Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Similarity scores of vibrational spectra reveal the atomistic structure of pentapeptides in multiple basins

Supporting Information

Hiroki Otaki,[1] Shun-ichi Ishiuchi,[2] Masaaki Fujii,[3] Yuji Sugita[4,5,6] and Kiyoshi Yagi[,4]*

[1]Center for Bioinformatics and Molecular Medicine, Graduate School of Biomedical Sciences, Nagasaki University, 1-14 Bunkyo, Nagasaki, Nagasaki 852-8521, Japan.

[2]Department of Chemistry, School of Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan.

[3]School of Life Science and Technology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8503, Japan.

[4]Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

[5]Computational Biophysics Research Team, RIKEN Center for Computational Science, 7-1-26 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

[6]Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, 1-6-5 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

Table of Contents

S1. Supporting Method

Figure S1 summarizes the computational schemes of the present study. Our method was based on an enhanced conformational sampling with REMD and an anharmonic vibrational analysis with VQDPT2. After the REMD simulation, the conformers for VQDPT2 calculations were selected in three different ways. The first scheme (Scheme 1) was the same as the one we used in the previous study to determine the structure of SIVSF-NH_2.[1] The second scheme (Scheme 2) is an extension of Scheme 1, where the clustering process is improved. Finally, Scheme 3 is newly developed in this work, which exploits the structure-spectrum correlation.

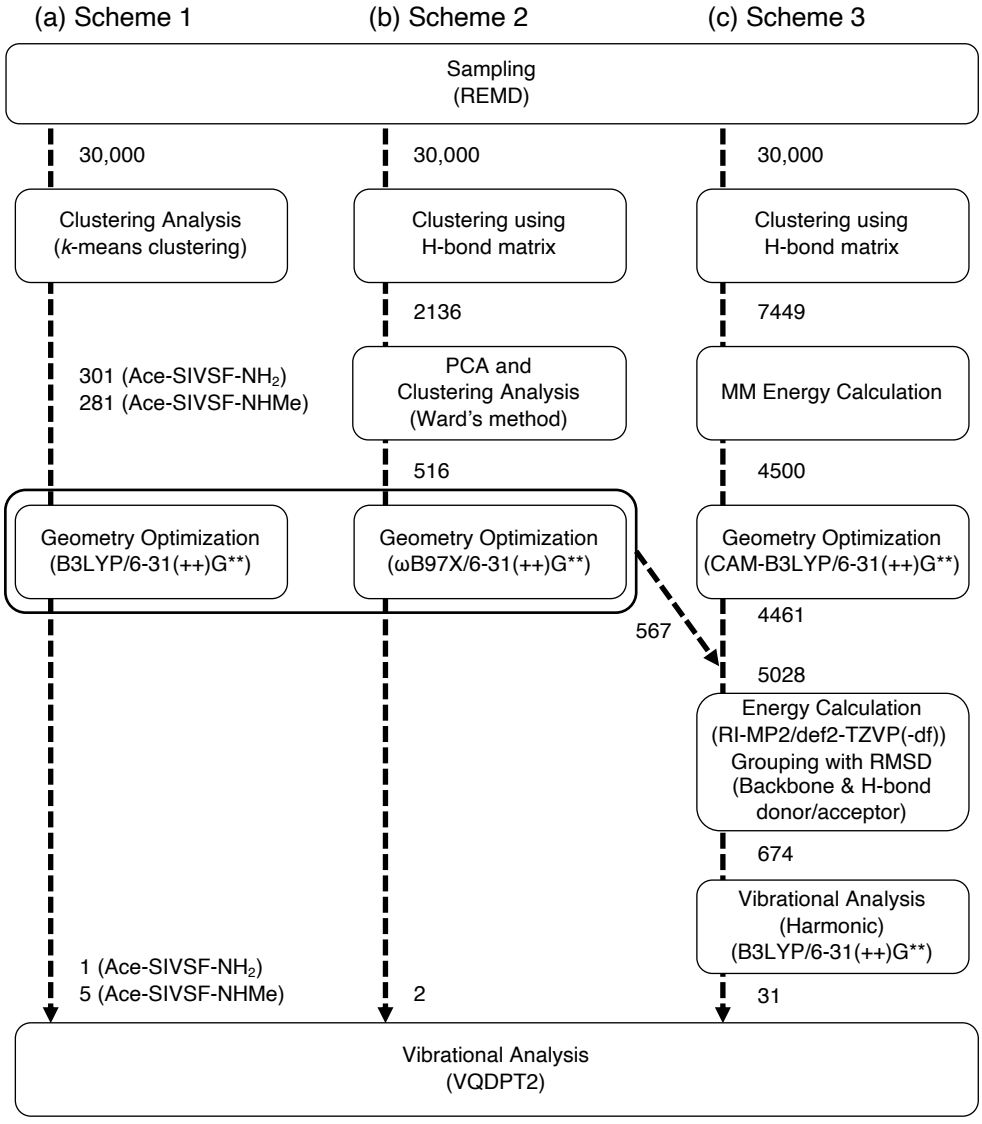


Figure S1. Schemes the conformational search. The numbers in the figure indicate the number of the candidate structures in each process. Scheme 1 was used for Ace-SIVSF-NH_2 and Ace-SIVSF-NHMe. Schemes 2 and 3 were used for Ace-SIVSF-NHMe.

S1.1. Replica exchange molecular dynamics (REMD) simulations

The target pentapeptides, Ace-SIVSF-NH_2 and Ace-SIVSF-NHMe, were set up in vacuum to mimic the gas-phase experiment. CHARMM36 force field with CMAP correction[2-3] was employed for the peptides. After the energy minimization (20 steps), the system was equilibrated for 1 ns (300 - 1300 K, NVT). Then, REMD simulations were performed for 60 ns with the replica-exchange trial of every 2 ps. Twelve replicas were distributed between 300 and 1300 K.

S1.2. Conformational search

S1.2.1. Scheme 1 (Ace-SIVSF-NH_2, Ace-SIVSF-NHMe)

After the REMD simulation, the *k*-means clustering analysis was performed based on the root-mean-square deviation (RMSD) of all the atoms. For each cluster, the conformer closest to the centroid was selected as a representative structure. Then, the representative structures were geometry optimized by the DFT method. The DFT calculations were performed using the B3LYP hybrid functional[4-5] and mixed basis sets of 6-31G** and 6-31++G** [6-8] [denoted as 6-31(++)G**]. The diffuse functions were applied only to the nitrogen and oxygen atoms and the hydrogen atoms bound to them. The lowest-energy structures were used for the VQDPT2 calculation.

H-bond acceptor

H-bond donor	CO_{Ace}	OH_{Ser1}	CO_{Ser1}	CO_{Ile}	CO_{Val}	OH_{Ser4}	CO_{Ser4}	CO_{Phe}	Ph_{Phe}
NH_{Ser1}	0	0	0	0	0	0	0	0	0
OH_{Ser1}	0	0	0	0	1	0	0	0	0
NH_{Ile}	0	0	0	0	0	0	0	1	0
NH_{Val}	0	0	1	0	0	0	0	0	0
NH_{Ser4}	0	0	0	0	0	0	0	0	0
OH_{Ser4}	0	0	0	0	0	0	0	0	1
NH_{Phe}	0	0	0	1	0	0	0	0	0
$NH_{Cter/NHMe}$	0	0	0	0	0	0	1	0	0

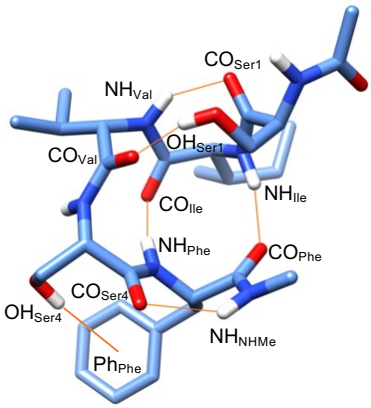


Figure S2. Schematic illustration of the H-bond matrix. The OH, NH, and CO groups of residue R are labeled OH_R, NH_R, and CO_R, respectively. The subscripts Cter and NHMe in the last row are used for the Ace-SIVSF-NH_2 and Ace-SIVSF-NHMe, respectively. Each element is set to 1 or 0, which depends on the existence of the HB. The interaction with the phenyl ring of Phe (Ph_{Phe}, the last column) was considered in Scheme 3. Matrix elements of the conformer in the right figure are represented as an example. In this figure, hydrogen bonds are shown in orange lines, and hydrogen atoms except for OH and NH are omitted for clarity.

S1.2.2. Scheme 2 (Ace-SIVSF-NHMe)

Let us define a H-bond matrix, which represents the HB network. The H-bond matrix lists HB donors and acceptors in row and column, respectively, and its element is set to 1 when the donor and acceptor meet criteria of HB, but 0 otherwise (Figure S2). The criteria of HB are defined by[9]

$$r_{HA} < 3.2\ \text{Å},\ r_{DA} < 3.2\ \text{Å},\ \text{and}\ \theta_{DHA} > 130^{\circ}, \quad \text{(S1)}$$

where D and A are the donor (either N or O) and acceptor (O) atoms, and r_{HA}, r_{DA} and θ_{DHA} denote the distance between H···A, distance between D···A and the angle of D–H···A, respectively.

The 30,000 snapshot structures obtained from the REMD simulation were grouped based on the H-bond matrix yielding 2136 groups. For each group of the H-bond matrix, the structure closest to the centroid was selected as a representative structure. Then, the principal component analysis (PCA) was carried out using heavy atoms of the peptide. The resulting PCs were used for the hierarchical clustering by the Ward's method,[10] yielding 516 clusters. For each cluster thus obtained, the centroid was calculated. Finally, the geometry optimization and the harmonic vibrational analysis were carried out starting from a structure closest to the centroid by the DFT method at the level of ωB97X functional[11] and 6-31(++)G** basis sets.

S1.2.3. Scheme 3 (Ace-SIVSF-NHMe)

D–H···π interaction is added to the H-bond matrix to account for the interaction of Phe with D-H bonds (the last column in Figure S2). Malone's criteria[12] are used to judge the formation of D–H···π interaction. The criteria of HB are loosened compared to Scheme 2,

$$r_{HA} < 3.2 \text{ Å}, r_{DA} < 4.0 \text{ Å and } \theta_{DHA} > 120^{o}, \quad \text{(S2)}$$

to account for weak HBs, in which D-A distance may become close after the geometry optimization.

As in Scheme 2, the 30,000 snapshot structures were grouped based on the modified H-bond matrix, yielding 7449 groups. The energies of all conformers were calculated using the CHARMM36/CMAP force field, and the lowest-energy structure was selected as a representative structure in each group. The geometry of the 4500 lowest-energy representative structure was optimized at the level of CAM-B3LYP functional[13] and 6-31(++)G** basis sets, and 4461 of them were successfully optimized. In addition, the 567 structures obtained from Schemes 1 and 2 excluding the duplication were optimized at the same level and used in the subsequent process.

The energy of 5028 candidate conformers was calculated at the RI-MP2/def2-TZVP(-df) level of theory.[14-15] Then, the conformers were classified in the following way:

(1) Set the lowest energy conformer among the non-classified conformers to a reference structure.
(2) Align the non-classified conformers to the reference structure with backbone atoms.
(3) Calculate the RMSD of backbone atoms and sidechain atoms of HB-donor/acceptor. Sidechain atoms of residues that do not form HB (i.e., Ile, Val) were excluded from the RMSD calculation, because the conformation of these residues has little effect on the IR spectrum in a range of interest (i.e., 3100–3700 cm^{-1}).
(4) Create a group that consists of conformers with RMSD $\leq$ 0.4 Å.

The alignment and RMSD were calculated using AmberTools17.[16] These steps [(1) - (4)] were iterated until all the conformers were classified into a group. The lowest-energy structure at the RI-MP2/def2-TZVP(-df) level of theory was selected as a representative structure of each group. This process reduced the number of candidate conformers to 674.

Then, the geometry optimization and the harmonic vibrational analysis were carried out for the 674 representative structures at the level of B3LYP/6-31(++)G**. Consequently, 666 structures were successfully optimized and found without imaginary frequencies. These candidate conformers were enumerated in increasing order of the RI-MP2 energy. Then, the structure was rejected when the harmonic spectrum was obviously different from the experimental one. The experimental spectrum showed a peak at

ca. 3600 cm^{-1}, indicating one weakly hydrogen-bonded (or free) OH stretching mode of Ser. Also, several peaks were observed between 3200 cm^{-1} and 3400 cm^{-1}. From this observation, the conformer was rejected when one of the following criteria was met:

- More than one peak appears in the region above 3700 cm^{-1}.
- The highest frequency peak appears lower than 3700 cm^{-1}.
- The frequency difference between the first and the second highest frequency peak is less than 50 cm^{-1}.
- No peak is observed in a region of 3350 - 3430 cm^{-1}.

Note that the criteria were applied to unscaled harmonic frequencies. Among the 56 low-energy structures, 25 were rejected, and the remaining 31 were selected for the anharmonic vibrational analysis.

S1.3. VQDPT2 calculation

In the previous work, we suggested the anharmonic vibrational analysis with selected target modes and applied to SIVSF-NH_2.[1] We showed that this scheme effectively reduce the computational cost and calculate a vibrational spectrum accurately. In this work, we also applied this scheme to Ace-SIVSF-NH_2 (86 atoms and 252 vibrational modes) and Ace-SIVSF-NHMe (89 atoms and 261 vibrational modes). The target modes of the peptides are the nine (Ace-SIVSF-NH_2) or eight (Ace-SIVSF-NHMe) NH/OH stretching modes. We considered only active modes that were strongly coupled with the target modes, and kept other modes frozen.

Here, we briefly summarize the procedure for anharmonic vibrational calculation. For details, see ref [1].

(1) Calculate the third- and fourth-order derivatives of the potential in normal coordinates and construct a cubic force field (CFF). The derivatives were calculated with numerical differentiations of Hessian matrices.

(2) Select 80 active modes for vibrational calculations using mode-coupling strength (MCS) proposed by Seidler et al.,[17] which contains nine (Ace-SIVSF-NH_2) or eight (Ace-SIVSF-NHMe) target modes and 71 or 72 active modes, respectively.

(3) Optimize the selected vibrational coordinates with the oc-VSCF method[18-19] using the CFF. We used a pair selection scheme with a threshold value of 200 cm^{-1}.[19] In order to avoid divergence of the VSCF calculation, the normal modes with frequency lower than 500 cm^{-1} were kept frozen.

(4) Construct a hybrid PES in terms of the optimized coordinates.[20] We calculated the derivatives of the potential up to the fourth order in optimized coordinates and generated quartic force field (QFF). Using the QFF, MCS was evaluated for two- and three-mode coupling terms. Subsequently, a grid potential was generated for all the one-mode terms using 11 grid points, and for strongly coupled (MCS larger than 10 cm^{-1}) two- and three-mode terms using seven grid points. We also constructed a dipole moment surface by using the dipole moment obtained at the grid points.

(5) Execute VQDPT2 calculation with the hybrid PES. We set the parameters to construct the degenerate space, $k = 4$ and $N_{gen} = 1$ (see refs [19, 21] for more detail).

(6) Using the frequency and intensity obtained from the VQDPT2 calculation, spectrum was constructed using a Lorentzian line shape function with a full width at half-maximum of 5 cm^{-1}.

All the quantum chemistry calculations for the PES generation were carried out at the B3LYP/6-31(++)G** level of theory. Note that we have changed the MCS threshold from 1.0 cm^{-1} to 10.0 cm^{-1} from

the previous study[1] in order to reduce the grid points and calculate anharmonic vibrational spectra of about 40 conformers. Figure S3 compares the computed anharmonic spectra of one conformer using the VQDPT2 method with the MCS cutoff changed. The shape of the spectra is almost converged when the MCS threshold is lower than 20 cm^{-1}. Thus, we chose 10.0 cm^{-1} as a threshold in this work, which could reduce the number of calculations for grid points by ~80% compared with the threshold of 1.0 cm^{-1}.

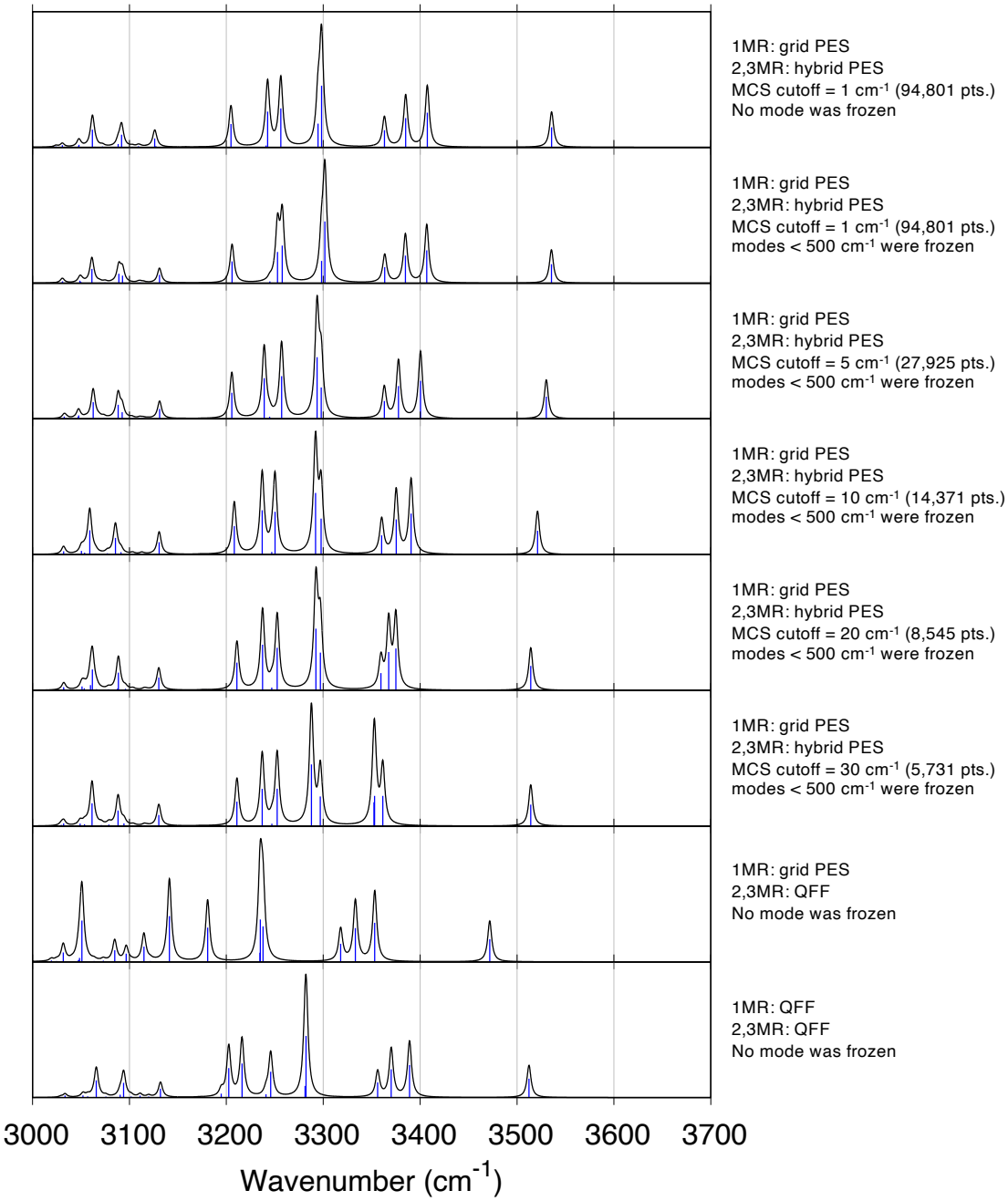


Figure S3. Convergence of spectrum with respect to resolution of potential energy surface (PES). f19375 of Ace-SIVSF-NHMe was used for the calculations. All the PESs were calculated at B3LYP/6-31(++)G** level of theory. The number in the parentheses means the number of grid points to generate PES. The spectra are sorted in decreasing order of accuracy of the PES.

S1.4. Assessment of similarity scores

Using the spectra for Conformer P, Conformer Q, Conformer A', and f28623, we assessed the similarity scores. Figure S4 represents the change of the similarity score S_1 with respect to α (see Eq. (3)). When α is less than or equal to 50 cm^{-1}, the score S_1 suggests that Conformer A' and f28623 are assigned to Conformer P and Conformer Q, respectively. However, when α is greater than 50 cm^{-1}, the score S_1 of f28623 was lower than that of Conformer A' both in Conformer P and Conformer Q.

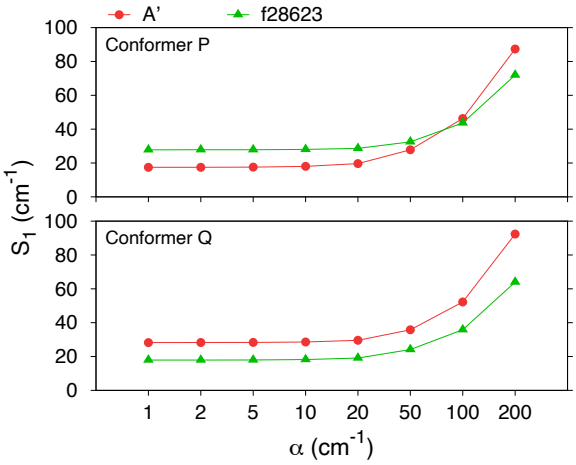


Figure S4. Dependencies of similarity score S_1 (in cm^{-1}) with respect to α.

The results of the peak assignment in our scheme are summarized in Table S1. The value of α affected the frequency region where some peaks were close to each other. In the assignment of Conformer P and Conformer A', the small peak at 3449.0 cm^{-1} in the VQDPT2 calculation was assigned to the peak at 3438.3 cm^{-1} at α = 1 and 20 cm^{-1}. Also, the relative intensity to a neighboring peak shows opposite trend from the experimental spectrum (see the rows of 3457.0 and 3438.3 cm^{-1} in the experimental spectrum in Table S1a). However, for α = 50 cm^{-1}, the peak at 3452.8 cm^{-1} in the VQDPT2 computation was assigned to 3438.3 cm^{-1} in the experimental one, and the relative intensity shows the same trend as the experimental one. Likewise, in the assignment of Conformer Q and f28623, although the small peak at 3339.4 cm^{-1} in the VQDPT2 calculation was assigned to the peak at 3349.1 cm^{-1} at α = 1 and 20 cm^{-1}, the large peak at 3330.1 cm^{-1} was assigned instead at α = 50 cm^{-1} (see Table S1b). Notably, these assignments at α = 50 cm^{-1} are identical to the those obtained with manual assignment (Table S3). From these results, we decided to set α at 50 cm^{-1} in this study.

The similarity scores are also listed in Table S1. The score ΔS_2 is the relative value to the lowest S_2 of all the combinations for peak correspondence (see Methods section in the main text). In the assignment of Conformer P and Conformer A', the score ΔS_2 was close to zero, whereas ΔS_2 was fairly large in that of Conformer Q and f28623 (see Table S1). Figure S5 compares the peak assignment of Conformer Q and f28623 with minimum S_1 and minimum S_2. In the assignment with minimum S_1 (Figure S5a), the smallest band at 3339.4 cm^{-1} was discarded. In the assignment with minimum S_2 (Figure S5b), all the dashed lines showing correspondence of the peaks are almost parallel, which indicates that the peak intervals are well-represented in accordance with the definition of S_2. However, the band at 3447.6 cm^{-1} was not assigned and

the band at 3339.4 cm^{-1} was assigned to the strong peak at 3360.2 cm^{-1}. We thus evaluated peak assignment using S_1 prior to S_2.

Table S1. Peak assignment of experimental and computed spectra of Ace-SIVSF-NH_2 with similarity scores S_1 and S_2 (both in cm^{-1}): ν, frequency (in cm^{-1}); I_{rel}, relative intensity (in arb. unit); α, weight factor for S_1 (in cm^{-1})

(a) Conf.P – Conf.A'

Exp.		VQDPT2					
		assignment (α = 1)		assignment (α = 20)		assignment (α = 50)	
ν	I_{rel}	ν	I_{rel}	ν	I_{rel}	ν	I_{rel}
3591.9	0.71	3563.2	0.22	3563.2	0.22	3563.2	0.22
3504.0	0.49	3471.9	1.00	3471.9	1.00	3471.9	1.00
3457.0	0.40	3454.1	0.11	3452.8	0.26	3454.1	0.11
3438.3	0.79	3449.0	0.07	3449.0	0.07	3452.8	0.26
3353.2	0.20	3370.2	0.10	3370.2	0.10	3370.2	0.10
3309.0	1.00	3324.2	0.27	3324.2	0.27	3324.2	0.27
3300.6	0.92	3307.7	0.56	3307.7	0.56	3307.7	0.56
3254.7	0.80	3241.4	0.64	3241.4	0.64	3241.4	0.64
3187.8	0.29	3184.4	0.02	3184.4	0.02	3184.4	0.02
	S_1	17.4		19.6		27.8	
	S_2	17.1		17.1		17.5	
	ΔS_2	0.1		0.1		0.5	

(b) Conf.Q – f28623

Exp.		VQDPT2					
		assignment (α = 1)		assignment (α = 20)		assignment (α = 50)	
ν	I_{rel}	ν	I_{rel}	ν	I_{rel}	ν	I_{rel}
3596.8	0.72	3561.3	0.24	3561.3	0.24	3561.3	0.24
3446.1	0.52	3447.6	0.11	3447.6	0.11	3447.6	0.11
3423.6	0.66	3427.4	0.26	3427.4	0.26	3427.4	0.26
3413.8	0.47	3420.8	0.21	3420.8	0.21	3420.8	0.21
3360.2	0.54	3384.5	0.24	3384.5	0.24	3384.5	0.24
3349.1	0.48	3339.4	0.03	3339.4	0.03	3330.1	0.57
3317.8	1.00	3318.1	1.00	3318.1	1.00	3318.1	1.00
3250.3	0.32	3245.3	0.64	3245.3	0.64	3245.3	0.64
3219.1	0.36	3214.4	0.67	3214.4	0.67	3214.4	0.67
3196.8	0.39	3162.9	0.52	3162.9	0.52	3162.9	0.52
	S_1	17.9		19.1		24.1	
	S_2	17.1		17.1		17.6	
	ΔS_2	4.5		4.5		5.0	

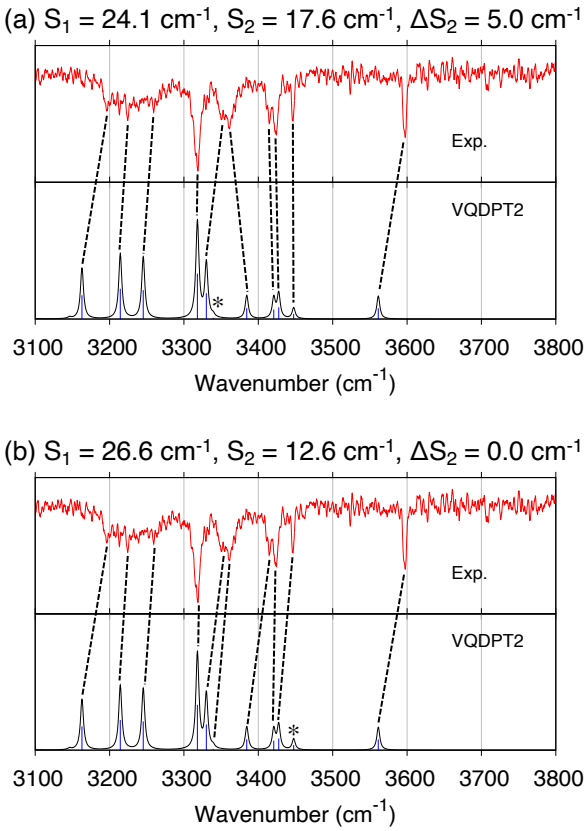


Figure S5. Peak assignment of theoretical (f28623) and experimental (Conformer Q) IR spectra of Ace-SIVSF-NH_2 using the similarity scores with (a) minimum S_1 and (b) minimum S_2. The score S_1 was evaluated with $\alpha = 50$ cm^{-1}. The bands with asterisk were not used for the assignment.

S1.5. Computational programs

Ace-SIVSF-NH_2 and Ace-SIVSF-NHMe in vacuum were prepared using the CHARMM program (version c36b2).[22] The N-terminal of SIVSF-NH_2 was replaced with an acetyl group using Avogadro software (version 1.2.0). [23] The MD and REMD simulations were carried out using the NAMD software (version 2.9).[24] The *k*-means clustering analysis was performed using the MMTSB Toolset.[25] The PCA and the hierarchical clustering were carried out using Bio3D package (version 2.3).[26-28] The geometry optimization and the harmonic vibrational analysis at the DFT level were carried out using Gaussian09.[29] The single point energy at the RI-MP2/def2-TZVP(-df) level was calculated using ORCA program package (version 3.0.3).[30] The generation of PES, oc-VSCF calculation, and VQDPT2 calculation were carried out with SINDO program.[31]

UCSF Chimera (version 1.12)[32] was used for molecular drawing. HB diagrams were drawn using Cytoscape (version 3.8.0).[33]

S1.6. Computational costs

The software used and the computational costs in each process in the case of Ace-SIVSF-NHMe (89 atoms and 261 vibrational modes) are summarized in Table S2. The time in the "for one conformer" column indicates the typical time of the calculation or processing. The costs for Schemes 1 and 2 in Ace-SIVSF-NHMe and Ace-SIVSF-NH_2 (86 atoms and 252 vibrational modes) are comparable with those for the corresponding procedure in this table.

Table S2. A list of the software used and the computational cost in the computational steps undertaken for Ace-SIVSF-NHMe.

Process	Software	Notes	Number of conformers	Wall-clock time		CPU
				for one conformer	total	
REMD simulation	NAMD	12 replicas 1 fs × 60,000,000 steps		—	4 hrs	a
Conformational Search						
H-bond analysis	in-house script		30,000	< 1 sec	~ 2 hrs	a
MM energy calculation	CHARMM	CHARMM36/CMAP	30,000	< 1 sec	~ 1 hr	a
Grouping with H-bond matrix	in-house script	Reducing candidate conformers from 30,000 to 7449		—	< 1 hr	a
Geometry optimization	Gaussian09	CAM-B3LYP/6-31(++)G**	5067	8 hrs	40,536 hrs	b
Single-point energy calculation	ORCA	RI-MP2/def2-TZVP(-df)	5028	30 min	2514 hrs	c, d, e
RMSD-based grouping	in-house script + AmberTools17	Reducing candidate conformers from 5028 to 674		—	24 hrs	a
Geometry optimization and harmonic vibrational analysis	Gaussian09	B3LYP/6-31(++)G**	674	11 hrs	7414 hrs	c, d, e
Anharmonic Vibrational Analysis						
Construction of CFF in normal coordinates	Gaussian09 (+ SINDO)	B3LYP/6-31(++)G** 135 min/geometry × 523 (2 × 261 + 1) geometries	31	1177 hrs	36,487 hrs	f
oc-VSCF calculation	SINDO		31	1 hr	31 hrs	a
Construction of QFF in optimized coordinates	Gaussian09 (+ SINDO)	B3LYP/6-31(++)G** 135 min/geometry × 161 (2 × 80 + 1) geometries	31	363 hrs	11,253 hrs	f
Single-point energy calculation for construction of grid potential of hybrid PES	Gaussian09 (+ SINDO)	B3LYP/6-31(++)G** MCS cutoff for two- and three-mode coupling terms: 10.0 cm^{-1} 8 min/geometry × 29,000 geometries	31	3867 hrs	119,877 hrs	f
VQDPT2 calculation	SINDO		31	1 min	31 min	a
Band calculation	in-house program		31	< 1 sec	~ 30 sec	a

a. Intel Core i7-3960X EE (3.3 GHz, 15 MB, 6 cores/12 threads), b. Fujitsu SPARC64 XIfx (1.975 GHz, 32 cores), c. Intel Xeon E5-2680v3 (2.5 GHz, 30MB, 12 cores/24 threads), d. Intel Xeon E5-2630v3 (2.4 GHz, 20MB, 8 cores/16 threads), e. Intel Xeon E5-2620v4 (2.1 GHz, 20MB, 8 cores/16 threads), f. Intel Xeon Gold 6148 (2.4 GHz, 27.5 MB, 20 cores/40 threads)

S2. Experiment

IR spectra of SIVSF-NH_2[34] and Ace-SIVSF-NHMe[35] have already been reported in the previous papers. The IR spectrum of Ace-SIVSF-NH_2 was measured as follows. Prior to measuring the conformer-selected IR spectrum, a resonance-enhanced 2-photon ionization (R2PI) spectrum of jet-cooled Ace-SIVSF-NH_2 which was evaporated by a laser desorption technique was measured (Figure S6). The experimental setup was the same as that reported in the previous papers.[34-35] In the spectrum, two intense bands were observed at 37654 and 37660 cm^{-1} which are labeled as P and Q. By probing them, IR-UV dip spectra were measured (Figure S7b,c). We also tried to measure the IR-UV dip spectrum by probing band R (37546 cm^{-1}), however, it was not successful because of its weak intensity. For comparison, the IR spectra of the other capped SIVSFs are also presented in Figure S7.

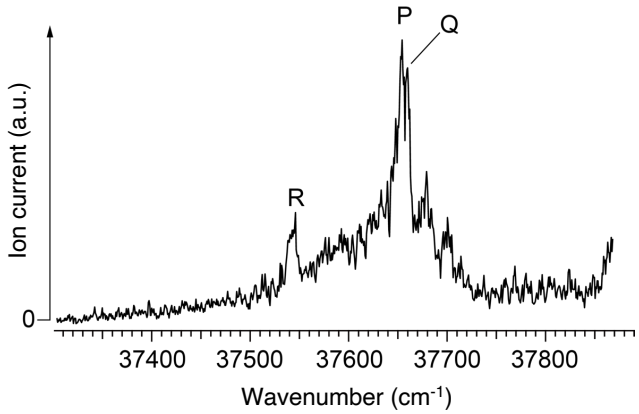


Figure S6. R2PI spectrum of Ace-SIVSF-NH_2.

In Ace-SIVSF-NH_2, Conformer P showed similar spectrum to SIVSF-NH_2 (Figure S7a,b). These spectra were fitted with Lorentz functions (Figure S8), and the peak frequencies and intensities thus obtained were used to calculate the similarity scores.

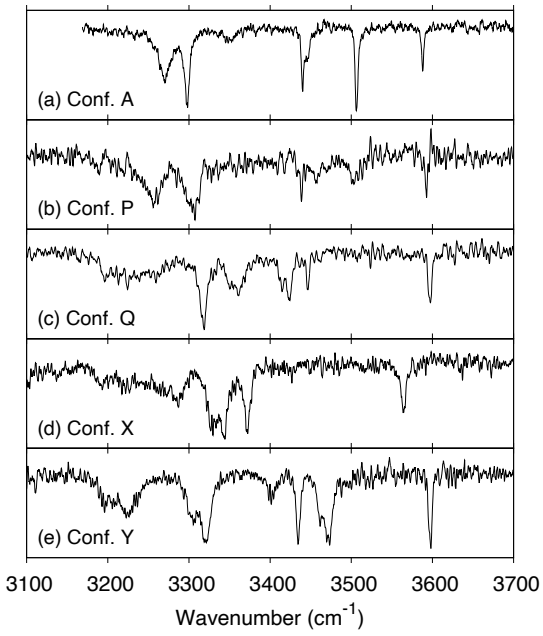


Figure S7. Experimental IR spectra of (a) SIVSF-NH_2, (b-c) Ace-SIVSF-NH_2, and (d-e) Ace-SIVSF-NHMe.

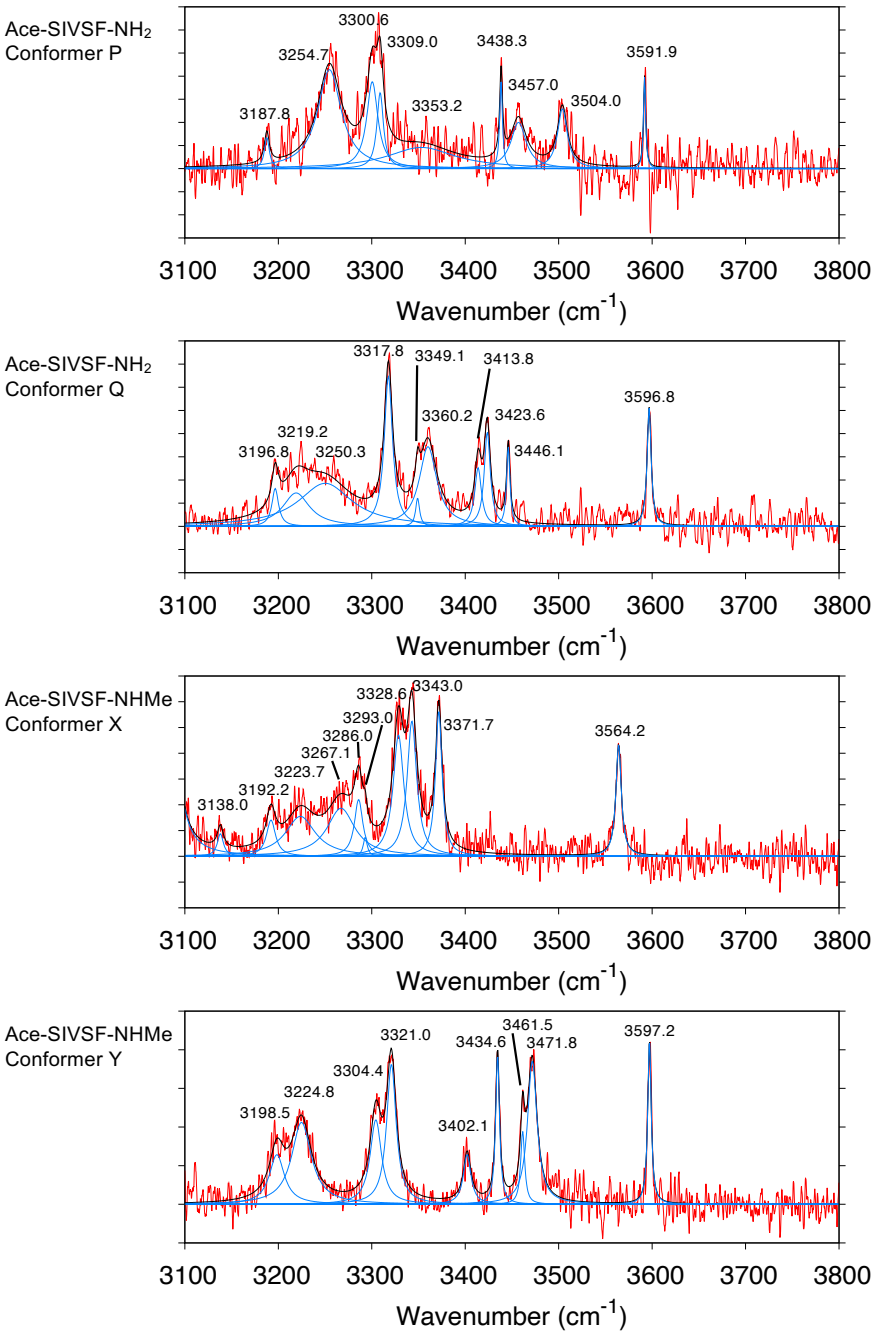


Figure S8. Fitting of the observed spectra (red) to Lorentz functions (black). The components of the Lorentz functions are shown in blue lines. The numbers in the figures indicate peak frequencies (in cm^{-1}).

S3. Supporting results of Ace-SIVSF-NH_2

S3.1. REMD simulations

Results of REMD simulations are shown in Figure S9. We observed a random walk in replica, temperature, and potential energy space, which corroborates that the conformational sampling was performed properly with the REMD simulation. The acceptance ratios were 27% on average. By using the trajectories obtained from the REMD simulations, we calculated potential of mean force (PMF) as a function of backbone dihedral angles φ and ψ of each residue. The PMF was calculated from,

$$E(\phi,\psi) = -k_{\mathrm{B}}T \ln\left(\frac{N_{\phi\psi}}{N}\right) \quad \text{(S3)}$$

where k_{B} is the Boltzmann constant, T is the temperature, N is the total number of snapshots (30,000), and $N_{\varphi\psi}$ is the number of snapshots whose backbone dihedral angles belong to bins of φ and ψ. Figure S9e shows the PMFs of Ace-SIVSF-NH_2 at 300 and 1300 K. The PMFs at 1300 K demonstrates that the REMD simulation samples a sufficiently wide conformational space including the α-helix region (φ ~ -80° and ψ ~ -30°) and the β-sheet region (φ ~ -100° and ψ ~ 120°).

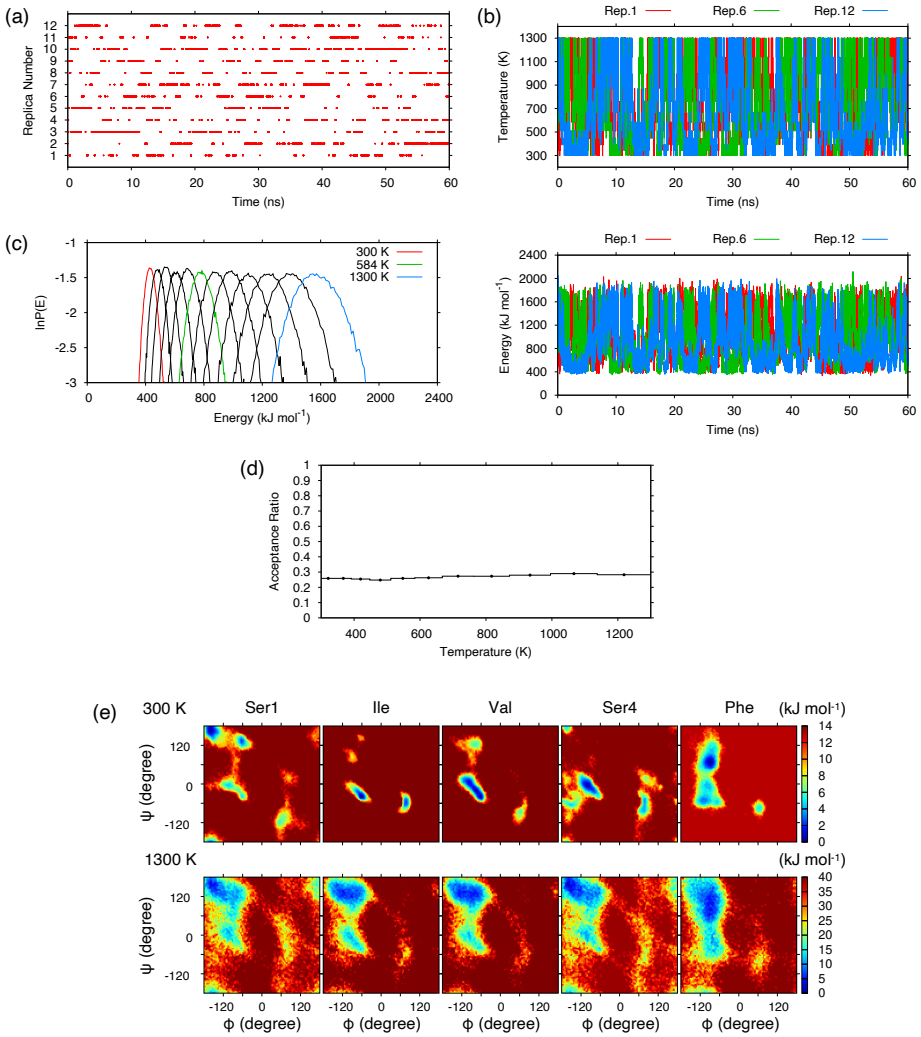


Figure S9. Results of REMD simulations of Ace-SIVSF-NH_2. (a) Replica number of the trajectory at 300 K. (b) Temperature and potential energy of replicas 1, 6, and 12 as a function of time. (c) Canonical probability distribution of the potential energy with different temperature (d) Acceptance ratio of adjacent replicas. (e) Potential of mean force at 300 and 1300 K.

S3.2. Clustering analyses with Scheme 1

In Scheme 1, *k*-means clustering analysis is used to find representative structures from the trajectory. Figure S10a plots the number of representative structures as a function of a clustering radius. 301 and 146 structures were found by setting the clustering radius to 1.6 and 1.9 Å, respectively. Then, these structures were geometry optimized by the DFT method at the level of B3LYP/6-31(++)G**. Figure S10b shows the relative energies of the optimized structures. It is notable that the two sets of structures both yield a conformer, f28623, to be the most stable one. Therefore, f28623 is a strong candidate of the experimentally observed conformer.

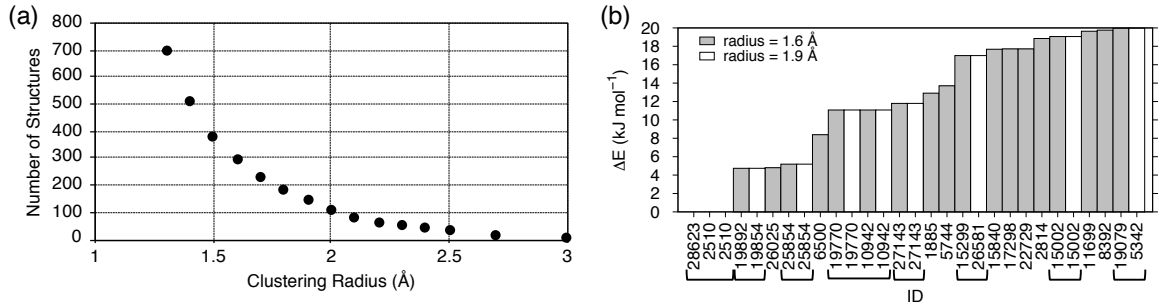


Figure S10. Results of conformational search with Scheme 1 for Ace-SIVSF-NH_2. (a) The number of clusters as function of clustering radius. (b) Relative energy of optimized geometries at the B3LYP/6-31(++)G** level without the zero-point energy correction (up to 20 kJ mol^{-1}). The conformer IDs of the same structure are bracketed.

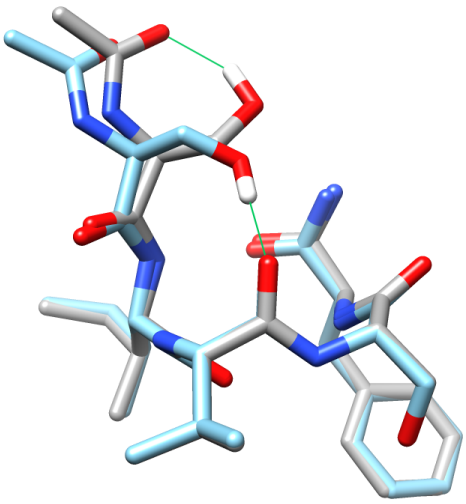


Figure S11. Molecular structures of Conformer A' (light blue) and f28623 (gray). Hydrogen atoms except for OH_{Ser1} are omitted for clarity. Hydrogen bonds of OH_{Ser1} are shown in green lines.

S3.3. Assignment of the peaks

The assignment of peaks is summarized in Table S3. Table S3 shows the ratio between the anharmonic and harmonic frequencies (*f*). In Conformer A', the ratio varies in range of 0.94–0.96 except for the symmetric and asymmetric NH_2 stretching mode of C-terminal ($\nu_s(NH_{2,Cter})$ and $\nu_{as}(NH_{2,Cter})$, respectively). Usually, frequency of asymmetric NH_2 stretching mode is higher than that of symmetric NH_2 stretching mode. However, the assignment of these two modes is interchanged in the VQDPT2 calculation. In the result of the VQDPT2 calculation, the vibrational states of 3370.2 cm^{-1} ($\nu_{as}(NH_{2,Cter})$) and 3449.0 cm^{-1} ($\nu_s(NH_{2,Cter})$) are represented as

$$\Psi(3370.2\ \mathrm{cm^{-1}}) = 0.60|250_1\rangle + 0.55|247_1\rangle - 0.43|202_2\rangle + 0.38|207_1202_1\rangle \quad \text{(S4)}$$

$$\Psi(3449.0\ \mathrm{cm^{-1}}) = 0.65|247_1\rangle - 0.58|250_1\rangle - 0.34|248_1\rangle + 0.33|251_1\rangle \quad \text{(S5)}$$

where $|m_1\rangle$, $|m_2\rangle$, and $|m_1n_1\rangle$ represent the VSCF configuration function of a fundamental tone of mode *m*, overtone of mode *m*, and combination tone of modes *m* and *n*, respectively. The modes 251, 250, 248, 247, 207, and 202 are $\nu(OH_{Ser1})$, $\nu_{as}(NH_{2,Cter})$, $\nu(NH_{Ser1})$, $\nu_s(NH_{2,Cter})$, $\nu(CO_{Phe})$, and $\delta(NH_{2,Cter})$, respectively, where $\nu(XX_R)$ is XX stretching mode of residue R, and $\delta(XX_R)$ is XX bending mode of residue R. Equations S4 and S5 show that the weights of $\nu_{as}(NH_{2,Cter})$ (mode 250) and $\nu_s(NH_{2,Cter})$ (mode 247) for these states are comparable, which implies that this interchange is accidental because of the accuracy of the PES. An assignment of these states without the interchange yields the ratio values of 0.95 (= 3449.0/3530.2) and 0.96 (= 3370.2/3510.4) for $\nu_{as}(NH_{2,Cter})$ and $\nu_s(NH_{2,Cter})$, respectively, which are within the range of the other modes.

In f28623, the band positions of $\nu(OH_{Ser1})$, and those of $\nu(NH_{Val})$ and $\nu(NH_{Phe})$ are interchanged. The ratio for $\nu(OH_{Ser1})$ is 0.93, which is smaller than those of the other modes. This implies that the HB of OH_{Ser1} to CO_{Ace}, which makes a seven-membered ring (Figure S11), is stronger than the others.

In f28623, the computed spectrum shows two Fermi resonances in the region of 3000–3350 cm^{-1}. Table S4 lists the vibrational states and weights of each mode obtained with the VQDPT2 calculation. One Fermi resonance is composed of $\nu_s(NH_{2,Cter})$, a combination tone of $\delta(NH_{2,Cter})$ and $\nu(CO_{Phe})$, and an overtone of $\delta(NH_{2,Cter})$. These three modes are well-mixed, and the energy levels are at 3339.4 cm^{-1}, 3245.3 cm^{-1}, and 3162.9 cm^{-1}. The peak at 3339.4 cm^{-1} is small and observed as a weak shoulder of the band in the region 3320 cm^{-1}; however, the other two peaks are large. This is likely due to the contribution of the fundamental mode, $\nu_s(NH_{2,Cter})$. The smaller value of the ratio f = 0.93 for $\nu_s(NH_{2,Cter})$ at 3162.9 cm^{-1} could be attributable to the energy shift due to the Fermi resonance.

The other resonance consists of $\nu(NH_{Ile})$, a combination tone of the CH_3 symmetric bending mode of N-terminal ($\delta_s(CH_{3,Ace})$) and $\delta(NH_{Ile})$, and an overtone of $\delta(NH_{Ile})$. The energy levels after the Fermi resonance are at 3214.4 cm^{-1}, 3061.8 cm^{-1}, and 3049.0 cm^{-1}. Although the two peaks with lower frequencies could not be compared with the experimental spectra because the region below 3100 cm^{-1} was not observed in the experiment, our calculation suggests that there is relatively strong band in the region of ~3050 cm^{-1}.

Table S3. Peak assignment of experimental and computed spectra of Ace-SIVSF-NH_2. Vibrational frequency calculated using the VQDPT2 method and the harmonic approximation is listed with experimental frequency. The modes with asterisk were discarded in the assignment: ν, frequency (in cm^{-1}); δν = ν(VQDPT2) - ν(Exp.); *I*, intensity (in km mol^{-1}); *f* = ν(VQDPT2)/ν(harmonic); MAD, mean absolute deviation.

(a) Conf.P – Conf.A' (S_1: 27.8 cm^{-1}, S_2: 17.5 cm^{-1}, ΔS_2: 0.5 cm^{-1})

		harmonic	VQDPT2				Exp.
	assignment	ν	ν	δν	*f*	*I*	ν
	$\nu(OH_{Ser4})$	3775.2	3563.2	-28.7	0.944	124.7	3591.9
	$\nu(OH_{Ser1})$	3670.5	3471.9	-32.1	0.946	566.1	3504.0
	$\nu(NH_{Ser1})$	3620.0	3454.1	-2.9	0.954	61.0	3457.0
	$\nu(NH_{Ser4})$	3603.5	3452.8	14.5	0.958	148.6	3438.3
*	$\nu_s(NH_{2,Cter})$	3510.4	3449.0		0.983	40.1	
	$\nu_{as}(NH_{2,Cter})$	3630.2	3370.2	17.0	0.928	58.1	3353.2
*	$\delta(NH_{2,Cter}) + \nu(CO_{Phe})$		3324.5			14.7	
	$\nu(NH_{Val})$	3491.8	3324.2	15.2	0.952	150.2	3309.0
	$\nu(NH_{Phe})$	3483.8	3307.7	7.1	0.949	315.1	3300.6
	$\nu(NH_{Ile})$	3426.0	3241.4	-13.3	0.946	360.5	3254.7
	$2\delta(NH_{2,Cter})$		3184.4	-3.4		11.1	3187.8
			MAD	14.9			

(b) Conf.Q – f28623 (S_1: 24.1 cm^{-1}, S_2: 17.6 cm^{-1}, ΔS_2: 5.0 cm^{-1})

		harmonic	VQDPT2				Exp.
	assignment	ν	ν	δν	*f*	*I*	ν
	$\nu(OH_{Ser4})$	3782.3	3561.3	-35.5	0.942	106.2	3596.8
	$\nu(NH_{Ser4})$	3612.2	3447.6	1.5	0.954	50.6	3446.1
	$\nu_{as}(NH_{2,Cter})$	3601.2	3427.4	3.8	0.952	115.0	3423.6
	$\nu(NH_{Ser1})$	3590.2	3420.8	7.0	0.953	94.2	3413.8
	$\nu(NH_{Val})$	3548.1	3384.5	24.3	0.954	108.1	3360.2
*	$\delta(NH_{2,Cter}) + \nu(CO_{Phe})$		3339.4			11.7	
	$\nu(NH_{Phe})$	3496.1	3330.1	-18.9	0.953	254.4	3349.1
	$\nu(OH_{Ser1})$	3560.5	3318.1	0.3	0.932	447.6	3317.8
	$\delta(NH_{2,Cter}) + \nu(CO_{Phe})$		3245.3	-5.0		285.0	3250.3
	$\nu(NH_{Ile})$	3371.4	3214.4	-4.8	0.953	299.1	3219.2
	$\nu_s(NH_{2,Cter})$	3409.2	3162.9	-33.9	0.928	234.4	3196.8
			MAD	13.5			

Table S4. Vibrational states contributing to Fermi resonances in f28623 of Ace-SIVSF-NH_2

ν	*I*	assignment	weight of dominant vibrational modes							
(cm^{-1})	(km mol^{-1})		weight	mode	weight	mode	weight	mode	weight	mode
3339.4	11.7	$\delta(NH_{2,Cter})$ + $\nu(CO_{Phe})$	0.48	$\delta(NH_{2,Cter})$ + $\nu(CO_{Phe})$	0.32	$2\delta(NH_{2,Cter})$	0.11	$\nu_s(NH_{2,Cter})$		
3245.3	285.0	$\delta(NH_{2,Cter})$ + $\nu(CO_{Phe})$	0.40	$\delta(NH_{2,Cter})$ + $\nu(CO_{Phe})$	0.30	$\nu_s(NH_{2,Cter})$	0.26	$2\delta(NH_{2,Cter})$		
3162.9	234.4	$\nu_s(NH_{2,Cter})$	0.52	$\nu_s(NH_{2,Cter})$	0.32	$2\delta(NH_{2,Cter})$	0.10	$\nu_{as}(NH_{2,Cter})$	0.03	$\delta(NH_{2,Cter})$ + $\nu(CO_{Phe})$
3214.4	299.1	$\nu(NH_{Ile})$	0.52	$\nu(NH_{Ile})$	0.44	$2\delta(NH_{Ile})$				
3061.8	85.1	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ile})$	0.52	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ile})$	0.22	$2\delta(NH_{Ile})$	0.18	$\nu(NH_{Ile})$		
3049.0	95.1	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ile})$	0.45	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ile})$	0.20	$\nu(NH_{Ile})$	0.19	$2\delta(NH_{Ile})$	0.11	$\delta(NH_{Ser1})$ + $\delta(NH_{Ile})$

S4. Supporting results of Ace-SIVSF-NHMe

S4.1. REMD simulations

Results of the REMD simulations are shown in Figure S12. As in the case of Ace-SIVSF-NH_2, a random walk in replica, temperature, and potential energy space was observed. The acceptance ratios were 26% on average. The PMFs at 1300 K demonstrates that the REMD simulation samples a sufficiently wide conformational space. At 300 K, the diagram for each residue is similar to that of Ace-SIVSF-NH_2 except for Phe. This difference in Phe is attributed to the *N*-methyl cap of C-terminal in Ace-SIVSF-NHMe.

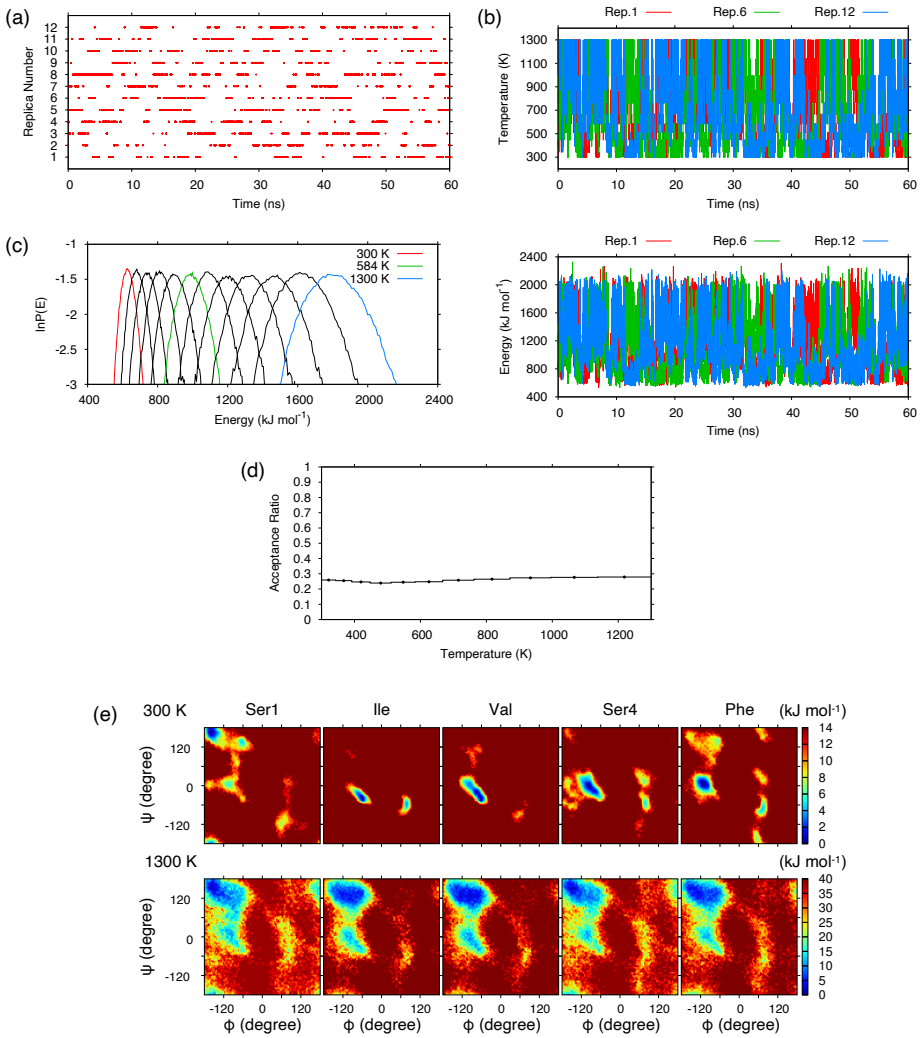


Figure S12. REMD simulations of Ace-SIVSF-NHMe. (a) Replica number of the trajectory at 300 K as a function of time. (b) Temperature and potential energy of replicas 1, 6, and 12 as a function of time. (c) Canonical probability distribution of the potential energy with different temperature. (d) Acceptance ratio of adjacent replicas. (e) Potential of mean force at 300 and 1300 K.

S4.2. Clustering analyses with Scheme 1

Figure S13a plots the number of representative structures obtained from the *k*-means clustering analysis. 281 and 138 structures were found by setting the clustering radius to 1.6 and 1.9 Å, respectively. The relative energy of the optimized structures, shown in Figure S13b, demonstrates that f19375 is the most stable conformer in both cases.

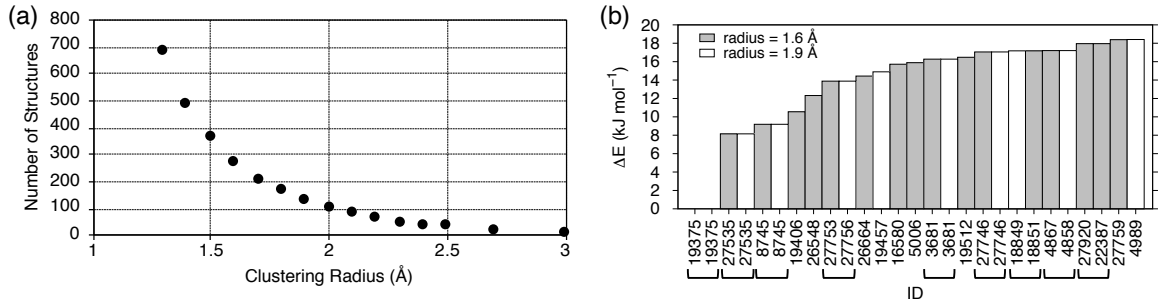


Figure S13. Results of conformational search with Scheme 1 for Ace-SIVSF-NHMe. (a) The number of clusters as function of clustering radius. (b) Relative energy of optimized geometries at the B3LYP/6-31(++)G** level without the zero-point energy correction (up to 20 kJ mol^{-1}). The conformer IDs of the same structure are bracketed.

S4.3. Conformational search with Scheme 2

The 30,000 snapshot structures obtained by REMD were classified into 2136 groups based on a H-bond matrix. The representative structures of each group were used for the PCA. The result of the PCA is shown in Figure S14 indicating that the first five PCs explain over 70% of variance. We therefore used five PCs for the hierarchical clustering with the Ward's method.[10] The hierarchical clustering yielded a dendrogram shown in Figure S15. From the dendrogram, 516 clusters were obtained by cutting at $h = 3$.

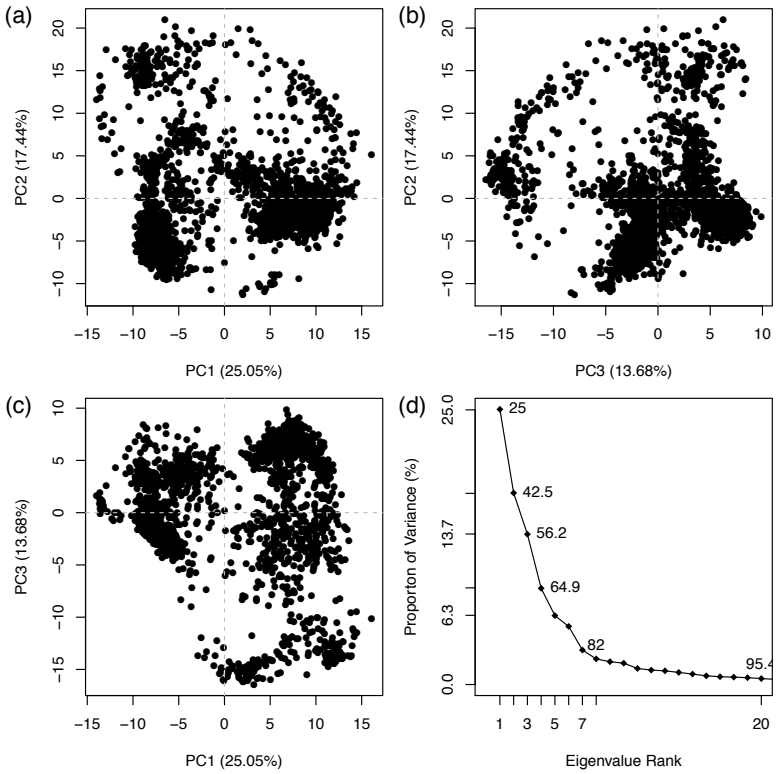


Figure S14. Result of the PCA for Ace-SIVSF-NHMe. Scatter plots of (a) PC1 vs PC2, (b) PC2 vs PC3, and (c) PC1 vs PC3. Proportion of variance is shown in parentheses (in %). (d) Eigenvalue contribution of PCs to variance of the dataset. The numbers in the frame indicate cumulative proportion of variance.

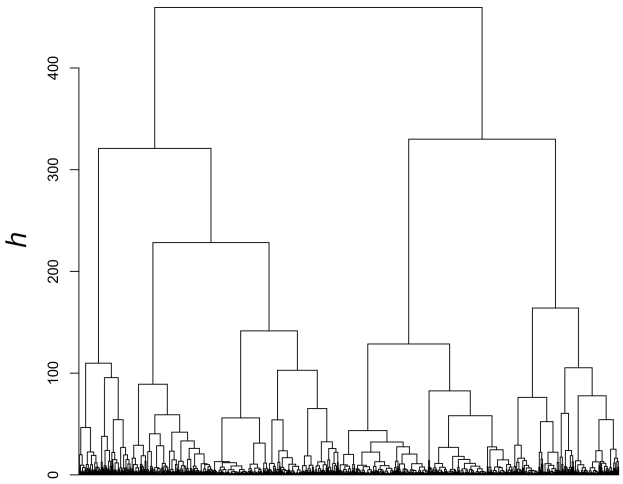


Figure S15. Dendrogram obtained with Ward's method of hierarchical clustering.

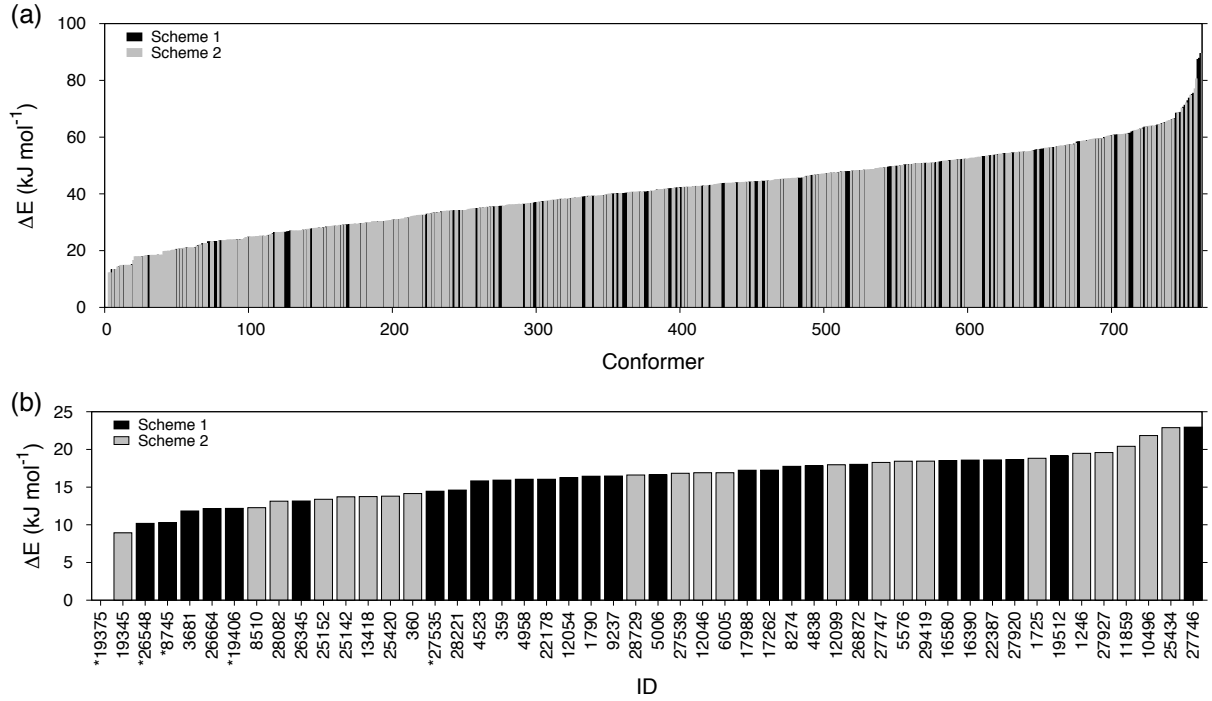


Figure S16. Relative energy of the Ace-SIVSF-NHMe conformers at the ωB97X/6-31(++)G** level of theory (a) without the ZPE correction and (b) with the ZPE correction for 50 lowest energy conformers. Relative energies of conformers found in Scheme 1 are shown in black, and those newly found in Scheme 2 are shown in gray. The conformers labeled with asterisk were selected for the anharmonic vibrational analysis in Scheme 1.

Figure S16a shows the relative energy of the optimized structure obtained at the ωB97X/6-31(++)G** level of theory. For comparison, we performed geometry optimization at the same level of theory for the 281 structures obtained from *k*-means clustering at 1.6 Å in Scheme 1, and the results are compared therein. Harmonic vibrational analyses were carried out for conformers with the relative energy less than 25 kJ mol^{-1} eliminating structural duplications, and 50 conformers were obtained without imaginary frequency. Figure S16b represents the relative energy of the conformers with the ZPE correction, indicating that more than half of the lowest-energy conformers are found in Scheme 1. Although the energy of the five structures selected for anharmonic vibrational analysis in Scheme 1 are relatively low, Scheme 2 has found several conformers

of comparable energy. Interestingly, f19375 is the energetically most stable conformer despite the different computational level used in Scheme 1. It is notable that f19375 is found in Scheme 2 as well.

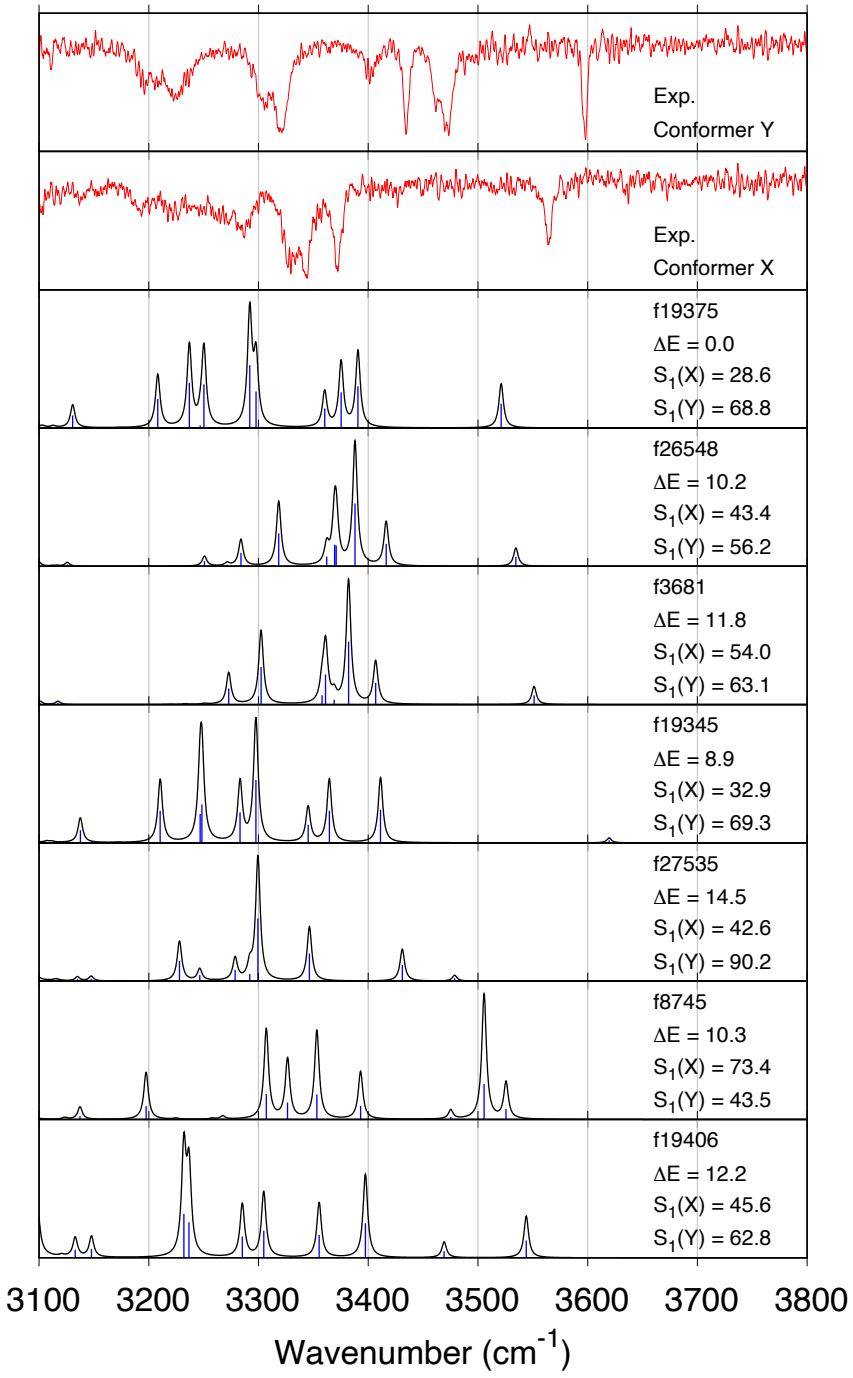


Figure S17. Theoretical IR spectra of the conformers of Ace-SIVSF-NHMe obtained by Schemes 1 and 2 compared with experimental ones. ΔE (in kJ mol^{-1}) is the relative energy from the most stable conformer (f19375) obtained at the ωB97X/6-31(++)G** level with the ZPE correction. The similarity scores S_1 (in cm^{-1}) with respect to Conformer X and Conformer Y are written as S_1(X) and S_1(Y), respectively.

S4.4. VQDPT2 spectra of conformers found with Schemes 1 and 2

VQDPT2 calculations were carried out for seven lowest-energy structures, f19375, f27535, f8745, f19406, and f26548 obtained by Scheme 1 and f19345 and f3681 obtained by Scheme 2. Figure S17 shows VQDPT2 spectra of the seven conformers and the experimental IR spectra of Conformer X and Conformer Y. The similarity of the calculated spectrum with the experimental spectrum of Conformer X, denoted S_1(X), is also shown. f8745, which gives three distinct peaks around 3500 cm^{-1}, is found with the largest S_1(X) = 73.4 cm^{-1}. Although S_1(X) of f27535, f19406, f26548, and f3681 are decreased to 42.6, 45.6, 43.4 and 54.0 cm^{-1}, those of f19375 and f19345 are smaller, 28.6 and 32.9 cm^{-1}, respectively.

f19345 has the second lowest energy at the level of ωB97X/6-31(++)G**, and has the second smallest and smallest value of S_1(X) and S_2(X), respectively. The spectrum of f19345 has a peak over 3600 cm^{-1} implying a free OH stretching mode, whereas the experimental spectra of Conformer X does not exhibit such a peak. In Figure S18, f19375 and f19345 are superimposed. f19345 has a similar structure to f19375 but different only in the interaction of OH_{Ser4}. OH_{Ser4} of f19375 forms a six-membered-ring HB with CO_{Ser4} but that of f19345 is free. We therefore conclude that Conformer X is assigned to f19375 from both energetic and spectroscopic points of view.

On the other hand, none of the calculated conformers can be assigned to the other experimentally observed conformer, Conformer Y. The score S_1(Y) of the five conformers is calculated in a range of 40 – 90 cm^{-1}, because the shape of the spectrum is quite different from that of Conformer Y shown in Figure S17. The result has prompted us to carry out further extensive conformational search, i.e., Scheme 3.

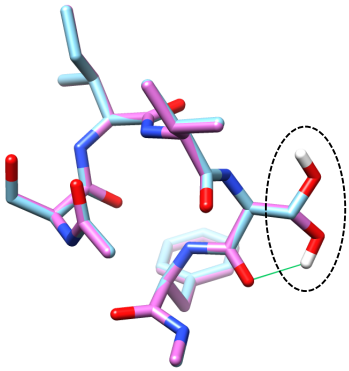


Figure S18. Molecular structures of Ace-SIVSF-NHMe conformers: purple, f19375; light blue, f19345. Hydrogen atoms except for OH_{Ser4} (rounded by dashed circle) are omitted for clarity. OH_{Ser4} of f19375 forms a six-membered-ring HB with CO_{Ser4} (green line), whereas that of f19345 is free.

S4.5. VQDPT2 spectra of all calculated conformers

The VQDPT2 spectra of all the 38 conformers are shown in Figure S19.

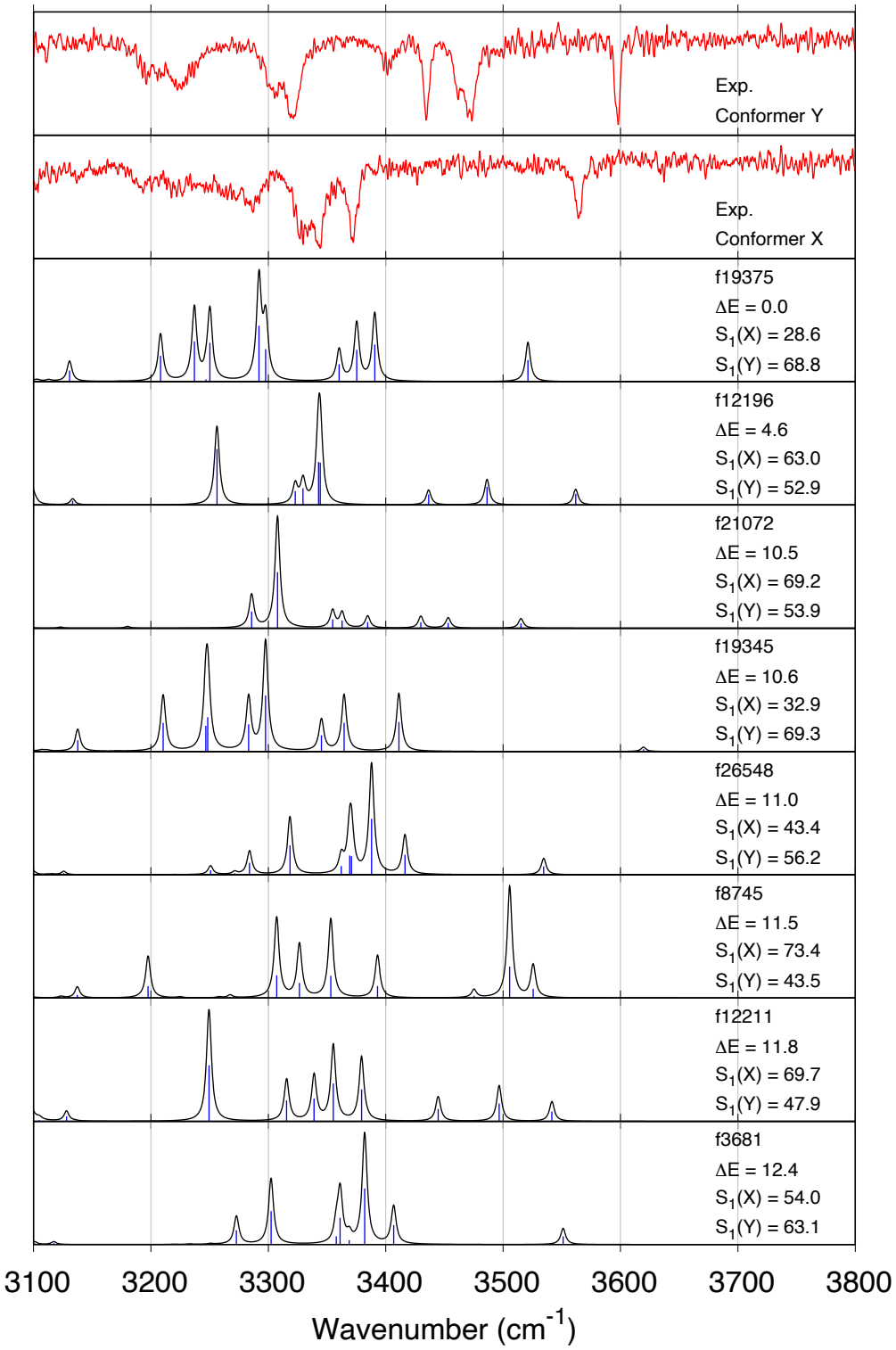


Figure S19. Calculated spectra with VQDPT2 method compared with experimental spectra for Conformer X and Conformer Y. ΔE (in kJ mol^{-1}) is the relative energy from the most stable conformer (f19375) obtained at the RI-MP2/def2-TZVP(-df) level. The computed spectra are sorted in increasing order of the relative energy. The similarity scores S_1 (in cm^{-1}) with respect to Conformer X and Conformer Y are written as S_1(X) and S_1(Y), respectively.

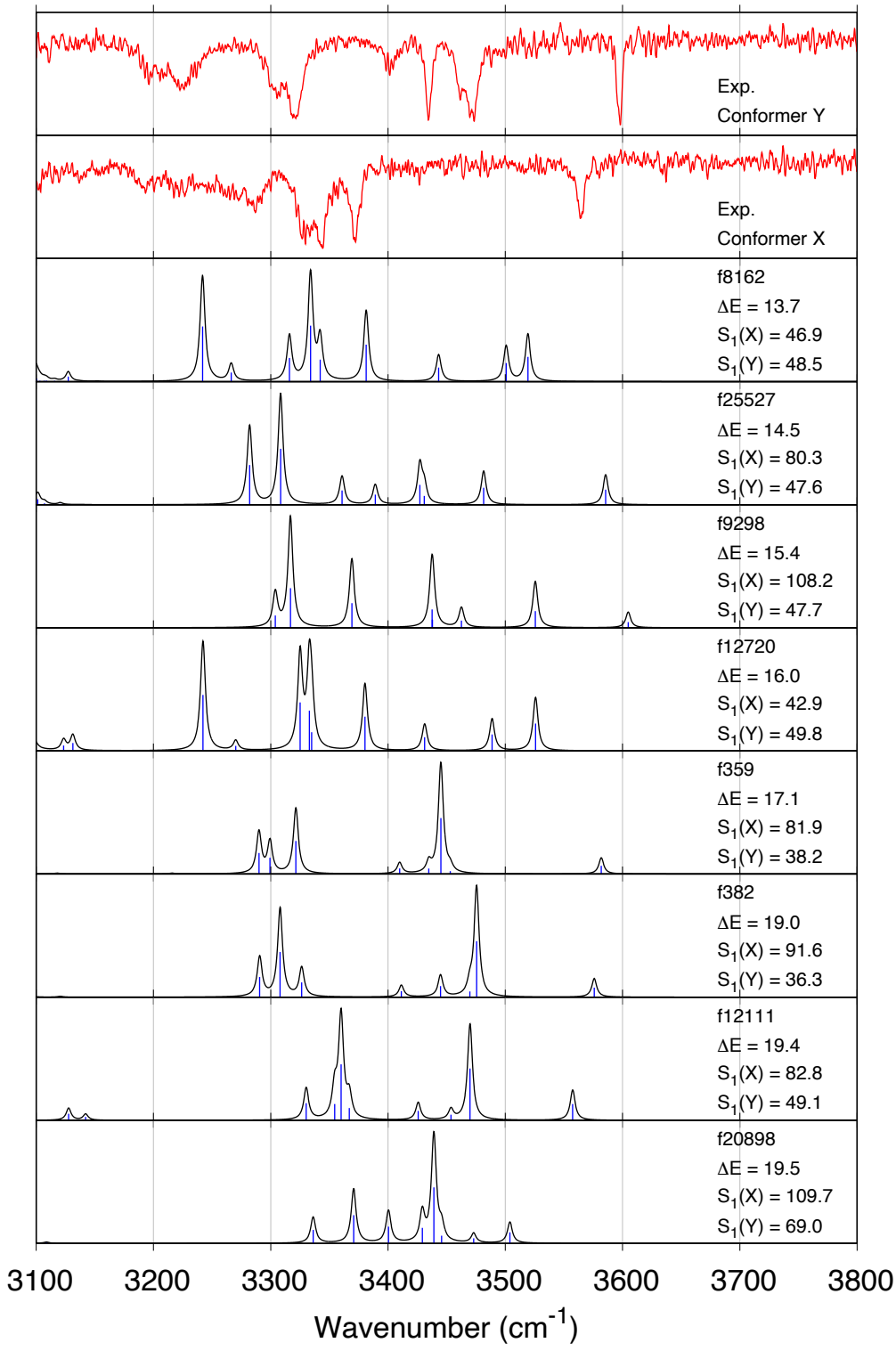


Figure S19 (cont.)

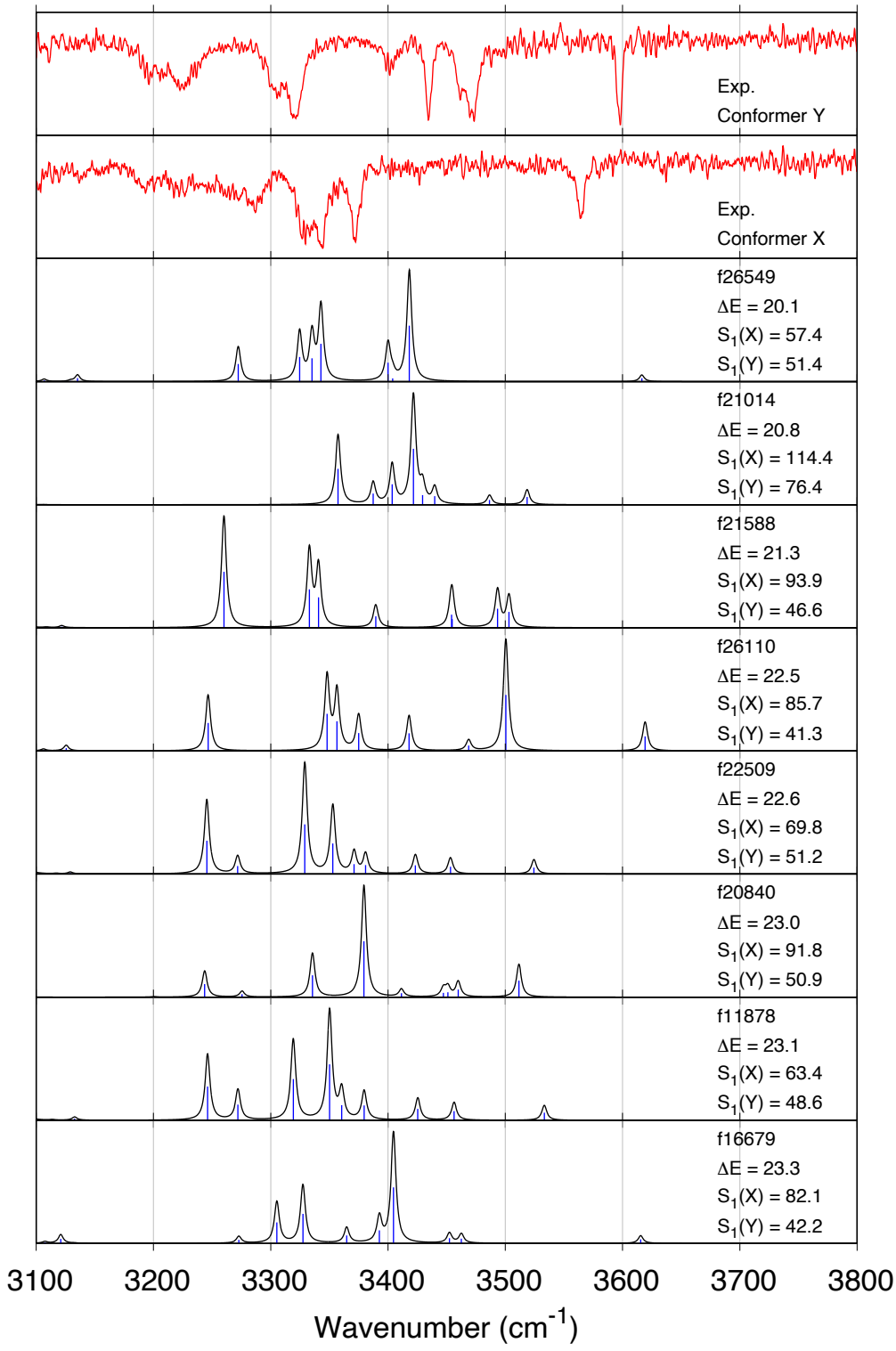


Figure S19 (cont.)

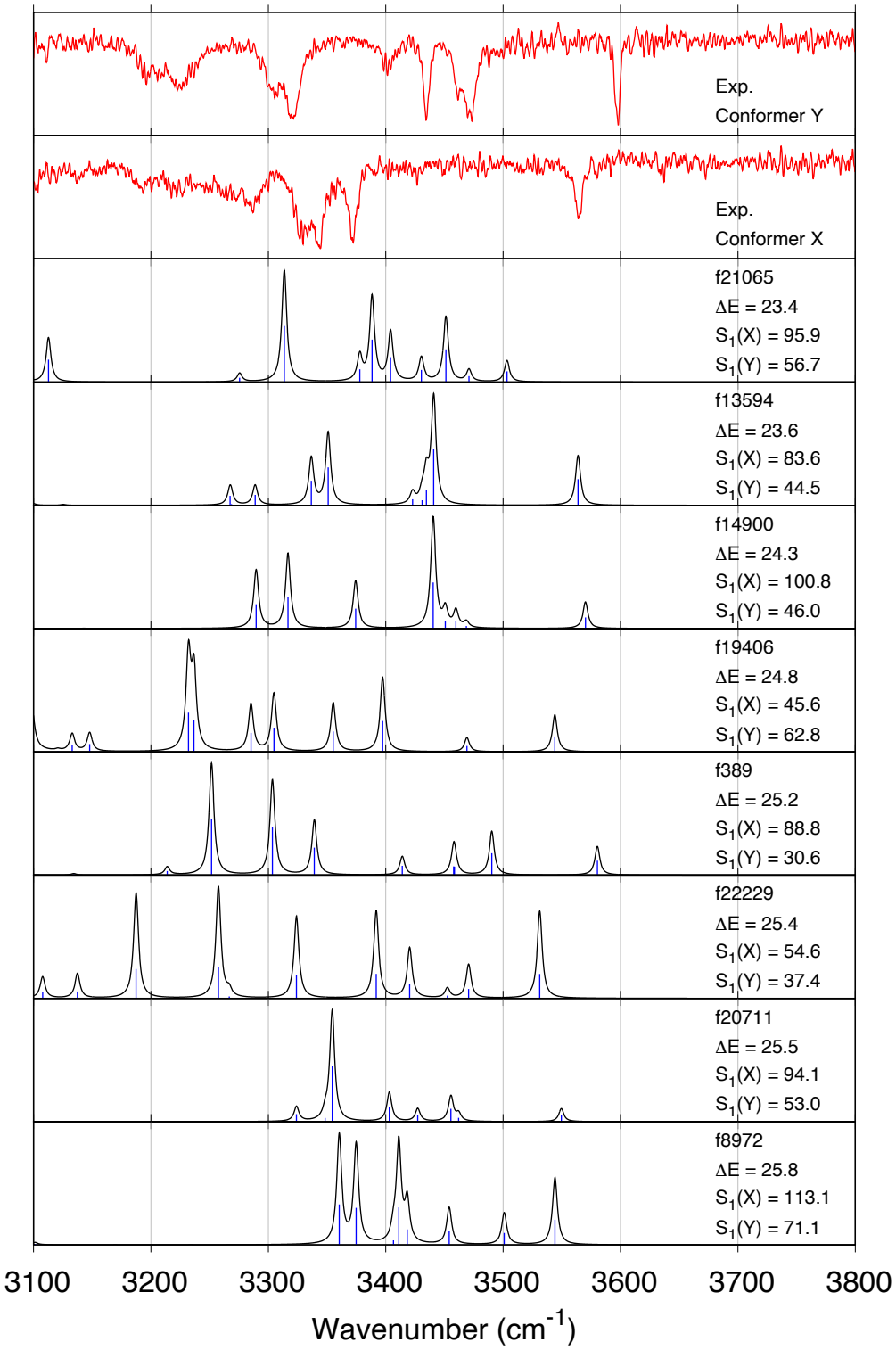


Figure S19 (cont.)

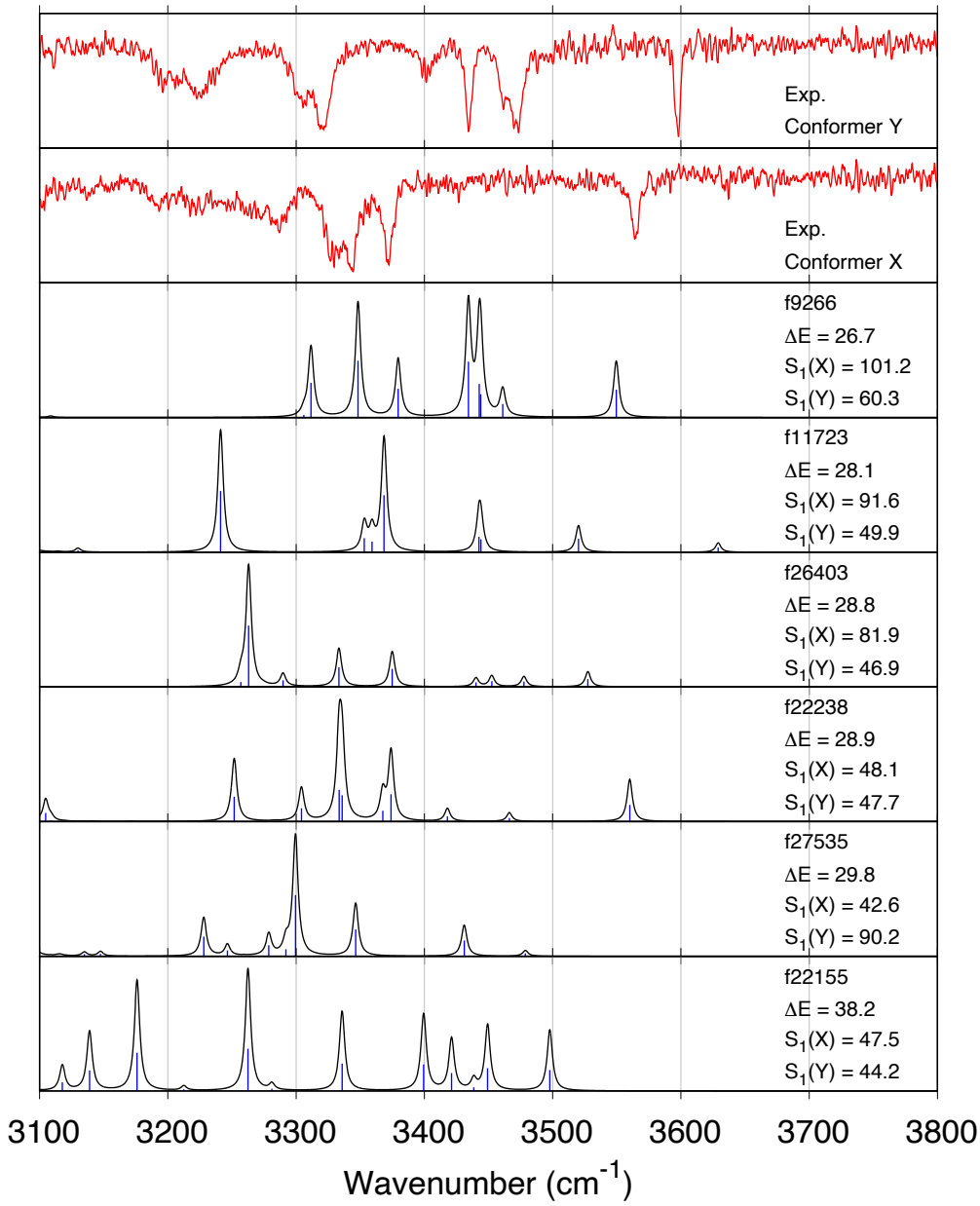


Figure S19 (cont.)

S4.6. Assignment of the peaks

The results of the peak assignment of the experimental and calculated spectra are summarized in Figure S20 and Table S5. The assignment based on S_1 shows reasonable correspondence of the peaks between spectra obtained with the experiment and the VQDPT2 method. As in the assignment for Ace-SIVSF-NH_2, a markedly small peak ("$\delta(NH_{NHMe}) + \nu(CO_{Ser4})$" at 3246.8 cm^{-1} in the VQDPT2 spectrum of f19375) was discarded by setting α in S_1, which might contribute to the reasonable assignment. The values of ΔS_2 are close or equal to zero, which indicates that the assignments are (almost) the best representations of peak intervals. MAD between spectra X and f19375 is 19.0 cm^{-1}, and that between Y and f389 is 15.0 cm^{-1}. These values are comparable to MAD for SIVSF-NH_2 and Ace-SIVSF-NH_2.

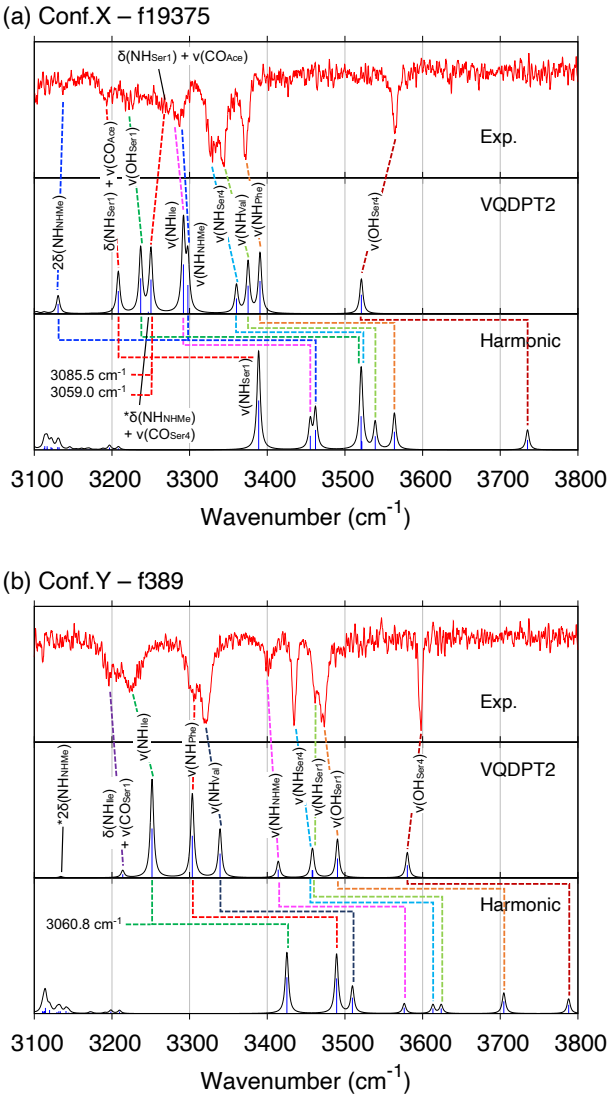


Figure S20. Theoretical IR spectra of (a) f19375 and (b) f389 of Ace-SIVSF-NHMe compared with experimental ones. "$2\delta(NH_{NHMe})$" is an overtone of the NH bending mode of C-terminal. "$\delta(NH_{Ser1}) + \nu(CO_{Ace})$" is a combination tone of the NH bending mode of Ser1 and the CO stretching mode of the acetyl cap in N-terminal. "$\delta(NH_{NHMe}) + \nu(CO_{Ser4})$" is a combination tone of the NH bending mode of C-terminal and the CO stretching mode of Ser4. "$\delta(NH_{Ile}) + \nu(CO_{Ser1})$" is a combination tone of the NH bending mode of Ile and the CO stretching mode of Ser1. The modes labeled with asterisk were not used for the assignment. The computed intensity of $2\delta(NH_{NHMe})$ in (b) was less than 10 km mol^{-1} (4.3 km mol^{-1}); thus, this mode was not used for assignment.

The ratios between the anharmonic and harmonic frequencies are summarized in Table S5. In f389, the ratio varies in the range of 0.94 to 0.96, and peak positions shift lower in the VQDPT2 calculation compared to the harmonic spectrum without interchanging. In f19375, however, the band position of $\nu(OH_{Ser1})$ is interchanged with $\nu(NH_{Ile})$ and $\nu(NH_{NHMe})$. The ratio *f* of $\nu(OH_{Ser1})$ is 0.92, which is smaller than those of the other modes (0.94–0.95) and close to that of $\nu(OH_{Ser1})$ in f28623 of Ace-SIVSF-NH_2 (0.93, see Table S3b). The HB diagrams in Figure 7 show that both f19375 of Ace-SIVSF-NHMe and f28623 of Ace-SIVSF-NH_2 have the HB of OH_{Ser1} to CO_{Ace} in N-terminal, which makes up a seven-membered ring (for example, see f28623 in Figure S11).

Both in the VQDPT2 spectra of f19375 and f389, strong peaks of overtones or combination tones were observed, which contribute to qualitative description of the broad bands in the region below 3300 cm^{-1}. Fermi resonances are seen both in f19375 and f389. Table S6 summarizes the vibrational states and weights of each mode obtained with the VQDPT2 calculation. In f19375, our results show two resonances: One resonance is involving $\nu(NH_{NHMe})$ and an overtone of $\delta(NH_{NHMe})$. The other resonance consists of $\nu(NH_{Ser1})$, a combination tone of $\delta(NH_{Ser1})$ and $\nu(CO_{Ace})$, a combination tone of $\delta_s(CH_{3,Ace})$ and $\delta(NH_{Ser1})$, and an overtone of $\delta(NH_{Ser1})$. The peak of $\nu(NH_{Ser1})$ is the strongest in the harmonic spectrum, whereas the contribution of this mode is almost equally distributed to the four vibrational states and these peaks have large intensities in the VQDPT2 analysis, although the position of the two peaks is out of range of the experimental spectrum. Because of the difference in this region, f19375 might not be assigned to Conformer X with the harmonic analysis.

In f389, the VQDPT2 calculation suggests a Fermi resonance involving $\nu(NH_{Ile})$ and an overtone of $\delta(NH_{Ile})$. These vibrational states are split into 3251.6 cm^{-1} and 3060.8 cm^{-1}. Although the contribution of the fundamental mode to the latter is rather small, the intensity is relatively large. This is likely due to the strong intensity of $\nu(NH_{Ile})$.

Table S5. Identical to Table S3 but for Ace-SIVSF-NHMe.

(a) Conf.X – f19375 (S_1: 28.6 cm^{-1}, S_2: 21.9 cm^{-1}, ΔS_2: 0.6 cm^{-1})

		harmonic	VQDPT2				Exp.
	assignment	ν	ν	δν	*f*	*I*	ν
	$\nu(OH_{Ser4})$	3735.1	3521.2	-43.0	0.943	122.7	3564.2
	$\nu(NH_{Phe})$	3563.7	3390.7	18.9	0.951	210.9	3371.7
	$\nu(NH_{Val})$	3539.0	3375.3	32.3	0.954	180.8	3343.0
	$\nu(NH_{Ser4})$	3521.8	3360.3	31.7	0.954	98.7	3328.6
	$\nu(NH_{NHMe})$	3462.1	3297.8	4.7	0.953	184.9	3293.0
	$\nu(NH_{Ile})$	3455.4	3292.0	6.1	0.953	317.1	3286.0
	$\delta(NH_{Ser1}) + \nu(CO_{Ace})$		3250.2	-16.8		220.3	3267.1
*	$\delta(NH_{NHMe}) + \nu(CO_{Ser4})$		3246.8			13.4	
	$\nu(OH_{Ser1})$	3520.7	3237.0	13.3	0.919	228.1	3223.7
	$\delta(NH_{Ser1}) + \nu(CO_{Ace})$		3208.2	16.0		146.8	3192.2
	$2\delta(NH_{NHMe})$		3130.7	-7.3		63.6	3138.0
			MAD	19.0			

(b) Conf.Y – f389 (S_1: 30.6 cm^{-1}, S_2: 13.5 cm^{-1}, ΔS_2: 0.0 cm^{-1})

	harmonic	VQDPT2				Exp.
assignment	ν	ν	δν	*f*	*I*	ν
$\nu(OH_{Ser4})$	3788.0	3580.3	-16.9	0.945	86.2	3597.2
$\nu(OH_{Ser1})$	3704.6	3490.4	18.6	0.942	131.5	3471.8
$\nu(NH_{Ser1})$	3624.0	3458.6	-2.9	0.954	50.6	3461.5
$\nu(NH_{Ser4})$	3613.4	3457.8	23.2	0.957	51.9	3434.6
$\nu(NH_{NHMe})$	3576.4	3414.1	12.1	0.955	56.0	3402.1
$\nu(NH_{Val})$	3509.7	3339.2	18.2	0.951	165.4	3321.0
$\nu(NH_{Phe})$	3489.3	3303.5	-0.9	0.947	286.6	3304.4
$\nu(NH_{Ile})$	3425.3	3251.6	26.8	0.949	336.5	3224.8
$\delta(NH_{Ile}) + \nu(CO_{Ser1})$		3213.9	15.4		24.3	3198.5
		MAD	15.0			

Table S6. Identical to Table S4 except for (a) f19375 and (b) f389 of Ace-SIVSF-NHMe.

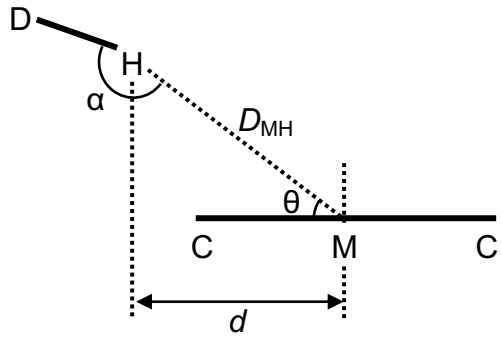
(a) Ace-SIVSF-NHMe, f19375

ν	*I*	assignment	weight of dominant vibrational modes					
(cm^{-1})	(km mol^{-1})		weight	mode	weight	mode	weight	mode
3297.8	184.9	$\nu(NH_{NHMe})$	0.62	$\nu(NH_{NHMe})$	0.30	$2\delta(NH_{NHMe})$		
3130.7	63.6	$2\delta(NH_{NHMe})$	0.67	$2\delta(NH_{NHMe})$	0.30	$\nu(NH_{NHMe})$		
3250.2	220.3	$\delta(NH_{Ser1})$ + $\nu(CO_{Ace})$	0.46	$\delta(NH_{Ser1})$ + $\nu(CO_{Ace})$	0.32	$\nu(NH_{Ser1})$	0.21	$2\delta(NH_{Ser1})$
3208.2	146.8	$\delta(NH_{Ser1})$ + $\nu(CO_{Ace})$	0.53	$\delta(NH_{Ser1})$ + $\nu(CO_{Ace})$	0.24	$2\delta(NH_{Ser1})$	0.21	$\nu(NH_{Ser1})$
3085.5	85.6	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ser1})$	0.50	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ser1})$	0.24	$2\delta(NH_{Ser1})$	0.14	$\nu(NH_{Ser1})$
3059.0	125.1	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ser1})$	0.47	$\delta_s(CH_{3,Ace})$ + $\delta(NH_{Ser1})$	0.21	$2\delta(NH_{Ser1})$	0.20	$\nu(NH_{Ser1})$

(b) Ace-SIVSF-NHMe, f389

ν	*I*	assignment	weight of dominant vibrational modes					
(cm^{-1})	(km mol^{-1})		weight	mode	weight	mode	weight	mode
3251.6	336.5	$\nu(NH_{Ile})$	0.72	$\nu(NH_{Ile})$	0.16	$2\delta(NH_{Ile})$	0.11	$\delta(NH_{Ile})$ + $\nu(CO_{Ser1})$
3060.8	83.6	$2\delta(NH_{Ile})$	0.71	$2\delta(NH_{Ile})$	0.21	$\nu(NH_{Ile})$		

S5. Supporting data



criteria	Ace-SIVSF-NH_2 f28623	Ace-SIVSF-NHMe f389
$D_{MH} \leq 4.0$	3.11	3.13
$\theta \leq 90$	62.1	63.8
$90 \leq \alpha \leq 180$	163.3	158.9
$d > 1.4$	1.45	1.39

Figure S21. Results of the hydrogen bond analysis of OH_{Ser4}–Ph_{Phe} for f28623 (Ace-SIVSF-NH_2) and f389 (Ace-SIVSF-NHMe) based on Malone's criteria for D–H···π interactions (Type V). D, donor atom; H, hydrogen atom; C, carbon atom of aromatic ring; M, aromatic centroid; D_{MH}, distance between M and H; θ, angle of approach of H···M to the plane of aromatic ring; α, angle formed by H–D and H···M; *d*, distance of orthogonal projection of H···M onto the plane of aromatic ring. Distances and angles are given in angstroms and degrees, respectively.

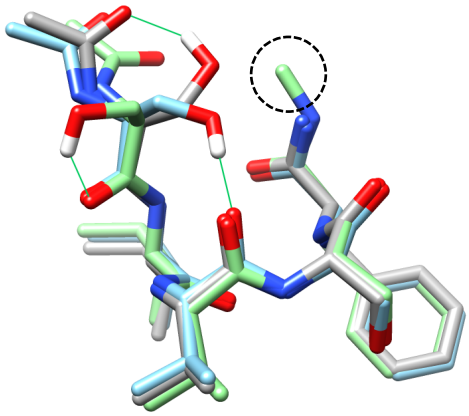


Figure S22. Molecular structures of the SIVSF conformers: gray, f28623 of Ace-SIVSF-NH_2; light blue, Conformer A' of Ace-SIVSF-NH_2; green, f389 of Ace-SIVSF-NHMe. The C-terminal methyl group of f389 is rounded with a dashed circle. Hydrogen atoms except for OH_{Ser1} are omitted for clarity. Hydrogen bonds of OH_{Ser1} are shown in green lines.

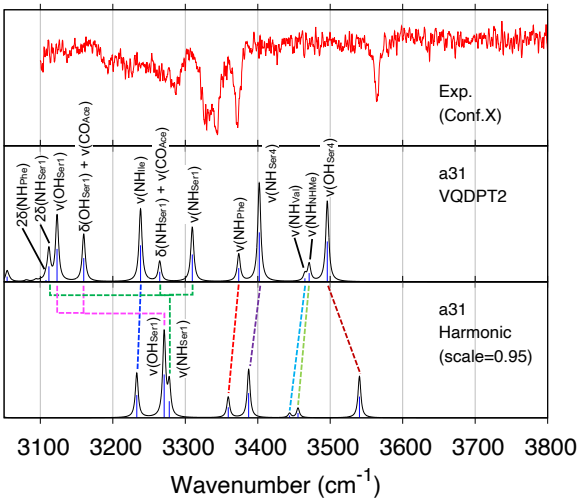


Figure S23. Calculated IR spectra of a31 of Ace-SIVSF-NHMe, compared with experimental spectrum of Conformer X. The harmonic calculation and PES generation for the anharmonic calculation were performed at the B3LYP/6-31(++)G** level. The harmonic frequencies are scaled with a linear scaling factor of 0.95.

Table S7. Vibrational frequency calculated using VQDPT2 method and harmonic approximation for a31 of Ace-SIVSF-NHMe: ν, frequency (in cm^{-1}); *I*, intensity (in km mol^{-1}); *f* = ν(VQDPT2)/ν(harmonic)

assignment	harmonic		VQDPT2		
	ν	scaled (0.95)	ν	*f*	*I*
$\nu(OH_{Ser4})$	3726.6	3540.3	3495.9	0.938	283.8
$\nu(NH_{NHMe})$	3637.4	3455.5	3470.9	0.954	60.9
$\nu(NH_{Val})$	3625.1	3443.9	3465.0	0.956	24.3
$\nu(NH_{Ser4})$	3565.8	3387.5	3402.1	0.954	347.6
$\nu(NH_{Phe})$	3536.1	3359.3	3373.8	0.954	96.3
$\nu(NH_{Ser1})$	3450.5	3278.0	3309.7	0.959	191.9
$\delta(NH_{Ser1}) + \nu(CO_{Ace})$			3264.7		70.6
$\nu(NH_{Ile})$	3403.2	3233.1	3238.3	0.952	256.2
$\delta(OH_{Ser1}) + \nu(CO_{Ace})$			3159.9		166.9
$\nu(OH_{Ser1})$	3443.1	3270.9	3123.1	0.907	231.2
$2\delta(NH_{Ser1})$			3111.8		108.6
$2\delta(NH_{Phe})$			3107.8		12.9

Table S8. Vibrational states contributing to Fermi resonances in a31 of Ace-SIVSF-NHMe

ν	*I*	assignment	weight of dominant vibrational modes					
(cm^{-1})	(km mol^{-1})		weight	mode	weight	mode	weight	mode
3309.7	191.9	$\nu(NH_{Ser1})$	0.48	$\nu(NH_{Ser1})$	0.31	$\delta(NH_{Ser1}) + \nu(CO_{Ace})$	0.20	$2\delta(NH_{Ser1})$
3264.7	70.6	$\delta(NH_{Ser1}) + \nu(CO_{Ace})$	0.68	$\delta(NH_{Ser1}) + \nu(CO_{Ace})$	0.18	$\nu(NH_{Ser1})$	0.14	$2\delta(NH_{Ser1})$
3111.8	108.6	$2\delta(NH_{Ser1})$	0.64	$2\delta(NH_{Ser1})$	0.30	$\nu(NH_{Ser1})$		
3159.9	166.9	$\delta(OH_{Ser1}) + \nu(CO_{Ace})$	0.58	$\delta(OH_{Ser1}) + \nu(CO_{Ace})$	0.36	$\nu(OH_{Ser1})$		
3123.1	231.2	$\nu(OH_{Ser1})$	0.52	$\nu(OH_{Ser1})$	0.41	$\delta(OH_{Ser1}) + \nu(CO_{Ace})$		

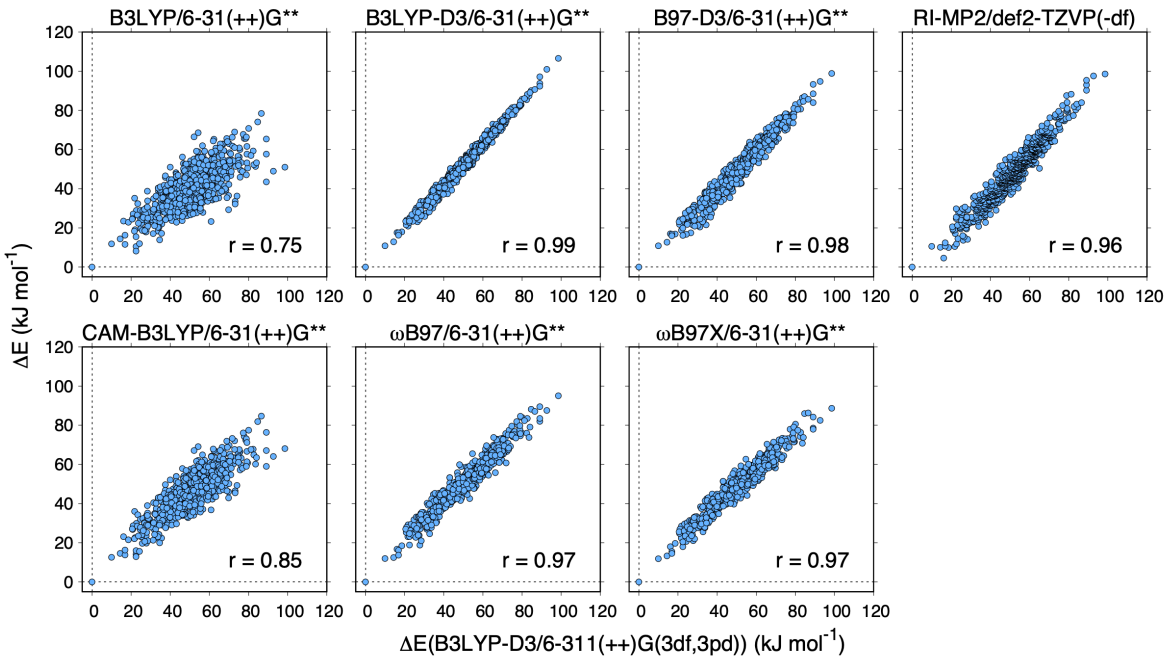


Figure S24. Relative energies of conformers of Ace-SIVSF-NHMe used in the harmonic vibrational analysis in the new scheme. The structures were optimized at the B3LYP/6-31(++)G** level, and single-point energies were calculated at each level. The relative energies obtained at the B3LYP-D3/6-311(++)G(3df,3pd) level are compared with those obtained by various levels of electronic structure theory. The correlation coefficient (*r*) is shown in the inset.

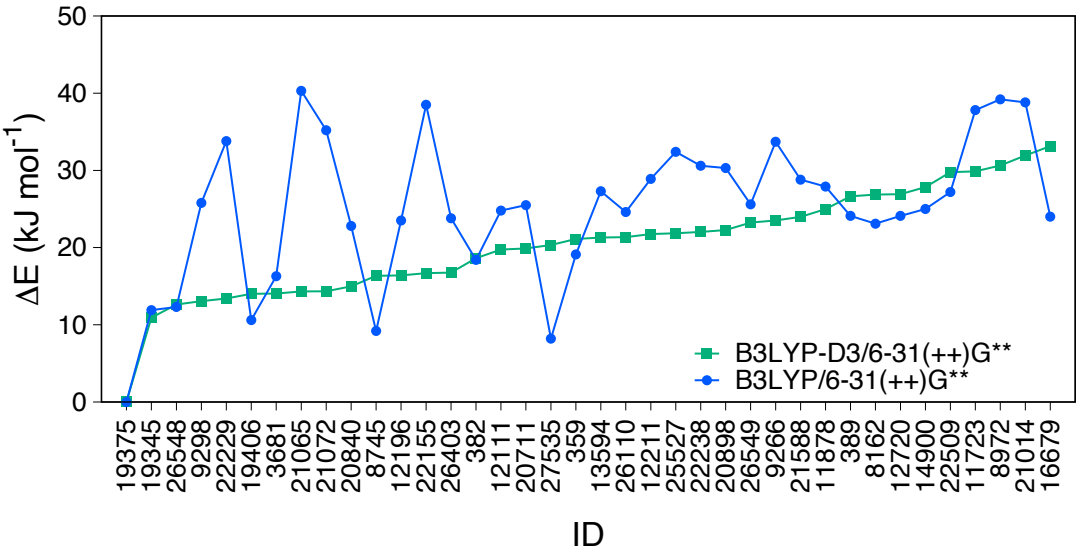


Figure S25. The relative energies of low-energy conformers calculated at the B3LYP/6-31(++)G** and B3LYP-D3/6-31(++)G** level of theory. The structures were optimized by the respective method.

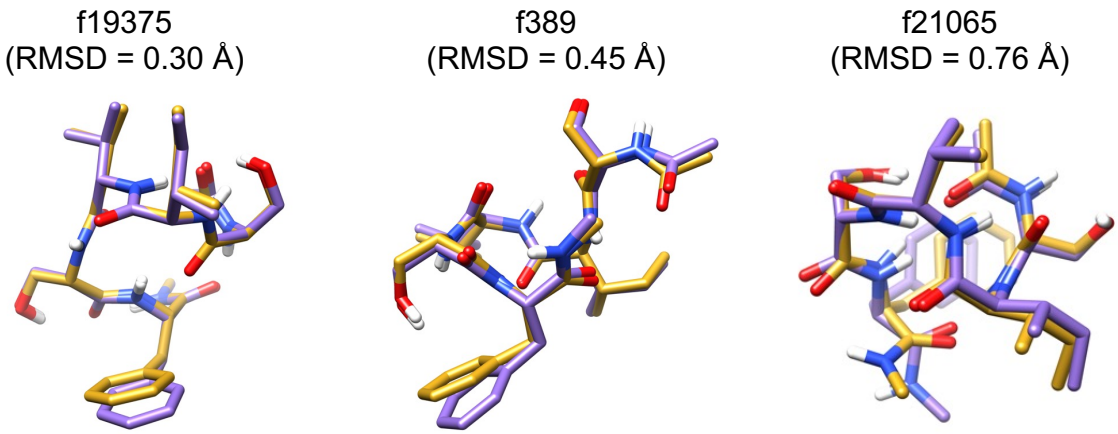


Figure S26. Comparison of the structure of f19375, f389, and f21065 optimized at the level of B3LYP-D3/6-31(++)G** (orange) and B3LYP/6-31(++)G** (purple). The RMSD is calculated from the backbone and sidechain atoms of HB donors and acceptors after the alignment of the molecule with backbone atoms.

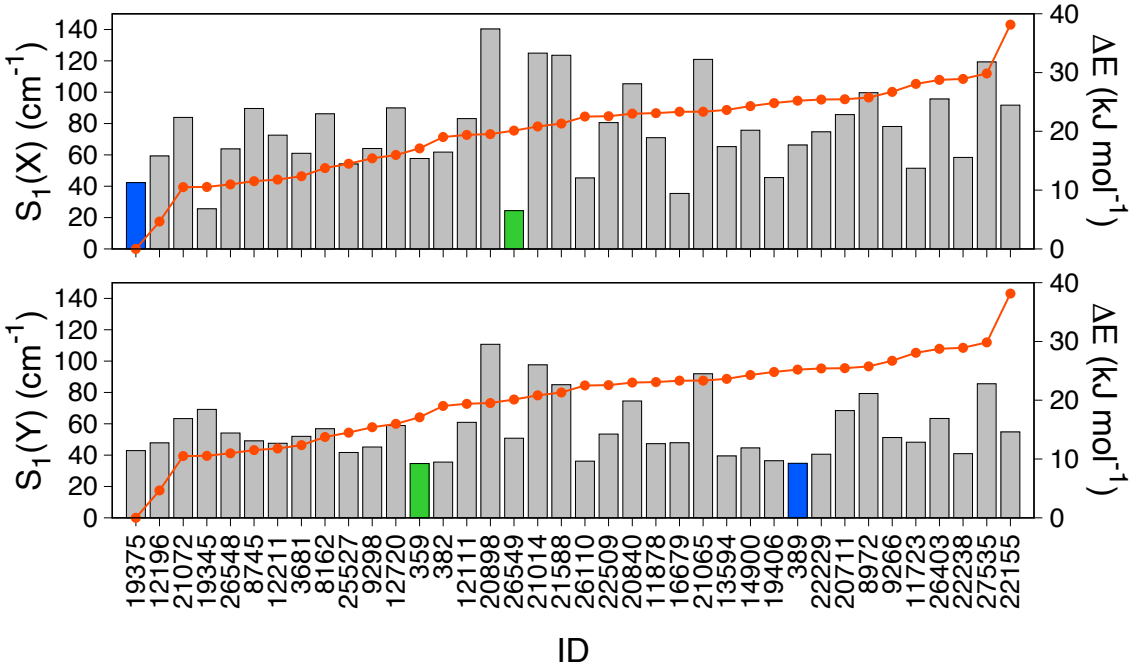


Figure S27. Similarity scores S_1 (left axis) with respect to Conformer X and Conformer Y obtained from scaled harmonic frequencies and intensities, and the relative energy ΔE (right axis) from the most stable conformer, f19375, obtained at the RI-MP2/def2-TZVP(-df) level. The smallest S_1 values are indicated in green, while f19375 and f389 are in blue.

References

[1] H. Otaki, K. Yagi, S. Ishiuchi, M. Fujii, Y. Sugita, *J. Phys. Chem. B* **2016**, *120*, 10199-10213.
[2] J. Huang, A. D. MacKerell Jr, *J. Comput. Chem.* **2013**, *34*, 2135-2145.
[3] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.
[4] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785-789.
[5] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5652.
[6] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213-222.
[7] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257-2261.
[8] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294-301.
[9] G. R. Desiraju, T. Steiner, *The Weak Hydrogen Bond*, Oxford University Press, New York, **2001**.
[10] J. H. Ward, *J. Am. Stat. Assoc.* **1963**, *58*, 236-244.
[11] J.-D. Chai, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 084106.
[12] J. F. Malone, C. M. Murray, M. H. Charlton, R. Docherty, A. J. Lavery, *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 3429-3436.
[13] T. Yanai, D. P. Tew, N. C. Handy, *Chem. Phys. Lett.* **2004**, *393*, 51-57.
[14] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
[15] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143-152.
[16] D. A. Case, D. S. Cerutti, T.E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, P. A. Kollman, AmberTools17, University of California, San Francisco, **2017**,
[17] P. Seidler, T. Kaga, K. Yagi, O. Christiansen, K. Hirao, *Chem. Phys. Lett.* **2009**, *483*, 138-142.
[18] K. Yagi, M. Keçeli, S. Hirata, *J. Chem. Phys.* **2012**, *137*, 204118.
[19] K. Yagi, H. Otaki, *J. Chem. Phys.* **2014**, *140*, 084113.
[20] K. Yagi, S. Hirata, K. Hirao, *Theor. Chem. Acc.* **2007**, *118*, 681-691.
[21] K. Yagi, S. Hirata, K. Hirao, *Phys. Chem. Chem. Phys.* **2008**, *10*, 1781-1788.
[22] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545-1614.
[23] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminform.* **2012**, *4*, 17.
[24] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781-1802.
[25] M. Feig, J. Karanicolas, C. L. Brooks III, *J. Mol. Graph. Model.* **2004**, *22*, 377-395.
[26] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. D. Caves, *Bioinformatics* **2006**, *22*, 2695-2696.
[27] L. Skjærven, X.-Q. Yao, G. Scarabelli, B. J. Grant, *BMC Bioinformatics* **2014**, *15*, 399.
[28] B. J. Grant, L. Skjærven, X.-Q. Yao, *Protein Sci.* **2021**, *30*, 20-30.
[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09 Revision D.01, Gaussian, Inc., Wallingford CT, **2009**,
[30] F. Neese, *WIREs Computational Molecular Science* **2012**, *2*, 73-78.
[31] K. Yagi, SINDO 4.0, **2019**, (https://tms.riken.jp/en/research/software/sindo/).
[32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.
[33] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498-2504.
[34] S. Ishiuchi, K. Yamada, H. Oba, H. Wako, M. Fujii, *Phys. Chem. Chem. Phys.* **2016**, *18*, 23277-23284.
[35] T. Sekiguchi, M. Tamura, H. Oba, P. Çarçarbal, R. R. Lozada-Garcia, A. Zehnacker-Rentien, G. Grégoire, S. Ishiuchi, M. Fujii, *Angew. Chem. Int. Ed.* **2018**, *57*, 5626-5629.