Accounting for Conformational Variability in NMR Structure of Cyclopeptides: Ensemble Averaging of Interproton Distance and Coupling Constant Restraints

Philippe Cuniasse,*,† Isabelle Raynal,†,§ Athanosios Yiotakis,‡ and Vincent Dive†

Contribution from the Département d'Ingénierie et d'Etudes des Protéines D.S.V., CEA/Saclay, 91191 Gif sur Yvette Cedex, France, and Department of Organic Chemistry, Laboratory of Organic Chemistry, University of Athens, Panepistimiopolis Zografou, Greece

Received October 23, 1996[⊗]

Abstract: The application of an ensemble-averaging (EA) protocol to highlight conformational variability and to determine the interconverting conformations in NMR structure of cyclopeptides is described. Most of the NMR-based conformational studies of cyclopeptides reported in the literature rely on protocols that basically assume the existence of a single structure. This is sometimes referred as the one NOE (or ROE)/one distance hypothesis. In contrast, the EA protocol used in this work relies on a model that explicitly takes into account the averaging in NMR data and tests the significancy of the results which is very often disregarded in structure determination by NMR. This EA method was applied to the conformational analysis of the peptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) in DMSO by NMR. Qualitative analysis of the ROEs observed for this peptide indicates that it adopts the well-known double reverse turn structure. However, certain interproton distances derived from a set of ROESY experiments, as well as some coupling constants, are not compatible with the existence of a unique conformation but reflect the presence of several conformers in fast exchange on the NMR time scale. Therefore, structures consistent as ensemble with the NMR-derived restraints were determined using a restrained molecular-dynamics-based ensemble-averaging protocol which explicitly takes multiconformers into account and treats the restraints as ensemble-averaged quantities. The NMR-derived data used as input restraints in this EA protocol include the distance restraints (DR), the homonuclear coupling constants (*J*), and a large set of unambiguous antidistance restraints (ADR) that are generally disregarded in conformational analysis of cyclopeptides. The number of interconverting conformers was determined from the significance of the fit of the DR and ADR using the complete cross-validation method. The results shows that pairs of conformers give a satisfactory and significant fit of all NMR data. The conformational analysis of the interconverting partners reveals that the hexapeptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) exists in solution either as a βVIII-βII/$i\gamma$-βI or a βII-βII/βI-βI equilibrium.

Introduction

Cyclic peptides of four to ten residues have been extensively used in the field of biological chemistry to probe the conformational requirements for biological activity.[1] Cyclic peptides are used since they are thought to be rigid, or at least to adopt a small number of conformations. These cyclic peptides have a reduced entropy contribution upon binding to their targets as compared to their linear counterparts. Thus, if the conformational restriction brought by the cyclization does not introduce additional unfavorable interactions, it should increase the free energy of binding of the peptide to its target. The solution conformation of the constrained peptide is used to define the conformational requirements for its activity. However, to be relevant, such a conformation−activity relationship requires proper determination of the conformation(s) that the peptide adopts in solution.

Solution structures of peptides are usually determined by NMR. Interproton distances are derived from cross-relaxation experiments (NOESY (nuclear Overhauser spectroscopy) or ROESY (rotating frame Overhauser spectroscopy)). Torsional angles are estimated from *J*-coupling constants using empirical Karplus-type relations.[2] These geometrical restraints are used to define the conformation(s) of the molecule under study by means of various protocols including distance geometry (DG)[3,4] and/or simulated annealing (SA).[5] For the most part, these protocols basically assume the existence of a unique conformation. However, the NMR data are collected as time- and ensemble-averaged quantities. In the case where several conformers are in fast interconversion on the NMR time scale, both NOE/ROE and *J*-couplings correspond to population-weighted averaged values. It should be noticed that the derivation of structural parameters from the related population-weighted NMR data leads to unrealistic conformations or at least to misconclusions.[6,7]

Ensemble-averaging (EA) represents an attractive way to take account of the conformational variability in NMR-based structure determination.[8−12] The EA method consists in averaging the NMR parameters over a set of conformations that exist

* Author to whom correspondence should be addressed.
† CEA/Saclay.
‡ University of Athens.
§ Present address: Guerbet G. C. A. B P 50400, 95943 Roissy CDG Cedex France.
⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1997.
(1) Rizo, J.; Gierash, L. M. *Annu. Rev. Biochem*. **1992**, *61*, 387−418.
(2) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870−2871.
(3) Havel, T. F.; Wüthrich, K. *Bull. Math. Biol.* **1984**, *46*, 673−698.
(4) Braun, W.; Go, N. *J. Mol. Biol.* **1985**, *186*, 611−626.
(5) Nilges, M.; Clore, M.; Gronenborn, A. *FEBS Lett.* **1988**, *239*, 129−136.
(6) Jardetsky, O. *Biochim. Biophys. Acta* **1980**, *621*, 227−232.
(7) Van Gunsteren, W. F.; Brunne, R. M.; Gros, P.; van Schaik, R. C.; Schieffer, C.; Torda, A. E. *Methods Enzymol*. **1994**, *239*, 619−654.
(8) Sheek, R. M.; Torda, A. E.; Kemmink, J.; van Gunsteren, W. F. *Computational Aspects of the Study of Biological Macromolecules by NMR*; Hoch, J. C., et al., Eds.; Plenum Press: New York, 1991; pp 209−217.
(9) Kemmink, J.; van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993**, *230*, 312−322.

S0002-7863(96)03681-5 CCC: $14.00

simultaneously to match the NMR restraints. In the presence of conformational variability, the number of interconverting conformers as well as their particular topology are a priori unknown. Therefore, this number of conformers should be treated as an adjustable parameter during the derivation of the structures by means of the EA protocol and should only be determined from the content of the NMR data set.

One drawback of the EA approach is that it increases the number of degrees of freedom of the system. As already pointed out by several authors,[12−14] one should retain a cautious pespective when better satisfying the NMR restraints by increasing the number of adjustable parameters. In other terms, better agreement of the NMR restraints by averaging over several conformers should not be considered as evidence of conformational averaging. Thus, the quality of the fit of the NMR restraints must be evaluated as significant as compared to the NMR data set in order to avoid overfitting the experimental data. To this end, it has been proposed to use the complete cross-validation method.[12,15,16]

The content and size of the NMR data set play a crucial role in determining the threshold of the overfitting. For the most part, the EA protocols reported in the literature considers only the distance restraints (DR) derived from the observed NOE/ROE,[10,12,16−18] whereas the other NMR-derivable restraints are disregarded. Two other sources of conformational information must be taken into account in the derivation of the structures. One must use the so-called antidistance restraints (ADR) corresponding to unobserved ROE/NOE.[19] The ADR are based on the fact that the absence of dipolar correlation between two protons means that there is no significantly populated conformation for which the distance separating these two protons is lower than a threshold. The coupling constants (*J*) also contain very important conformational information. Indeed, torsional angles can be estimated from certain *J*-couplings from suitable Karplus-type relations. However, the observed *J*-coupling constant often corresponds to multiple solutions through this relation. The direct application of *J*-coupling restraints for the derivation of NMR structures can be used to overcome this problem.[20] Thus, in addition to the DR, the use of both ADR and coupling constants in the determination of the solution conformation(s) of peptides increases the size of the data set and therefore enhances the significance of the resulting conformation(s).

The problem we face when studying the solution structure of a cyclopeptide is to determine *all combinations* of *different conformers* that give a *satisfactory and significant* fit of *all NMR restraints* (DR, ADR, and *J*-couplings) using an explicit *averaging* model of the corresponding quantities over these conformers. As a test case, we have investigated the solution conformation of the hexapeptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) in DMSO at 298 K by ^{1}H NMR spectrocopy. A set of DR, unambiguous ADR, and *J*-couplings was determined for this peptide. These data were used as input restraints in a simulated-annealing-based ensemble-averaging protocol. The significance of the resulting sets of conformation was estimated by complete cross-validation.

Experimental Section

NMR Measurements. NMR spectra of cyclo(Gly1-Pro2-Phe3-Gly4-Pro5-Nle6) were recorded on Brüker AMX500 spectrometer. The sample concentration was 30 mM in DMSO solution. All spectra were recorded at 298 K. Chemical shifts were measured relative to internal reference sodium 2,2,3,3-tetradeutero-3-(trimethylsilyl)propionate (TSP). The spectral width in both dimensions was 5050 Hz. All 2D NMR spectra were recorded with quadrature detection in F2. Quadrature detection in F1 was made by time-proportional phase incrementation.[21] The NMR data were processed on an SGI indigo R4000 workstation with Felix 2.3 software.[22]

TOCSY spectra[23,24] were recorded with either a 20 or 80 ms Waltz16 sequence for the isotropic mixing,[25] a 1.5 s relaxation delay, 32 scans, 2048 complex data points in F2, and 256 experiments in F1.

The *J*(NH-Hα)-coupling constants were determined in the 1D resolution-enhanced spectrum recorded with 32K complex data points and a spectral width of 6024 Hz to give a final resolution of 0.18 Hz/Pt. *J*(NH-Hα′Hα″) of the Gly residues and *J*(Hα-Hβ′Hβ″) were determined using a DQF-COSY spectrum[26,27] and E-COSY spectrum[28−30] recorded with 8192 data points in the F2 dimension and a 5050 Hz spectral width to give a final resolution of 0.73 Hz/Pt.

Interproton distances were obtained from analysis of 2D off-resonance ROESY spectra.[31−33] It has been shown that the use of the off-resonance radio frequency field for spin lock in ROESY experiments leads to suppression of HOHAHA transfers and reduction of offset effects. These experiments were recorded with 128 scans, a 2s relaxation delay, 2048 complex data points in F2, and 256 experiments in F1. The mixing sequence was achieved by adiabatic rotation and irradiation at two opposite offsets.[34] Five sets corresponding to different Θ angles (0, 5, 10, 40, and 54.7°) of five mixing times (30, 50, 70, 90, and 120 ms) of off-resonance ROESY experiments were recorded. For integration, apodization with a sinebell function shifted by 90° was used in both dimensions. After Fourier transform, all ROESY spectra were corrected using T1 noise reduction and/or local baseplane correction routines written in the macrolanguage of the Felix software. The volumes of the cross peaks were integrated. For each correlation, the volumes of the corresponding diagonal peaks were also measured and a corrected intensity was obtained by normalization of the cross-peak volume.[35,36] In the case where one of the diagonal peaks was not

(10) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143−149.
(11) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042−1049.
(12) Bonvin, M. J. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.
(13) Kim, Y.; Prestegard, J. H. *Biochemistry* **1989**, *28*, 8792−8797.
(14) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960−10974.
(15) Brünger, A. T . *Nature* **1992**, *355*, 472−475.
(16) Brünger, A. T.; Clore, M. G.; Gronenborn, A. M.; Safrich, R.; Nilges, M. *Science* **1993**, *261*, 328−331.
(17) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. *J. Biomol. NMR* **1995**, *6*, 163−170.
(18) Bonvin, M. J. J.; Brünger, A. T. *J. Biomol. NMR* **1996**, *7*, 72−76.
(19) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR.* **1991**, *1*, 3−11.
(20) Kim, Y.; Prestegard, J. H. *Proteins* **1990**, *8*, 377−385.
(21) Marion, D.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967−974.
(22) Biosym Technologies Inc., 10065 Barnes Canyon Rd, San Diego, CA 92121.
(23) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528.
(24) Davis, D. G.; Bax, A. *J. Am. Chem. Soc.* **1985**, *107*, 2820−2821.
(25) Shaka, A. J.; Keeler, J. K.; Freeman, R. *J. Magn. Reson.* **1983**, *53*, 313−340.
(26) Piantini, U.; Sørensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800−6801.
(27) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(28) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1985**, *107*, 6394−6397.
(29) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Chem. Phys.* **1986**, *85*, 6837−6852.
(30) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Magn. Reson.* **1987**, *75*, 474−492.
(31) Desvaux, H.; Berthault, P.; Birlirakis, N.; Goldman, M. *C. R. Acad. Sci. Paris Ser. 2* **1993**, *317*, 19.
(32) Desvaux, H.; Berthault, P.; Birlirakis, N.; Goldman, M. *J. Chem. Phys.* **1994**, *91*, 646.
(33) Desvaux, H.; Berthault, P.; Birlirakis, N.; Goldman, M. *J. Magn. Reson. A* **1994**, *108*, 219.
(34) Desvaux, H.; Berthault, P.; Birlirakis, N.; Goldman, M.; Piotto, M. *J. Magn. Reson. A* **1995**, *113*, 47−52.
(35) Macura, S.; Farmer, B. T.; Brown, L. R. *J. Magn. Reson.* **1986**, *70*, 493−499.
(36) Kazmierski, W. M.; Yamamura, H. I.; Hruby, V. *J. Am. Chem. Soc.* **1991**, *113*, 2275−2283.

sufficiently resolved for integration, the cross-peak was scaled with respect to one of them. The cross-relaxation rates σ were determined for each value of Θ in the two-spin approximation by linear fitting of the buildup curves. Least square fitting of σ versus the angle Θ gives the longitudinal ($\Theta = 0°$) and transversal ($\Theta = 90°$) cross-relaxation rates for each pair.[33,3437] Interproton distances were determined from the longitudinal and transverse cross-relaxation rates assuming Lorentzian spectral density functions. This method has the advantage that interproton distances are determined without calibration and thus should give more accurate data than the standard procedure.

The stereospecific assignment of the Hβ proton resonances was achieved by standard analysis of the NH/Hβ, Hα/Hβ ROEs in conjunction with the J(Hα/Hβ)-coupling constants. The unambiguous assignment of the Gly4 Hα^{ProR} and Gly4 Hα^{ProS} proton resonances was possible based on the ROEs observed for these resonances. The stereospecific assignment of the Gly1 Hα^{ProR} and Gly1 Hα^{ProS} proton resonances was not achieved.

Modeling of the NMR Data. Ensemble-averaging. A modified version of the X-PLOR 3.1 software has been used to run the ensemble-averaging protocol.[38] The NOE.s routine has been modified, and an ensemble-averaging option was written in a similar way as suggested by Bonvin and Brunger.[12] This option takes advantage of the standard $\langle$r-6$\rangle$ (or $\langle$r-3$\rangle$) averaging mode but avoids the intermolecular interproton distance calculations. This, together with the SELE facility of the standard X-PLOR, allows use of multicopies of the structure under study, averaging of the interproton distances over the ensemble of conformers, and evaluation of intramolecular energy terms only. The experimental interproton distances were restrained to the $\langle$r-6$\rangle$ averaged over the multiple conformations using a square well potential of form identical to that of the standard X-PLOR 3.1.

A routine was written in FORTRAN 77 to treat the coupling constants directly as restraints in the building or refinement of structure from NMR data. This J-coupling restraint routine includes an ensemble-averaging mode which allows the calculation of the coupling constants as an average over the multiconformations according to $J^{calcd} = {}^1/_N\sum_i J_i$, where J_i corresponds to the coupling constant calculated for the individual conformation and N is the number of conformations in the ensemble. The individual coupling constants are deduced from dihedral torsion angles related to the ϕ and χ_1 angles using Karplus-type relations.[2] The ^3J(HN-Hα)-coupling constants for non-glycine residues were calculated from the ϕ value using a Bystrov−Karplus equation,[39] and for glycine residues, the ^3J(HN-Hα'Hα'') constants were calculated using the coefficient of DeMarco and co-workers.[40] ^3J(Hα-Hβ'Hβ'') couplings were calculated from the χ_1 angles according to Cung and Maraud.[41]

The ensemble-averaged coupling constants are restrained to the observed values using a harmonic potential in the form defined by $V_J = \sum_i k_J(J_i^{expt} - J_i^{calcd})^2$, where J_i^{expt} and J_i^{calcd} correspond, respectively, to the experimental and ensemble-averaged values of coupling constant i.

The NMR data set contains 31 distance restraints (DR) corresponding to the observed cross-relaxation effects and 151 antidistance restraints (ADR) corresponding to unambiguous nonobserved cross-relaxation effects. In addition, eight coupling constants were used as input restraints. The uncertainty used for the distance restaints was ±10%. The antidistance restraints were treated to give no energy contribution if the corresponding averaged distance was greater than 3.7 Å.[19] This value was choosen because it corresponds to the lower bound of the longest distance determined in this work (4.00 Å, see Results). When the ±10% uncertainty on the DR assumed in this study is taken into account, any pair of protons separated by a distance shorter or equal to 3.6 Å must give rise to an observable ROE under the experimental conditions. Furthermore, two protons that do not give rise to an observable ROE must be separated by a distance longer than 3.6 Å. The potential used for the ADRs actually allows the corresponding pair of proton to take any value greater than 3.7 Å and give an unfavorable energy contribution only if this value is lower than the threshold. Violations of the ADRs were counted from 0.1 Å below the threshold, i.e., from 3.6 Å.

Ensemble-Averaging Protocol. An ensemble of structures compatible overall with the NMR data (effective interproton distances and homonuclear coupling constants) was determined using a protocol of random simulated annealing. During all steps of the EA protocol, when used, the restraints are averaged and incorporated into the potential energy of the system using the restraining potential as reported above.

The initial step consists in generating a set of n initial structures ($n = 1-4$). Starting from random coordinates of the atoms, each structure was minimized in the parmallhdg force field of X-PLOR 3.1.[5] No experimental restraint terms were used during this initial step. These n structures were then associated to give an ensemble of random initial conformations. This ensemble was submitted to a simulated annealing protocol from 1000 to 100 K in the parmallhdg force field of X-PLOR. During this step, the total potential energy contains terms corresponding to bond stretching, angle bending, improper torsion angle twisting, NMR distance restraints, and J-coupling restraints. The van der Waals interactions were treated with a purely repulsive term, and no electrostatic term was taken into account. During the cooling, the force constants on the distance restraints (K_{NOE}) and J-coupling (K_J) constants were linearly increased from 0.1 to 50 kcal mol^{-1} (Å or Hz)$^{-2}$. The van der Waals radii were linearly increased to reach their standard parmallhdg value at 100 K. At the end of the cooling, the structures were then minimized in the parmallhdg force field.

The resulting ensemble was submitted to a refinement in the CHARMM22 force field of X-PLOR 3.1. For all distance restraints, the force constants K_{NOE} were set to 50 kcal mol^{-1} Å^{-2}. The force constants K_J were set to 5 and 0.5 kcal mol^{-1} Hz^{-2} for the J(NH-Hα) and J(Hα-Hβ), respectively. During this refinement, the potential contains terms corresponding to bond stretching, angle bending, dihedral torsion angle twisting, and NMR restraint terms (DR, ADR, and J) and nonbonded interactions. These nonbonded interactions include a Lennard−Jones potential, and the electrostatic interactions were taken into account with a C.D.I.E. treatment ($\epsilon = 48$). All calculations were carried out in vacuo.

Cross-Validation. The significance of the fit of the NMR data by the sets of n conformers was evaluated by complete cross-validation.[12] Twenty working sets containing 80% of the total distance restraints (DR and ADR) were obtained by random trial. For each value of n (1−4), the number of violations and the root mean square deviations (Rmsds) of the working and test sets were evaluated for each trial and subsequently averaged over the 20 trials to give the m_Rmsd criteria. The mean number of violations and the mean root mean square deviation of the DR and ADR of the test sets were used as criteria to evaluate the quality of the fit. This methods determines the optimum number (n) of conformers which produces a significant and satisfactory fit of all NMR distance restraints (DR and ADR). No cross-validation was applied to the J-coupling restraints. The statistical significance of the difference of mean values calculated for the criteria used in the complete cross-validation were determined by Student-t tests at a significance level of 5%.

When the optimal number of conformations was determined, the ensemble-averaging protocol described above was reiterated 50 times with the complete NMR data set (DR, ADR, and J) to produce 50 sets of n conformers compatible with these data. The energies of the resulting structures were calculated in the CHARMM22 force field. Sets for which the total energy of at least one of the structures was 8 kcal mol^{-1} above the minimum energy found in the 50 sets were rejected. For this selection, the NMR restraint terms were excluded of the potential energy. The resulting structures were clustered using a Rmsd(Φ,Ψ) criterion.[42] The threshold value for this clustering procedure was set to 30°.

Results

The assignment of all proton resonances of the peptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) was achieved using the standard

(37) Desvaux, H.; Berthault, P.; Birlirakis, N. *Chem. Phys Lett.* **1995**, *233*, 545−549.

(38) Brünger, A. T. X-PLOR version 3.1; Yale University Press: New Haven, CT, 1993.

(39) Bystrov, V. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, 41−81.

(40) De Marco, A.; Llinas, M.; Wuthrich, K. *Biopolymers* **1978**, *17*, 637−650.

(41) Cung, M. T.; Marraud, M. *Biopolymers* **1982**, *21*, 953−957.

(42) Karpen, M. E.; de Haseth, P. L.; Neet, K. E. *Proteins* **1989**, *6*, 155−167.

Table 1. ^{1}H NMR Chemical Shifts of Cyclo(Gly^1-Pro^2-Phe^3−Gly^4-Pro^5-Nle^6) in DMSO at 298 K

	chemical shift (ppm)		chemical shift (ppm)
Gly^1-HN	7.41	Gly^4-Hα^{ProR}	4.13
Gly^1-Hα1	3.85	Gly^4-Hα^{ProS}	3.89
Gly^1-Hα2	3.99	Pro^5-Hα	4.22
Pro^2-Hα	4.01	Pro^5-Hβ^{ProS}	2.18
Pro^2-Hβ^{ProS}	1.84	Pro^5-Hβ^{ProR}	1.75
Pro^2-Hβ^{ProR}	1.37	Pro^5-Hγ^{ProR}	1.93
Pro^2-Hγ^{ProR}	1.77	Pro^5-Hγ^{ProS}	1.93
Pro^2-Hγ^{ProS}	1.77	Pro^5-Hδ^{ProR}	3.49
Pro^2-Hδ^{ProR}	3.51	Pro^5-Hδ^{ProS}	3.64
Pro^2-Hδ^{ProS}	3.47	Nle^6-HN	8.14
Phe^3-HN	8.28	Nle^6-Hα	4.13
Phe^3-Hα	4.25	Nle^6-Hβ^{ProR}	1.61
Phe^3-Hβ^{ProR}	2.99	Nle^6-Hβ^{ProS}	1.96
Phe^3-Hβ^{ProS}	3.37	Nle^6-Hγ	1.18
Phe^3-Hδ	7.18	Nle^6-Hδ	1.24
Phe^3-Hϵ	7.30	Nle^6-Hδ	1.32
Phe^3-Hζ	7.23	Nle^6-Hϵ	0.87
Gly^4-HN	7.61		

analysis of the TOCSY and ROESY experiments. The NH/aromatic region of the 1D ^{1}H NMR spectrum of this cyclohexapeptide exhibits two sets of resonances which give rise to negative exchange cross-peaks in the ROESY spectra and corresponds to cis/trans isomerization of proline residue. However, the integration of the corresponding resonances in the 1D proton spectrum indicates that the minor components represent less than 20% of the major one. Thus, only the major species was further analyzed. The chemical shifts of all proton resonances of the molecule are reported in Table 1. The qualitative analysis of the ROESY connectivities (data not shown) is consistent with the existence of a double reverse turn structure involving residues Pro^2-Phe^3 and Pro^5-Nle^6 for this peptide under the experimental conditions. Thirty-one effective interproton distances obtained for the peptide under study from off-resonance ROESY experiments are reported in Table 2. Note that the interproton distances corresponding to geminal proton pairs exhibit a very small dispersion around the expected value.

Several interproton distances can be used to determine the types of the two reverse turns. Distances separating the NH proton of residue $i + 2$ and the Hα protons of residues in positions $i + 1$ and $i + 2$ are characteristic of the particular conformation adopted by the reverse turn. The d(NH$_{i+2}$-Hα_{i+2}) distance is 2.9 and 2.3 Å for type I and type II turns, respectively, and the distance d(NH$_{i+2}$-Hα_{i+1}) is 3.5 and 2.1 Å for type I and type II turns, respectively.[43,44] For the peptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle), the observed d(NH$_{i+2}$-Hα_{i+2}) and d(NH$_{i+2}$-Hα_{i+1}) distances for the Pro2-Phe3 reverse turn are 2.47 and 2.63 Å, respectively (Table 2). These values are not compatible with the standard values reported for canonical reverse turns. The same observation hold for the Pro5-Nle6 reverse turn where the d(NH$_{i+2}$-Hα_{i+2}) and d(NH$_{i+2}$-Hα_{i+1}) distances are 2.71 and 2.73 Å, respectively. Furthermore, for the two reverse turns, these d(NH$_{i+2}$-Hα_{i+2}) and d(NH$_{i+2}$-Hα_{i+1}) are not compatible with low-energy conformations. In addition, the J(NH$_{i+2}$-Hα_{i+2}) which depends on the Φ angle of this residue is sometimes used to determine or confirm the type of reverse turn.[45] The observed value for the J(NH-Hα) is 8.2 and 8.8 Hz for residue Phe3 and Nle6, respectively (Table 3). In both cases, these values correspond to multiple solutions through the Karplus relation using the coefficients of Bystrov.[39] For residue Phe3, the Φ angles compatible with the observed J(NH-Hα) are −151, −89, +47, and +72°, while for residue Nle6, the solutions are −147 and −93°. Only the values around −90° are compatible with standard reverse turns (type I or II′).[46] However, the interproton distances do not correspond to one of these conformations, and therefore, it is likely that the observed J(NH-Hα), as well as these effective interproton distances, correspond to average values over several interconverting conformations.

In order to take conformational averaging explicitly into account for the derivation of structures compatible with the NMR data, we carried out an ensemble-averaging protocol (EA). In this protocol, the DR, ADR, and coupling constants are treated as averaged quantities. The time scale of the motion involved determine the way the distances are averaged.[48,49] As reported above, the values observed for the characteristic

Table 2. Interproton Distances Used as Input Restraints (DR) in the EA Protocol for Cyclo(Gly^1-Pro^2-Phe^3-Gly^4-Pro^5-Nle^6) (Å)

DR id no.	proton 1	proton 2	d_{eff} (Å)
1	Gly^1-HN	Gly^1-Hα1	2.90
2	Gly^1-HN	Gly^1-Hα2	2.90
3	Gly^1-HN	Pro^5-Hα	3.75
4	Gly^1-HN	Nle^6-Hα	3.13
5	Gly^1-Hα1	Pro^2-Hδ^{ProR}	2.25
6	Gly^1-Hα2	Pro^2-Hδ^{ProR}	2.25
7	Phe^3-HN	Pro^2-Hα	2.63
8	Phe^3-HN	Pro^2-Hδ^{ProR}	3.35
9	Phe^3-HN	Phe^3-Hα	2.47
10	Phe^3-HN	Phe^3-Hβ^{ProR}	3.10
11	Phe^3-Hα	Phe^3-Hβ^{ProS}	2.40
12	Phe^3-Hδ*	Pro^2-Hβ^{ProR}	3.30
13	Phe^3-Hδ*	Pro^2-Hβ^{ProS}	3.30
14	Phe^3-Hδ*	Phe^3-Hβ^{ProR}	2.70
15	Phe^3-Hδ*	Phe^3-Hβ^{ProS}	2.90
16	Gly^4-HN	Pro^2-Hα	4.00
17	Gly^4-HN	Phe^3-Hα	2.95
18	Gly^4-HN	Gly^4-Hα^{ProR}	2.77
19	Gly^4-HN	Gly^4-Hα^{ProS}	2.92
20	Gly^4-Hα^{ProR}	Pro^5-Hδ^{ProR}	2.20
21	Gly^4-Hα^{ProS}	Pro^5-Hδ^{ProR}	2.70
22	Gly^4-Hα^{ProS}	Pro^5-Hδ^{ProS}	2.30
23	Nle^6-HN	Gly^4-Hα^{ProR}	3.82
24	Nle^6-HN	Pro^5-Hα	2.73
25	Nle^6-HN	Pro^5-Hβ^{ProR}	3.46
26	Nle^6-HN	Pro^5-Hg*	3.42
27	Nle^6-HN	Pro^5-Hδ^{ProR}	3.44
28	Nle^6-HN	Nle^6-Hα	2.71
29	Nle^6-HN	Nle^6-Hβ^{ProR}	3.27
30	Nle^6-Hα	Nle^6-Hβ^{ProS}	2.47
31	Nle^6-Hα	Nle^6-Hγ*	3.00
	Pro^2-Hβ^{ProR}	Pro^2-Hβ^{ProS}	1.90
	Phe^3-Hβ^{ProR}	Phe^3-Hβ^{ProS}	1.76
	Gly^4-Hα^{ProR}	Gly^4-Hα^{ProS}	1.84
	Pro^5-Hβ^{ProR}	Pro^5-Hβ^{ProS}	1.81
	Nle^6-Hβ^R	Nle^6-Hβ^S	1.80

Table 3. Experimental Coupling Constants (Hertz) Obtained for Peptide Cyclo(Gly^1-Pro^2-Phe^3-Gly^4-Pro^5-Nle^6) in DMSO

	Gly^1	Phe^3	Gly^4	Nle^6
^3J(NH-Hα)		8.20		8.80
^3J(NH-Hα^{ProR})	3.70[a]		5.30	
^3J(NH-Hα^{ProS})	4.70[b]		2.80	
^3J(Hα-Hβ^{ProR})		12.50		8.90
^3J(Hα-Hβ^{ProS})		3.60		4.50

[a] Downfield. [b] Upfield.

(43) Narasinga Rao, B. N.; Kumar, A.; Balaram, H.; Ravi, A.; Balaram, P. *J. Am. Chem. Soc.* **1983**, *105*, 7423−7428.
(44) Rose, G. D.; Gierash, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1−109.
(45) Aubry, A.; Cung, M. T.; Marraud, M. *J. Am. Chem. Soc.* **1985**, *107*, 7640−7647.
(46) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Biochim. Biophys. Acta* **1973**, *303*, 211−229.
(47) Wilmot, C. M., Thornton, J. M. *Protein Eng.* **1990**, *3*, 479−493.
(48) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035−6043.

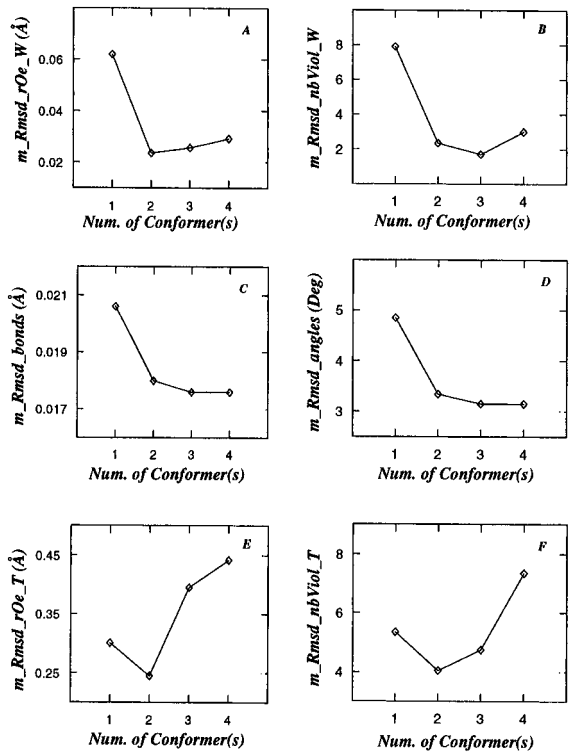

Figure 1. Complete cross-validation as a function of the number of conformers: (A) mean rmsd for DR and ADR (m_Rmsd_ROE_W) in angstroms, (B) average number of violations (m_nbViol_W), (C) average rmsd for the bonds (m_Rmsd_bonds) in Å, (D) average rmsd for the angles (m_Rmsd_angles) in degrees calculated for the 20 working sets, (E) mean rmsd for the DR and ADR (m_Rmsd_ROE_T), and (F) mean number of violations (m_nbViol_T) calculated on the 20 test sets.

distances for the two reverse turns suggest that several types of turns may interconvert on the NMR time scale. The energy barrier for βI/βII interconversion is on the order of 5−10 Kcal mol^{-1}.[50] This value corresponds to a time scale of motion that is slow as compared to the overall correlation time of a molecule of this size (ca. 0.5 ns under the experimental conditions). Assuming that this slow interconversion dominates the motion which take place in the molecule, the ⟨r-6⟩ averaging mode is used during the EA protocol.[49]

For each number of conformers in the ensemble (*n*), 20 EA runs were carried out with a distance set containing 80% of the DR and ADR (the so-called working set). These 20 working sets were obtained by random trial in the total distance database. Eight coupling constants were taken into account as input restraints in the 20 runs (Table 3). These correspond to the *J*(NH-Hα) of residue Phe3, Nle6 and both *J*(NH-H$\alpha^{\mathrm{Pro}R}$) and *J*(NH-H$\alpha^{\mathrm{Pro}S}$) of residue Gly1. The *J*(NH-Hα^1) and *J*(NH-Hα^2) of residue Gly4 were not used as input restraints because the prochirality of the Hα of this residue was not assigned. The *J*(Hα-H$\beta^{\mathrm{Pro}R}$) and *J*(Hα-H$\beta^{\mathrm{Pro}S}$) of residues Phe3 and Nle6 were taken into account as input restraints. The number of conformations ranged from 1 to 4. Several criteria were evaluated as function of the number of conformations in the ensemble. The averaged root mean square deviation of the DR and ADR of the working set (m_Rmsd_ROE_W), and the average number of violations greater than 0.1 Å in the working set (m_nbViol_W) over the 20 runs as a function of the number of conformers in the ensemble, are shown in Figure 1A,B. The statistical significance of the differences between the criteria calculated for each number of conformer(s) was evaluated by Student-t tests at a significance level of 5% (Table 4). As expected, the m_Rmsd_ROE_W and m_nbViol_W sharply decrease from a single conformation to a pair. Therefore, the agreement of the DR and ADR is significantly improved by modeling the conformation of the molecule under study as a pair of interconverting conformers. An indication of the quality of the structures is given by the mean Rmsd of the bonds (m_Rmsd_bonds) and bond angles (m_Rmsd_angles) over the 20 runs as compared to the standard geometry (Figure 1C,D and Table 4). These two criteria display the same behavior as m_Rmsd_ROE_W and m_nbViol_W, significantly decreasing from a unique conformation to the two-conformers model (Figure 1C,D). These observations show that the DR and ADR are not compatible with a unique conformation, and thus, the attempt to fit these restraints simultaneously in a single conformation tends to distort the standard covalent geometry. When increasing the number of conformers from $n = 2$ to $n = 3$ and from $n = 3$ to $n = 4$, the m_Rmsd_ROE_W does not display a large variation. The m_nbViol_W displays a slight but significant decrease from $n = 2$ to $n = 3$ and a slight but significant increase from $n = 3$ to $n = 4$. The same behavior is found for the covalent geometry criteria (Figure 1C,D and Table 4) that are slightly improved by increasing the number of conformers from $n = 2$ to $n = 3$ but not significantly improved from $n = 3$ to $n = 4$.

The cross-validation method gives unbiased criteria to evaluate the significance of the fit. For each of the EA runs, the root mean square deviation of the DR and ADR and the number of violations greater than 0.1 Å were determined for the test set (the omitted data during each run) and subsequently averaged over the 20 runs to give m_Rmsd_ROE_T and m_Rmsd_nbViol_T (Figure 1E,F and Table 4). The idea behind these criteria is that they give an estimate of the predictability of the restraints of the test set by those of the working set.[12,15] Figure 1E,F shows that the m_Rmsd_ROE_T and m_Rmsd_nbViol_T criteria significantly decrease from the unique conformation to

Table 4. Statistical Significance of the Cross-Validation Criteria Reported in Figure 1[a]

	n			
n	1	2	3	4
		Work Data		
1		S	S	S
2	S		NS	S
3	S	S		NS
4	S	S	S	
		Covalent Geometry Data		
1		S	S	S
2	S		S	S
3	S	S		NS
4	S	S	NS	
		Test Data		
1		S	S	S
2	S		S	S
3	S	S		NS
4	S	S	S	

[a] The difference in the average values of the criteria used in the complete cross-validation was estimated by Student-t tests at a significance level of 5%. NS: there is insufficient evidence to reject the hypothesis that the means differ. S: the hypothesis that the means are equal can be rejected at this level of significancy. n: number of conformer(s) in the ensemble. Work data: upper diagonal, m_Rmsd_ROE_W; lower diagonal, m_nbViol_W. Covalent geometry data: upper diagonal, m_Rmsd_bonds; lower diagonal, m_Rmsd_angles. Test data: upper diagonal, m_Rmsd_ROE_T; lower diagonal, m_nbViol_T.

(49) Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396.
(50) Stradley, S. J.; Rizo, J.; Bruch, M. D.; Stroup, A. N.; Gierash, L. M. *Biopolymers* **1990**, *29*, 263−287.

the two-conformers model. This observation shows that the test set is on average better predicted by the two-conformers model than by the single-conformer model. However, increasing the number of conformers from two to three and four results in a significant increase in the values obtained for m_Rmsd_ROE_T and m_Rmsd_nbViol_T. This situation corresponds to overfitting of the data for $n = 3$ or more[12] and indicates that the data of the test sets are on average less well predicted by a model of three or more interconverting conformers than by the two-conformers model. Thus, application of the EA protocol to the NMR-derived restraints obtained for the peptide cyclo(Pro-Phe-Gly-Pro-Nle-Gly) shows that these data give a significant and satisfactory fit for a two-conformers model.

No cross-validation was applied to the *J*-coupling restraints. However, these data were used as input restraints in the 20 runs that were carried out for each number of conformers. Taking into account the experimental uncertainty on these restraints as well as a 1 Hz uncertainty in the Bystrov−Karplus equation, the average Rmsd on the *J*(NH-Hα) restraints does not give significant differences as a function of the number of conformers in the cross-validation protocol. However, preliminary runs of the cross-validation protocol carried out whitout these ensemble-averaged *J*-coupling restraints (data not reported) have shown that some of the experimental coupling constants were poorly reproduced by the application of the distance restraint set alone. This observation prompted us to explicitly take into account the *J*-coupling constants as ensemble-averaged restraints.

In order to define the interconverting conformations, 50 runs of the EA protocol were performed with the complete database (all DR, ADR, and coupling constants) for pairs of conformers. The energies of the resulting 100 structures were calculated in the CHARMM22 force field. Pairs for which the total energy of one structure (or both) was 8 kcal mol^{-1} above the minimum energy found in the 100 structures were rejected. It should be noted that the energy used to select the conformers does not include the NMR restraint terms. This resulted in the selection of 23 pairs which were further analyzed. The (Φ,Ψ) plot for the six residues of the 46 selected structures is reported in Figure 2. The Ψ2, Φ3, Ψ5, and Φ6 dihedral angles exhibit several values indicating that the peptide planes in the reverse turns adopt several orientations. Ψ3 adopts different values between −60 and +100°, whereas Ψ6 spans a shorter range from −10 to +60°. These values differ from the standard value reported for type I and type II reverse turns (0°)[46] and are constistent with the fact that neither Phe3 NH/Gly4 NH ROE nor Nle6 NH/Gly1 NH ROE was observed. The distribution of the values found for the Φ1 and Φ4 angles (Figure 2) indicates that in addition to the extended conformation (Φ,Ψ around 180°), these residues may adopt nonextended conformations.

The 46 conformers (23 pairs) were clustered using a Rmsd-(Φ,Ψ) criterion[42] with a threshold of 30°. From this criterion, the 46 structures belong to four distinct classes (Cl1, Cl2, Cl3, and Cl4). The average (Φ,Ψ) angles are reported for each cluster (Table 5). Cl1 and Cl2 contain 19 structures, whereas Cl3 and Cl4 contain four structures (Table 5). It should be pointed out that Cl1 structures interconvert with Cl2 structures, whereas Cl3 structures interconvert with Cl4 structures. The stereoplots of the Cl1/Cl2 and Cl3/Cl4 conformers are reported in Figure 3A,B, respectively. It can be seen that the conformational spread of each cluster is small, leading to well-defined structures of these four conformers. In Cl1, the segment Pro2-Phe3 adopts a type VIII reverse turn conformation,[47] whereas the segment Pro5-Nle6 adopts a type II reverse turn conformation (Table 5 and Figure 3).[46] In Cl2, the interconverting partner of Cl1, the Pro2 adopts an inverse γ turn,[46] whereas the

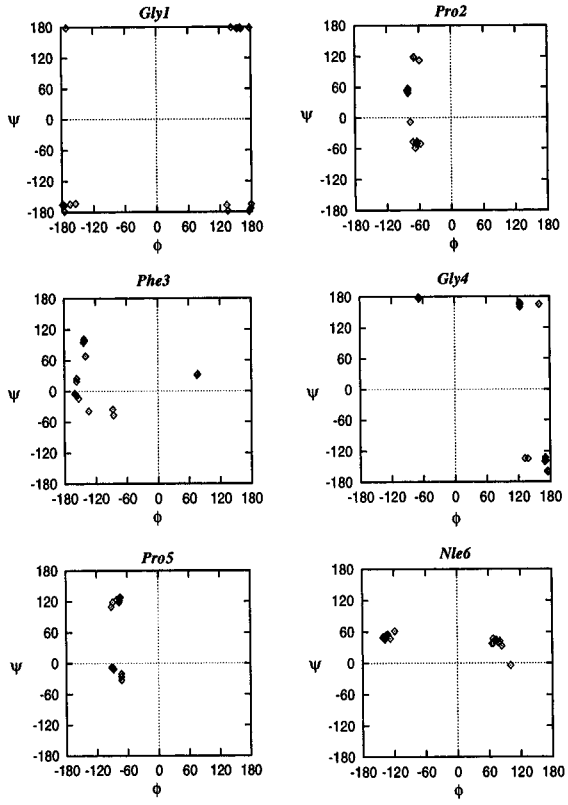

Figure 2. A (Φ, Ψ) plot for the six residues of the 23 selected pairs of structures for cyclo(Gly-Pro-Phe-PGly-Pro-Nle) obtained with the complete NMR data set (DR, ADR, J).

Table 5. Cluster Analysis of the Structures Obtained by Application of the EA Protocol to the Peptide Cyclo(Gly1-Pro2-Phe3-Gly4-Pro5-Nle6)[a]

	cluster			
	1	2	3	4
Φ_1	−178	151	−175	−172
Ψ_1	−166	178	−172	−176
Φ_2	−64	−81	−66	−72
Ψ_2	−49	54	116	−30
Φ_3	−141	−157	76	−109
Ψ_3	+94	−3	31	−39
Φ_4	129	165	174	−67
Ψ_4	164	−137	−159	178
Φ_5	−74	−89	−87	−72
Ψ_5	126	−8	117	−24
Φ_6	74	−137	75	−131
Ψ_6	42	49	27	52
no. structures	19	19	4	4
conf 2−3	VIII	*i*γ	II	I
conf 5−6	II	I	II	I

[a] Mean (Φ,Ψ) values (in degrees) for the six residues and number of structures found for each cluster. The conformation (conf) of segment Pro2-Phe3 and Pro5-Nle6 are indicated.

conformation of the segment Pro5-Nle6 is close to a type I reverse turn but with discrepancies in Ψ5 (ca. 20°) and Φ6 (ca. 40°).[46] Thus, classification of the conformation of segment Pro5-Nle6 as a type I turn is somewhat arbitrary but the observed value for the Ψ6 above −60° precludes classification of this conformation as a type VIII turn. In Cl3, the conformations of both segments Pro2-Phe3 and Pro5-Phe6 correspond to canonical type II turn conformations,[46] whereas the conformations of these segments in Cl4 correspond to canonical type I reverse turns.[46]

In order to check the fit of the NMR restraints for the two-conformers models Cl1-Cl2 and Cl3-Cl4, several criteria were

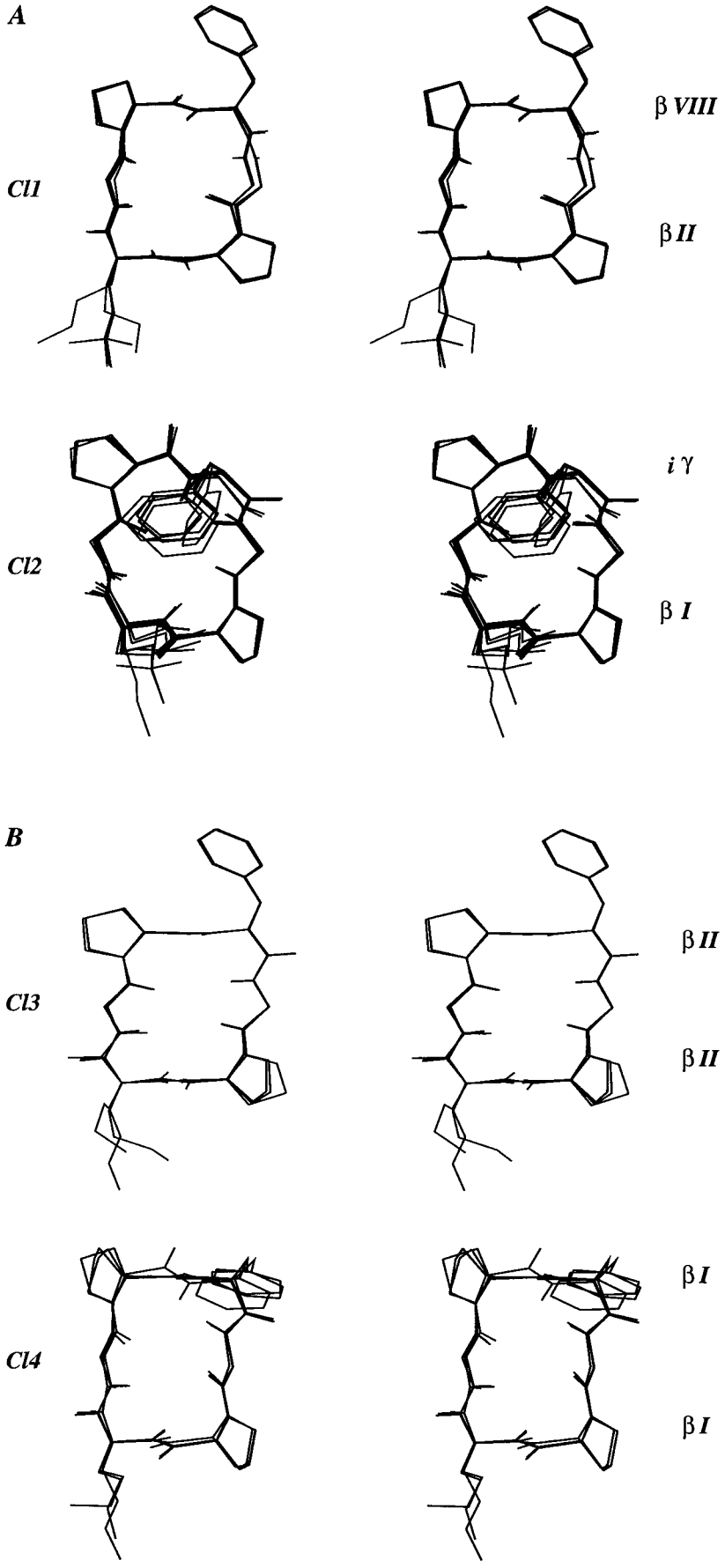

Figure 3. Stereoview of the interconverting structures of cyclo(Gly-Pro-Phe-Gly-Pro-Nle) obtained by ensemble-averaging with the complete NMR data set. (A) Two-conformers model 1, Cl1 and Cl2 structures. (B) Two-conformers model 2, Cl3 and Cl4 structures. The protons bound to carbons are omitted.

evaluated. The mean number of violations greater than 0.1 Å (m_nbViol), and the mean Rmsd of the DR and ADR (m_Rmsd_ROE) calculated over each set Cl1-Cl2 and Cl3-Cl4 are reported in Table 6. The m_nbViol for the Cl1-Cl2 pairs and

Table 6. Restraints Analysis for the Two Pairs of Conformers Cl1-Cl2 and Cl3-Cl4 Obtained from the EA Protocol[a]

	two-conformers model	
	Cl1-Cl2	Cl3-Cl4
m_nbViol	3.00 ± 0.33 (DR 2.0; ADR 1.0)	0.75 ± 0.50 (DR 0.75; ADR 0)
m_Rmsd_ROE	0.0250 ± 0.0008 (DR 0.052; ADR 0.014)	0.0140 ± 0.0018 (DR 0.035; ADR 0.003)
m_Rmsd_*J*all	0.810 ± 0.035	0.820 ± 0.060
m_Rmsd_*J*phi	0.10 ± 0.01	0.096 ± 0.022
m_Rmsd_*J*pg	0.280 ± 0.085	0.270 ± 0.068
m_Rmsd_*J*chi	1.12 ± 0.05	1.14 ± 0.09

[a] m_nbViol: mean number of violations of the DR and ADR greater than 0.1 Å and corresponding values calculated for the DR and ADR separately into parenthesis. m_Rmsd_ROE: mean Rmsd of the DR and ADR (Å) and corresponding values calculated for the DR and ADR separately into parenthesis; m_Rmsd_*J*all: mean Rmsd of the violations greater than 0.5 Hz calculated for all *J*-couplings. m_Rmsd_*J*phi: mean Rmsd of the violations greater than 0.5 Hz calculated for the *J*(NH-Hα) of residues Phe3 and Nle6. m_Rmsd_*J*pg: mean Rmsd of the violations greater than 0.5 Hz calculated for the *J*(NH-Hα^{ProR}) and *J*(NH-Hα^{ProS}) of residue Gly4. m_Rmsd_*J*chi: mean Rmsd of the violations greater than 0.5 Hz calculated for the *J*(Hα-Hβ^{ProR}) and *J*(Hα-Hβ^{ProS}) of residues Phe3 and Nle6. The standard deviations are given.

the Cl3-Cl4 pairs display comparable values (respectively 3 and 0.75). This holds for both DR and ADR (Table 6). It should be noted that the distance data set contains 182 restraints, and therefore, the mean number of violations is small compared with the number of restraints for the Cl1-Cl2 pairs as well as the Cl3-Cl4 pairs. The m_Rmsd_ROE are very similar for the Cl1-Cl2 and Cl3-Cl4 sets (0.025 Å for Cl1-Cl2 pairs and 0.014 Å for the Cl3-Cl4 pairs (Table 6)). It can be noted that no significant difference is observed for these criteria calculated for the DR and for the ADR. The mean Rmsd of the *J*-coupling restraints are reported for all *J* restraints, as well as for different subsets of the *J* database (*J*(NH-Hα) of Phe3 and Nle6, *J*(NH-Hα^{ProR}Hα^{ProS}) of Gly4, *J*(Hα-Hβ^{ProR}Hβ^{ProS}) of Phe3 and Nle6) calculated over Cl1-Cl2 and Cl3-Cl4 pairs (Table 6). The m_Rmsd_*J* criteria exhibit comparable values for the two sets of two-conformers. The standard deviations on the m_Rmsd criteria (*J* and ROE) reported in Table 6 are small, showing that the pairs found in Cl1-Cl2 and those found in Cl3-Cl4 are homogeneous with regard to the fit of the NMR data (Table 6). For both sets of two conformers, the m_Rmsd_*J* calculated for backbone restraints for Gly4 as well as for non-glycine residues are below the experimental uncertainty in the determination of these values (respectively, ±0.73 and ±0.18 Hz; see Experimental Section) (Table 6). The m_Rmsds on the χ_1-related *J*-coupling exhibit values above the experimental uncertainty in these values (Table 6) for Cl1-Cl2, as well as for Cl3-Cl4. This probably reflects one limitation of the method when the overall content of the NMR data set leads to a significant fit of these data by pairs of structures. For the case of the staggered rotamers of the side chain, sampling the three states of the side chains to fit the DR, ADR, and *J*-couplings is not possible, leading to a poorer agreement of the χ_1-related *J*-couplings as compared to the backbone-related *J*-couplings.

From the above considerations, it follows that none of the two two-conformers models returned (Cl1-Cl2 and Cl3-Cl4) can be excluded on the basis of experimental evidence. It is important to note that this result does not mean that the four conformers interconvert in solution. Indeed, we have demonstrated that only pairs of conformers can be derived to give a significant and satisfactory agreement of the NMR data. Therefore, one must conclude that the Cl1-Cl2 pair and the Cl3-Cl4 pair represent two alternative solutions to significantly and satisfactorily reproduce the NMR data set.

Discussion

Conformation. Structural analysis of a number of cyclic hexapeptides has shown that these peptides generally adopt a two-β reverse turn structure.[51,52] The β reverse turns have been shown to play a crucial role in biologically active peptides by determining the topology of side chains which contribute to their activity.[53] The two-β-turn backbone conformation of cyclohexapeptides has been previously well documented both by NMR and crystal studies.[54] Some examples have been reported where the crystal conformation of a cyclohexapeptide is present in solution but in rapid exchange with other conformations.[55] Furthermore, there are a number of NMR studies where the experimental data do not fit a unique conformation.[49,56−58] Interproton distances obtained from cross-relaxation experiments that are incompatible with each other have been used to infer the presence of interconverting conformers.[50] However, the methods used in most NMR studies of cyclohexapeptides do not take conformational variability explicitly into account. In contrast, the present work describes the application of an ensemble-averaging protocol to determine interconverting conformations of the peptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) in solution from a complete set of NMR restraints. Two possible models giving a satisfactory and significant fit of the NMR data are defined, suggesting that this peptide may exist as a βII-βII/βI-βI equilibrium or a βVIII-βII/$i\gamma$-βI double reverse turn equilibrium.

It is well-known that proline residues impose a conformational restriction on the Φ angle which corresponds to the preferred range for Φ_{i+1} of β turns.[44] In this respect, the observation of double reverse turn structure was expected for the peptide under study. The average values observed for the (Φ,Ψ) angles defining the two reverse turns for Cl3 (βII,βII) are almost canonical.[44,46] The two glycine residues are in extended conformations (Φ,Ψ around 180°), as already observed in crystal structures of double type II reverse turn containing glycine residues as linkers of the two reverse turns.[59]

In Cl4, the βI-βI conformation, the average (Φ,Ψ) dihedral angles of reverse turn Pro2-Phe3 correspond to almost canonical values for a type I turn,[44] whereas for the segment Pro5-Nle6, the Φ6 (−131°) displays a significant deviation as compared to the standard value for this angle in a type I turn (−90°). However, such deviation has already been found for the Φ_{i+2} angle, i.e., in the crystal structure of cyclo[(Gly-Pro-Gly)2] where the Φ_{i+2} of the reverse turn I is −115°.[60] The most striking feature of the conformation of Cl4 concerns the Ψ3 and Φ4 dihedral angles found to be respectively −39 and −67° (Table 5). This has one important topographical consequence: the NH of residue Gly4 points on the opposite side to the

(51) Schwyzer, R.; Sieber, P.; Gorup, B. *Chimia* **1958**, *12*, 90−91.
(52) Gierasch, L. M.; Deber, C. M.; Madison, V.; Niu, C.; Blout, E. R. *Biochemistry* **1981**, *20*, 4730−4738.
(53) Smith, J. A.; Pease, L. G. *C.R.C. Crit. Rev. Biochem.* **1980**, *8*, 315.
(54) Bean, J. W.; Kopple, K. D.; Peishoff, C. E. *J. Am. Chem. Soc.* **1992**, *114*, 5328−5334 and references therein.
(55) Yang, C.-H.; Brown, J. N.; Kopple, K. D. *J. Am. Chem. Soc.* **1981**, *103*, 17715−1719.
(56) Scarsdale, J. N.; Yu, R. K.; Prestegard, J. H. *J. Am. Chem. Soc.* **1986**, *108*, 6778−6784.
(57) Kim, Y.; Ohlrodgge, J. B.; Prestegard, J. H. *Biochem. Pharm.* **1991**, *40*, 7−13.
(58) Kopple, K. D.; Baure, P. W.; Bean, J. W.; D'Ambrosi, C. A.; Huges, J. L.; Eggleston, D. S. *J. Am. Chem. Soc.* **1992**, *14*, 9615−9623.
(59) Kostansek, E. C.; Lipscomb, W. N.; Thiessen, W. E. *J. Am. Chem. Soc.* **1979**, *101*, 834−837.
(60) Kostansek, E. C.; Thiessen, W. E.; Schomburg, D.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1979**, *101*, 5811−5815.

carbonyl oxygen of residue Gly1 (Figure 3B), therefore prohibiting the formation of a 1 → 4 hydrogen bond in the Pro2-Phe3 segment in this conformation. However, it is now well recognized from analysis of protein structures,[61] and from crystallographic studies of model peptides,[62] that the formation of the 1 → 4 hydrogen bond is not required to stabilize β turns.

Structures found in Cl1 correspond to the βVIII-βII double reverse turn conformation. The βVIII turn resembles the βI turn, one difference being that Φ_{i+2} is −90° for a canonical βI turn whereas this dihedral angle is −120° for a βVIII turn. The main difference concerns the Ψ_{i+2} which is 0° in a type I turn and 120° in a βVIII turn.[47] These authors proposed a 60° cutoff for Ψ_{i+2} to distinguish type I from type VIII reverse turn. It should be noted that the βVIII turn is the most represented reverse turn in proteins after the type I and type II reverse turns. In a previous publication, we investigated the conformation of the phosphinic analog (cyclo[Gly-Pro-PheΨ[PO_2-CH_2]Gly-Pro-Nle]) of the peptide analyzed in the present study.[63] In aqueous solution the preferred conformation of this phosphinic analog is a double βVIII reverse turn. In Cl1, the segment Pro5-Nle6 adopts an almost canonical βII reverse turn conformation. Only the Ψ6 value (+42°) differs from the canonical value (0°). It should be noted that Ψ_{i+2} density in type II reverse turns in protein structures is maximal, about 50°.[47]

Structures found in Cl1 are proposed to be in conformational interconversion with structures in Cl2 (Table 5, Figure 3B). In the Cl2 structures, the segment Pro5-Nle6 displays a somewhat distorted β I reverse turn, but Pro2-Phe3 does not fall into any classical β reverse turn conformation. The (Φ,Ψ) dihedral angles of residue Pro2 (−81°,54°) correspond to an almost ideal inverse γ turn.[46,47] The narrow range of (Φ,Ψ) dihedral angles allowed for proline residues includes values corresponding to an inverse γ turn, the proline being at position $i + 1$. Inverse γ turns have been reported in cyclopentapeptides around a proline residue.[64,65] To our knowledge, inverse γ turns in cyclohexapeptides have not been reported either in crystal structures or in solution studies. However, a simulation study of the hexapeptide cyclo[(Ala-Pro-DPhe)$_2$] has shown that this peptide, which displays a double β turn structure in the crystal, was able to adopt some γ turn containing structures during the simulation.[66] This observation was shown to be consistent with NMR and vibrational data. However, when the simulation includes the intermolecular forces present in the crystal, the molecule adopts a double β turn, as observed in the crystal structure. These observations may suggest that, in cyclohexapeptides, γ turns may be more populated in solution than in the crystal, where intermolecular forces tends to deviate the equilibrium toward the β turn structures. These observations suggest that cyclohexapeptides possess an intrinsic conformational variability. The present work demonstrates that the cyclohexapeptide under study exhibit conformational variability in solution. It also shows that the interconverting partners differ not only in the orientation of the peptide planes of the reverse turns but also in the topography of the backbone as well as in the orientation of the side chains. In this respect it can be seen in Figure 3A,B that in Cl1 and Cl3 structures the side chain of Phe3 adopts a ($g-$) rotamer, whereas in Cl2 and Cl4 structures this residue adopts a ($g+$) rotamer. This observation indicates that for both Cl1-Cl2 and Cl3-Cl4 there is a correlation between the conformation adopted by the side chain of residue Phe3 and the conformation of the backbone in this part of the peptide and in the rest of the molecule. Therefore, the conformations obtained by means of the EA protocol may suggest that the transition between the interconverting partners Cl1-Cl2 (or Cl3-Cl4) of peptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle) can be a concerted process that simultaneously involves the conformation of the two reverse turns as well as the conformations of the side chains.

Ensemble-Averaging Protocol. Several strategies have been proposed to take into account conformational variability in structure determination from NMR data. The use of time-averaged interproton distances (TAR)[67−69] and time-averaged *J*-couplings[70,71] in restrained molecular dynamics has been shown to increase the "mobility" allowed the molecule under study and improve the agreement between the calculated and measured NMR parameters.[63,68,72,73] This averaging method relies on the assumption that, during the simulation, the molecule will visit all conformers which contribute to the experimental NMR data. However, if all conformers are to be visited, high-energy barriers must be crossed, and this is not possible on the time scale of the simulation. During the simulation, the molecule can be artificially forced by the restraints to shuttle between conformers that are separated by high-energy barriers. This may generate relatively large forces and tends to increase the temperature of the system during the course of the simulation.[68,71,74] In addition, this sometimes creates large conformational fluctuations which hamper analysis of the resulting ensemble of conformers.

In principle, one advantage of the EA approach is that, in contrast to molecular dynamics simulation of a single molecule, several conformers separated by high-energy barriers at room temperature may be included to reproduce the NMR restraints. Fennen and co-workers have objected that the use of identical starting structures in the EA protocol may severely limit this approach due to the existence of insurmountable energy barriers between the conformers that are needed to reproduce the NMR data.[17] This remark does not take into account the fact that the studies in question take advantage of high-temperature simulated annealing to optimize the structures in the course of the refinement.[10,12,18] This may allow the molecules to visit conformations that are separated by high-energy barriers at room temperature. In the present work, in addition to the use of high-temperature SA, 20 runs were carried out in order to start with different peptide structures. The conformers randomly generated at the beginning of the present EA procedure display an marked conformational heterogeneity prior to refinement of these sets of structures by application of the NMR-derived restraints (data not shown). There are only two pairs of conformers returned by the EA protocol, and their structures do not depend on the starting conformations. This indicates that the high-temperature

(61) Nemethy, G.; Scheraga, H. A. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 320−327.
(62) Karle, I. L. In *The Peptides*; Gross, E., Meienhofer, J., Eds.; Academic Press: New York, 1981; Vol. 4, pp 1−54.
(63) Cuniasse, P.; Raynal, I.; Lecoq, A.; Yiotakis, A.; Dive, V. *J. Med. Chem.* **1995**, *38*, 553−564.
(64) Pease, L. G.; Watson, C. *J. Am. Chem. Soc.* **1978**, *100*, 1279−1286.
(65) Karle, I. L. *J. Am. Chem. Soc.* **1978**, *100*, 1286−1289.
(66) Kitson, D. H.; Hagler, A. T. *Biochemistry* **1988**, *27*, 5246−5257.
(67) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(68) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(69) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 467−479.
(70) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. *J. Biomol. NMR* **1993**, *3*, 55−66.
(71) Pearlman, D. A. *J. Biomol. NMR* **1994**, *67*, 28−41.
(72) Kessler, H.; Matter, H.; Gemmecker, G.; Kling, A.; Kottenhahn, M. *J. Am. Chem. Soc.* **1991**, *113*, 7550−7563.
(73) Kessler, H.; Matter, H.; Gemmecker, G.; Kottenhahn, M.; Bats, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 4805−4818.
(74) Nanzer, A. P.; van Gunsteren, W. F.; Torda, A. E. *J. Mol. Biol.* **1995,** *6*, 313−320.

simulated annealing refinement applied in this EA study to a cyclohexapeptide allows the conformers to cross the energy barriers necessary to fit the NMR restraints. When several trials of random structures are used as the starting point of the EA protocol, the ability of the high-temperature SA to allow these conformers to cross high-energy barriers affects the yield of the process. For the study of protein structures by EA protocols, it is more likely than for cyclohexapeptides that certain energy barriers remain insurmountable under high-temperature SA. It should be noted that random conformations can easily be generated for small peptides like the one studied here but this is not feasible for larger molecules like proteins. For proteins, it is conceivable to create a set of different starting structures using random subsets of the NMR data by DG or SA methods in order to create sufficient conformational diversity in the starting ensembles prior to EA refinement.

One common drawback of the TAR and EA approaches is that they increase the number of adjustable parameters. Indeed, the basic idea of the TAR approach is to treat NOEs like quantities that must be satisfied on average over the course of a restrained molecular dynamics trajectory.[67,68] In the EA approach, the experimental NMR restraints must be satisfied on average for a considered ensemble of structures. Thus, these two methods, when successful, improve the fit of the NMR restraints by averaging the corresponding distances, and this is paralleled by an increase in the number of degrees of freedom of the system. This raises the question of the significance of the fit. This has not been addressed either in reported TAR studies or in most EA studies. Mierke and co-workers have reported the application of an EA approach to a pentapeptide where a very large ensemble (about 300 structures) was used with a data set containing 20 DR and 8 *J*-couplings.[11] It is likely that the excellent agreement reported by these authors for both the NOE and *J*-coupling restraints only reflects the increase in the number of adjustable parameters and is therefore meaningless. To avoid this situation, the complete cross-validation methods can be used to estimate the significance of the fit of the NMR restraints in EA studies. Bonvin and Brunger demonstrated a correlation between the cross-validated measure of the fit and the number of conformers that best reproduce the conformational variability in solution in a synthetic case.[12] Using this method for the protein IL8, these authors proved that a two-conformers model significantly increases the fit of the distance restraints, demonstrating the presence of conformational variability in solution for a loop region of this protein.[12] In the present study, the application of the complete cross-validation method led to a similar conclusion for the cyclohexapeptide studied. However, two different pairs of interconverting conformers are defined and no experimental evidence can be used to discriminate between the two models.

In the present EA protocol, in contrast to previous studies,[10,17,20,75] the number of interconverting conformers is not arbitrarily chosen but is treated as an adjustable parameter. This feature makes this EA protocol very general, as it is therefore not restricted to cases where conformational variability is present but can also be used to demonstrate that a single conformer gives the best fit of the NMR restraints. This situation has been encountered in the complete cross-validation study of IL-4, where it has been shown that a single-conformer model gives the most accurate representation of the NMR data.[12] Note that the optimum number determined by the EA protocol for a given molecule may clearly depend on the number of conformers which actually interconvert on the NMR time scale. However, the defined optimum number will depend on the size of the NMR dataset as compared to the number of degrees of freedom of the system, as well as on the accuracy of these data. In this respect, Bonvin and Brunger have shown that, once multiple conformers were identified, cross-validation was unsuccessful in assessing the relative population of multiconformer structures from a set of qualitative distance restraints.[18] They show that the use of tight DR, i.e., more precise and possibly accurate restraints, may allow evaluation of these relative populations, but even in this case, the result must be interpreted with caution. This led us to take the maximum restraints into account and to use the ADR and *J*-couplings in addition to the DR. The set of DR was derived from off-resonance ROESY experiments to give accurate restraints. However, the possibility remains that the intrinsic density of NMR-derived restraints obtainable in a given structure, i.e., limited by the proton density for a ^{1}H homonuclear study, will restrict the number of conformers that can be involved to give a significant fit of these restraints.

(75) Scarsdale, J. N.; Ram, P.; Yu, R. K.; Prestegard, J. H. *J. Comput. Chem.* **1988**, *9*, 133−147.

Conclusions

Constrained peptides or pseudopeptides have been widely used to probe the biological activity of these molecules. The aim of introducing conformational restrictions into a peptide is to obtain a unique conformation for two reasons. First, the solution conformation of the constrained molecule can be easily determined in the free state in solution. Second, this conformation can be expected to be retained in the bound state. The present work demonstrates that for the cyclohexapeptide cyclo(Gly-Pro-Phe-Gly-Pro-Nle), there is conformational variability in solution, as depicted by the cross-validated EA approach of a complete and accurate NMR data set. In the different interconverting conformers found for this molecule, the orientation of the peptide planes can differ by 180°, giving very different positions of the carbonyl oxygens and amide protons which can be implicated in interactions with a target. The relative orientations of the side chains can also differ between one conformer and its interconverting partner. Moreover, we stress that the intrinsic conformational variability revealed by the present approach suggests that such a cyclohexapeptide may undergo a conformational transition from the free state to the bound state, the so-called induced-fit. In this case, the different conformations depicted by the cross-validated EA approach may be irrelevant to probe the conformational requirements for activity. The overall conclusions of this work should preclude strategies based on the single-structure hypothesis to determine the conformation of cyclopeptides or cyclopseudopeptides from NMR data particularly when there is evidence that several conformations occur. Where the EA approach demonstrates the presence of conformational variability, more constrained and/or smaller size molecules can be developed to probe the conformational requirements for activity from free-state conformational studies.

Acknowledgment. Drs. P. Berthault and H. Desvaux are gratefully acknowledged for their assistance in recording the off-resonance ROESY experiments. We thank M. Guenneugues who wrote the software used to extract the interproton distances from ROESY experiments.

JA9636810