Generating Multiple Conformations of Flexible Peptides in Solution Based on NMR Nuclear Overhauser Effect Data: Application to Desmopressin[†,‡]

Jianjun Wang, Robert S. Hodges, and Brian D. Sykes*

Contribution from the Protein Engineering Network of Centres of Excellence, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2S2, Canada

Received September 7, 1994[⊗]

Abstract: A computational procedure, PEPFLEX-II, for calculating multiple peptide solution conformations from NMR data is reported. It combines restrained simulated annealing for structural calculations, an ensemble averaged full relaxation matrix approach for NOE back calculation, and an iterative NOE restraint modification procedure into a protocol whose goal is to calculate the maximum range of peptide conformations consistent with the experimental data. The distance restraint modification is achieved by systematically increasing the upper boundary of the distance restraints when NOE violations occur. By using looser distance restraints and using a relatively small NOE "force constant" in the simulated annealing, the procedure allows the molecule to search over a wider conformational space. The result is an improved fit of the calculated NOE's for the ensemble to the original experimental NOE data. Two convergence criteria are used to test whether both the best fit of the experimental NOE data and the maximum coverage of conformational space are achieved. The PEPFLEX-II procedure has been applied to the nanopeptide desmopressin in aqueous solution, for which several well-defined structural families were generated. The structures satisfy all the criteria for "good NMR structures", and their ensemble average fits the experimental NOE data better than any one individual structure.

Introduction

A number of NMR techniques have been successfully developed to elucidate the solution structures of small proteins and peptides during the past decade.[1,2] These techniques make use of NMR parameters of several types: NOE data for internuclear distance information; scalar coupling constants for torsion angle information; and temperature coefficients or amide proton exchange rates for H-bond information.[3] More recently, chemical shift parameters have also been used to provide structural information.[4-6] NMR structure generation methods commonly take this structural information and construct parameter-based penalty functions which are used in restrained molecular dynamics and restrained energy minimization calculations or use distance geometry methods to determine a structure based on the experimental data.[7] In addition, approaches such as the iterative relaxation matrix approach (IRMA) are often used to refine these structures.[8,9] For proteins or peptides which are rigid and have only a single conformation in solution, these methods work well and lead to the generation of a reasonable structure.

The basic premise of these NMR approaches is that a single NOE or coupling constant is associated with a single interproton distance or angle, respectively. This is valid for rigid proteins. However, proteins or peptides are not always rigid. Indeed some, especially small peptides, are extremely flexible and exhibit multiple conformations in solution.[10-13] If these multiple conformation are interconverting rapidly on the NMR time scale, they give rise to a set of averaged experimental NMR parameters, and the distances derived from the NOE's, for example, correspond to weighted average distances.

Several approaches have been taken to address the problem of flexibility in structural calculations. Torda and co-workers have developed a time-averaged NOE/coupling constant restraint approach.[14-16] The use of time-averaged NOE's allows each spatially separated proton pair to be allotted a memory of its history during molecular dynamics simulation. Distance restraints are then required only to be satisfied over the course of a single time-averaged trajectory. This approach works well for flexible side chains on a rigid protein.[15] Ernst and

[†] This work was supported by the Protein Engineering Network of Centres of Excellence and funded by the goverment of Canada. Preliminary results were presented at the Keystone Symposium in Taos, NM, on March 11, 1993. J.W. is supported by a PENCE postdoctoral fellowship.

[‡] Abbreviations: IRMA, iterated relaxation matrix approach; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; H-bond, hydrogen bond; NOESY, nuclear Overhauser enhancement spectroscopy; DQF-COSY, double-quantum filtered correlated spectroscopy; TOCSY, total correlation spectroscopy; Mcpa, mercaptopropionic acid.

* Author to whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts*, August 1, 1995.

(1) Bax, A. D. *Ann. Rev. Biochem.* **1989**, *58*, 223–256.
(2) Wüthrich, K. *Methods Enzymol.* **1989**, *177* (Part B), 125–131.
(3) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.
(4) Gao, Y.; Veitch, N. C.; Williams, R. J. P. *J. Biomol NMR* **1991**, *1*, 457–471.
(5) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436–9444.
(6) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311–333.
(7) Clore, G. M.; Gronenborn, A. M. *Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 4479–4564.
(8) Boelens, R.; Koning, T. M. G.; Kaptein, R. J. *Mol. Struct.* **1988**, *173*, 299–310.
(9) Boelens, R.; Koning, T. M. G.; van der Marel, G. A; Van Boom, H.; Kaptein, R. *J. Magn. Reson.* **1989**, *82*, 290–299.
(10) Kessler, H.; Griesinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393–3396.
(11) Inman, W.; Crews, P. *J. Am. Chem. Soc.* **1989**, *111*, 2822–2829.
(12) Schmidt, J. M.; Ohlenschlager, O.; Ruterjans, H.; Grzonka, Z.; Kojro, E.; Pavo, I.; Fahrenholz, F. *Eur. J. Biochem.* **1991**, *201*, 355–371.
(13) Wang, J.; Hodges, R. S.; Sykes, B. D. *Int. J. Pept. Protein Res.* **1995**, *45*, 471–481.
(14) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289–294.
(15) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223–235.
(16) Torda, A. E.; Brunne, R. H.; Huber, T.; Kessler, H. van *J. Biomol. NMR* **1993**, *3*, 55–66.

0002-7863/95/1517-8627$09.00/0

co-workers have developed an algorithm called MEDUSA and applied it to cyclic decapeptide antamanide.[17-18] This algorithm is used to generate conformers which partially fulfill the NOE distance restraints and have an energy below a set threshold. The conformers are then combined in pairs, triples, or larger clusters to fulfill all restraints. The algorithm CPA[19] creates a predefined collection of possible conformers by structure generation methods such as Monte Carlo methods, calculates the cross-relaxation rates for each conformer, and then uses a cluster analysis routine to fit and refine the conformer population on the basis of the experimental NOE data. The final step of the algorithm is significance testing to obtain the best fit to the experimental NOE data using the smallest number of significant conformers. This algorithm was applied to the nucleoside derivative 2',3'-isopropylideneinosine.[19] The approach taken by Kessler and co-workers starts with a large ensemble of possible structures and applies the distance restraints as averages over multiple copies of the molecule.[20] Recently, Cicero et al.[21] described a method called NAMFIS which starts from an ensemble of structures generated by MEDUSA that encompass the possible conformational states compatible with the experimental data, then analyzes the population of these states such that they best reproduce the experimental data.

The goal of all NMR techniques for dealing with the multiple conformations of flexible peptides and proteins is to determine the full ensemble of structures whose averaged structural properties best fit the experimental NMR data. We wanted a procedure that is easy to apply and utilizes the iterative aspects of approaches such as IRMA which are useful in refining structures. The goal was to generate an ensemble of structures which sampled the widest possible conformational space and for which calculated averaged NOE's best fit the experimental data. In the present study, a computational procedure (PEP-FLEX-II) is reported which combines simulated annealing to generate an ensemble of structures for a peptide, an ensemble full-relaxation matrix analysis[22] to compare the ensemble averaged structural data with the experimental data, and an iterative NOE restraint modification approach to allow the molecule to sample more conformational space in regions of the structure for which the experimental data are not satisfied. We describe how this hybrid method was applied to desmopressin, a nine-residue peptide (Mcpa-Tyr-Phe-Gln-Asn-Cys-Pro-d-Arg-Gly-NH_2) with strong antiduiretic and antibleeding activities, in which the first six residues form a disulfide-bonded loop. Previous studies suggested that this peptide is very flexible in aqueous solution and exhibits an ensemble of conformations undergoing fast exchange on the NMR time scale.[13]

Theory

The full relaxation matrix approach can be extended to multiple-component systems. If a flexible peptide has multiple conformations and these conformers are in fast exchange on the NMR chemical shift time scale, only one set of signals can be detected. Under these conditions, the time evolution of the longitudinal magnetization for a homonuclear N-spin system takes the following form:[23]

$$d\mathbf{M}/d\tau_m = -\mathscr{R}\mathbf{M} \tag{1}$$

where

$$\mathscr{R} = \sum f_i \mathbf{W}_i$$
$$\mathbf{M} = \sum \mathbf{m}_i \tag{2}$$

For each conformer i, $\mathbf{W}_i$ is a N-dimensional relaxation matrix, $\mathbf{m}_i$ is the magnetization vector, and f_i is the fractional population. τ_m is the mixing time for the noesy experiment. The diagonal elements of $\mathbf{W}_i$ have the form

$$W^{kk}{}_i = (1/10)\sum_{l \neq k}[\gamma^4\hbar^2/(r_i^{kl})^6][J_0{}^i(\omega) + 3J_1{}^i(\omega) + 6J_2{}^i(\omega)] \tag{3}$$

and the off-diagonal elements have the form

$$W^{kl}{}_i = (1/10)[\gamma^4\hbar^2/(r_i^{kl})^6][6J_2{}^i(\omega) - J_0{}^i(\omega)] \tag{4}$$

where r_i^{kl} is the distance between nucleus k and l for conformer i and $J_p{}^i(\omega)$ are the spectral density functions:[24,25]

$$J_p{}^i(\omega) = \tau_{ci}/[1 + (p\omega_o\tau_{ci})^2] \tag{5}$$

where τ_{ci} is the correlation time for isotropic tumbling for conformer i. Equation 1 can be solved as[24,26]

$$\begin{aligned}\mathbf{M}(\tau_m) &= \exp[-\mathscr{R}\tau_m]\mathbf{M}(0)\\ &= \chi \exp(-\lambda\tau_m)\chi^{-1}\mathbf{M}(0)\\ &= \mathbf{a}(\tau_m)\mathbf{M}(0)\end{aligned} \tag{6}$$

where $\mathbf{M}(0)$ is the magnetization at time 0, χ is the matrix of eigenvectors of the averaged relaxation matrix $\mathscr{R}$, λ is its diagonal matrix of eigenvalues, and $\mathbf{a}(\tau_m)$ is the matrix of mixing coefficients which are proportional to the measured NOE intensities. In the limit that all conformers have the same correlation time τ_c and the exchange rate is slow on the τ_c time scale but fast on the chemical shift time scale[27]

$$\mathbf{I}(\text{expt}) = \sum f_i \cdot \mathbf{I}_i \tag{7}$$

where $\mathbf{I}_i$ is the NOE intensity for conformer i which is in proportional to $(1/r_i^{kl})^6$.

For any set of multiple conformations, the above theory allows us to simulate the experimental NOE's. The rms deviation between calculated and experimental NOE's can be obtained using the following equation:[28]

$$\text{rmsd} = \{[\sum_q(\text{NOE}_q{}^{\text{obsd}} - \text{NOE}_q{}^{\text{calcd}})^2]/[\sum_q(\text{NOE}_q{}^{\text{obsd}})^2 + \sum_q(\text{NOE}_q{}^{\text{calcd}})^2]\}^{1/2} \tag{8}$$

(17) Brüschweiler, R.; Blackledge, M.; Emst, R. R. *J. Biomol. NMR* **1991**, *1*, 3-11.
(18) Blckledge, M. J.; Bruschweiler, R.; Griesinger, C.; Schmidt, J. M.; Ping Xu; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960-10974.
(19) Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493-9499.
(20) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Amer. Chem. Soc.* **1994**, *116*, 1042-1049.
(21) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Amer. Chem. Soc.* **1995**, *117*, 1027-1033.
(22) Wang, J.; Sönnichsen, F. D.; Boyko, R.; Hodges, R. S.; Sykes, B. D. *Techniques in Protein Chemistry;* Angeletti, R. H., Ed.; Academic Press Inc.: San Diego, CA, 1993; Vol. IV, pp 569-576.
(23) Landy, S. B.; Rao, B. D. N. *J. Magn. Reson.* **1989**, *81*, 371-377.
(24) Campbell. A. C.; Sykes, B. D. *J. Magn. Reson.* 1991, *93*, 77-92.
(25) Campbell. A. C.; Sykes, B. D. *Ann. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 99-122.
(26) Bull, T. E. *J. Magn. Reson.* **1987**, *72*, 397-413.
(27) Neuhaus, D.; Williamson, M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis;* VCH Publishers, Inc.: New York, 1989; pp 141-182.
(28) Withka, J. M.; Srinivasan, J.; Bolton, P. H. *J. Magn. Reson.* **1992**, *98*, 611-617.

where the sum is over all NOE's. The rms deviation gives an indication of how well the ensemble of conformers fit the experimental NOE data. The difference between calculated and experimental values for any particular NOE can also be calculated as

$$\mathrm{diff}_q^{\mathrm{NOE}} = \{(\mathrm{NOE}_q^{\mathrm{obsd}} - \mathrm{NOE}_q^{\mathrm{calcd}})^2/(\mathrm{NOE}_q^{\mathrm{obsd}})^2 + (\mathrm{NOE}_q^{\mathrm{calcd}})^2\}^{1/2} \quad (9)$$

This measure can be used to indicate which proton pairs in the multiple conformations correspond best to the experimental NOE's. This provides one of the guidelines for the modification of the distance restraints during the PEPFLEX-II procedure.

In order to estimate the effect of experimental error on the calculated rmsd, as well as to evaluate the statistical significance of the improvement of the NOE rmsd during the PEPFLEX-II calculation, the standard deviations of the NOE rmsd's are also calculated using the following:

$$\mathrm{STD}_{\mathrm{NOE}} = (|\mathrm{rmsd}(\mathrm{NOE}^+) - \mathrm{rmsd}(\mathrm{NOE})| + |\mathrm{rmsd}(\mathrm{NOE}^-) - \mathrm{rmsd}(\mathrm{NOE})|)/2.0 \quad (10)$$

where rmsd(NOE^+) and rmsd(NOE^-) are calculated according to eq 8, and NOE^+ and NOE^- are calculated as

$$\mathrm{NOE}^+_q = \mathrm{NOE}_q^{\mathrm{obsd}} + \mathrm{error}_q$$

$$\mathrm{NOE}^-_q = \mathrm{NOE}_q^{\mathrm{obsd}} - \mathrm{error}_q \quad (11)$$

the error_q in eq 11 is the estimated experimental error for each particular NOE_q.

Computational Algorithm

PEPFLEX-II is an extension of PEPFLEX-I which was originally written to assist the evaluation of ensembles of structures by comparison of the calculated ensemble averaged NOE's with the experimental NOE's.[22] It differs from PEPFLEX-I in that it is a structure generation procedure capable of dealing with multiple-conformational systems. PEPFLEX-II is based upon the premise that an experimental NOE cross peak is an average distance representing the population-averaged distance between proton pairs among several possible conformations. The philosophy of PEPFLEX-II is that the generated structures should sample a maximum conformational space while the differences between the calculated and experimental NOE's are minimized. Therefore, the convergence of the procedure is defined by two criteria which are the requirement that the NOE rms deviation is at a minimum, while the structural rms deviation is at a maximum.

The PEPFLEX-II algorithm can be summarized as follows:

(1) Average distance restraints generated from experimental NOE intensities with upper and lower boundaries are used as the initial distance restraints for the first iteration.

(2) Simulated annealing is used with the initial distance restraints to generate a large number (e.g., 50) of structures. Any structure, such as crystal structure, can be used as a starting structure. During the first iteration, the initial distance restraints are used; in all subsequent iterations, the most recently updated distance restraints are used as input for the simulated annealing.

(3) Structures generated from simulated annealing are fed into the ensemble full relaxation matrix calculation (PEPFLEX-I) to calculate the NOE intensities. This procedure calculates the NOE rms deviation between calculated and experimental NOE's for all proton pairs using eq 8, as well as the difference of the calculated and experimental NOE for each individual proton pair using eq 9; the procedure also calculates the atomic rms deviation (structural RMSD) of all generated structures; finally a routine of the procedure is executed to find out the violation of the NOE distance restraints in terms of both the size of violation and the number of structures violated for each particular NOE in the total generated structures.

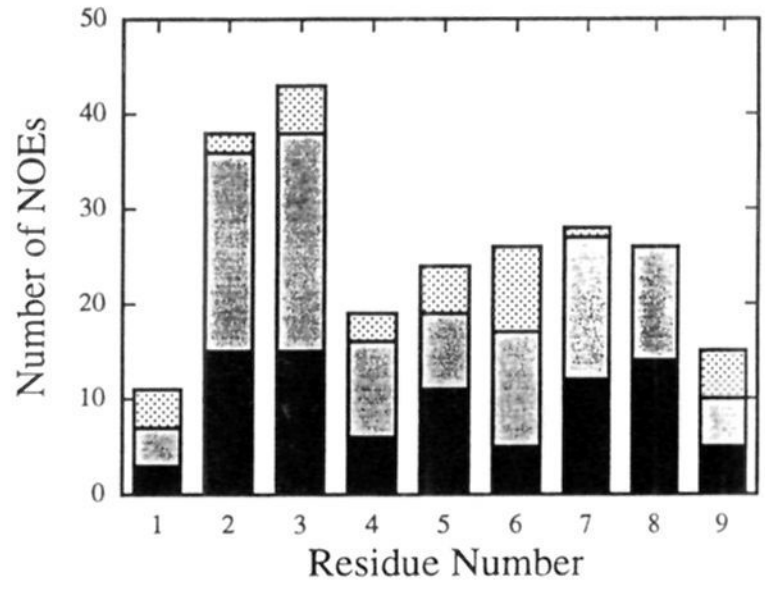

Figure 1. Number of NOE's for each amino acid residue. The black columns represent the distribution of the intraresidue NOE's or distance restraints (i, i) per residue; the dark shaded columns show the distribution of the sequential NOE's or distance restraints (i, $i + 1$) per residue; the light shaded columns stand for the distribution of the medium- and long-range NOE's or distance restraints ($\geq i$, $i + 2$) per residue. The NOESY experiments were done in 20%TFE aqueous solution at 5 °C, pH 5.0. The distance restraints were generated from a NOESY spectrum with a mixing time of 300 ms.

(4) The NOE rms deviation and the structural rms deviation from the present iteration are checked for convergence. If the convergence criteria are not satisfied, the procedure modifies the distance restraints automatically (see below), then returns to step 2, and repeats the procedure until convergence is achieved.

Two criteria are used to monitor which distance restraints should be modified during each iteration. One is the difference between calculated and experimental NOE intensities from eq 9. The other is the number and the size of violations of the distance restraints. In other words, if the difference between the calculated and experimental NOE intensities or the size and/or number of the NOE violations are larger than the allowed violation limit, the distance restraints are modified. One threshold is used for the detection of the distance restraint violations, while a second threshold is used for detecting the difference between the calculated and experimental NOE's.

For any violations, the distance restraints are automatically modified by adding the average distance violation to the upper boundary (shown in eq 10). Here the violation correction factor is the ratio of the number of violated structures to the total number of generated structures in this iteration.

$$d^+(i,j) = d^+(i,j) + (\mathrm{num}_{\mathrm{viol}}/\mathrm{num}_{\mathrm{total}})d_{\mathrm{viol}}(\mathrm{av}) \quad (12)$$

where d^+ is the upper boundary of the distance restraint, $\mathrm{num}_{\mathrm{viol}}$ is the number of structures violated to a particular NOE, $\mathrm{num}_{\mathrm{total}}$ is the number of the total generated structures, and $d_{\mathrm{viol}}(\mathrm{av})$ is the averaged distance restraint violation. In cases where the calculated NOE intensity is smaller than the experimental NOE with no NOE violation detected, this distance restraint is not modified.

The structural rms deviation is defined by the averaged pairwise rms deviation of structures. We have found that this is a reasonable measurement of the conformational space

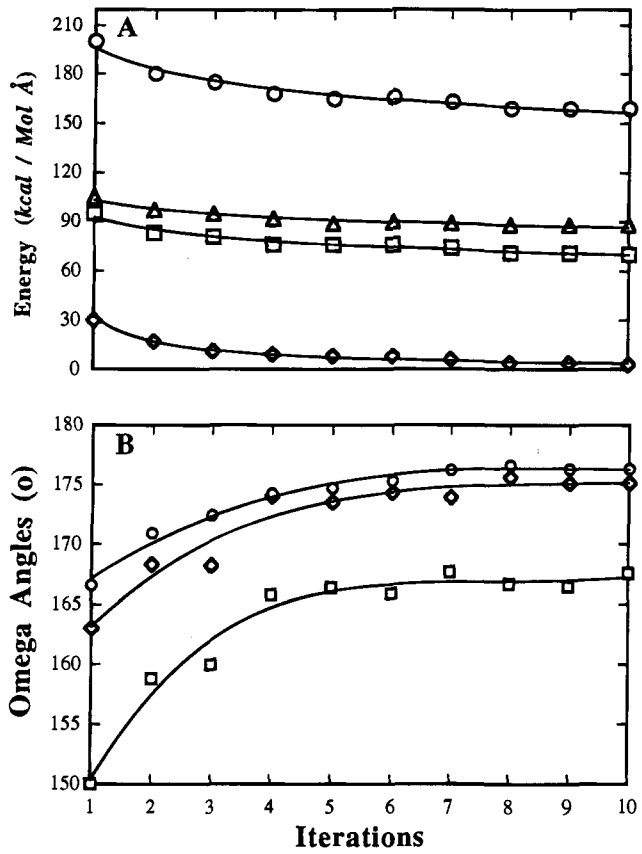

Figure 2. Energy terms (panel A) and average ω dihedral angles (panel B) in different iterations. In panel A, open circles are the total energies in different iterations; open triangles are the VDW energies in different iterations; open boxes are the covalent energies, including bond energy, bond angle energy, and out of plane energy, in different iterations; open diamonds are the NOE energies in different iterations. In panel B, open circles stand for the average ω dihedral angles of residue Cys1; open boxes stand for the average ω dihedral angles of residue Phe3; open diamonds stand for the average ω dihedral angles of residue Gln4. The average ω torsion angles are obtained by averaging the torsion angles from all the structures generated in that iteration.

occupied by the ensemble of structures. Once the PEPFLEX-II procedure has converged, the structures are analyzed for classification and identification of different structure families. A cluster analysis routine is included in PEPFLEX-II for this purpose which is based on pairwise structural rmsd's. The results of the cluster analysis are further tested by a dihedral angle analysis routine in which ϕ and ψ variations are calculated for each family of structures.

Materials and Methods

Materials and Sample Preparations. Desmopressin was synthesized by solid-phase peptide synthesis method and was purified by reverse-phase HPLC. Trifluoroethanol-d_3 (TFE) was obtained from Cambridge Isotope Laboratories. NMR samples (2 mM) were prepared in 20% D_2O/80% H_2O and 20% TFE/80% H_2O. The pH values for all NMR samples were adjusted to a value of 5.0 ± 0.1 by small aliquots of NaOH and HCl.

NMR Experiments. NMR experiments were performed on a Varian VXR-500 NMR spectrometer at 5 °C. The methyl signal of 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was used as the chemical shift reference at 0.0 ppm. Two-dimensional double-quantum filtered correlated spectroscopy (DQF-COSY),[29] nuclear Overhauser enhancement spectroscopy (NOESY),[30] and total correlation spectroscopy (TOCSY)[31] were acquired according to standard procedures at 5 °C. The spectral width was 6000 Hz in both dimensions with 1024 complex points in the t_2 dimension and 512−600 FIDs in the t_1 dimension. Typically 32−48 scans were collected for each experiment.

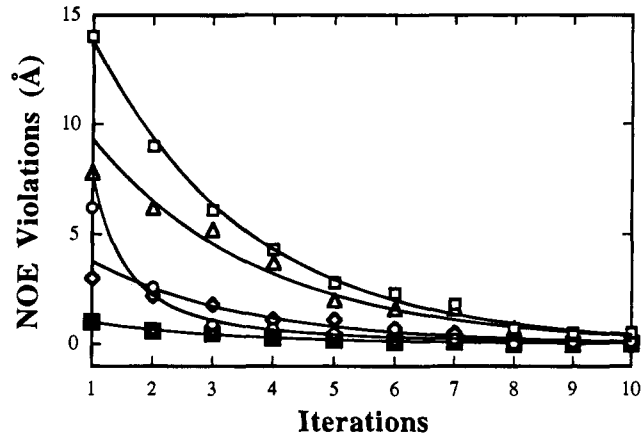

Figure 3. NOE violations *vs* different iterations. Open boxes are the total NOE violations per residue in different iterations; open triangles are the interresidue NOE violations per residue in different iterations; open circles are the intraresidue NOE violations per residue in different iterations; open diamonds are the medium-range NOE violations per residue in different iterations; black boxes are the long-range NOE violations per residue in different iterations.

Distance Constraints. The peak intensities were integrated from a NOESY spectrum at 300 ms mixing time. The volume integrals were translated into distances using the average 2.55 Å for $\beta-\delta$ proton pairs in Tyr2 and Phe3 as a calibration. Distances were grouped into strong, medium, and weak which corresponded to the upper boundaries of distance restraints of 2.7, 3.3, and 5.0 Å. The lower boundaries of distance restraints were set to 1.8 Å for strong and medium NOE's and to 2.3 Å for weak NOE's. The appropriate distance corrections were made to aromatic ring protons and the pseudo-atoms of nonstereospecifically assigned methylene protons to allow for centroid averaging.[3]

Structure Calculations. PEPFLEX-II is used for the structural calculation. The restrained dynamic simulated annealing approach was based on that by Nilges et al.[32] with some modifications. It was achieved using Discover (Biosym, San Diego, CA) running on a Silicon Graphics Indigo computer. In the restrained simulated annealing, a small NOE force constant was used with its maximum value of 10 kcal/(mol Å^2). Also, after the temperature was cooled to 300 K, another 10 ps of restrained molecular dynamics was carried out, then followed by the restrained energy minimizations. A cutoff distance of 8.0 Å was used in the calculation with the chiral restraints to retain the right chiralities for all residues. Ω torsion angle restraints were also used to constrain all the peptide bonds to be of trans geometry. Meanwhile both charge−charge interaction and cross term interaction (CVFF forcefield) were switched off during the simulation. In each PEPFLEX-II iteration, a total of 50 structures were generated and analyzed. A combination of a C-program with a set of BIOSYM INSIGHTII macros was used to connect the simulated annealing with the evaluation of the generated structures (PEPFLEX-I).[22] This contains the structural rms deviation calculation, the detection of NOE violations, the automatic adjustment of the distance restraints, the cluster analysis, and the procedure of checking if the iteration is converged.

Results and Discussion

In a previous study, we reported that desmopressin in aqueous solution and in solutions with less than 20% TFE exhibited similar conformations.[13] The conformations are flexible and in fast exchange on the NMR time scale. However, a significant population of a conformation which contains two flexible fused

(29) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(30) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546−4553.
(31) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 393−402.
(32) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 129−324.

Table 1. Number and Size (Å) of NOE Violations in Different Iterations

	iteration									
	1	2	3	4	5	6	7	8	9	10
no. of viol[a]	66	38	31	29	31	32	22	12	10	13
viol/struct[b]	30 ± 5	9 ± 2	4.3 ± 3	3.3 ± 3	2.8 ± 2	2.6 ± 2	1.8 ± 1	0.7 ± 1	0.5 ± 1	0.5 ± 1
$viol_{max}$[c]	0.96	0.89	0.61	0.64	0.52	0.46	0.32	0.24	0.19	0.20
$viol_{SD}$[d]	±0.12	±0.06	±0.25	±0.38	±0.19	±0.39	±0.26	±0.12	±0.10	±0.10
viol (>0.25 Å)[e]	20	12	8	5	4	4	2	0	0	0

[a] Total number of NOE's which are violated (>0.2 Å). [b] Number of NOE violations per structure (>0.2 Å). [c] Maximum average NOE violations (Å) among all structures. [d] Standard deviation of the maximum average NOE violat (Å). [e] Number of the NOEs whose violations are >0.25 Å.

Table 2. Statistics of Fit of the Calculated NOE's to the Experimental Data for Desmopressin in 20% TFE/80% H_2O at 5 °C, pH 5.0

				PEPFLEX iterations									
	extend[a]	EM1[a]	EM2[a]	1	2	3	4	5	6	7	8	9	10
rmsd	0.913	0.259	0.253	0.242	0.220	0.212	0.202	0.190	0.177	0.171	0.168	0.165	0.167
±SD[b]	0.000	0.001	0.001	0.010	0.011	0.011	0.012	0.012	0.013	0.013	0.014	0.014	0.014

[a] Extend is the starting structure which is an extended structure; EM1 is a structure obtained by unrestrainted energy minimization of the starting structure; EM2 is a structure obtained by restrained energy minimization of the starting structure with the NOE distance restraints from the last iteration. [b] Standard deviations were obtained (see methods) by assuming the experimental NOE measurement errors were as follows: ±50% for weak cross peaks; ±15% for medium cross peaks; ±10% for strong cross peaks.

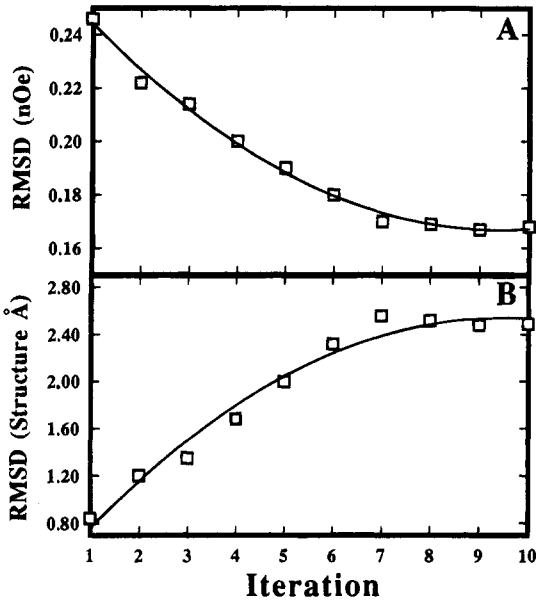

Figure 4. Rmsd's between the calculated and experimental NOE's (panel A) and the structural rmsd's in different iterations (panel B). In panel A, the calculated NOE's are obtained by the ensemble full relaxation analysis (PEPFLEX-I); in panel B, the structural rmsd's are the average value of the backbone pairwise rmsd matrix which shows a better measure of the conformational space the generated structures covered.

β-turns with weak H-bonds can be observed. One turn is in Phe3-Gln4-Asn5-Cys6 and the other is in Cys6-Pro7-d-Arg8-Gly9.[13] The multiple conformations of desmopressin provide a good model system to test the PEPFLEX-II procedure.

Distance Restraints. A total of 158 distance restraints were generated from measured NOE data, among which there were 72 interresidue distance restraints including 55 sequential distance restraints and 14 medium-ranged distance restraints. Three distance restraints were also found between residue 1 and residue 6 due to the disulfide bond between these two residues. Figure 1 shows the distribution of the NOE's along the sequence. The average number of distance restraints per residue is 17.5, which is relatively large for a small flexible peptide. However, there are only 1.9 medium and long-range restraints per residue, which is a small number compared with that of rigid peptides or proteins, indicating the flexibility of the peptide. Several distance restraints were found between residues 6 and 9, implying that the three-residue tail is not independent from the disulfide-bonded loop and that the interaction between the tail and the loop may play an important role in stabilizing the conformations of the peptide. This result is different from that of Zieger and co-workers for the two homologues of [Lys^8]-vasopressin in dimethyl sulfoxide (DMSO) which indicated that the loop showed "rigid" conformation with a reverse γ-turn containing residues Phe3-Gln4-Asn5; the tail was in rapid motion and had no preferred conformation.[33]

PEPFLEX-II Calculations. A total of 10 iterations were carried out. The last three iterations satisfied the two convergence criteria. In addition to testing whether the generated ensemble of conformations better fit the experimental NOE data, the structures should also satisfy current criteria for good structures in terms of ideal bond geometry, good van der Waals contacts, and ϕ and ψ dihedral angles in the allowed regions of the Ramachandran plot; and also these structures should cover the largest possible conformational space. Figure 2 shows the energy terms (panel A) and the averaged ω peptide bond dihedral angles (panel B) for each iteration. Panel A indicates that all of the energy terms, including van der Waals (VDW) energy, significantly decrease toward a minimum. Panel B shows that the ω dihedral angles of generated structures have much larger variations from ideal bond geometry in the earlier iterations but a good bond geometry in the final iteration. These data, together with the rms deviations of the bond lengths and bond angles (data not shown), indicate that the bond geometry is good for the generated structures.

Figure 3 shows that the NOE violations per residue reach a minimum in the last three iterations, indicating the generated structures in these iterations better fit the experimental NOE distance restraints. Table 1 lists the total number of violations of the NOE distance restraints and the average maximum size of NOE violations in different iterations. This shows that the NOE violations are much less in the final iteration than those in the earlier iterations. The structures also violated the distance restraints much more frequently in the earlier iterations than those in the final iteration.

Panel A in Figure 4 displays the rms deviations between calculated and experimental NOE's for each iteration. It demonstrates that the structures generated by later iterations not only satisfy the current criteria for good structures but also better

(33) Zieger, G.; Andreae, F.; Sterk, H. *Magn. Reson. Chem.* **1991**, *29*, 409–417.

Table 3. Cluster Analysis for Each Different Iteration

	iteration									
	1	2	3	4	5	6	7	8	9	10
no. of struct[a]	50	50	50	50	50	50	50	50	50	50
no. of bad[b]	4	3	3	3	2	2	2	2	2	2
no. of families[c]	1	2	2	3	3	4	4	5	5	5
ϱ (family 1)[d]	100%	90%	87%	79%	71%	59%	50%	47%	47%	46%
ϱ (family 2)		10%	13%	15%	57%	20%	21%	21%	22%	24%
ϱ (family 3)				6%	12%	15%	17%	14%	14%	14%
ϱ (family 4)						6%	12%	10%	10%	10%
ϱ (family 5)								8%	7%	6%
rmsd(f1)[e]	0.5 ± 0.1	0.7 ± 0.2	0.5 ± 0.2	0.5 ± 0.2	0.8 ± 0.2	0.7 ± 0.2	0.8 ± 0.2	0.8 ± 0.2	0.8 ± 0.2	0.8 ± 0.2
rmsd(f2)		0.7 ± 0.1	1.0 ± 0.1	0.8 ± 0.2	1.0 ± 0.3	0.8 ± 0.3	1.0 ± 0.3	0.8 ± 0.2	0.7 ± 0.1	0.7 ± 0.1
rmsd(f3)				0.5 ± 0.1	0.6 ± 0.2	0.9 ± 0.3	0.8 ± 0.2	0.3 ± 0.1	0.4 ± 0.1	0.5 ± 0.1
rmsd(f4)						0.5 ± 0.1	0.7 ± 0.3	0.4 ± 0.2	0.4 ± 0.1	0.5 ± 0.2
rmsd(f5)								0.9 ± 0.2	0.7 ± 0.2	0.8 ± 0.2

[a] Number of structures generated in each iteration. [b] Number of the bad structures which were eliminated from further discussion (see text). [c] Number of the structural families generated by cluster analysis. [d] ϱ (family 1) is the population of family 1. [e] rmsd(f1) is the structural rms deviation for family 1.

Table 4. Pairwise Structural Rms Deviations (Å) between the Different Families[a,b]

	family				
	1	2	3	4	5
		Backbone Atoms			
1		2.50	3.62	3.29	3.04
2	1.79		3.67	3.24	2.32
3	0.77	1.56		2.17	3.38
4	1.96	0.59	1.59		2.85
5	1.95	1.57	1.76	1.73	
		All Heavy Atoms			
1		4.47	4.28	4.26	5.73
2	3.56		4.93	3.87	4.56
3	2.01	3.46		3.78	5.34
4	3.66	0.81	3.53		5.92
5	3.30	3.43	3.10	3.47	

[a] The structural rms deviations in the table were obtained by superimposing the average structures for two different families. The average structure for each family from simulated annealing is followed by restrained energy minimization with NOE restraints. [b] The upper triangles contain the rms deviations for superimposing the whole peptide (residues 1−9); the lower triangles contain the rms deviations for superimposing the six-residue disulfide-bonded loop (residues 1−6).

fit experimental data. However, whether the improvement of the NOE fit is statistically significant remains to be verified. Table 2 shows the NOE rms deviations of different structures, e.g. the starting structure and the structures from different iterations, with their standard deviations. The standard deviations were obtained from the error estimation of our experimental NOE measurements. In Table 2, we assumed that the error in the NOE measurements were 50%, 15%, and 10% for weak, medium, and strong NOE cross peaks, respectively, which is a reasonable estimation of the NOE experimental error. Table 2 shows that the standard deviations are less than 10% of NOE rms deviations for all iterations. Therefore, we conclude that PEPFLEX-II calculation leads to a statistically significant improvement of the NOE fit for the generated structures.

Once the procedure found the best fit of experimental data, we expected that the generated structures should cover the maximum reasonable conformational space. Panel B of Figure 4 shows the structural rms deviation for each iterations. The backbone structural rms deviation of the first iteration was 0.84 Å, suggesting that these 50 structures can be grouped into one single structure family. On the other hand, the structures

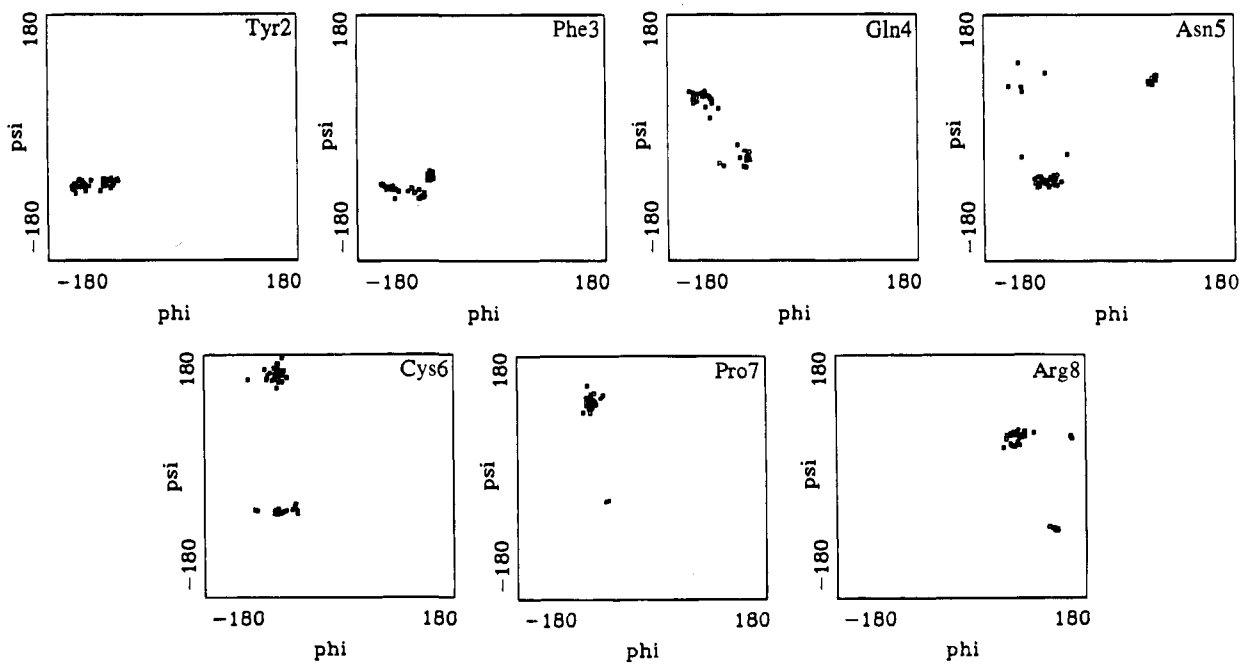

Figure 5. Ranmachandran plots of the total structures in the last iteration for each residue. Residue name and number are marked in each Ranmachandran plot. In each Ranmachandran plot, 48 structures are shown with the elimination of the two bad structures (see text).

generated in the final iteration reach a maximum backbone structural rms deviation (about 2.4 Å). Thus the ensemble of structures calculated by the PEPFLEX-II procedure cover a significantly larger conformational space where the calculated NOE's best fit to the experimental data. A similar result was also obtained by the James group for [d(GTATAATG]$_2$-[d(GTATAATG], which showed an structural rms deviation of 0.77 Å from standard MD simulation, whereas MD-tar simulation gave a much larger structural rmsd at 2.27 Å.[34]

It is interesting to ask if the generated structures in the final iteration can be clustered into different well-defined families and whether they are in the allowed Ramachandran ϕ and ψ regions. The results of cluster analysis are shown in Table 3. One structure family was found in the first iteration. In the next several iterations, more structural families were generated, and a total of five families are found in the last three iterations. It is worth noting that the results of the last three iterations are nearly identical in terms of the number of structural families, as well as the average structure and the population of each family. The population of each family is estimated by the number of structures in the family divided by the total number of structures in all families. Several bad structures in each iteration, with significant higher total potential energy as well as van der Waals energy, were eliminated from the discussion. The backbone structural rms deviation in each family is less than 0.8 Å, indicating well-defined structures in all five families. Table 4 shows the structural rmsd differences between the families. In general, the data indicate that these five family structures are significant different. However, the backbone in the disulfide-bonded loop of families 1 and 3 as well as the backbone in the disulfide-bonded loop of families 2 and 4 showed similar conformation with the backbone structural rms deviations less than 0.8 Å. Interestingly, for families 1 and 3, only the backbone in the disulfide-bonded loop exhibits a similar conformation (backbone structural rms deviation, 0.77 Å); the side chain in the loop between these two families shows a different orientation (structural rms deviation for all heavy atoms, 2.01 Å). On the other hand, not only the backbone atoms but also all heavy atoms in the disulfide-bonded loop in families 2 and 4 showed similar conformation (backbone rms deviation, 0.59 Å; rms deviation for heavy atoms, 0.81 Å). Only the C-terminal three-residue tails in these two families orientate differently. Figure 5 shows the ϕ and ψ Ramachandran plots of different residues from structures generated in the last iteration. The ϕ and ψ dihedral angles of residue 2 to residue 7 are all in the allowed regions. For residue 8, some of its ϕ and ψ dihedral angles are not in the normal allowed regions which is reasonable since this residue is a D-amino acid.

Solution Structure and Flexibility of Desmopressin. Figure 6 shows the stereoview of the average structures from the five family structures obtained from the last PEPFLEX-II calculation. Table 5 lists all the backbone dihedral angles of the average structures in these five families. The dihedral angles of residues PRO7 and ARG8 shown in Table 5 indicate that there is a type II β-turn in the fragment of residue 6 to residue 9 in all five families. This type II β-turn is flexible, and no hydrogen bond was found in the turn. The flexible type II β-turn in all five families generated by the PEPFLEX-II is consistent with our previously discussion based on the experimental data.[13] In the disulfide bond loop, a distorted type I β-turn in residue 3 to residue 6 was only found in families 2 and 4 with a total population about 31%. No hydrogen bond was found in this turn. The experimental data like amide proton temperature coefficients and exchange rates for residues 3, 6, and 9 (temperature coefficients: 4.0 ppb/°C for residue3, 3.9 ppb/°C for residues 6, and 4.9 ppb/°C for residue 9) of the peptide support the flexibility of these two β-turns.

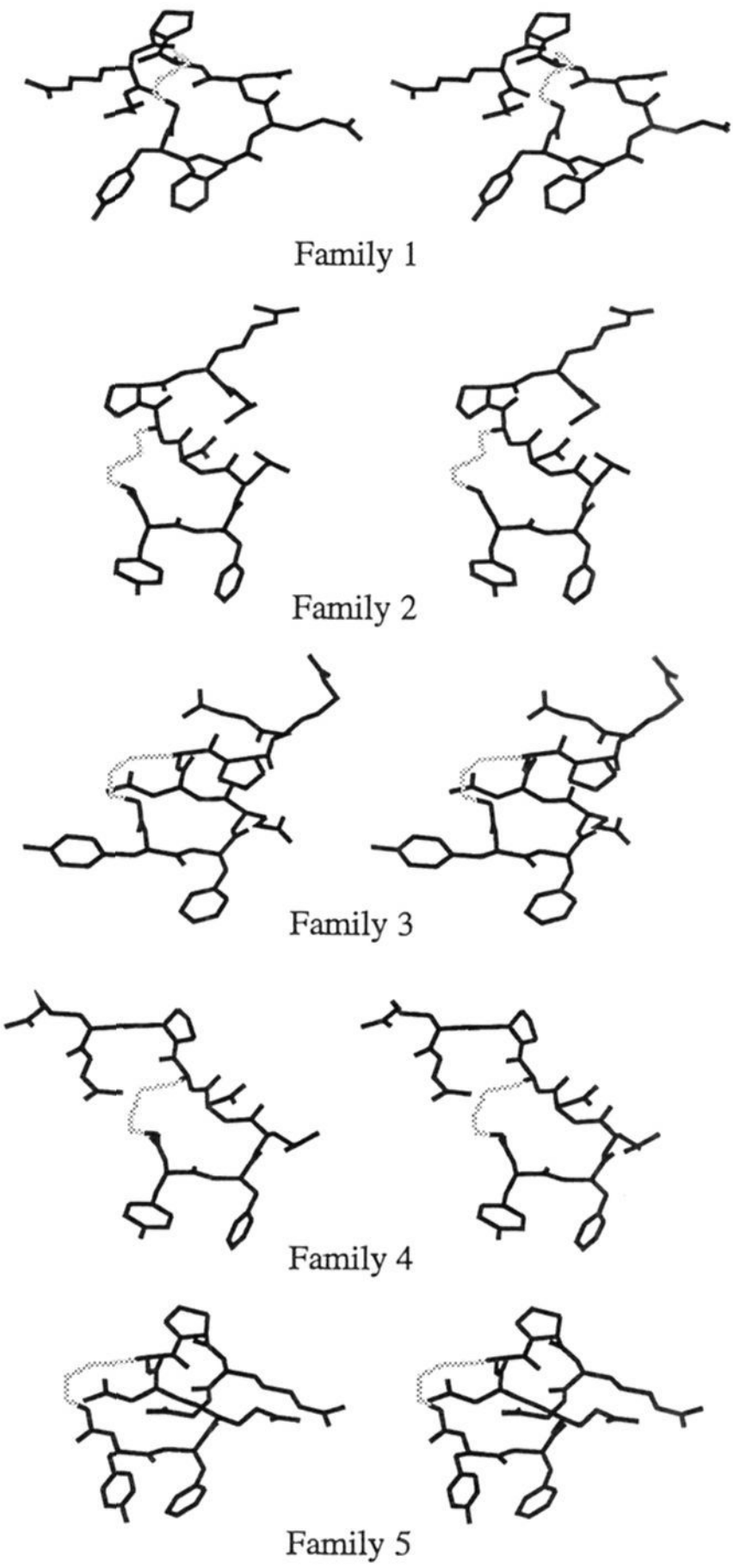

Figure 6. Stereoview of the average structures from the five families generated from the last iteration. The disulfide bonds are shown in light color. The average structure in each family shown in the figure is the average structure after restrained energy minimization.

In order to understand the structural differences between the families, the average structures between families 1 and 3 (Figure 7, panel A), as well as the average structures between families 2 and 4 (Figure 7, panel B), were superimposed. In panel A, backbone atoms of the disulfide-bonded loop of both families 1 and 3 show a similar conformation, but the side chain atoms of Tyr2, Phe3, and Gln4 orient differently, especially Tyr 2. Panel B suggests that, in families 2 and 4, the disulfide-bonded loop, both backbone and sidechain, shows the same conformation with a distorted type I β-turn in the fragment of residue 3

(34) Schmitz, U.; Kumar, A.; James, T. L. *J. Am. Chem. Soc.* **1992**, *114*, 10654−10656.

Table 5. Backbone Dihedral Angles of the Average Structures for the Five Families Generated by PEPFLEX-II[a]

	Mcpa1	Tyr2	Phe3	Gln4	Asn5	Cys6	Pro7	Arg8	Gly9
					Family 1				
ϕ		−128.0	−144.7	−109.9	−131.0	−86.8	−69.7	90.0	79.4
Ψ		−83.9	−72.2	60.7	−76.3	155.9	111.8	61.9	
Ω	179.7	180.0	178.0	179.6	179.1	−179.7	−178.2	−179.6	
					Family 2				
ϕ		−92.4	−81.3	−67.7	5.80	−66.8	−68.5	94.0	75.4
Ψ		−62.9	−60.0	−23.9	83.4	157.0	115.2	57.8	
Ω	179.7	−179.8	178.2	179.3	−179.3	−179.6	−178.3	179.9	
					Family 3				
ϕ		−93.8	−134.0	−116.4	−75.7	−70.5	−69.8	90.2	148.6
Ψ		−68.4	−72.9	20.7	−74.2	−55.4	105.1	63.2	
Ω	180.0	179.8	177.7	179.8	178.5	176.8	−176.1	−178.7	
					Family 4				
ϕ		−93.6	−74.9	−67.5	59.7	−38.8	−75.6	82.4	153.3
Ψ		−64.9	−63.3	−28.5	86.5	−54.1	103.9	62.3	
Ω	179.5	−179.6	178.2	179.3	−179.4	177.2	−177.0	−178.9	
					Family 5				
ϕ		−90.4	−94.1	−153.0	−129.4	−57.7	−73.9	87.3	149.6
Ψ		−67.7	−68.8	−17.7	112.6	−50.4	108.9	57.9	
Ω	179.9	178.7	179.9	177.5	178.1	177.9	176.8	178.5	

[a] The average structure is the average of all structures in a family, followed by restrainted energy minimization with NOE restraints.

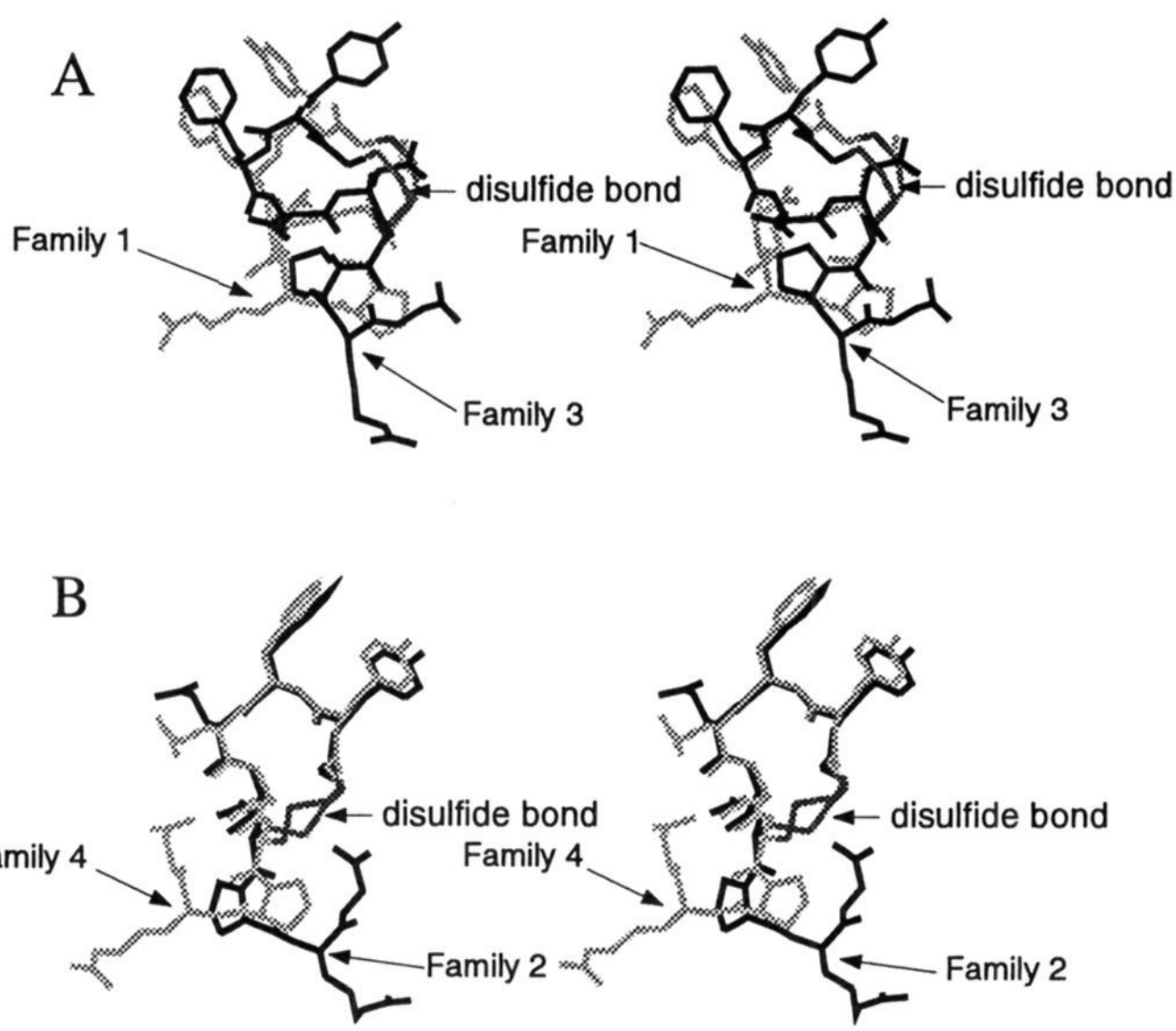

Figure 7. Stereoview of the average structures from family 1 and family 3 (panel A), as well as from family 2 and family 4 (panel B). In both panels, the disulfide bond is indicated and each average structure is marked by its family number.

to residue 6 (Table 5). By combining panel A and panel B, one can find that the disulfide bond in all four families is flexible and takes different conformations. The flexibility of the disulfide bond in the peptide (particularly the ψ dihedral angle of Cys6) causes the C-terminal three-residue tail to orient to the different direction in different families. The similar result about the flexibility of the disulfide bond in desmopressin was also found in the crystal structures of a similar hormone oxytocin,[35] which shows two different conformations in the disulfide-bond region of the peptide.

Acknowledgment. The authors wish to thank Leigh Willard, Tim Jellard, and Robert Boyko for computer assistance. We also wish to thank Drs. David S. Wishart, Frank D. Sönnichsen, Key-Sun Kim, and Krishina Rajarathnam for many helpful discussions and comments on earlier versions of the manuscript.

JA942968Z

(35) Wood, S. P.; Tickle, I. J.; Treharne, A. M.; Pitts, J. E.; Mascarenhas, Y.; Li, J. Y.; Husain, J.; Cooper, S.; Blundell, T. L.; Hruby, V. J.; Buku, A.; Fischman, A. J.; Wyssbrod, H. R. *Science* **1986**, *232*, 633−636.