Structural and Dynamic Properties of a β-Hairpin-Forming Linear Peptide. 1. Modeling Using Ensemble-Averaged Constraints

Keith L. Constantine,*,† Luciano Mueller,† Niels H. Andersen,§ Hui Tong,§ Charles F. Wandler,§ Mark S. Friedrichs,† and Robert E. Bruccoleri‡

Contribution from the Departments of Macromolecular NMR and Macromolecular Modeling, Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543-4000, and the Department of Chemistry, University of Washington, Seattle, Washington 98195

Received June 22, 1995⊗

Abstract: The linear peptide Y-Q-N-P-D-G-S-Q-A (one letter amino acid code) displays a high population of β-hairpin conformations in aqueous solution at 5 °C (F. J. Blanco et al. *J. Am. Chem. Soc.* **1993**, *115*, 5887−5888), indicating that it should be a useful model system for elucidating local interactions that induce and stabilize β-hairpins. Against this background, we have performed a detailed study of the peptide's conformational and dynamic properties using 2D NMR and computational modeling. Using the linear component of NOE buildup curves, 122 nuclear Overhauser effect (NOE) distance constraints were derived for the major (*trans*-Pro-4) isomer. These distance constraints and three dihedral angle constraints were used in conjunction with simulated annealing (X-PLOR program) to produce a well-converged set of 24 solution structures. The individual structures all contain significant NOE constraint violations. The average RMS violation is 0.25 Å, and the average maximum violation is 1.02 Å. This result prompted a further evaluation of the NOE distance constraints using the DISCON program, which iteratively removes spin diffusion contributions. Reduced, but still significant, violations were observed after reminimizing with reduced repulsive interactions and with loosened distance constraints corresponding to the widest bounds derived from the DISCON and NOE buildup calculations. To investigate whether the residual constraint violations reflect conformational averaging, selected sets of constraints were treated in an ensemble-averaged fashion within the CONGEN molecular modeling program. In addition to the 122 NOE constraints, 41 ^3J coupling constant constraints and 55 "no NOE" constraints were also used in the CONGEN calculations. Constraint satisfaction and physical energies improved significantly when ensemble-averaging was applied. The predominant solution conformation contains a type I β-turn, with Pro-4 and Asp-5 occupying the corners. Major stabilizing interactions include backbone-backbone hydrogen bonds involving Asn-3, Gly-6, and Ser-7, hydrogen bonds involving the Asn-3 side chain, and contacts between Tyr-1 and Ala-9. Thus, the peptide's conformation space is restricted by hydrogen bonding interactions with varying degrees of occupancy and by the presence of a proline residue. Hydrophobic, hydrogen bonding, and electrostatic interactions occur between Tyr-1 and Ala-9, but the ensemble-averaged calculations indicate that these interactions are transient. These conclusions are supported by ^{13}C relaxation measurements and a water-solvated unrestrained molecular dynamics simulation [M.S. Friedrichs et al. *J. Am. Chem. Soc.* **1995**, *117*, 10855−10864].

Introduction

The propensities of small protein fragments and model peptides to manifest detectable conformer populations with elements of secondary structure has attracted a great deal of attention in recent years.[1,2] Studies of the conformational properties of unconstrained peptides are of interest for two primary reasons. First, their behavior may resemble initiation events in protein folding, such as the formation of nascent helices[1b] and reverse turns.[1c] Second, the ordered states may sample bioactive conformations. Compared to globular proteins, most small peptides cannot bury a large fraction of their hydrophobic surface area upon adopting an ordered conformation under nonaggregating conditions. Therefore, well-ordered hydrophobic packing and the "hydrophobic effect", which play major roles in cooperative folding and stability of globular proteins,[3] may not always be as important for stabilizing the ordered conformations of small peptides. In addition, high conformational entropy favors the disordered states of unconstrained peptides. Hence, significant populations of ordered states require that the contributions of the local noncovalent

† Department of Macromolecular NMR, Bristol-Myers Squibb.
‡ Department of Macromolecular Modeling, Bristol-Myers Squibb.
§ Department of Chemistry, University of Washington.
⊗ Abstract published in *Advance ACS Abstracts*, November 1, 1995.

(1) (a) Jimenez, M. A.; Nieto, J. L.; Herranz, J.; Rico, M.; Santoro, J. *FEBS Lett.* **1987**, *221*, 320−324. (b) Dyson, H. J.; Rance, M.; Houghteen, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161−200. (c) Dyson, H. J.; Rance, M.; Houghteen, R. A.; Wright, P. E.; Lerner, R. A. *J. Mol. Biol.* **1988**, *201*, 201−217. (d) Reed, J.; Hull, W. E.; Lieth, C. W.; Kubler, D.; Subai, S.; Kinzel, V. *Eur. J. Biochem.* **1988**, *178*, 141−154. (e) Goodman, E. M.; Kim, P. S. *Biochemistry* **1989**, *28*, 4343−4347. (f) Blanco, F. J.; Jimenez, M. A.; Rico, M.; Santoro, M.; Herranz, J.; Nieto, J. L. *Eur. J. Biochem.* **1991**, *200*, 345−351. (g) Bruch, M. D.; Dhingra, M. M.; Gierasch, L. M. *Proteins: Struct. Funct. Genet.* **1991**, *10*, 130−139. (h) Dyson, H. J.; Wright, P. E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519−538. (i) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95−118. (j) Constantine, K. L.; Mapelli, C.; Meyers, C. A.; Friedrichs, M. S.; Krystek, S.; Mueller, L. *J. Biol. Chem.* **1993**, *268*, 22830−22837. (k) Merutka, G.; Morikis, D.; Bruschweiler, R.; Wright, P. E. *Biochemistry* **1993**, *32*, 13089−13097. (l) Aruidsson, K.; Jaruet, J.; Allard, P.; Ehrenberg, A. *J. Biomolec. NMR* **1994**, *4*, 653−672.

(2) (a) Tobias, D. J.; Mertz, J. E.; Brooks, C. L., III *Biochemistry* **1991**, *30*, 6054−6058. (b) Tobias, D. J.; Brooks, C. L., III *Biochemistry* **1991**, *30*, 6059−6070.

0002-7863/95/1517-10841$09.00/0

interactions are sufficient to overcome the loss of conformational entropy. In general, the balance of forces that stabilize the folded states of small peptides may differ from the balance typical of globular proteins in their native states.

Investigations of the solution conformations of unconstrained peptides have been performed using experimental[1] and theoretical[2] methods. Peptides have been observed to adopt helical and β-turn structures. In general, these data have been interpreted in terms of transient, fractionally populated ordered conformations. Among the important issues raised by these studies are questions regarding the residual flexibility and multiplicity of the ordered states. A striking example of multiple ordered states has been reported for a 17-residue model peptide in aqueous solution.[1k] In that case, the data indicate "the occurrence of multiple conformations involving a helical backbone structure and at least one alternative backbone conformation, allowing the clustering of Leu-9 with the N-terminal side chains". One of the goals of the present work is to further develop modeling protocols capable of handling cases of conformational equilibria.

As noted above, numerous peptides display helical and β-turn conformations. Much less common is the occurrence of full β-hairpins.[1h,4,5] A 21 residue peptide corresponding to the N-terminal β-hairpin of ubiquitin has been observed to form a β-hairpin in 30% and 60% methanol.[4a] Transient β-hairpins have been observed for 16 and 18 residue fragments of the protein-G B1 domain in aqueous trifluoroethanol and aqueous solutions.[4b,c] The design and characterization of a linear nine residue peptide that adopts a highly populated β-hairpin conformation in aqueous solution at 5 °C has been reported.[5] The sequence of this peptide is Y-Q-N-P-D-G-S-Q-A (one letter amino acid code). Using ^{1}H NMR techniques, the authors demonstrated that this peptide is monomeric up to concentrations of at least 15 mM in aqueous solution. H^N chemical shift temperature coefficients and 2D NOESY experiments were used to evaluate the peptide's conformational preferences. The authors noted that the observed NOEs "cannot be rigorously interpreted in terms of a unique structure because of the existence of conformational averaging". Nevertheless, 19 structures, having an average root-mean-square deviation (RMSD) of 1.4 Å, were calculated using 37 qualitative NOE distance constraints. These structures displayed a fully formed β-hairpin, with a central β-turn involving Asn-3, Pro-4, Asp-5, and Gly-6.

The reported results[5] demonstrate that the Y-Q-N-P-D-G-S-Q-A peptide can serve as an important model system for investigating the structural and dynamic properties of nascent β-hairpins. We have undertaken an extensive characterization of this peptide using ^{1}H homonuclear and ^{1}H-^{13}C heteronuclear NMR methods. Distance constraints were derived by the quantitative fitting of NOE buildup curves and by the application of a modified version of the DISCON algorithm,[6] which iteratively removes second-order spin diffusion contributions to the NOEs. Structures were calculated by simulated annealing with the X-PLOR program[7] using "standard constraints and by simulated annealing with the CONGEN program[8] using both standard and ensemble-averaged constraints. The resulting structures are analyzed in terms in β-hairpin stabilizing interactions, and the conformational distribution is characterized by the ensemble-averaged computations. The strengths and limitations of the methods employed are discussed.

Background and Overview for Ensemble-Averaged Calculations

Conformational averaging affects the cross-relaxation rate R_{ij} between two protons i and j.[9] The effective functional form of the motional correlation functions that mediate relaxation processes depends on the time scale of the motion.[10] For motions occurring on time scales that are long relative to the overall rotational correlation time (τ_r), radial fluctuations in the internuclear separation d_{ij} lead to the following dependence for the intergroup magnetization transfer rate (σ_{ij})

$$\sigma_{ij} = -R_{ij} \propto \langle d_{ij}^{-6} \rangle \tag{1}$$

In this regime, the contributions made by the shortest approachs between i and j are heavily weighted. Motions occurring on time scales shorter than τ_r lead to

$$\sigma_{ij} \propto \langle d_{ij}^{-3} \rangle^2 \tag{2}$$

The enhancement of σ_{ij} due to radial fluctuations is not as great in the case of fast motions. In addition, the enhancement can be offset by angular fluctuations of the interproton vector.[9]

Vicinal J coupling constants are averaged over all conformational fluctuations that are fast on the NMR time scale.[11] For a static conformation, the coupling constants depend on the spanned dihedral angle θ according to the Karplus equation[12]

$$J(\theta) = A\cos^2(\theta) + B\cos(\theta) + C \tag{3}$$

where A, B, and C are empirical parameters specific for spin pairs in a particular type of chemical moiety.[13] If θ fluctuates,

(3) (a) Privalov, P. L.; Gill, S. J. *Adv. Prot. Chem.* **1988**, *39*, 191–234. (b) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155. (c) Pace, C. N. *Trends Biochem. Sci.* **1990**, *15*, 14–17. (d) Freire, E.; Murphy, K. P. *J. Mol. Biol.* **1991**, *222*, 687–698. (e) Murphy, K. P.; Freire, E. *Adv. Prot. Chem.* **1992**, *43*, 313–361. (f) Sneddon, S. F.; Tobias, D. J. *Biochemistry* **1992**, *31*, 2842–2846.

(4) (a) Cox, J. P. L.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Wollfson, D. N. *J. Mol. Biol.* **1993**, *234*, 483–492. (b) Blanco, F. J.; Jimenez, M. A.; Pineda, A.; Rico, M.; Santoro, J.; Nieto, J. L. *Biochemistry* **1994**, *33*, 6004–6014. (c) Blanco, F. J.; Rivas, G.; Serrano, L. *Nature Struct. Biol.* **1994**, *1*, 584–590.

(5) Blanco, F. J.; Jimenez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. *J. Am. Chem. Soc.* **1993**, *115*, 5887–5888.

(6) Lai, X.; Chen, C.; Andersen, N. H. *J. Magn. Reson.* **1993**, *B 101*, 271–288.

(7) Brünger, A. T. *X-PLOR (Version 3.1) Manual*; Yale University Press: New Haven, CT, 1992.

(8) (a) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217. (b) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137–168. (c) Bruccoleri, R. E.; Haber, E.; Novotny, J. *Nature* **1988**, *335*, 564–568. (d) Bruccoleri, R. E. *CONGEN (Version 2.0) Manual*; Bristol-Myers Squibb Pharmaceutical Research Institute: Princeton, NJ, 1992. (e) Bruccoleri, R. E. *Molecular Simulation* **1993**, *10*, 151–174.

(9) (a) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289–2302. (b) Chandrasekar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. *J. Mol. Biol.* **1992**, *226*, 239–250. (c) Palmer, A. G.; Case, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 9059–9067. (d) Post, C. B. *J. Mol. Biol.* **1992**, *224*, 1087–1101. (e) Brüschweiler, R.; Case, D. A. *Prog. NMR Spec.* **1994**, *26*, 27–58.

(10) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH Publishers: New York, 1989; pp 510–511.

(11) (a) Pople, J. A. *Mol. Phys.* **1958**, *1*, 3–8. (b) Kessler, H.; Griesinger, C.; Müller, A.; Lautz, J.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393–3396. (c) Constantine, K. L.; DeMarco, A.; Madrid, M.; Brooks, C. L., III; Llinas, M. *Biopolymers* **1990**, *30*, 239–256. (d) Karimi-Nejad, Y.; Schmidt, J. M.; Rüterjans, H.; Schwalbe, H.; Griesinger, C. *Biochemistry* **1994**, *33*, 5481–5492.

(12) (a) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11–15. (b) Karplus, M. *J. Chem. Soc.* **1963**, *85*, 2870–2871.

(13) (a) Bystrov, V. F. *Prog. NMR. Spec.* **1976**, *10*, 41–81. (b) DeMarco, A.; Llinas, M.; Wüthrich, K. *Biopolymers* **1978**, *17*, 617–636. (c) DeMarco, A.; Llinas, M.; Wüthrich, K. *Biopolymers* **1978**, *17*, 617–636. (d) Fischman, A. J.; Live, D. H.; Wyssbrod, H. R.; Agosta, W. C.; Cowburn, D. *J. Am. Chem. Soc.* **1980**, *102*, 2533–2539.

eq 3 is replaced by

$$\langle J(\theta)\rangle = A\langle\cos^2(\theta)\rangle + B\langle\cos(\theta)\rangle + C \qquad (4)$$

Several methods that incorporate the effects of conformational averaging into molecular modeling protocols have been developed. Ernst and co-workers have devised the MEDUSA procedure.[14a,b] In this approach, a large number of different conformations are generated, each of which satisfies a subset of the constraints. All pairs (or triplets) of conformers are considered as possible conformational exchange systems, with those pairs (or triplets) showing the best overall constraint satisfaction being regarded as the most likely conformers. This approach becomes very computationally expensive as the number of structures and/or constraints increases. Recently, a method of clustering conformers produced by MEDUSA and estimating their populations has been described.[14c]

A second approach involves the use of time-averaged constraints.[15] In a typical structure determination using restrained molecular dynamics (RMD), distance constraints are implemented as potential energy functions that depend on the instantaneous value of $d_{ij}(t)$. In the time-averaged approach, the instantaneous $d_{ij}(t)$ is replaced by a time-averaged quantity, $\langle d_{ij}^{-x}\rangle^{-1/x}(t)$, where $x = 3$ or 6. The average is not calculated over the entire trajectory, since $\langle d_{ij}^{-x}\rangle^{-1/x}(t)$ stabilizes and the system becomes unresponsive. Instead, an exponential decay term, with a user-defined characteristic decay rate, is introduced.[15a] This tends to drive conformational transitions at rates comparable to the characteristic decay rate,[15b,d,e] which can be much faster than the actual rate at which the real system samples different conformations. Therefore, the choice of $x = 3$ or 6 should be made on the basis of experimental evidence regarding the time scale of the motional processes under investigation. Time-averaged coupling constant constraints have been implemented in an analogous fashion.[15g,j]

Ensemble-averaged constraints offer a third approach to the treatment of conformational averaging.[15h,16] In this approach, multiple structures are refined "in parallel"; e.g., in a RMD simulation, all members of the ensemble advance in a lockstep fashion along their respective trajectories. While the conformations, and hence the physical forces, are different for each structure in the ensemble, the individual conformational parameters (NOE distance or derived J value) used in the restraint potential are replaced by the appropriate ensemble-averaged values at each update. Thus, all structures are subjected to the same restraining forces, and constraint satisfaction is gauged by how well the *ensemble as a whole* reproduces the experimental distance and J coupling constraints. This approach has several advantages. Since the method employs ensemble rather than time-averages, it is also suitable for restrained energy minimization, conformational search, and Monte Carlo calculations. The number of ensemble members, i.e., the number of alternate conformations, is under direct user control. Also, the starting conformations can be as similar or different as deemed appropriate. Therefore, this is the approach that we have selected to use.

An ensemble-averaging facility has been incorporated into the CONGEN program.[8] All constraints, or any subset of constraints, can be ensemble-averaged. For a distance constraint between protons (or groups) i and j, a term of the following form is included in the potential energy function[8]

$$V_{NOE} = \begin{cases} 0.0 & \text{if } u_{ij} \geq \langle r_{ij}\rangle \geq l_{ij} \\ K_N(\langle r_{ij}\rangle - l_{ij})^2 & \text{if } \langle r_{ij}\rangle < l_{ij} \\ K_N(u_{ij} - \langle r_{ij}\rangle)^2 & \text{if } u_{ij} < \langle r_{ij}\rangle < r_{sw} \\ K_N(r_{plat} - \langle r_{ij}\rangle)^2 + 2E_{max} & \text{if } r_{sw} < \langle r_{ij}\rangle < r_{out} \\ K_N \text{ SLOPE}(\langle r_{ij}\rangle - r_{out}) + E_{max2} & \text{if } \langle r_{ij}\rangle \geq r_{out} \end{cases} \qquad (5)$$

where l_{ij} and u_{ij} are the experimental lower and upper bounds, respectively, and $\langle r_{ij}\rangle$ is the ensemble-averaged distance. At the point r_{sw}, the function switches to an inverted harmonic, and at the point r_{plat}, the derivative of the inverted harmonic is zero. E_{max} is the energy where the inverted harmonic switches on. E_{max2} is the energy where the function goes linear (with slope = K_N SLOPE) at point r_{out}. For an ensemble of N members, $\langle r_{ij}\rangle$ is given by

$$\langle r_{ij}\rangle = [(1/N)\sum_{n=1}^{N}(r_{ij})_n^{-x}]^{-1/x}, \quad x = 3 \text{ or } 6 \qquad (6)$$

For single proton−single proton NOEs, e.g., a sequential H^N−H^N interaction, r_{ij} in eq 6 is just the distance d_{ij} observed in structure n. For NOEs involving equivalent or nonstereospecifically assigned prochiral groups, e.g., a methyl−methyl interaction, r_{ij} is the effective distance given by

$$r_{ij} = [\sum(d_{ij})^{-6}]^{-1/6} \qquad (7)$$

where the sum is over all proton pairs between groups i and j; e.g., there are nine terms for a methyl−methyl NOE. This treatment of equivalent and nonstereospecifically assigned prochiral groups is used both for standard and ensemble-averaged distance constraints. The gradient of the NOE potential is computed analytically using the chain rule.

Coupling constants can be incorporated directly as constraints in molecular modeling algorithms.[17] For normal coupling constant constraints, terms of the following form are included in the potential energy function

$$\begin{gathered} K_J(\langle J\rangle - u_J)^2, \quad \langle J\rangle > u_J \\ V_J = 0, \quad u_J \geq \langle J\rangle \geq l_J \\ K_J(l_J - \langle J\rangle)^2, \quad \langle J\rangle < l_J \end{gathered} \qquad (8)$$

(14) (a) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3−11. (b) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960−10974. (c) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027−1033.

(15) (a) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294. (b) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235. (c) Kessler, H.; Matter, H.; Gemmecker, G.; Kling, A.; Kottenhahn, M. *J. Am. Chem. Soc.* **1991**, *113*, 7550−7563. (d) Pearlman, D. A.; Kollman, P. A. *J. Mol. Biol.* **1991**, *220*, 457−479. (e) Brunne, R. M.; Leibfritz, D. *Int. J. Peptide Protein Res.* **1992**, *40*, 401−406. (f) Kessler, H.; Geyer, A.; Matter, H.; Kock, M. *Int. J. Peptide Protein Res.* **1992**, *40*, 25−40. (g) Torda, A. E.; Brunne, R. M.; Huber, T.; Kessler, H.; van Gunsteren, W. F. *J. Biomol. NMR* **1993**, *3*, 55−66. (h) Bonvin, A. M. M. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143−149. (i) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 1−16. (j) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 279−299. (k) Abseher, R.; Ludemann, S.; Schreiber, H.; Steinhauser, O. *J. Am. Chem. Soc.* **1994**, *116*, 4006−4018. (l) Nanzer, A. P.; Poulson, F. M.; van Gunsteren, W. F.; Torda, A. E. *Biochemistry* **1994**, *33*, 14503−14511.

(16) (a) Scheek, R. M.; Torda, A. E.; Kemmink, J.; van Gunsteren, W. F. *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance*; Plenum Press: New York, 1991; pp 209−217. (b) Kemmink, J.; van Mierlo, C. P. M.; Scheek, R. M.; Creighton, T. E. *J. Mol. Biol.* **1993**, *230*, 312−322. (c) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042−1049. (d) Mierke, D. F.; Scheek, R. M.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563. Smith, L. J.; Alexandrescu, A. T.; Pitkeathly, M.; Dobson, C. M. *Structure* **1994**, *2*, 703−712.

(17) (a) Kim, Y.; Prestegard, J. H. *Proteins: Struc. Funct. Genet.* **1990**, *8*, 377−382. (b) Eberstadt, M.; Mierke, D. F.; Kock, M.; Kessler, H. *Helv. Chim. Acta* **1992**, *75*, 2583−2592. (c) Mierke, D. F.; Kessler, H. *Biopolymers* **1993**, *33*, 1003−1017.

where u_J and l_J are experimental upper and lower bounds on J, respectively, and K_J is the J constraint force constant. $\langle J \rangle$ is the ensemble-averaged J given by eq 4. The gradient of the coupling constant potential is computed using the chain rule. J-coupling constants involving nonstereospecifically assigned prochiral methylene protons were incorporated using potential energy functions that do not require the stereospecific assignments. These functions are based on the sums and the magnitudes of the differences of the J-coupling constants. The implementation and testing of these functions within CONGEN have been described elsewhere.[18]

Materials and Methods

Sample Preparation. The linear peptide Y-Q-N-P-D-G-S-Q-A was purchased from Chiron Mimotopes, Clayton, Victoria, Australia. Analytical high performance liquid chromatography data supplied with the peptide indicated a purity of 99%. The peptide was used without further purification. A ~12 mM sample (0.60 mL) was prepared in 90% ^1H_2O, 10% ^2H_2O (Isotec, Inc, Miamisburg, OH), and a ~15 mM sample (0.60 mL) was prepared in 99.999% ^2H_2O (Isotec, Inc, Miamisburg, OH). The peptide does not aggregate at these concentrations.[5] The pH was adjusted to 5.5 (uncorrected for deuterium isotope effects) by μL additions of dilute ^{2}HCl and NaO^{2}H solutions. No additional salts or buffers were added. Samples were purged and sealed under argon.

NMR Spectroscopy. All spectra were recorded on Varian Unity 600 NMR spectrometers operating at 599.91 or 598.46 MHz proton frequencies. The temperature was set to 5 °C for all experiments. ^{1}H chemical shifts were referenced to the H_2O resonance, which is at 5.00 ppm relative to an external TSP ((trimethylsilyl)[2,2,3,3-$^2H_4$]propionate) standard. The ^{13}C chemical shifts were referenced indirectly to TSP.[19]

A number of homonuclear 2D ^{1}H−^{1}H experiments were performed. DQF-COSY[20] and clean TOCSY[21] spectra were recorded for the sample in 90% ^1H_2O/10% ^2H_2O. For the clean TOCSY experiments, a mixing time (τ_m) of 62 ms was used. The radio frequency (rf) field strengths for hard pulses and TOCSY pulses were set to 37.3 and 12.1 kHz, respectively. For both samples, NOESY spectra[22] with τ_m = 50, 150, 250, 350, and 450 ms were collected. For the 100% ^2H_2O sample, a PE-COSY spectrum[23] was recorded. Water suppression was achieved by low power continuous-wave irradiation during the relaxation delays and the NOESY mixing periods. Spectral widths of 7.0 kHz in each proton dimension were used for all ^{1}H−^{1}H 2D spectra except the PE-COSY, where spectral widths of 5.0 kHz were used. The acquisition time was set to 0.293 s followed by a 1.6 s recovery delay for all experiments except the PE-COSY, where a 0.410 s acquisition time was used. For all ^{1}H−^{1}H 2D experiments, 16 or 32 scans were accumulated per t_1 increment. The number of t_1 increments recorded were 512, 512, 700, and 1024 for the DQF-COSY, TOCSY, NOESY and PE-COSY spectra, respectively.

Cosine-bell and shifted skewed cosine-bell apodizations were employed prior to Fourier transformation of the TOCSY t_1 and t_2 data, respectively, and 45° shifted sine-bell apodization was applied to the DQF-COSY t_1 and t_2 data. For the NOESY spectra, the t_1 and t_2 data were subjected to exponential-to-Gaussian conversion using an exponential broadening of −1 Hz and a Gaussian broadening of 6 Hz. In addition, the t_1 data were multiplied by a cosine function to minimize truncation errors. Low-frequency deconvolution[24] was applied to the t_2 data for all spectra aquired with the 90% ^1H_2O/10% ^2H_2O sample. The PE-COSY data were processed as described previously.[23] Zero-filling was applied to yield 4096 × 4096 spectral matrices for the NOESY, TOCSY, and DQF-COSY spectra. The PE-COSY data matrix was zero filled to 2048 × 4096.

Heteronuclear ^{1}H−^{13}C coupling constants were extracted from 2D heteronuclear multiple quantum coherence (HMQC) spectra, which were collected with the following pulse sequence

$$^1\text{H: } 90°_{\phi 1}-\Delta- \quad |-t_1/2-180°-t_1/2-|-\Delta-\text{acquire } (t_2)$$

$$^{13}\text{C: dec}(\Delta-\tau)-\tau-90°_{\phi 2}- \; t_1- \quad 90°_{\phi 3}-\tau-\text{dec}$$

The phase cycling employed was as follows: $\phi 1 = \{x, y, -x, -y\}$; $\phi 2 = \{4x, 4(-x)\}$; $\phi 3 = \{8x, 8(-x)\}$; and receiver = $\{x, -y, -x, y, 2(-x, y, x, -y), x, -y, -x, y\}$. Two datasets were acquired. For the first dataset, Δ was set to 68.2 ms, and spectra were recorded with σ delays of 36, 44, 52, 60, and 68 ms. For the second dataset, Δ was set to 90.2 ms, and spectra were recorded σ delays of 30, 45, 60, 75, and 90 ms. Broadband ^{13}C decoupling during the intervals $(\Delta - \tau)$ and $(\Delta - \tau + t_2)$ was achieved with the modulated sequence MPF7,[25] with the ^{13}C rf field strength set to 4.27 kHz. ^{13}C decoupling was also turned on for a period of 2τ prior to the first proton excitation pulse to ensure constant decoupler power dissipation for all τ values. The ^{13}C carrier frequency was set to 94.7 ppm. The heteronuclear ^{1}H-^{13}C couplings are active during the σ delays before and after the t_1-evolution period. The amplitudes $I(\tau)$ of ^{1}H−^{13}C crosspeaks are proportional to $\mathbf{I}_{max}\sin^2(\pi J\tau)$. A related pulse sequence for measuring ^{1}H−^{13}C coupling constants has recently been published.[26] Our method relies on nonlinear least squares fitting of time series data (see below), rather than measuring intensity ratios.

Constraint Derivation. Conformational constraints were derived from ^3J coupling constants and NOE buildup rates. $^3J_{HNH\alpha}$ coupling constants were estimated from simulations of a 1D ^{1}H NMR spectrum recorded with a digital resolution of 0.44 Hz/pt. These calculations were performed using Varian's LAME spin simulation program. $^3J_{H\alpha H\beta}$ coupling constant estimates were obtained from H^α−H^β crosspeak splitting patterns observed in the DQF-COSY and PE-COSY spectra. Heteronuclear ^{1}H−^{13}C coupling constants were derived by least squares fitting of the crosspeak intensites observed in the long range ^{1}H−^{13}C HMQC spectra. These data were converted to magnitude spectra, and intensities were derived from crosspeak heights and volumes. The intensities $I(\tau)$ were fit to $\mathbf{I}_{max}\sin^2(\sigma J\tau)$. $\mathbf{I}_{max}$ was derived from the most intense peak in a given column. Error estimates were obtained by repeating the fits for both datasets (see above) and by Monte Carlo simulations.

Distance constraints were derived from NOESY data using two different procedures. Both account for spin diffusion effects mediated by one intervening spin or by one intervening group of equivalent spins. The first method involved fitting of the NOESY times series data using diagonal peak-normalized NOE volumes,[6,27] using an expression modified to properly treat equivalent or degenerate protons, as described previously.[28] (In ref 28, the sign convention of σ_{ij} is the opposite of that used here.) With the exceptions of the spin population treatment and the intensity normalization procedure, this procedure is analogous

(18) Constantine, K. L.; Friedrichs, M. S.; Mueller, L.; Bruccoleri, R. E. *J. Magn. Reson.* **1995**, *B108*, 176−184.
(19) Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Reizer, J.; Saier, M. H., Jr.; Wright, P. E. *Biochemistry* **1992**, *31*, 4413−4425.
(20) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(21) (a) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528. (b) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355−360. (c) Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870−7872.
(22) (a) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1−6. (b) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95−117. (c) Bodenhausen, G.; Kolger, H.; Ernst, R. R. *J. Magn. Reson.* **1984**, *58*, 370−388.
(23) Mueller, L. *J. Magn. Reson.* **1987**, *72*, 191−197.
(24) Marion, D.; Ikura, M.; Bax, A. J. *Magn. Reson.* **1989**, *84*, 425−428.
(25) Fujikawa, T.; Anal, T.; Kuriha, N.; Nagayama, K. *J. Magn. Reson.* **1993**, *A104*, 103−105.
(26) Zhu, G.; Renwick, A.; Bax, A. *J. Magn. Reson.* **1994**, *A110*, 257−261.
(27) (a) Macura, S.; Farmer, B. T., II; Brown, L. R. *J. Magn. Reson.* **1986**, *70*, 493−499. (b) Andersen, N. H.; Eaton, H. L.; Lai, X. *Magn. Reson. in Chem.* **1989**, *27*, 515−528. (c) Andersen, N. H.; Lai, X.; Hammen, P. K.; Marschner, T. M. *NMR Applications in Biopolymers*; Plenum Press: New York, 1990; pp 95−134.
(28) Constantine, K. L.; Friedrichs, M. S.; Detlefsen, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; Lin, P.-F.; Mueller, L. *J. Biomolec. NMR* **1995**, *5*, 271−286.

to the method of Baleja et al.[29] The Tyr-1 $H^{\delta}-H^{\epsilon}$ interaction was used as the reference effective magnetization transfer rate, which is 0.48 s^{-1} for the ^2H_2O data. The correct calibration distance for this interaction is the $(\sum r^{-6})^{-1/6}$ effective distance (2.19 Å) and not the distance between individual H^{δ} and H^{ϵ} protons on the same side of the ring (2.46 Å). This calibration was found to be consistent with the range of effective cross rates found for other intraresidue crosspeaks. For example, the average of effective cross rates between methylene protons is 1.56 s^{-1}. Note that $(0.48/1.56)^{1/6} \times (2.19)$ Å = 1.80 Å, which is within the expected range of 1.75 to 1.80 Å for a geminal pair of protons. Adjustments of +10% and −20% were applied to the r_{ij} in order to establish upper and lower bound distance constraints, respectively, with no lower bounds being allowed to exceed 3.0 Å. These adjustments are introduced to account for possible perturbations of the NOEs due to higher order spin diffusion effects and damping of the NOEs due to fast angular fluctuations of the interproton vectors.

In the second approach used for deriving distance constraints, a new automated version of the DISCON algorithm[6] was employed, which incorporates populations for equivalent spins. The details of these calculations are given in the supporting information. The minimum and maximum DISCON distance estimates were subsequently compared to the NOE buildup-derived distances to obtain the final, most conservative (i.e., with the widest distance bounds) merged constraint file—see Results.

Modeling with Standard (Nonensemble-Averaged) Constraints. A "standard" structure determination of Y-Q-N-P-D-G-S-Q-A was performed using X-PLOR.[7] Initially, distance constraints were derived from fitting the NOE buildup curves. Constraints involving equivalent and nonstereospecifically assigned protons were incorporated as $(\sum r^{-6})^{-1/16}$ effective distances.[30] Where feasible, homonuclear coupling constants were used as explicit dihedral angle constraints; most of the measured coupling constants were not used as constraints for the X-PLOR nonensemble-averaged calculations. A simulated annealing protocol, modified slightly from the standard procedure,[31] was used to refine a set of X-PLOR-generated random structures. Briefly, stage 1 consists of 2000 steps of restrained Powell minimization. Stage 2 consists of 20 ps of restrained dynamics at 800 K using "soft" potentials for the distance constraints and reduced repulsive interactions. Stage 3 consists of an additional 10 ps of restrained dynamics over which the "soft" potentials are converted to harmonic square-well potentials and the repulsive interactions are increased. Stage 4 consists of 5.6 ps of restrained dynamics over which the temperature is lowered to 100 K. During stage 5, a final 2000 steps of restrained Powell minimization is performed. A timestep of 1 fs was used for all dynamics calculations. Experimental dihedral angle constraints were incorporated in all stages with force constants of 50 kcal/(mol·rad^2). A distance constraint force constant of 50 kcal/(mol·Å^2) was used for stages 1−3, while a constant 30 kcal/(mol·Å^2) was used for stages 4 and 5. Simple repulsive van der Waals interactions were incorporated using the X-PLOR repel function, with the repel parameter initially set to 1.0 for the final minimization. Electrostatic, attractive van der Waals and empirical dihedral angle terms were excluded from all stages. The constraint set with the widest distance bounds (see above) was subsequently used to reminimize the resulting structures, with the repel parameter set to 0.8.

Modeling with Ensemble-Averaged Constraints. The merged, conservative constraint set was used for standard (i.e., one-member ensemble) and multimember ensemble-averaged simulated annealing calculations performed by CONGEN.[8] Selected sets (see **Results**) of distance constraints were ensemble-averaged according to eqs 5−7. Both homonuclear and heteronuclear coupling constant data were incorporated as constraints via eqs 4 and 8, and as described.[18] The following Karplus parameters (eqs 3 and 4) were used: for $^3J_{HNH\alpha}$, A = 6.4 Hz, B = −1.4 Hz, C = 1.9 Hz;[32] for $^3J_{H\alpha C'(i-1)}$, A = 9.0 Hz, B = −4.4 Hz, C = −0.8 Hz;[13a] for $^3J_{H\alpha H\beta}$, A = 9.5 Hz, B = −1.6 Hz, C = 1.8 Hz;[13b] and for $^3J_{C'H\beta}$, A = 7.2 Hz, B = −2.0 Hz, C = 0.6 Hz.[13d]

Briefly, the CONGEN simulated annealing protocol consisted of the following stages. High temperature (1000 K) dynamics were run for 36 ps. During this stage, the NOE and J-coupling force constants were gradually increased to 15 kcal/(mol·Å^2) and 150 kcal/(mol·Hz^2), respectively. (CONGEN force constants are not divided by 2; see eq 5). Also during this stage, the weights of the nonbonded interactions were reduced to 1% of their full values, and SLOPE (eq 5) was increased to its maximum value (60). During the second stage, the temperature was gradually lowered to 100 K over 12 ps, and the weights of the nonbonded interactions were gradually increased to 100% of their full values. During the third stage, 13 ps of dynamics at 100 K were run, and the average coordinates of the trajectory were computed over the last 12 ps. These coordinates were then subjected to 2000 steps of restrained adopted basis Newton−Raphson (ABNR) minimization to produce the final structure. A 1 fs timestep was used for the molecular dynamics. The CHARMM potential energy function[8a] was used for all CONGEN calculations.

For the ensemble-averaged calculations, sets of 120 structures were partitioned into ensembles of various sizes: 60 two-member ensembles, 40 three-member ensembles, etc. Three different treatments of the constraints were examined: all nonensemble-averaged, all ensemble-averaged, and mixed nonensemble-averaged/ensemble-averaged. The CONGEN ensemble-averaging facility allows mutually exclusive subsets of constraints to be either ensemble-averaged or treated in the standard manner. In order to compare the root-mean-square (RMS) violations of the NOE restraint distances among the various treatments, the following procedure is used. For a particular structure A in ensemble B, we have

$$\mathrm{RMS\_NOE_A} = [(1/N_r)(\sum_{\text{non-ens}} (\Delta r_{ij}^{A})^2 + \sum_{\text{ens}} (\Delta r_{ij}^{B})^2)]^{1/2} \quad (9)$$

where N_r is the total number of NOE constraints, both nonensemble-averaged and ensemble-averaged, and Δr_{ij} is the difference between the experimental constraint and modeled distance. The first sum on the right hand side of eq 9 is over the nonensemble-averaged constraints and is unique to each structure A, since the modeled distance is derived from structure A exclusively. The second sum is over the ensemble-averaged constraints and is the same for every structure in ensemble B, since the modeled distance is derived from all of the members of ensemble B *via* eq 6. The average rms NOE violation is then given by

$$\langle \mathrm{RMS\_NOE} \rangle = (1/N)(\sum_{i=1}^{N} \mathrm{RMS\_NOE}_i) \quad (10)$$

where N is the total number of structures over all ensembles. Analogous equations are applied to compute the average RMS violation of the J-coupling constraints.

Results

Assignments, NOESY Spectra, and Conformational Constraints. The ^1H NMR spectra of both the major (*trans*Pro-4, ~93%) and the minor (*cis*Pro-4, ~7%) isomers of Y-Q-N-P-D-G-S-Q-A in aqueous solution (5 °C, pH 5.5) were assigned by analyzing ^2D $^1H-^1H$ NMR spectra in the usual manner.[32] These assignments are reported in the supporting information. Our ^1H assignments closely match those reported by Nieto and co-workers[5] for the peptide in aqueous solution at 5 °C, pH 4.3. For the major isomer, the largest ^1H chemical shift difference between our results and those of Nieto and co-workers[5] is that of the Asp-5 H^N resonance, which shifts 0.08 ppm downfield upon increasing the pH. The ^{13}C resonances of the major isomer were assigned by analysis of the natural

(29) Baleja, J. D.; Muolt, J.; Sykes, B. D. *J. Magn. Reson.* **1990**, *87*, 375−384.

(30) (a) Constantine, K. L.; Madrid, M.; Banyai, L.; Trexler, M.; Patthy, L.; Llinas, M. *J. Mol. Biol.* **1992**, *223*, 281−298. (b) Constantine, K. L.; Freidrichs, M. S.; Metzler, W. J.; Wittekind, M.; Hensley, P.; Mueller, L. *J. Mol. Biol.* **1994**, *236*, 310−327.

(31) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein. Eng.* **1988**, *2*, 27−38.

(32) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.

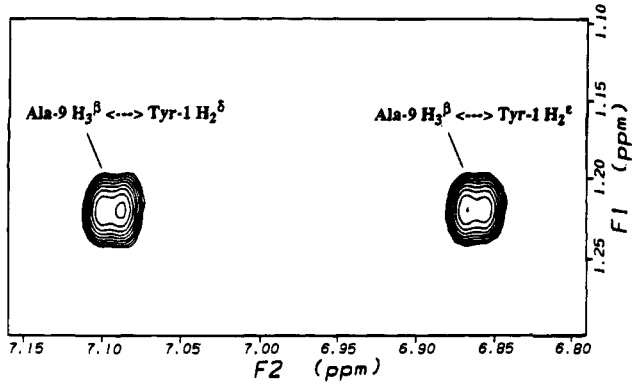

Figure 1. NOESY spectrum of Y-Q-N-P-D-G-S-Q-A: region showing interresidue crosspeaks between the side chains of Tyr-1 and Ala-9. Crosspeaks between the Ala-9 β-methyl group (F1) and the Tyr-1 H^δ and H^ϵ protons (F2) are labeled. Recorded with $\tau_m = 450$ ms at 5 °C. The peptide concentration was approximately 15 mM in 100% ^2H_2O, pD 5.5.

abundance $^1H-^{13}C$ HMQC spectra. Henceforth, all results presented and discussed will refer to the major isomer only.

In general, we observe the same qualitative pattern of sequential, medium and long-range NOEs as that reported by Nieto and co-workers;[5] therefore, a detailed discussion of the qualitative aspects of the NOESY spectra is not warranted here. However, some features are worth noting. Figure 1 shows a small region of a NOESY spectrum ($\tau_m = 450$ ms) recorded with the 100% ^2H_2O sample. This region contains prominent crosspeaks between the β-methyl group of Ala-9 and the H^δ and H^ϵ ring protons of Tyr-1, providing evidence for hydrophobic interactions between the N- and C-terminal residues and suggesting the presence of β-hairpin conformers. Also, nearly all of the observed NOESY crosspeaks are in phase with the diagonal peaks, demonstrating negative NOE enhancements. The only exceptions are intraresidue NOEs involving the side chains of Gln-2 and Gln-8. Figure 2 shows a small region of a NOESY spectrum ($\tau_m = 450$ ms) recorded with the 90% ^1H_2O/10% ^2H_2O sample. Relatively weak intraresidue crosspeaks involving the $H^{\epsilon 1}$ and H^γ/H^β protons of Glu-2 and Glu-8, reflecting positive intraresidue NOE enhancements, are shown with broken contours. In contrast, strong negative enhancements are observed between the Asn-3 $H^{\delta 1}$ and H^β protons. This indicates that the side-chain amides of Glu-2 and Glu-8 are much more mobile on the ps-ns time scale than the side-chain amide of Asn-3.[33]

Distance constraints were derived from fitting NOE buildup curves using the 50, 150, 250, 350, and 450 ms NOESY spectra. All constraints involving exchangeable protons were derived from the 90% ^1H_2O/10% ^2H_2O data, while most of the constraints involving only nonexchangeable protons were derived from the 100% ^2H_2O data. In total, 122 NOE distance constraints were obtained. These include 22 intraresidue constraints, 44 sequential constraints, 33 medium-range ($4 \geq |res_j - res_i| \geq 2$) constraints, and 23 long-range constraints. Homonuclear and heteronuclear ^3J coupling constants are reported in Table 1. For use in the standard modeling protocol with X-PLOR, the ϕ dihedral angle of Gln-2 was constrained to $-120° \pm 40°$ on the basis of $^3J_{HNH\alpha} = 8.0$ Hz and weak intraresidue $H^\alpha-H^N$ NOEs. Also, the χ^1 dihedral angles of Asn-3 and Gln-8 were constrained to $-120° \pm 100°$ (i.e., $\chi^1 \sim 60°$ was excluded) based on the $^3J_{H\alpha H\beta}$ coupling constants.

Structures Obtained Using Standard Constraints with X-PLOR. Using the 122 distance constraints derived from the

(33) (a) Uma, K.; Balaram, H.; Raghothama, S.; Balaram, P. *Biochem. Biophys. Res. Commun.* **1988**, *151*, 153-1257. (b) Krishnan, V. V.; Shekar, S. C.; Kumar, A. *J. Am. Chem. Soc.* **1991**, *113*, 7542-7550.

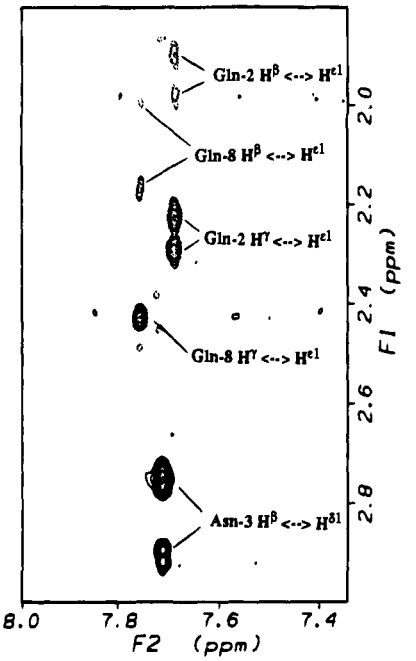

Figure 2. NOESY spectrum of Y-Q-N-P-D-G-S-Q-A: region showing intraresidue crosspeaks involving the Asn and Gln residues. Crosspeaks between the Asn $H^{\delta 1}$ and Gln $H^{\epsilon 1}$ (F2) with their respective C-bound protons (F1) are labeled. Crosspeaks involving Asn-3 (in phase with the diagonal) are drawn with solid contours, while crosspeaks involving Gln-2 and Gln-8 (180° out of phase with the diagonal) are drawn with broken contours. Recorded with $\tau_m = 450$ ms at 5 °C. The peptide concentration was approximately 15 mM in 10% ^2H_2O/90% ^1H_2O, pH 5.5.

NOE buildup fit, and the three dihedral angle constraints described above, 80 random structures were generated and then refined using simulated annealing with X-PLOR. The 24 structures with the lowest RMS distance constraint violations were selected as a representative set of solution conformations. Overall structural statistics are given in Table 2, and residue-specific information is reported in Table 3. These structures are similar to those produced by CONGEN using nonensemble-averaged constraints (Figure 3, see below).

The peptide backbone is quite well defined, as indicated by the low overall average backbone N, C′, and C^α atom RMSD to the mean structure of 0.26 Å (Table 2) and low ϕ/ψ dihedral angle standard deviations (Table 3). The somewhat higher values observed for Ser-7 ψ and Gln-8 ϕ are due primarily to two structures. Of the 24 structures, 22 have Ser-7 $\psi \sim 120° \pm 15°$ and Gln-8 $\phi \sim -40° \pm 20°$. Structure 13 has Ser-7 $\psi = -5.5°$ and Gln-8 $\phi = 58.0°$, and structure 19 has Ser-7 $\psi = 20.2°$ and Gln-8 $\phi = 50.9°$. The Gln-8 ϕ dihedral angle adopts an energetically unfavorable conformation in these latter two structures.

The high degree of backbone atom resolution obtained is somewhat surprising for a linear nine residue peptide, even one that displays strong evidence for a highly populated β-hairpin. Also, the residual NOE constraint violations are relatively high (Table 2), with an average RMS constraint violation of 0.250 Å. The average maximum violation in each structure is 1.02 Å. Of the 122 NOE constraints, 45 are violated by at least 0.30 Å in at least one structure, and 17 of these 45 constraints are violated by at least 0.30 Å in 12 or more structures. These 17 constraints include six long-range, four medium-range, six sequential, and one intraresidue constraints.

The residual NOE constraint violations could be due to a number of factors, including the particular modeling parameters employed, spin diffusion effects or multiple conformations in

Table 1. Vicinal Coupling Constants[a] for the Major Isomer of Y-Q-N-P-D-G-S-Q-A at 5 °C, pH 5.5

residue	$^3J_{HNH\alpha}$	$^3J_{H\alpha C'(i-1)}$	$^3J_{H\alpha H\beta}$	$^3J_{C'H\beta}$
Tyr-1			6.6 ± 0.3, 7.8 ± 0.3	4.7 ± 0.7, 4.0 ± 0.7
Gln-2	8.0 ± 0.2	3.5 ± 0.4	6.5 ± 0.4, 8.5 ± 0.4	−, 4.2 ± 0.6[b]
Asn-3	6.4 ± 0.2	2.0 ± 0.3	5.4 ± 0.3, 9.9 ± 0.3	2.7 ± 0.4, 3.5 ± 0.4
Pro-4		1.6 ± 0.3	7.7 ± 1.0, 7.6 ± 1.0	3.2 ± 0.8, 4.8 ± 0.5
Asp-5	7.2 ± 0.2	2.5 ± 0.4	6.3 ± 0.3, 4.2 ± 0.3	1.6 ± 0.4, 6.5 ± 0.3
Gly-6	8.1 ± 0.4, 4.5 ± 0.4	3.6 ± 0.4, 5.4 ± 0.3		
Ser-7	6.6 ± 0.2	2.3 ± 0.5	5.4 ± 0.3, 4.5 ± 0.3	2.4 ± 0.6, 6.0 ± 1.0
Gln-8	7.4 ± 0.2	2.7 ± 0.3	5.0 ± 0.4, 9.5 ± 0.4	1.7 ± 1.3[b]
Ala-9	7.1 ± 0.2	2.0 ± 0.5		

[a] All coupling constants are given in Hz. For the H^β and Gly-6 H^α resonances, the first ^3J value corresponds to the downfield resonance, and the second value refers to the upfield resonance (supporting information). [b] Not used as a constraint for the CONGEN calculations.

Table 2. Structural Statistics for the 24 Y-Q-N-P-D-G-S-Q-A Structures Produced Using NOE Buildup-Derived Standard Constraints by X-PLOR Simulated Annealing

⟨backbone N, C′, and C^α atom RMSD to mean⟩	0.26 ± 0.08 Å
⟨all heavy atom RMSD to mean⟩	0.78 ± 0.19 Å
⟨RMS distance constraint violation⟩	0.250 ± 0.017 Å
⟨number of distance constraint violations > 0.3 Å⟩	18.7 ± 3.9
⟨maximum distance constraint violation⟩	1.02 ± 0.07 Å
⟨RMS dihedral angle constraint violation⟩	0.45°
⟨RMS bond length deviation from ideality[a]⟩	0.019 Å
⟨RMS bond angle deviation from ideality⟩	3.19°
⟨RMS improper torsion deviation from ideality⟩	2.57°

[a] Deviations from ideal values refer to those in the X-PLOR parallhsa.pro parameter set.

Table 3. Average ϕ and ψ Angles and Local RMSDs for the 24Y-Q-N-P-D-G-S-Q-A Structures Produced Using NOE Buildup-Derived Standard Constraints by X-PLOR Simulated Annealing

residue	ϕ (deg)	ψ (deg)	backbone RMSD[a] (Å)	heavy-atom RMSD (Å)
Tyr-1			0.31	1.05
Gln-2	−148 ± 5	130 ± 7	0.20	0.97
Asn-3	−72 ± 5	161 ± 4	0.18	0.58
Pro-4	−58 ± 9	−21 ± 4	0.22	0.32
Asp-5	−76 ± 4	−12 ± 4	0.17	1.33
Gly-6	77 ± 6	31 ± 19	0.30	0.68
Ser-7	−147 ± 25	119 ± 36	0.40	0.96
Gln-8	−45 ± 31	111 ± 11	0.26	1.31
Ala-9			0.29	0.99

[a] Local RMSDs were computed after global fitting of the backbone atoms of residues 1−9.

Figure 3. Stereoview of the 60 best (lowest energy) CONGEN nonensemble-averaged structures showing backbone HN, N, C′, C^α, and O (including the Ala-9 carboxyl oxygens) atoms. The N-terminus is located at the lower left region of the figure.

equilibrium. The latter possibility was considered after the first two were examined, as follows. First, the NOE buildup-derived structures described above were reminimized (2000 steps) with reduced atom sizes, i.e., with the X-PLOR repel parameter set to 0.8. This yielded a set of structures with an average RMS violation of 0.167 Å, and an average maximum violation of 0.77 Å. While improved, this level of constraint satisfacation is still rather poor. Therefore, a conservative, merged bounds file was created. The maximum and minimum distances produced by DISCON were compared to the upper and lower bounds established previously by fitting the NOE buildups. For the majority of the 122 NOE buildup distance constraints, the original limits encompassed the DISCON distances. For 20 distance constraints, the upper bounds were increased on the basis of the DISCON results. In addition, 55 "no NOE" lower bound constraints of 3.0 Å were added; these were established on the basis of no observable intensity above the noise floor in the expected location in the 450 ms NOESY spectra. The final merged (conservative) set of distance constraints is available as supporting information. The merged NOE constraints and the 55 "no NOE" lower bound constraints were used to reminimize the 24 NOE buildup-derived structures (2000 steps, repel = 0.8). This yielded a set of structures with an average RMS violation of 0.150 Å for the 122 NOE constaints, an average RMS violation of 0.125 Å for all 177 distance constraints, and an average maximum violation of 0.69 Å. Thus, even after extensive precautions against spin diffusion have been taken, the constraint violations are still significant.

In a number previous studies,[11b-d,14-16] failure to achieve a high level of constraint satisfacation in any single structure has been attributed to the occurrence of multiple conformations. To examine this possibility, the peptide's conformational behavior was studied with ensemble-averaged calculations in the CONGEN program.

Benchmark Nonensemble-Averaged Structures Obtained Using CONGEN. In what follows, 120 starting structures are refined using the CONGEN ensemble-averaging facility. Four copies of the 24 final NOE buildup-derived structures calculated with X-PLOR (see above), and 24 extended models, were used as starting structures. Nonensemble, full ensemble, and mixed nonensemble/ensemble calculations were performed using all 177 distance constraints (the conservative merged NOE constraint set) and 41 J-coupling constraints. For the latter two cases, three sets of ensembles were generated: a set of 60 two-member ensembles, a set of 40 three-member ensembles, and a set of 30 four-member ensembles. As a control, in order to verify that the residual constraint violations were not caused by the 55 "no NOE" lower bounds, a set of 120 nonensemble-averaged structures was also computed with the lower bound restraints left out. The average RMS NOE constraint violation for this set was 0.125 Å, demonstrating that the lower bound restraints are not the cause of the high residual constraint violations.

In order to establish benchmarks using the full set of constraints, the 120 starting structures were subjected to nonensemble-averaged simulated annealing within CONGEN.

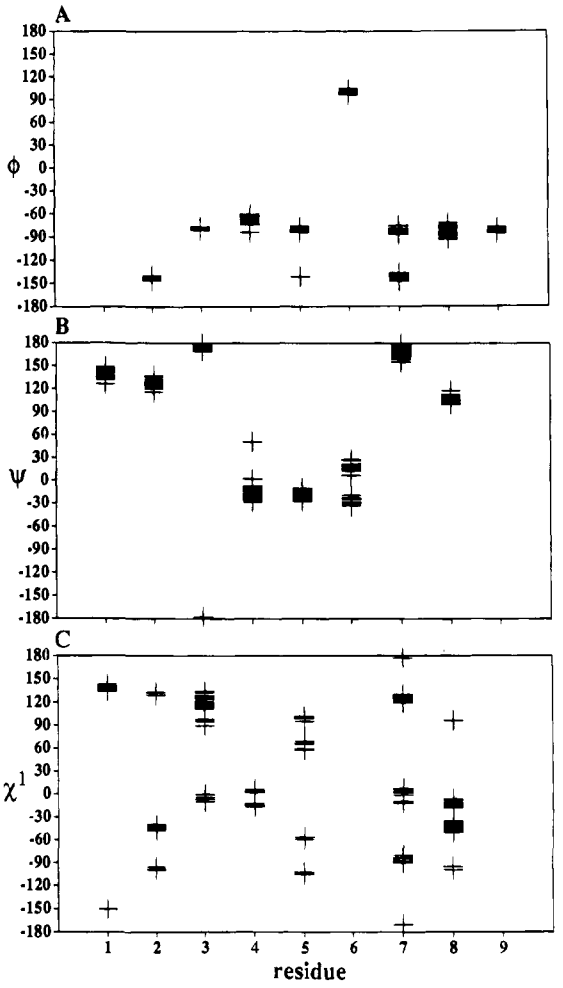

Figure 4. Residue by residue dihedral angle distributions for the 60 best (lowest energy) CONGEN nonensemble-averaged structures. The dihedral angle values are indicated by crosses (+). (A) Backbone ϕ, angles for residues 2−9. (B) Backbone ψ angles for residues 1−8. (C) Side-chain χ^1 angles for residues 1−5, 7, and 8.

The average backbone and all heavy atom RMSDs to the mean coordinates are 0.22 and 0.90 Å, respectively. As expected, the resulting structures are quite similar to those produced with X-PLOR. Energy and RMS violation statistics for the full set of 120 nonensemble-average CONGEN structures are given in Table 4. These data serve as a benchmark for the results of the ensemble-averaged calculations (see below). Figure 3 shows a stereoview (all backbone C, C^α, N, H^N, and O atoms) of the 60 nonensemble-averaged structures with the lowest total potential energies (E_{tot}). Figure 4 shows a residue-by-residue distribution of the ϕ, ψ, and χ^1 angles for the 60 lowest-energy structures. The higher energy structures were excluded from the characterization of the conformer distribution. When restricted to the lower energy subset, the average backbone and all heavy atom RMSDs to the mean coordinates reduce to 0.18 and 0.84 Å, respectively. These correspond to an overly precise set of structures, as indicated by a comparison of experimental and structure-derived C^α−H^α order parameters (following paper in this issue). Table 5 lists interresidue H-bonds observed in the subset of 60 lowest energy nonensemble-averaged structures. H-bonds linking Asn-3 O with Gly-6 H^N, and Asn-3 H^N with Ser-7 O, occur in all 60 structures. A number of less populated H-bonds also occur, including H-bonds between the Asn-3 side-chain and backbone atoms of residues Asp-5, Ser-7, and Ala-9. The H-bonding interactions stabilize a type I β-turn[34] involving residues 3−6. In total, 14 H-bonds with populations of 10% or higher were observed for these structures.

Table 4. Average, Minimum, and Maximum Energies[a] and RMS Violations[b] Obtained for the Set of 120 CONGEN Nonensemble-Averaged Structures

energies, violations	av	min.	max.
E_{tot}	63.7	39.9	101.5
E_{bond}	3.4	2.6	4.4
E_{ang}	49.5	31.5	76.4
E_{dihe}	22.8	17.1	28.4
E_{impr}	3.6	1.5	7.3
$E_{L\text{-}J}$	−20.4	−26.6	−13.6
E_{elec}	−16.3	−23.1	−9.6
$E_{H\text{-}bond}$	−17.9	−24.8	−10.9
$E_{con.\ dihe}$[c]	0.4	0.1	0.9
E_{NOE}	28.4	20.9	38.2
$E_{J\text{-}coupling}$	10.3	6.6	15.4
RMS NOE violation	0.103	0.091	0.118
RMS J violation	0.36	0.30	0.42

[a] Energies are in kcal/mol. The cutoff used for electrostatic interactions was 40 Å, with a distance dependent dielectric of $4r$. The H-bond donor heavy atom−acceptor heavy atom cutoff used was 4.5 Å, and the H-bond donor heavy atom−hydrogen−acceptor heavy atom angle cutoff used was 90°. [b] NOE violations are in Å, and J violations are in Hz. [c] Dihedral angle constaints were applied to maintain *trans* amide bonds.

The backbone ϕ (Figure 4A) and ψ (Figure 4B) angles are tightly clustered in the 60 lowest energy nonensemble-averaged structures. Two distinct clusters are observed for the ϕ angle of Ser-7 at $\sim -80°$ and $\sim -145°$. A single outlier with $\phi = -142°$ is observed for Asp-5; for the remaining structures, the ϕ of Asp-5 is $\sim -80°$. The ψ angle of Gly-6 shows two distinct clusters near −25° and +25°, with −25° associated with Ser-7 ϕ near −80° and +25° associated with Ser-7 ϕ near −145°. The side-chain χ^1 (Figure 4C) angles also tend to show overly tight clustering and energetically unfavorable conformations. This is particularly evident for χ^1 of Tyr-1, where all structures but one have χ^1 near +140°; the single outlier has $\chi^1 = -150°$. Overall, the set of nonensemble-averaged CONGEN structures is unrealistically precise and energetically strained, again demonstrating the need for ensemble-averaged calculations.

Calculations Using Full Ensemble-Averaged Constraints. Full ensemble-averaged computations were performed with x in eq 6 set to 3. This choice for x was based on ^{13}C relaxation data (following paper in this issue), which indicate that picosecond−nanosecond motions dominate the relaxation-active dynamics. Qualitatively similar results to those discussed below were obtained with x set to 6 (data not shown). Sets of 60 two-member ensembles, 40 three-member ensembles, and 30 four-member ensembles were calculated. Respectively, the average total (physical + constraint) energies for the structures in these ensembles are −23.8, −37.1, and −43.7 kcal/mol, the RMS NOE violations are 0.041, 0.029, and 0.021 Å, and the RMS J violations are 0.24, 0.19, and 0.19 Hz. Energy and violation statistics for the 60 two-member ensembles are given in Table 6. The average energy and RMS violations of the structures in the set of 60 two-member ensembles are improved dramatically over the nonensemble-averaged structures (Table 4). The energies and RMS violations improve incrementally on going from the two-member to the three-member ensembles and on going from the three-member to four-member ensembles. Thus, the energies and RMS violations continue to improve incrementally with increasing ensemble size. However, recent complete cross-validation studies[35] of ensemble-averaged mod-

(34) (a) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167−339. (b) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1−109. (c) Robson, B.; Garner, J. *Introduction to Proteins and Protein Engineering*; Elsevier Science Publishers: New York, 1986; pp 86−87.

(35) Bonvin, A. M. M. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.

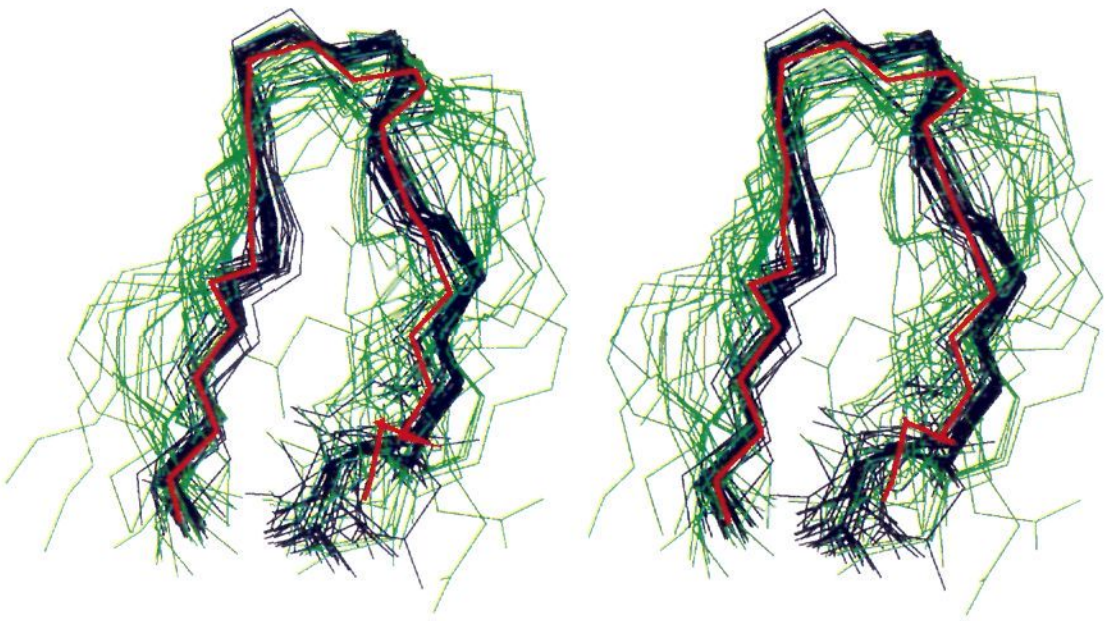

Figure 5. Stereoview of the 60 CONGEN structures from the 30 best two-member ensembles computed using full ensemble-averaged constraints. The backbone N, C′, C^{α}, and Ala-9 carboxyl oxygens atoms are shown. The mean backbone coordinates from the nonensemble structures (Figure 3) is shown in red heavy trace, and the members from each two-member ensemble nearest to the nonensemble-averaged mean structure are shown in blue. The remaining 30 structures are shown in green. The N-terminus is located at the lower left region of the figure.

eling have demonstrated that using ensembles of arbitrarily large size can result in an overfitting of the NMR distance and *J*-coupling data. Simply put, the information content of a typical NMR data set is generally not sufficient to warrant calculation of ensembles containing more than two members. This is not meant to imply that only two conformations are accessible. The range of accessible conformations must be gauged by applying cluster analysis to the set of ensembles (see below). Therefore, based on these considerations and the above results, we focus on the two-member ensembles.

As with the nonensemble case, the higher energy ensembles were excluded from the characterization of the conformer distribution. Figure 5 shows a stereoview (all backbone C, C^{α}, N, and C-terminal O atoms) of the 60 structures derived from 30 lowest-energy two-member ensembles. Figure 6 shows a residue-by-residue distribution of the ϕ, ψ, and χ^1 angles for the lowest-energy subset, and H-bond statistics are given in Table 7. The average backbone and all heavy atom RMSDs to the mean coordinates of the lowest-energy subset are 1.32 and 2.29 Å, respectively, demonstrating a significant increase in conformational sampling over the nonensemble case.

A total of 21 H-bonds with populations of 10% or higher were observed for the 60 structures derived from the 30 lowest-energy two-member ensembles (Table 7). This value is reduced to 17 when the chemical indistinguishability of the Asp-5 and C-terminal carboxylate oxygens is taken into consideration. A greater variety of H-bonds are sampled in the full-ensemble case relative to the nonensemble case. The populations of the H-bonds linking Asn-3 O with Gly-6 H^N and Asn-3 H^N with Ser-7 O, which were 100% populated in the nonensemble structures, are reduced to 65% and 55%, respectively.

The dihedral angle distributions obtained for the lowest-energy two-member ensembles (Figure 6) are much broader than those obtained from the nonensemble calculations (Figure 4). In general, the ensemble-averaged structures include members whose ϕ and/or ψ angles are within a few degrees of the ϕ and ψ angles of the nonensemble structures. The ϕ and ψ distributions expand asymmetrically relative to the nonensemble ranges, and there are a few notable outliers among the backbone dihedral angle distributions. In one structure, the Gln-2 ϕ is near 0°. A small subset of the structures have Gly-6 ϕ near −120°, and two structures have positive Ser-7 ϕ angles. The ψ angles of Ser-7 and Gln-8 are particularly disordered. The side-chain χ^1 angle distributions are also broadened relative to the nonensemble-averaged results, with many of the χ^1 angles

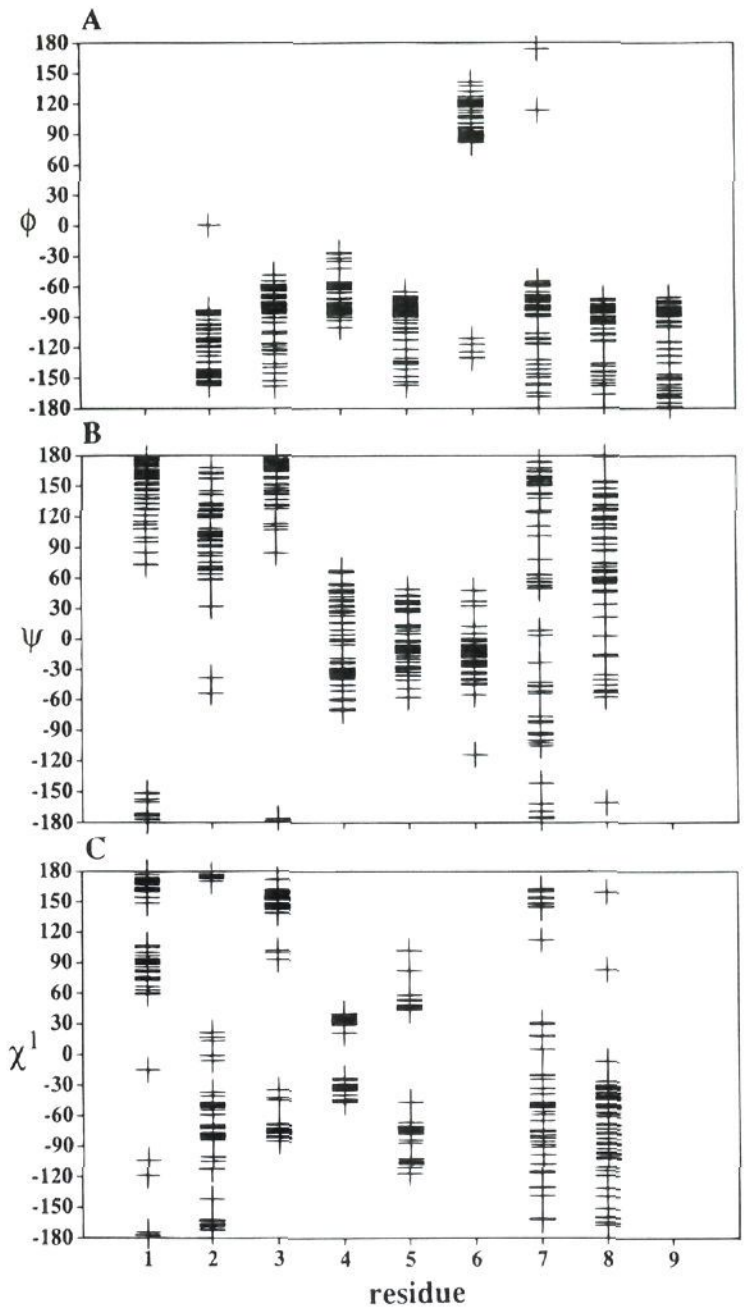

Figure 6. Residue by residue dihedral angle distributions of the 60 CONGEN structures from the 30 best two-member ensembles computed using full ensemble-averaged constraints. The dihedral angle values are indicated by crosses (+). (A) Backbone ϕ, angles for residues 2−9. (B) Backbone ψ angles for residues 1−8. (C) Side-chain χ^1 angles for residues 1−5, 7, and 8.

moving into more energetically favored conformations. Besides Pro-4, all of the side-chains are highly disordered.

The distribution of conformers obtained from the full ensemble-averaged calculations shows a noteworthy feature: at least one of the members of the two-member ensembles is

Figure 7. Stereoview of the 60 CONGEN structures from the 30 best two-member ensembles computed using the mixed nonensemble/ensemble-averaged constraints. The backbone, N, C′, C^{α}, and Ala-9 carboxyl oxygens atoms are shown. The N-terminus is located at the lower left region of the figure.

always relatively similar to the nonensemble-averaged structures. In Figure 5, the mean backbone coordinates from the nonensembled-averaged structures (Figure 3) are shown in red heavy trace. The 30 structures shown in blue denote the structure from each two-member ensemble that is nearest to the nonensemble-averaged structure; the remaining 30 structures are shown in green. For the 30 blue structures, the average backbone N, C, and C^{α} atom RMSD to the mean nonensemble-averaged structure (red trace) is 0.84 Å. By applying pairwise backbone atom RMSD cluster analysis[36] to the 60 structures with a cluster threshold of 1.83 Å, six clusters are obtained. The largest cluster contains 36 members, including all of the blue structures in Figure 5. The remaining clusters have 10, 6, 5, 2, and 1 members. These results indicate that the NMR data reflect one predominant global β-hairpin-like fold and that the dynamics of the system involves sampling of locally different conformations within the manifold of β-hairpin folds. Once one member of an ensemble satisfies most of the constraints, the other member is much less constrained. (The three-member and four-member ensemble-averaged calculations produced similar results, with at least one member always being close to the nonensemble averaged structures.) The "unconstrained" member of the ensemble may oversample conformation space, leading to an overly imprecise set of structural ensembles, as verified in the following paper in this issue.

Calculations Using Mixed Ensemble-Averaged/Nonensemble-Averaged Constraints. While the use of only ensemble-averaged constraints leads to lower energies and improved constraint satisfaction, the system becomes progressively less well determined as more structures are added to the ensembles (data not shown), with (in a two-member ensemble) one of the ensemble members becoming much less restrained, as described above. In essence, the effective number of constraints per residue decreases as the number of ensemble members increases. As suggested by the above results, these properties may result in an oversampling of the conformational space; i.e., conformers may be produced that, in reality, have vanishingly small populations. In an effort to address this problem, mixed nonensemble/ensemble calculations were performed.

A mixed ensemble-averaged/nonensemble-averaged set of distance constraints was constructed based on the following considerations. All of the available evidence, including ^{13}C relaxation data (see following paper in this issue), indicating that β-hairpin-like conformers predominate for Y-Q-N-P-D-G-S-Q-A at 5 °C. Therefore, it may be possible to achieve significantly improved constraint satisfaction (relative to the nonensemble case) without treating all of the distance constraints in the ensemble-averaged fashion. First, all of the "no NOE" lower bound distance constraints are treated as nonensemble-averaged constraints, since they are always well satisfied in the nonensemble-averaged calculations. Any distance constraint violated by at least 0.1 Å in at least one of the 60 best nonensemble-averaged CONGEN structures (see above) is treated as an ensemble-averaged constraint. All constraints involving Ala-9 or the side chains of Gln-2 or Gln-8 were ensemble-averaged, since the relaxation properties of these regions indicate that they are highly dynamic (see Figure 2 and the following paper in this issue). After this analysis, 45 NOE distance constraints (supporting information) were established as nonensemble-averaged constraints; these include 10 intraresidue constraints, 19 sequential constraints, and (most importantly) 14 medium range and two long range constraints. All J-coupling constraints were ensemble-averaged.

As with the full-ensemble calculations, sets of 60 two-member ensembles, 40 three-member ensembles, and 30 four-member mixed-constraint ensembles were calculated. The averaged total energies for structures in these ensembles are −16.4, −31.2, and −37.6 kcal/mol, the RMS NOE violations are 0.058, 0.046, and 0.042 Å, and the RMS J violations are 0.25, 0.22, and 0.21 Hz, respectively. Again, for the reasons discussed above for the full ensemble-averaged case, we focus on the two-member mixed-constraint ensembles. Energy and RMS violation statistics for all 60 two-member mixed-constraint ensembles are given in Table 8. A stereoview of the backbone atoms of 60 structures derived from the 30 lowest-energy two-member mixed constraint ensembles is shown in Figure 7, and the corresponding dihedral angle distributions are shown in Figure 8. H-bond statistics are given in Table 9. The averaged backbone and all heavy atom RMSDs to the mean coordinates of the lowest-energy subset are 0.91 and 1.57 Å, respectively.

Interestingly, the 30 lowest-energy mixed constraint ensembles show a greater variety of H-bonds than either the nonensemble or full ensemble cases, with 24 H-bonds having populations of 10% or higher (Table 9). The populations of the H-bonds linking Asn-3 O with Gly-6 H^N and Asn-3 H^N with Ser-7 O are 88.3% and 78.3%, respectively. The side chain of Asn-3 is involved in eight fractionally populated H-bonds.

The dihedral angle distributions obtained from the mixed ensemble/nonensemble calculations (Figure 8) are also much broader than the nonensemble-averaged case (Figure 4), but they are not quite as broad as those observed for the full ensemble-averaged case (Figure 6). The minor Gly-6 conformer (ϕ

(36) *QUANTA 4.0 Simulation, Search, and Analysis*; Molecular Simulations, Inc.: Burlington, MA, 1994; pp 165−172.

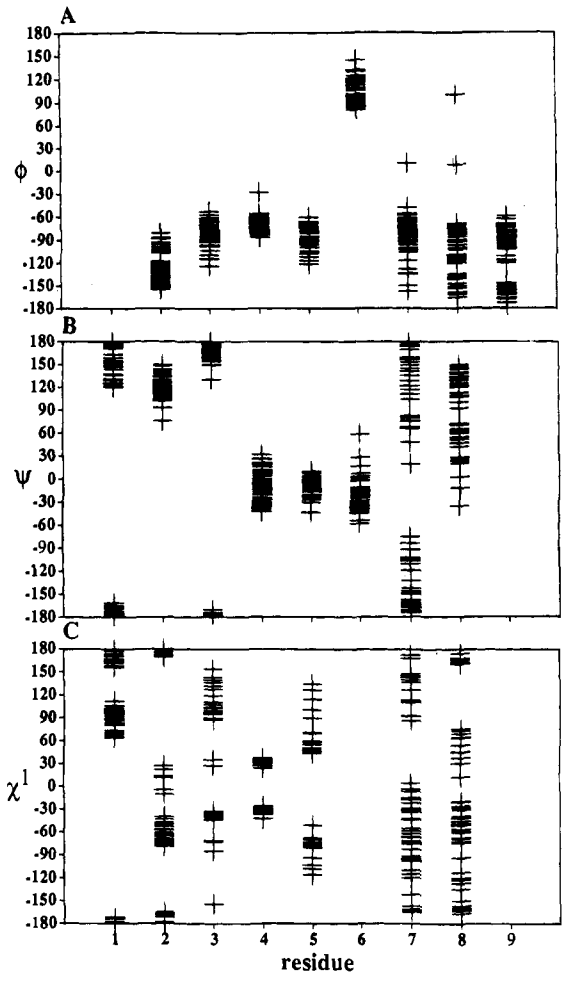

Figure 8. Residue by residue dihedral angle distributions of the 60 CONGEN structures from the 30 best two-member ensembles computed using mixed nonensemble-averaged constraints. The dihedral angle values are indicated by crosses (+). (A) Backbone ϕ, angles for residues 2−9. (B) Backbone ψ angles for residues 1−8. (C) Side-chain χ^1 angles for residues 1−5, 7, and 8.

$\sim -120°$) observed in the full ensemble case is not observed in any of the mixed ensemble/nonensemble structures. As with the full ensemble-averaged calculations, the ψ angles of Ser-7 and Gln-8 are highly variable, and, with the exception of Pro-4, all χ^1 angles are disordered.

The structures obtained by the mixed ensemble/nonensemble calculations are more evenly distributed than those produced with the full ensemble-averaged calculations. Six clusters are obtained using a backbone atom pairwise RMSD threshold of 1.04 Å. The clusters have 25, 20, 8, 3, 2, and 2 members. The residues displaying the greatest variability among the structures are Tyr-1 and Gly-6 through Ala-9. This is a physically reasonable result, since glycines and alanines are intrinsically flexible residues with either no side chain (Gly) or a relatively small side chain (Ala).

Discussion

Conformational Distributions and Hairpin-Stabilizing Interactions. The bounds on the distance constraints used to model Y-Q-N-P-D-G-S-Q-A have been thoroughly characterized, and they should be wide enough to account for any contributions made by spin diffusion to the observed NOEs. The fact that upper and lower bounds are used, rather than exact target distances and *J*-coupling constants, allows for structural variability even when a standard nonensemble-averaged structure determination is performed. Nevertheless, it is clear that the sets of structures produced by the nonensemble-averaged calculations are overly precise, energetically strained, and contain significant residual constraint violations. From this result, it was concluded that conformational dynamics affect the measured *J*-couplings and effective interproton distances, demonstrating the need for ensemble-averaged calculations.

Table 5. Interresidue Hydrogen Bonds[a] Observed in at Least 10% of the 60 Lowest Energy CONGEN Nonensemble-Averaged Structures

residue 1	acceptor 1	residue 2	donor 2	fractional occurrence	⟨distance (Å)⟩[b]
Tyr-1	O	Gln-8	HE21	0.100	1.988
Asn-3	OD1	Asp-5	HN	0.850	2.000
Asn-3	OD1	Ala-9	HN	0.133	2.116
Asn-3	O	Gly-6	HN	1.000	2.015
Asp-5	OD1	Asn-3	HD21	0.183	2.002
Asp-5	OD1	Ser-7	HN	0.200	2.085
Gly-6	O	Gln-2	HE21	0.383	2.012
Ser-7	OG	Asn-3	HD21	0.200	2.091
Ser-7	O	Asn-3	HN	1.000	1.899
Ser-7	O	Asn-3	HD21	0.317	2.181
Ser-7	O	Asn-3	HD22	0.200	2.193
Gln-8	OE1	Ser-7	HG	0.183	1.905
Ala-9	OT1	Tyr-1	HH	0.300	1.985
Ala-9	OT2	Tyr-1	HH	0.467	1.944

[a] A particular hydrogen bond is considered to be present if the donor H−acceptor O distance is less than 2.5 Å. [b] Average distance over those structures in which the hydrogen bond is present.

Using all constraints in the ensemble-averaged fashion, sets of two-member ensembles were computed that displayed much lower precision and much better physical energies and constraint satisfactions than the nonensemble-averaged structures. The blue subset of structures shown in Figure 5 is probably a better representation of the most populated class of β-hairpin conformers than are the structures presented in Figure 3, since the former have been largely freed of structural distortions induced in the latter by fitting mutually inconsistent NOE and *J*-coupling data to a single structure. Table 10 gives the average backbone ϕ and ψ angles for the full set of structures shown in Figure 3 and the subset of 30 blue structures shown in Figure 5. Most of the average backbone dihedral angles are similar between the two sets of structures. The largest difference occurs for the average ϕ angle of Ser-7, which is $-116°$ for the nonensemble-averaged structures, and is $-77°$ for the blue subset of ensemble-averaged structures. Recall that Ser-7 ϕ displayed two distinct clusters at $\sim -80°$ and $\sim -145°$ in the nonensemble averaged structures. The full ensemble-averaged structures retain a tight cluster about Ser-7 $\phi \sim -80°$ but not about Ser-7 $\phi \sim -145°$. The populations of the H-bonds linking Asn-3 O with Gly-6 H^N and Asn-3 H^N with Ser-7 O are 97% and 90%, respectively, in the blue subset of full ensemble-averaged structures, consistent with the high population observed for these H-bonds in the nonensemble-averaged structures (Table 5). A H-bond between Tyr-1 O and Ala-9 H^N is 53% populated in the blue subset of full ensemble-averaged structures but is not significantly populated in the nonensemble-averaged structures. This indicates that certain constraints, when treated in the standard fashion, preclude the formation of this H-bond. This H-bond was also significantly populated in an unrestrained, solvated molecular dynamics simulation of the peptide (following paper in this issue).

As argued above and in the following paper in this issue, the entire set of structures shown in Figure 5 appears to be overly imprecise. For example, the experimental order parameter for the Gln-8 $C^\alpha-H^\alpha$ vector, derived as a fitting parameter from an analysis of ^{13}C NMR relaxation data, is 0.67 ± 0.03 (following paper in this issue). The DISCON NOESY fitting

Table 6. Average, Minimum, and Maximum Energies[a] and RMS Violations[b] Obtained for the Set of 60 CONGEN Two-Member Ensembles Computed Using All Constraints in the Ensemble-Averaged Fashion

energies, violations	av	min.	max.
E_{lol}	−23.8	−35.4	−6.4
E_{bond}	2.7	2.1	3.5
E_{ang}	19.4	15.7	24.9
E_{dihe}	13.2	10.2	17.9
E_{impr}	1.3	0.6	2.2
$E_{L\text{-}J}$	−29.6	−33.6	−25.3
E_{elec}	−17.9	−24.7	−11.7
$E_{H\text{-}bond}$	−20.9	−25.8	−15.2
$E_{con.\ dihe}$[c]	0.3	0.1	0.9
E_{NOE}	4.5	2.4	9.3
$E_{J\text{-}coupling}$	3.3	2.1	6.9
RMS NOE violation	0.041	0.030	0.059
RMS J violation	0.24	0.17	0.43

[a] Energies are in kcal/mol. [b] NOE violations are in Å, and J violations are in Hz. [c] Dihedral angle constraints were applied to maintain *trans* amide bonds.

procedure[6] yielded an order parameter of 0.66 ± 0.16 for the Gln-8 $H^{\beta 1}-H^{\beta 2}$ vector, consistent with the ^{13}C relaxation results. The entire set of structures shown in Figure 5 produced with full ensemble-averaged constraints yielded an order parameter of 0.38 for the Gln-8 $C^{\alpha}-H^{\alpha}$ vector, which is significantly lower than the experimental value obtained from the ^{13}C relaxation data (following paper in this issue). We therefore explored the possibility of retaining a subset of the NOESY constraints as nonensemble-averaged constraints. Constraints that were essentially never violated in the nonensemble-averaged structures were retained as nonensemble-averaged constraints for calculations using mixed ensemble/nonensemble constraints. We reasoned that constraints that were essentially never violated already have wide enough bounds to account for both spin diffusion and dynamic effects. The structures produced using the mixed constraints also show much better physical energies and constraint satisfaction relative to the nonensemble structures, and a level of precision that is intermediate between the nonensemble and full ensemble structures. In general, it is difficult to judge which set of ensembles best represents the native state of the real peptide in solution. This problem can be addressed by additional experimental and computer simulation studies (following paper in this issue) of the peptide's dynamic behavior. The mixed ensemble/nonensemble distribution of structures is more consistent with ^{13}C relaxation and molecular dynamics simulation data than either the nonensemble or full ensemble structures (following paper in this issue).

Fractionally populated H-bonding interactions appear to be the primary determinants of the β-hairpin stability (Tables 5, 7, and 9). The most highly populated H-bonds occur between Asn-3 O and Gly-6 H^N, and between Asn-3 H^N and Ser-7 O. A H-bond between Tyr-1 O and Ala-9 H^N occurs in 40% of the mixed ensemble/nonensemble structures (Table 9), and numerous fractionally populated H-bonds occur that involve the Asn-3 side chain and Asp-5, Ser-7, Gln-8, and Ala-9. We have chosen a population cutoff of 10% for reporting H-bonds. While this may seem like a rather low cutoff, it is likely that any H-bonds populated by 10% or more contribute favorably to the stability of the β-hairpin manifold of conformers (the folded state). In a recent publication,[37] Sharp and Englander have clarified the contributions made by fractionally populated noncovalent interactions to the folded state. In general, a noncovalent interaction contributes a factor of $(1 + K_{b,F})/(1 + K_{b,U})$ to the overall folding equilibrium constant. $K_{b,F}$ and $K_{b,U}$ are the microscopic equilibrium constants for noncovalent bond formation in the folded and unfolded states, respectively. Thus, a noncovalent bond will increase the overall folding equilibrium constant as long as it is formed more often in the folded state than in the unfolded state. For a linear peptide, the latter is likely to be a random coil state with very few (if any) significantly populated H-bonds.

The proline at position 4 undoubtedly contributes to the stability of the type-I β-turn, since the *trans*-proline conformation restricts the backbone motions of adjacent residues.[38] Consistent with this apparently general property of *trans*-prolines, Gln-2 through Asp-5 constitute the most well-ordered section of the peptide (following paper in this issue). Tyr-1 is also involved in a number of interactions, including hydrophobic contacts with the Ala-9 methyl group and fractionally-populated H-bonding interactions involving the ring hydroxyl and the C-terminal carboxylate. These interactions, along with the electrostatic attraction between the N- and C-termini, serve to

Table 7. Interresidue Hydrogen Bonds[a] Observed in at Least 10% of the 30 Lowest Energy CONGEN Two-Member Ensembles COmputed Using the Full Set of Ensemble-Averaged Constraints

residue 1	acceptor 1	residue 2	donor 2	fractional occurrence	⟨distance (Å)⟩
Tyr-1	O	Ala-9	HN	0.383	2.007
Asn-3	O	Asp-5	HN	0.433	2.127
Asn-3	O	Gly-6	HN	0.650	2.134
Pro-4	O	Gly-6	HN	0.217	2.162
Asp-5	O	Ser-7	HN	0.183	2.181
Asp-5	OD1	Asn-3	HD21	0.167	2.043
Asp-5	OD1	Ser-7	HN	0.217	2.085
Asp-5	OD1	Ser-7	HG	0.167	1.923
Asp-5	OD1	Gln-8	HN	0.117	2.011
Asp-5	OD2	Asn-3	HD21	0.117	2.019
Asp-5	OD2	Ser-7	HN	0.217	2.119
Asp-5	OD2	Ser-7	HG	0.167	1.919
Gly-6	O	Gln-2	HE21	0.433	1.982
Gly-6	O	Gln-8	HN	0.167	2.063
Ser-7	O	Asn-3	HN	0.550	1.952
Ser-7	O	Asn-3	HD21	0.233	1.994
Ser-7	O	Ala-9	HN	0.267	2.115
Ser-7	OG	Asn-3	HD21	0.267	1.990
Gln-8	O	Asn-3	HD22	0.150	2.251
Ala-9	OT1	Tyr-1	HH	0.117	2.155
Ala-9	OT2	Tyr-1	HH	0.183	1.012

[a] See footnotes of Table 5 for additional details.

Table 8. Average, Minimum, and Maximum Energies[a] and RMS Violations[b] Obtained for the Set of 60 CONGEN Two-Member Ensembles Computed Using Mixed Ensemble-Averaged/ Nonensemble Averaged Constraints

energies, violations	av	min.	max.
E_{lol}	−16.4	−29.3	3.0
E_{bond}	2.8	2.4	3.2
E_{ang}	22.0	18.3	25.7
E_{dihe}	14.0	11.2	17.7
E_{impr}	1.4	0.9	3.0
$E_{L\text{-}J}$	−29.8	−34.3	−22.0
E_{elec}	−18.4	−22.8	−13.9
$E_{H\text{-}bond}$	−21.1	−28.3	−13.0
$E_{con.\ dihe}$[c]	0.3	0.2	0.6
E_{NOE}	9.0	6.2	11.5
$E_{J\text{-}coupling}$	3.4	2.0	6.6
RMS NOE violation	0.058	0.048	0.066
RMS J violation	0.25	0.18	0.37

[a] Energies are in kcal/mol. [b] NOE violations are in Å, and J violations are in Hz. [c] Dihedral angle constraints were applied to maintain *trans* amide bonds.

(37) Sharp, K. A.; Englander, S. W. *Trends Biochem. Sci.* **1994**, *19*, 526–529.

(38) Mikhailov, D.; Daragan, V. A.; Mayo, K. H. *Biophys. J.* **1995**, *68*, 1540–1550.

Table 9. Interresidue Hydrogen Bonds[a] Observed in at Least 10% of the 30 Lowest Energy CONGEN Two-Member Ensembles Computed Using Mixed Ensemble-Averaged/ Nonensemble-Averaged Constraints

residue 1	acceptor 1	residue 2	donor 2	fractional occurrence	⟨distance (Å)⟩
Tyr-1	O	Gln-9	HE21	0.117	2.020
Tyr-1	O	Ala-9	HN	0.400	2.013
Gln-2	OE1	Gln-8	HE21	0.117	1.985
Asn-3	O	Asp-5	HN	0.350	2.024
Asn-3	O	Gly-6	HN	0.883	2.121
Asn-3	OD1	Asp-5	HN	0.400	2.071
Asn-3	OD1	Ala-9	HN	0.117	2.034
Asp-5	O	Ser-7	HN	0.283	2.213
Asp-5	OD1	Asn-3	HD21	0.317	2.170
Asp-5	OD1	Ser-7	HN	0.267	2.054
Asp-5	OD1	Ser-7	HG	0.283	1.981
Asp-5	OD2	Asn-3	HD21	0.333	2.130
Asp-5	OD2	Ser-7	HN	0.250	2.065
Asp-5	OD2	Ser-7	HG	0.250	1.930
Gly-6	O	Gln-2	HE21	0.367	1.996
Gly-6	O	Gln-8	HN	0.117	2.052
Ser-7	O	Asn-3	HN	0.783	1.969
Ser-7	O	Asn-3	HD21	0.133	2.017
Ser-7	O	Ala-9	HN	0.217	2.036
Ser-7	OG	Asn-3	HD21	0.183	2.035
Ser-7	OG	Asn-3	HD22	0.217	2.106
Ser-7	OG	GLn-8	HN	0.200	2.069
Gln-8	O	Asn-3	HD22	0.183	2.054
Ala-9	OT2	Tyr-1	HH	0.167	1.978

[a] See footnotes of Table 5 for additional details.

Table 10. Average ϕ and ψ Angles for the 60 CONGEN Nonensemble-Averaged Structures (Figure 3, All Structures) and the 30 Structures Corresponding to the Predominant Conformational Family Obtained from the Two-Member Full Ensemble-Averaged Calculations (Figure 5, Blue Structures)

residue	$\phi_{\text{non-ens}}$[a] (deg)	$\psi_{\text{non-ens}}$[a] (deg)	$\phi_{\text{full-ens}}$[b] (deg)	$\psi_{\text{full-ens}}$[b] (deg)
Gln-2	−142 ± 2	128 ± 4	−126 ± 20	134 ± 16
Asn-3	−79 ± 1	176 ± 3	−72 ± 12	172 ± 7
Pro-4	−65 ± 4	−15 ± 10	−67 ± 13	−13 ± 30
Asp-5	−81 ± 8	−18 ± 5	−94 ± 26	−21 ± 14
Gly-6	100 ± 2	−2 ± 21	107 ± 14	−12 ± 10
Ser-7	−116 ± 31	171 ± 6	−77 ± 6	177 ± 42
Gln-8	−81 ± 5	106 ± 4	−98 ± 26	117 ± 32

[a] Average angle observed for the 60 CONGEN nonensemble-averaged structures. [b] Average angle observed for the 30 structures from the predominant conformational family obtained from the two-member full ensemble-averaged calculations.

complete the hairpin. It is noteworthy that a recent study[39] of residue−residue association propensities in known protein structures found that there is a "pervasive attraction for tyrosine by almost all residue types".

Adequacy of the Ensemble-Averaged Modeling Protocol. The modeling of conformationally flexible systems is a challenging, unresolved problem. As described in the **Background and Overview for Ensemble-Averaged Calculations** section, a number of approaches have been developed. As of this writing, no single approach has been clearly demonstrated to be superior. Time-averaged refinement over a single trajectory has the advantage of automatically producing a Boltzmann distribution of conformers for a sufficiently long trajectory, whereas the explicit enforcement of Boltzmann weights in ensemble-averaged calculations is rather involved.[40] We have made no effort to explicity enforce Boltzmann sampling; this issue is discussed further below. Ensemble-averaging is a more general approach than single-trajectory time-averaging, in that the problem of crossing large energy barriers can be avoided by a suitable choice of starting conformations.[40] Ensemble-averaged refinement provides a method for overcoming the potential problem of undersampling conformational space. Within the ensemble-averaged refinement method, one is still faced with a myriad of choices regarding the details of the simulation protocol. In the present study, we have opted to use a protocol that is very similar to what is routinely done in standard NMR structure determinations: *in vacuo* simulated annealing calculations with a high temperature phase, repeated for many different ensembles. (The latter is analogous to repeating standard structure determination calculations for many different structures.) A problem that can occur for full ensemble-averaged calculations is oversampling of conformational space, since if one member (or a significant fraction of members) of the ensemble satisfies most of the medium- and long-range constraints, the other members are much less restrained.

In cases where the conformational distribution consists of a family of closely related structures (e.g., β-hairpin-like conformers), the problem of oversampling can be alleviated by the judicious choice of a subset of nonensemble-averaged restraints for mixed nonensemble/ensemble calculations, as demonstrated here. For these cases, it may also be possible to perform full ensemble-averaged calculations using a set of similar starting structures near the average conformation, employing explicit solvent in conjunction with low-temperature-only molecular dynamics. The ability to define mixed nonensemble/ensemble constraints should also prove useful for applying ensemble-averaged calculations to selected regions (e.g., mobile loops) in globular proteins, wherein hydrogen-bonded constraints[35] and/or nonensemble-averaged distance constraints[15h] can maintain the highly ordered regions of the structure. In cases where two or more distinctly different global conformational families are indicated by the data (e.g., see ref 1k), it may be necessary to use full ensemble-averaged calculations, with the number of ensemble members carefully restricted to the number of distinct conformational families. The variability within each family can then be gauged by repeating the calculations for many different ensembles. Such a strategy requires *a priori* knowledge of the number of global conformers.[40]

We conclude this section with a discussion of the "Boltzmann problem" in ensemble-averaged refinement, which has recently been addressed by W. F. van Gunsteren and co-workers.[40] If one wants precise values for the populations of various classes of conformers, then Boltzmann weighting must be enforced. In practice, this is very problematic, since the relative free energies of particular conformations cannot be determined to a high degree of accurracy. This is due to inadequacies of empirical potential energy functions, even with explicit solvent included, and the commonly made assumption that the entropies among different, folded conformers are similar. An example of the possible errors associated with the latter problem has been revealed by a study of disulfide bond isomers of bovine pancreatic trypsin inhibitor (BPTI),[41] where the free energy difference between the pro-*R* and pro-*S* conformations of the disulfide bridge was found to be due primarily to entropic factors, with $\Delta H = -0.5 \pm 0.8$ kcal/mol and $\Delta S = 10.2 \pm 3.2$ cal/(mol·K). It is also difficult to accurately determine populations directly from NOE and *J*-coupling data, since knowledge of the exact number of conformers as well as precise values for the distances and dihedral angles associated with each

(39) Karlin, S.; Zuker, M.; Brocchieri, L. *J. Mol. Biol.* **1994**, *239*, 227−248.

(40) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. *J. Biomolec. NMR* **1995**, *6*, 163−170.

(41) Otting, G.; Liepinsh, E.; Wüthrich, K. *Biochemistry* **1993**, *32*, 3571−3582.

conformer is required. Therefore, we believe that it is not currently possible to obtain accurate population estimates for conformers that interconvert rapidly on the chemical shift time scale. In the present study, we simply excluded the highest-energy structural ensembles from the final conformer analysis. Ensembles of structures have been produced with relatively low physical energies that reproduce the NOE and *J*-coupling data better than structures computed by standard methods. The set produced by the mixed ensemble/nonensemble calculations agrees most closely with experimental dynamic data (following paper in this issue). For many purposes, a very qualitative estimate of conformer populations, at the level of "insignificantly populated" versus "significantly populated", is sufficient. For example, we may wish to know whether any of the significantly populated conformers of a bioactive peptide free in solution fit favorably into a protein binding site of interest. In such an application, it does not much matter if a particular class of conformer is 20% populated or 40% populated. Any significantly populated free-peptide conformer that fits well into the protein binding site would be able to do so with a minimal unfavorably intrapeptide strain enthalpy change. Of course, a theoretical estimate of the conformational entropy loss associated with binding would require accurate knowledge of the free peptide conformer populations. A rough experimental estimate of the conformational entropy loss can also be obtained by determining changes in order parameters.[42] In summary, explicit enforcement of Boltzmann weighting, while desirable in principle, is not essential to the extraction of meaningful information from ensemble-averaged calculations.

Conclusions

The conformational equilibrium of the linear peptide Y-Q-N-P-D-G-S-Q-A in aqueous solution at 5 °C is dominated by a family of closely related, β-hairpin conformers. These yield a dense network of NOE connectivities. The application of standard structure determination methods to this system proved to be inadequate, producing a family of overly precise, highly strained structures. A general and versatile ensemble-averaged molecular modeling facililty was incorporated into the CONGEN program[8] and used to generate sets of solution structure ensembles of the peptide. The results of these calculations are consistent with a highly populated manifold of β-hairpin conformers stabilized by transient noncovalent interactions, particularly H-bonds, and by the inherent rigidity of proline at position 4. The methods employed in this study should prove to be generally useful for the characterization of flexible biological macromolecules.

Acknowledgment. We thank Drs. Roberto Tejero and Gaetano Montelione for advice and providing scripts used to control the execution of the CONGEN runs and Dr. Donna Bassolino-Klimas for implementation of the NOE restraint function (eq 5) in CONGEN. We also thank Dr. A. M. M. J. Bonuin and Dr. W. F. van Gunsteren for providing copies of refs 35 and 40, respectively, prior to their publication. For information about obtaining the CONGEN program, please contact R. E. Bruccoleri at the above address or by e-mail at bruc@bms.com.

Supporting Information Available: The new DISCON iteration schemes, ^{1}H and ^{13}C NMR assignments, and a listing of the distance and *J*-coupling constraints used for the CONGEN calculations (9 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA9520373

(42) Akke, M.; Brüschweiler, R.; Palmer, A. G., III *J. Am. Chem. Soc.* **1993**, *115*, 9832−9833.