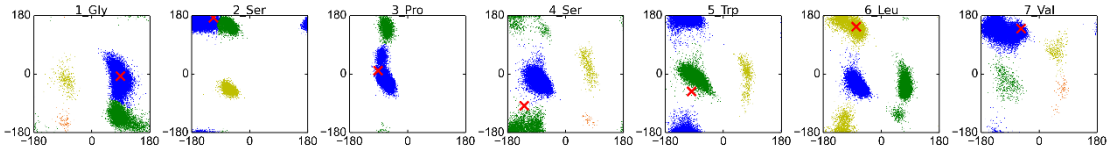


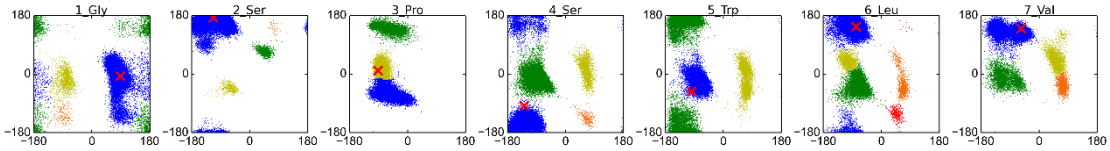
Figure S8. (*Continued*).

peptide 12, sequnce: GSPSWLV

RSFF1

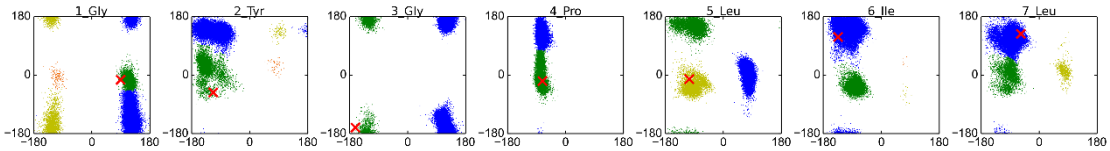


RSFF2



peptide 13, sequnce: GYGPLIL

RSFF1



RSFF2

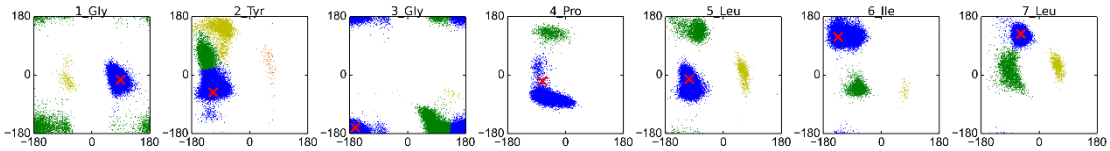
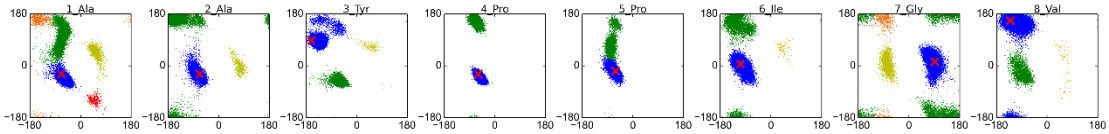


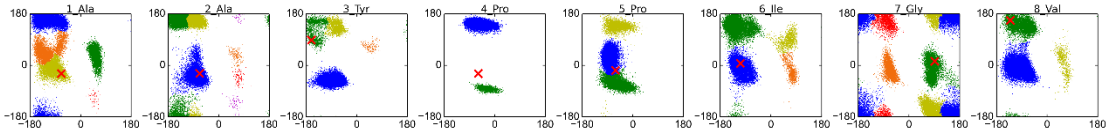
Figure S8. (*Continued*).

peptide 14, sequnce: AAYPPIGV

RSFF1

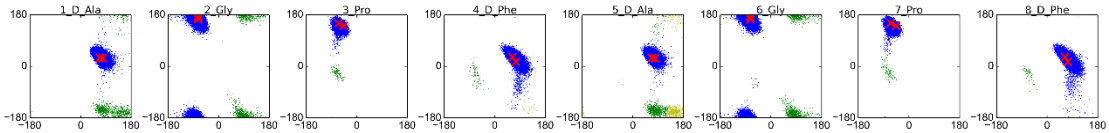


RSFF2



peptide 15, sequnce: aGPfaGPf

RSFF1



RSFF2

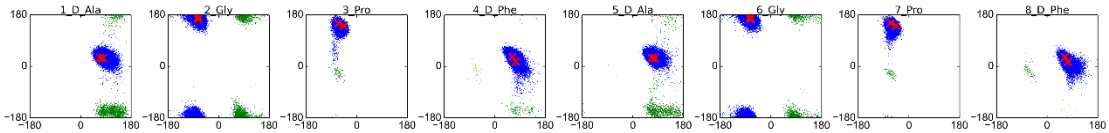
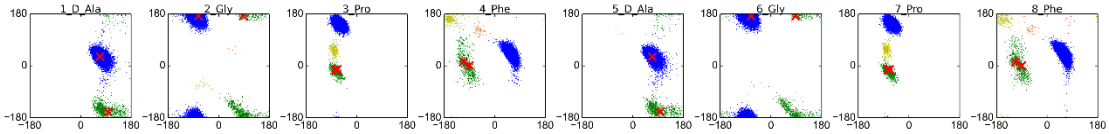


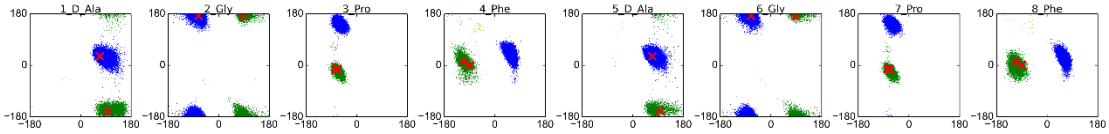
Figure S8. (*Continued*).

peptide 16, sequnce: aGPFaGPF

RSFFF1

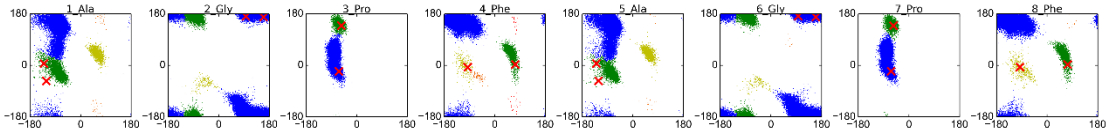


RSFF2



peptide 17, sequnce: AGPFAGPF

RSFF1



RSFF2

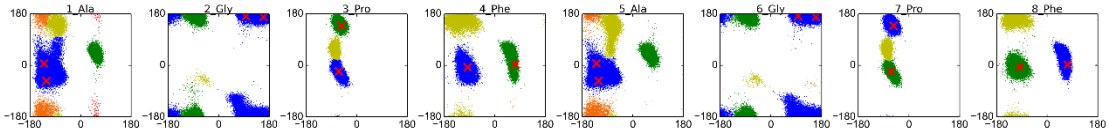
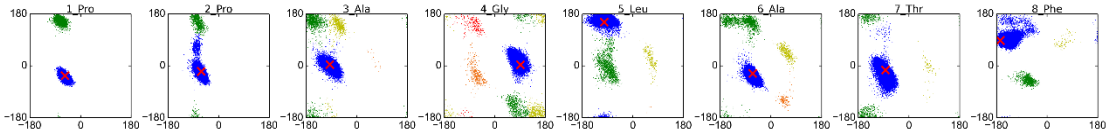


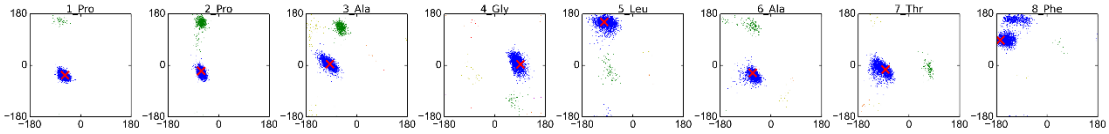
Figure S8. (*Continued*).

peptide 18, sequnce: PPAGLATF

RSFF1

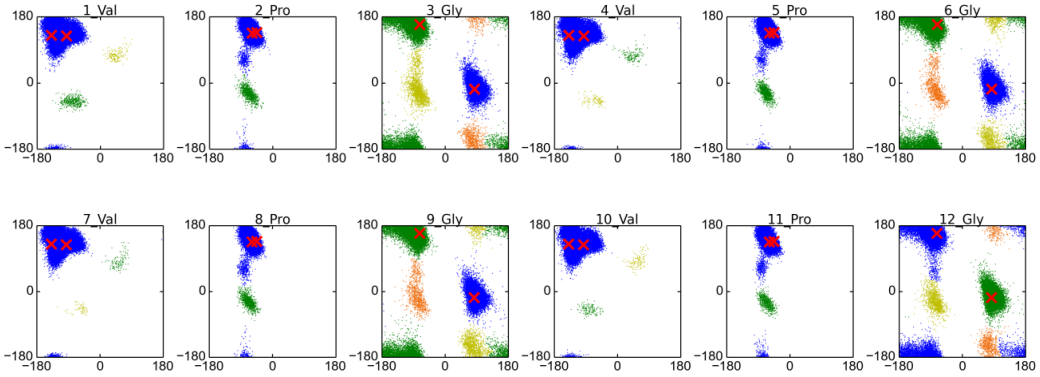


RSFF2



peptide 19, sequnce: VPGVPGVPGVPG

RSFF1



RSFF2

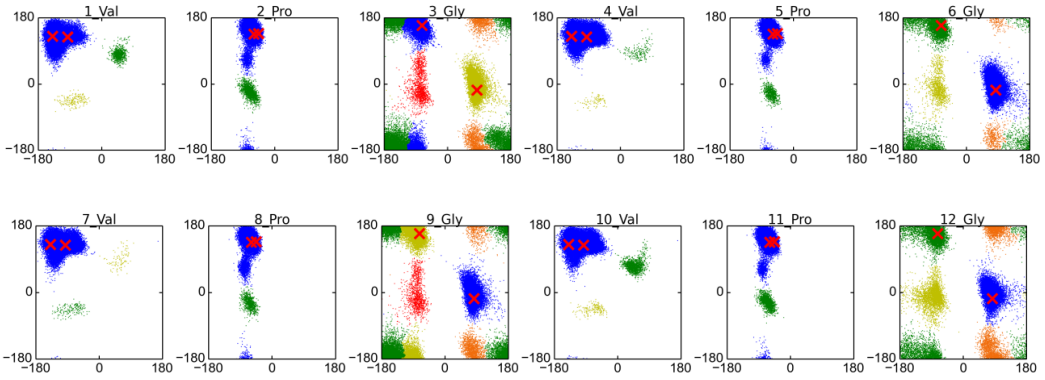
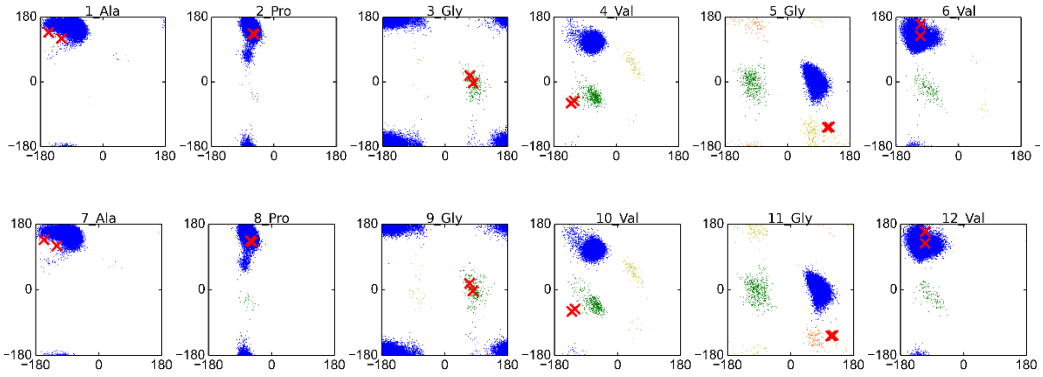


Figure S8. (*Continued*).

peptide 20, sequnce: APGVGVAPGVGV

RSFF1



RSFF2

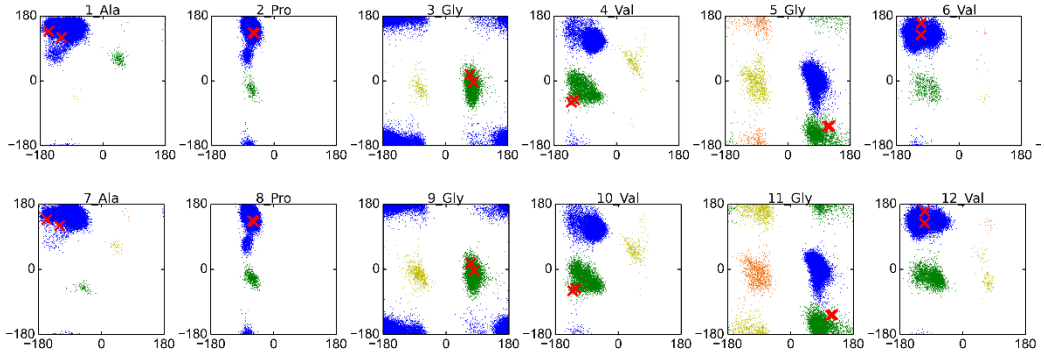


Figure S8. (*Continued*).