J|A|C|S
ARTICLES
Published on Web 11/19/2003

Conformational Studies of a Bombolitin III-Derived Peptide Mimicking the Four-Helix Bundle Structural Motif of Proteins

Elisabetta Schievano,[†] Stefano Mammi,[†] Luca Monticelli,[†,‡] Marta Ciardella,[†] and Evaristo Peggion*[,†]

Contribution from the Institute of Biomolecular Chemistry, CNR, Department of Organic Chemistry, University of Padova, Via Marzolo 1, 35131 Padova, Italy and Center for Biomolecular Interdisciplinary Studies and Industrial Applications, University of Milan, Italy

Received February 10, 2003; Revised Manuscript Received September 16, 2003; E-mail: evaristo.peggion@unipd.it

Abstract: Bombolitins are five structurally related heptadecapeptides originally isolated from the venom of a bumblebee. In aqueous solution, bombolitins at sufficiently high concentration form oligomeric aggregates with consequent conformational transition from a random coil to the α-helical structure. Previous studies suggested that oligomeric aggregates could mimic the four-helix bundle structural motif of proteins. In the present work, we synthesized the following peptide sequence formed by two bombolitin III sequences linked head-to-tail by the tetrapeptide bridge -Gly-Pro-Val-Asp-: I^{1}-K^{2}-I^{3}-M^{4}-D^{5}-I^{6}-L^{7}-A^{8}-K^{9}-L^{10}-G^{11}-K^{12}-V^{13}-L^{14}-A^{15}-H^{16}-V^{17}-G^{18}-P^{19}-V^{20}-D^{21}-I^{22}-K^{23}-I^{24}-M^{25}-D^{26}-I^{27}-L^{28}-A^{29}-K^{30}-L^{31}-G^{32}-K^{33}-V^{34}-L^{35}-A^{36}-H^{37}-V^{38}-NH_2. The tetrapeptide GPVD connecting the two helical peptide sequences was chosen to facilitate the formation of the helix−loop−helix structural motif. The conformational properties of the peptide were studied by CD, NMR, and molecular dynamics calculations. The results indicate the presence of a helix−loop−helix conformation at 10^{-5} M concentration. At higher concentrations, NOESY connectivities were detected which are compatible with the presence of dimers or higher aggregates of peptide molecules in the helix−loop−helix structure packed in an antiparallel fashion. Molecular dynamics simulation were run either with NOE distance restraints or without restraints in explicit solvent for extended time. The results of these simulations support the dimerization of the molecules in the helix−loop−helix structure with formation of the four-helix bundle motif.

Introduction

The study of *de novo* designed polypeptides capable of folding into well-defined tertiary structures is of great importance to understand the forces responsible for the conformational stability of peptides and proteins. Natural peptides forming amphipathic α-helices in aqueous solution tend to form oligomeric aggregates stabilized by long-range interactions, mimicking helix bundle motifs of proteins. A typical example is melittin,[1] initially isolated from the bee venom, which forms tetrameric aggregates of helical molecules in aqueous solution at basic pH and high ionic strength. In general, the aggregation process is governed by a delicate balance of electrostatic and hydrophobic interactions and depends on various factors such as sequence, side-chain functionality, and hydrophobic moment.[2] In previous work,[3−5] we have studied the conformational and aggregation properties of bombolitin I, II, and III, three of a series of five structurally related heptadecapeptides, originally isolated from bumblebee venom.[6] Bombolitins exhibit lytic properties toward erythrocytes and liposomes and enhance the activity of the calcium-dependent enzyme phospholipase A_2.[6] It has been shown that the biological activity is related to the capability of these peptides to fold into an amphipathic α-helical conformation in lipophilic or membrane-mimetic environments.[5] We have shown that bombolitins in aqueous solution and at sufficiently high concentration form oligomeric aggregates.[3−5] The aggregation process induces a conformational transition from the random coil to the α-helical structure, although it is sequence dependent and different bombolitins exhibit different behaviors. Bombolitin III at pH 4.5 is almost completely aggregated at 2.5 mM concentration, while bombolitin I requires a much higher concentration. The only significant difference between the two sequences is the presence of an Asp residue in position 5 in bombolitin III, and it was concluded that the carboxylate side chain is involved in a favorable electrostatic interaction stabilizing the oligomeric structure. Bombolitin II, which also contains an Asp residue in position 5 behaves exactly as bombolitin III, exhibiting a strong tendency toward aggregation.[4]

In a subsequent work,[7] we synthesized and characterized an analogue of bombolitin III in which the Ile residue at position

[†] University of Padova.
[‡] University of Milan.

(1) Hagihara, Y.; Kataoka, M.; Aimoto, S.; Goto, Y. *Biochemistry* **1992**, *31*, 11908−11914.
(2) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. *J. Mol. Biol.* **1984**, *179*, 125−142.
(3) Bairaktari, E.; Mierke, D. F.; Mammi, S.; Peggion, E. *Biochemistry* **1990**, *29*, 10097−10102.
(4) Battistutta, R.; Pastore, A.; Mammi, S.; Peggion, E. *Macromol. Chem. Phys.* **1995**, *196*, 2827−2841.
(5) Peggion, E.; Mammi, S.; Schievano, E. *Pept. Sci.* **1998**, *43*, 419−431.
(6) Argiolas, A.; Pisano, J. J. *J. Biol. Chem.* **1984**, 259, 10106−10111.

10.1021/ja0300970 CCC: $25.00

3 was replaced by L-*p*-benzoylphenylalanine containing the photoreactive benzophenone group in the side chain:

I^1-K^2-Bpa^3-M^4-D^5-I^6-L^7-A^8-K^9-L^{10}-G^{11}-K^{12}-V^{13}-L^{14}-A^{15}-H^{16}-V^{17}-NH_2 (**Bpa^3BIII**)

Upon irradiation, the benzophenone group has the property of cross-reacting with an adjacent aliphatic proton, forming a covalent bond according to a well-known photochemical reaction.[8] This approach was initially followed by Kim and co-workers[9] to determine the relative orientation of helices in the case of the leucine zipper. The aggregation properties of this analogue were identical to those of the native bombolitin III sequence. Irradiation of **Bpa^3BIII** aggregates yielded dimeric structures with the two covalently bound helices oriented in an antiparallel fashion. We have also shown that the formation of these interhelix bonds enhances the stability of the helical structure at low concentration. Further aggregation of the dimers occurs at higher concentration with possible formation of the four-helix bundle motif.

Recently, a number of *de novo* designed peptide sequences were reported to fold into the hairpin helix−loop−helix motif.[10−13] Baltzer and co-workers[14−16] showed that the formation of this structural motif is favored when two helical peptide sequences are connected head-to-tail by the tetrapeptide bridge Gly-Pro-Val-Asp. A *de novo* designed 42 amino acid peptide containing this bridge folds into a helix−loop−helix motif, which dimerizes to form a four-helix bundle.[15]

Following the same approach, we designed and studied the following peptide formed by two bombolitin III sequences linked head-to-tail by the tetrapeptide bridge -Gly-Pro-Val-Asp-:

I^1-K^2-I^3-M^4-D^5-I^6-L^7-A^8-K^9-L^{10}-G^{11}-K^{12}-V^{13}-L^{14}-A^{15}-H^{16}-V^{17}-**G^{18}-P^{19}-V^{20}-D^{21}**−I^{22}-K^{23}-I^{24}-M^{25}-D^{26}-I^{27}-L^{28}-A^{29}-K^{30}-L^{31}-G^{32}-K^{33}-V^{34}-L^{35}-A^{36}-H^{37}-V^{38}-NH_2 (**BIII-GPVD-BIII**)

In the present paper, we report the structural characterization of this 38 amino acid sequence in aqueous solution, by CD, NMR, and molecular dynamics simulations.

Experimental Section

Peptide Purification. A sample of crude peptide was a generous gift from Polypeptide Laboratories, Torrance, CA. The crude peptide was purified by HPLC using a Pharmacia Biotech P-900 chromatograph equipped with a Pharmacia Biotech UV-900 spectrophotometric detector. For analytical and semipreparative HPLC, Sigma Aldrich columns Vydac C-18 RF 300, 7.8 × 300 mm, and Vydac C-18 RF 300, 4.6 × 150 mm, respectively, were used. The purity of the final product isolated by lyophilization was assessed by analytical HPLC. Amino acid analysis of the acidic hydrolysate yielded the correct amino acid ratios. Electrospray mass spectrometry yielded the correct molecular weight of 4073 Da. The peptide content, determined by quantitative amino acid analysis, was 57%.

CD Experiments. CD measurements were carried out at room temperature on a JASCO J-715 spectropolarimeter interfaced with a PC. The CD spectra were acquired and processed using the J-700 program for Windows. All spectra were recorded using a bandwidth of 2 nm and a time constant of 2−8 s at a scan speed of 20 or 50 nm/min. The signal-to-noise ratio was improved by accumulating at least six scans. HELLMA quartz cells with suprasil windows with optical path lengths of 0.01−0.1 cm were used. The helix content of the peptide was estimated according to the method of Greenfield and Fasman.[17]

NMR Experiments. NMR experiments were performed on a Bruker Avance DMX 600 spectrometer. The measurements were carried out at 318 K on a 1.00 mM peptide in aqueous solution at pH 3.3, in the presence of 10% D_2O and 4% 2,2,2-trifluoroethanol (TFE). Tetramethylsilane was used as internal standard. The water signal was suppressed using the Watergate pulse sequence.[18] The complete assignment of the spin systems was achieved from DQF-COSY[19] and CLEAN-TOCSY spectra,[20−21] while NOESY and semisoft NOESY experiments[22−23] were used for sequential assignment. In the NOESY experiments, the mixing time was 200 ms. All spectra were acquired by collecting 512 experiments, each one consisting of 112−208 scans and 4096 data points in t_2. Chemical shift differences relative to the random coil values for each residue were averaged with those of its neighboring residues to minimize local anisotropic effects, as reported by Pastore and Saudek.[24]

Interproton distances were obtained by integration of the NOESY cross-peaks using the XEASY program. The calibration was based on the cross-peaks of the geminal γ-protons of Met^4 set to a distance of 1.78 Å. Addition and subtraction of 10% of the calculated distances yielded upper and lower bounds. The distance derivation from integration of NOESY cross-peaks is complicated by the presence of cis and trans peptide bonds in the GPVD bridge region (see the Results and Discussion section**)**. The isomer population is 75% trans and 25% cis. Only cross-peaks corresponding to the trans isomer were integrated and used without correction. Given that the reduction of the normalized intensity of these peaks is 25% relative to the peaks of other portions of the sequence, the corresponding uncertainty on the interproton distances is within the ±10% confidence interval utilized in the structure calculations.

Distance Geometry and Simulated Annealing Calculations. Distance geometry (DG) and restrained molecular dynamics (MD) calculations using the simulated annealing (SA) protocol were carried out with the X-PLOR 3.0 program.[25] For distances involving equivalent or nonstereoassigned protons, r^{-6} averaging was used. The general procedure of the MD calculations was the following. After a minimization step of 200 cycles, the calculation involved an SA and a refinement stage. The SA consisted of 48 ps of dynamics at 2000 K (16 000 cycles, in 3 fs steps) followed by 30 ps of cooling from 2000 to 100 K in 50 K decrements (15 000 cycles, in 2 fs steps). The SA, in which the weights of NOE and nonbonded terms were gradually increased, was followed by 200 cycles of energy minimization. In the refinement stage, the system was cooled from 1000 to 100 K in 50 K decrements (20 000

(7) Cabrele, C.; Furin, S.; Gurrath, M.; Mammi, S.; Schievano, E.; Peggion, E. *Biopolymers*, **1997**, *42*, 147−156.
(8) O′Neil, K. T.; Erickson-Viitanen, S.; DeGrado, W. F. *J. Biol. Chem*. **1989**, *24*, 14571−14578.
(9) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem*. **1990**, *59*, 631−660.
(10) Dolphin, G. T.; Brive, L.; Johansson, G.; Baltzer, L. *J. Am. Chem. Soc*. **1996**, *118*, 11297−11298.
(11) Brive, L.; Dolphin, G. T.; Baltzer, L. *J. Am. Chem. Soc*. **1997**, *119*, 8598−8607.
(12) Ho, P. S.; DeGrado, W. F. *J. Am. Chem. Soc*. **1987**, *109*, 6751−6758.
(13) Hill, B. R.; DeGrado, W. F. *J. Am. Chem. Soc*. **1998**, *120*, 1138−1145.
(14) Olofsson, S.; Johansson, G.; Baltzer, L. *J. Chem. Soc., Perkin Trans*. **1995**, *2*, 2047−2056.
(15) Olofsson, S.; Baltzer, L. *Protein Folding and Design* **1996**, *1*, 347−356.
(16) Broo, K. S.; Nilsson, H.; Nilsson, J.; Baltzer, L. *J. Am. Chem. Soc*. **1998**, *120*, 10287−10295.
(17) Greenfield, N.; Fasman, G. D. *Biochemistry* **1969**, *8*, 4108−4116.
(18) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson., Ser. A* **1993**, *102*, 241−245.
(19) Rance, M.; Sørensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun*. **1983**, *117*, 479−485.
(20) Bax, A.; Davis, D. G. *J. Magn. Reson*. **1985**, *65*, 355−360.
(21) Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc*. **1988**, *110*, 7870−7872.
(22) Macura, S.; Huang, Y.; Suter, D.; Ernst, R. R. *J. Magn. Reson*. **1980**, *43*, 259−281.
(23) Cavanagh, J.; Waltho, J. P.; Keeler, J. *J. Magn. Reson*. **1987**, *74*, 386-393.
(24) Pastore, A.; Saudek, V. *J. Magn. Reson*. **1990**, *90*, 165−176.
(25) Brunger, A. T. *X-PLOR Manual*, version 3.0; Yale University: CT, 1992.

cycles, in 2 fs steps). Finally, the calculation was completed with 200 cycles of energy minimization using an NOE force constant of 50 kcal/mol.

Molecular Dynamics in Explicit Solvent. Three unrestrained MD simulations were carried out in the presence of explicit water. The starting structure for all the simulations was the lowest energy structure resulting from the DG/SA calculations. In the first simulation, one peptide molecule was placed at the center of the simulation box, while, in the second and third simulations, two molecules were placed in the same box. The distance between the two peptide molecules was short in the second simulation and about 1.8 nm in the third; in both cases, the helical axes were parallel to each other.

The peptide was solvated with water in a cubic box large enough to contain about 1 nm of solvent around the peptide. The resulting system was composed of 5077 water molecules in the first simulation, 6688 in the second, and 7342 in the third. The simple point charge (SPC) water model was used.[27] Chloride ions replaced water molecules to yield an electrically neutral system; the calculation of electrostatic forces utilized the PME implementation of the Ewald summation method; a twin-range cutoff of 0.9−1.4 nm was used for the calculation of the van der Waals interactions. In all cases, periodic boundary conditions were used. All simulations were carried out using the GROMACS package[28,29] (version 3.0), with the GROMOS96 43A1 force field.[30] The peptide, water, and the counterions were coupled separately to a temperature bath at 300 K with $\tau_T = 0.1$ ps using the Berendsen algorithm.[31] The pressure was kept at 1 bar using weak pressure coupling with $\tau_P = 1.0$ ps.

The aromatic ring of His side chains was kept rigid using dummy atom constructions, according to a published procedure, to remove degrees of freedom from the system. According to Berendsen and coworkers,[32] removal of these degrees of freedom from the system allows the time step to be increased up to 7 fs with negligible influence on the thermodynamical and dynamical properties of the system. In our simulations, the time step was 2 fs during solvent equilibration and 6 fs in the successive steps; the neighbor list for the calculation of nonbonded interactions was updated every 10 time steps in the first case and every 3 in the latter. All bond lengths of nonwater molecules were constrained with the LINCS algorithm,[33] while the SETTLE algorithm was used[34] to constrain bond lengths and angles in the water molecules.

The system was energy minimized with a steepest descent method for 5000 steps. In all the simulations, the solvent was equilibrated in a 50 ps MD run with position restraints on the peptide. The force constant on the peptide atoms was 1000 kJ mol^{-1} nm^{-2}. The solvent equilibration run was followed by another 100 ps run without position restraints on the peptide, in which all atoms were given an initial velocity obtained from a Maxwellian distribution at the desired initial temperature. The production runs, after equilibration, were 100 ns long, and peptide structures were sampled every 0.5 ps.

Conformational clustering analysis was performed on a subset of the conformations sampled in the trajectory, containing 5000 structures taken at 20 ps intervals, using the method described by Daura and coworkers (as implemented in Gromacs[35]). The positional rmsd calculated on the backbone atoms was used as a similarity criterion, and a cluster radius of 0.1 or 0.15 nm was chosen. Calculation of the total and hydrophobic areas accessible to solvent was performed using a solvent probe radius of 0.14 nm. Secondary structure assignments were based on the DSSP algorithm.[36] Note that in the DSSP notation the difference between secondary structure elements is based only on the backbone geometry and the presence of hydrogen bonds.

The interproton distances were calculated from the simulations as $<d(t)^{-6}>^{-1/6}$ averages, and violations were derived as $d_{viol} = |<d(t)^{-6}>^{-1/6} - d_{exp}|$. Since in the GROMOS96 43A1 force field the aliphatic hydrogen atoms are treated within a united-atom model, their position at each time in the trajectory was back-calculated assuming an ideal geometry.

All calculations were performed on clusters of PCs, with a Linux operating system and LAM-MPI software for the parallelization of the MD algorithm (obtained free of charge from http://www.lam-mpi.org). Total CPU time for each of the three simulations was approximately 150 days on a two-processor Pentium III 500 MHz PC. The graphical representations of the peptide were realized with the programs VMD[37] and INSIGHT II (Accelrys Inc.).

Results and Discussion

CD Results. The CD spectra of 5×10^{-5} M **BIII-GPVD-BIII** in aqueous solution as a function of pH are shown in Figure 1a.

At pH 2.2, the CD spectrum is consistent with the presence of a largely disordered conformation, with a small amount of ordered structure revealed by the wide shoulder around 220 nm. When the pH is increased, the typical CD pattern of the α-helical structure is formed with two minima at 208 and 220 nm and a strong positive band at 193 nm. All spectra fit a well-defined isodichroic point at 203 nm consistent with the presence of the two-component coil-helix equilibrium system. The most significant spectral change occurs between pH 2.2 and 4.9. At pH 4.9, the structural change is almost complete: only small changes are observed between pH 4.9 and 7.0. At pH 7.0, the maximum helix content estimated according to the method of Greenfield and Fasman[17] is of the order of 65%. Upon increasing the pH from 2.2 to 4.9, where the major conformational change occurs, deprotonation of the carboxylate side chains of the three Asp residues at positions 5, 21, and 26 takes place. In our previous work,[3,4] we have shown that the carboxylate ions of the Asp residues contribute to the stabilization of aggregates of helical **BIII** molecules at peptide concentrations ≥2.5 mM. At concentrations of the order of 10^{-5} M, the structure of **BIII** remains essentially unordered and very little pH dependent.[3,4] The results of Figure 1a therefore might suggest that the helix formation upon increasing the pH is consequent to the folding of the molecule in a hairpin helix−loop−helix structure in which the carboxylate side chain is involved in stabilizing interhelix electrostatic interactions.

The CD spectra of **BIII-GPVD-BIII** recorded in aqueous solution at pH 4.5 and 5.6 at increasing peptide concentrations are reported in Figure 1b and c. At pH 4.5, there is an enhancement of helix content upon increasing the peptide concentration up to 1.06 mM. This behavior suggests that aggregation of **BIII-GPVD-BIII** molecules takes place at higher peptide concentrations with the consequent increase of helix content. At pH 5.6, the CD spectra are almost concentration independent and the

(26) Havel, T. F. *Biopolymers* **1990**, *29*, 1565−1585.
(27) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(28) Berendsen, H. J. C.; Spoel, D. V. D.; Drunen, R. V. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(29) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(30) van Gunsteren, W. F.; Daura, X.; Mark, A. E. GROMOS force field. *Encyclopedia of Computational Chemistry*; **1998**, *2*, 1211−1216.
(31) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(32) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786−798.
(33) Hess, B.; Bekker, H.; Fraaije, J.; Berendsen, H. J. C. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(34) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(35) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(36) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2576−2637.
(37) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

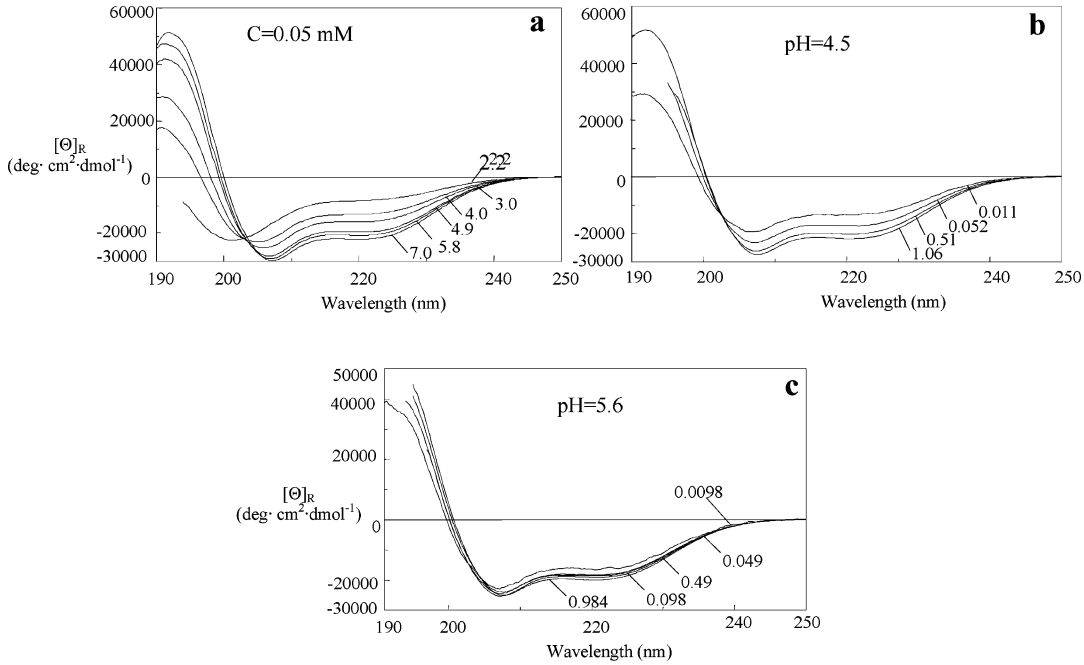

Figure 1. (a) CD spectra of **BIII-GPVD-BIII**, 5×10^{-5} M in aqueous solution at increasing pH values (indicated in the spectra); (b−c) CD spectra at pH 4.5 (b) and 5.6 (c) at increasing peptide concentration (mM, indicated in the spectra).

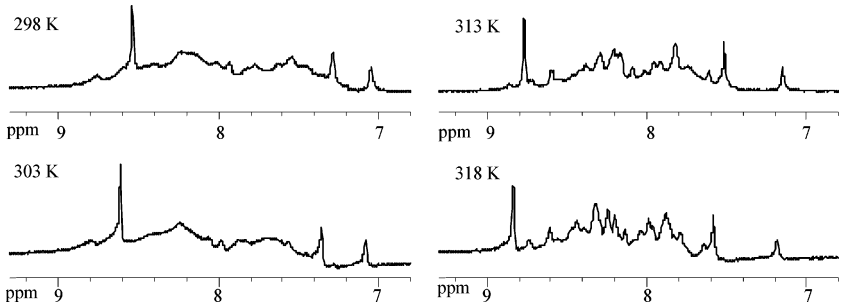

Figure 2. Amide region of 1D NMR spectra of 1mM peptide in aqueous solution, at pH 3.3 as a function of temperature (indicated in the spectra).

spectrum recorded at 0.98 mM peptide is practically identical to the one recorded at pH 4.5 at the same concentration. The different concentration dependences observed in the two cases can be explained taking into account that, at pH 5.6, where the carboxylate ions of the Asp side chains are completely ionized, the putative helix−loop−helix structure is almost completely formed already at a 5×10^{-5} M peptide concentration (Figure 1a). Thus, the aggregation process induced by increasing peptide concentration is not expected to increase appreciably the helix content. At a 1 mM peptide concentration, the CD spectra recorded in the pH range 3.3−7.0 are practically identical (data not shown) and indicate a helix content of ∼65%. At this concentration, the CD spectra are also temperature independent in the range 298−318 K (data not shown), indicating a remarkable structural stability in this temperature range.

NMR Results. The amide region of 1D NMR spectra in water recorded in the temperature range 298−318 K is shown in Figure 2.

The NMR spectrum shows broad lines that become narrower with increasing temperature. The large line width at 298 K and the temperature effect are similar to those found for other polypeptides known to fold in four-helix bundles. In the literature,[11,15] such a temperature effect has been attributed to the presence of a peptide system composed of several conformers in fast exchange. The reduction of line width at higher temperature could also be a consequence of either disruption of higher order aggregates or increased rate of tumbling. Our data do not allow us to distinguish among these possibilities. To improve resolution, small amounts of TFE were added to a water solution of **BIII-GPVD-BIII** at 318 K (Figure 3).

In the presence of 4% TFE, there is a significant reduction of line width, without an appreciable change of chemical shifts. A similar effect was observed with other polypeptides and has been attributed to a reduction of the exchange rate of amide protons with the solvent.[15]

The fingerprint region of the TOCSY (total correlation spectroscopy) spectrum is shown in Figure 4.

The presence of the Pro residue at position 19 induces cis−trans isomerism of the Gly^{18}-Pro^{19} peptide bond, which affects the resonances of Val^{17}, Gly^{18}, Val^{20}, and Asp^{21}, that is, of the two residues preceding and the two following the Pro residue. This is evident from doubled cross-peaks observed for these residues in the fingerprint region of the TOCSY spectrum (Figure 4). The relative amounts of trans and cis isomers resulting from the TOCSY spectrum are 75% and 25%, respectively. The fingerprint and the aliphatic regions of the NOESY spectrum are shown in Figure 5.

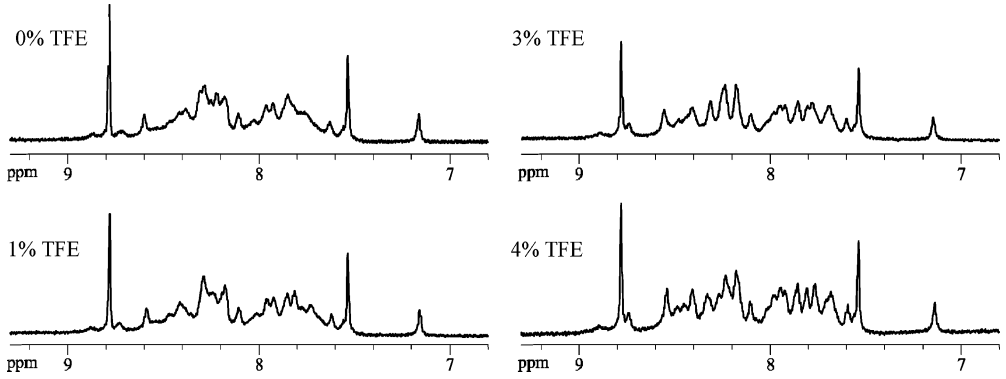

Figure 3. 1D NMR spectra at 318 K of 1 mM peptide in aqueous solution, at pH 3.3, as a function of TFE content (v/v), indicated in the spectra.

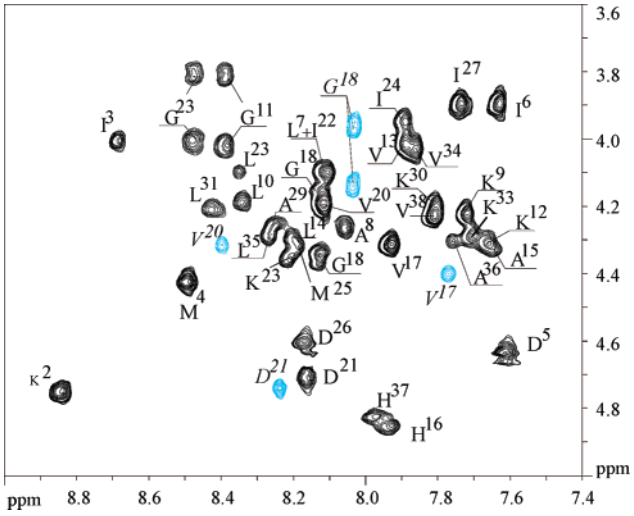

Figure 4. Fingerprint region of the TOCSY spectrum [1 mM peptide in water containing 4% (v/v) TFE], at 318 K; doubled cross-peaks due to cis−trans isomerism are indicated in blue.

The complete assignment of proton resonances resulting from these experiments is reported in Table 1 of the Supporting Information. The histogram representing the chemical shift differences of the αH resonances with respect to the random coil values is shown in Figure 6.

Negative values indicative of the presence of an α-helical structure[24] identify two helical segments in the sequences Met[4]-Leu[10] and Lys[23]-Leu[35]. The C-terminal helical sequence appears slightly more extended than the N-terminal one. The helical structure is clearly interrupted in the central part of the sequence comprising the tetrapeptide bridge.

The pattern of sequential, medium range, and long range NOESY connectivities essential for structural assignment is reported in Figure 7.

The ensemble of sequential NH(*i*)−NH(*i*+1) and medium range connectivities confirms the presence of two α-helical segments in the N-terminal and C-terminal segments. Specifically, medium range connectivities αH(*i*)−NH(*i*+3), αH(*i*)−*β*H(*i*+3), and αH(*i*)−NH(*i*+4) are observed in the stretches 2−12 and 22−37, that is, the central regions of the two **BIII** sequences. Helical NOEs are clearly missing in the central region of the sequence comprising the tetrapeptide bridge. There is a higher density and a longer stretch of helical NOEs in the C-terminal sequence, compared to the N-terminal one, indicating a higher stability of this helical segment. In general, the number of medium range connectivities found in the helical segments is rather low, suggesting relatively loose helices. A very important result from NOESY experiment is the presence of the long range connectivities *β*H(Val[13])−αH(Ile[24]), *γ*H(Val[13])−αH(Ile[24]), NH(Val[13])−*γ*H(Ile[27]), and *γ*H(Ile[1])−C-terminal-(Val[38])NH (Figure 7).

These connectivities between protons of amino acid residues far apart along the sequence are necessarily interhelical with the helical segments oriented in an antiparallel way. These findings are consistent with the hypothesis of folding of **BIII-GPVD-BIII** into a hairpin helix−loop−helix structure with interhelix interactions.

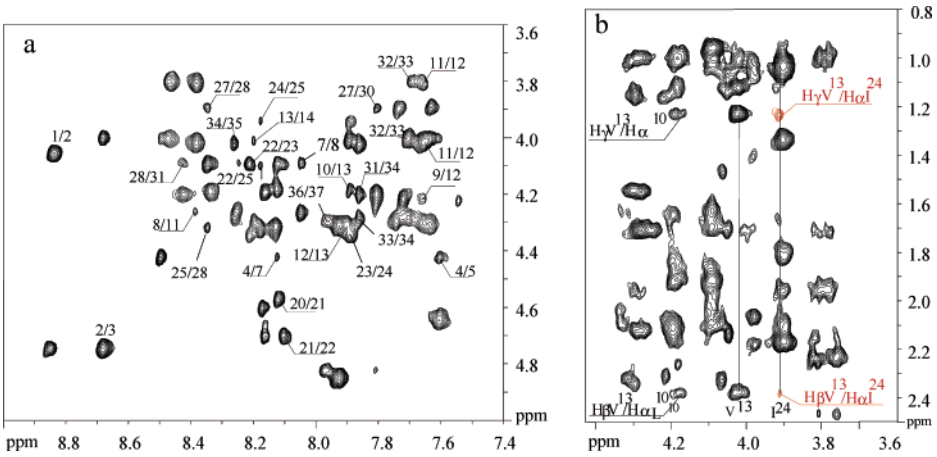

Figure 5. Fingerprint region (a) and aliphatic region (b) of the NOESY spectrum [1 mM peptide in water containing 4% (v/v) TFE)], at 318 K; intramolecular long range connectivities are indicated in red.

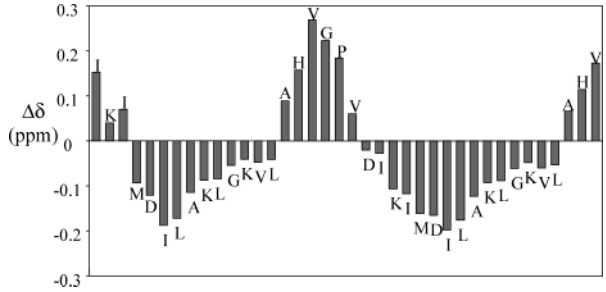

Figure 6. Chemical shift differences (ppm) of the αH protons relative to the random coil values.

Table 1. Average ϕ and ψ Backbone Dihedral Angles with Relative Standard Deviations and Havel Order Parameters

residue		ϕ_{av}	order parameter	ψ_{av}	order parameter
1	Ile			134 ± 39	0.80
2	Lys	−59 ± 12	0.98	160 ± 16	0.96
3	Ile	−59 ± 7	0.96	−25 ± 2	1.00
4	Met	−57 ± 13	0.98	−47 ± 8	0.99
5	Asp	−49 ± 6	0.99	−47 ± 2	1.00
6	Ile	−52 ± 1	1.00	−36 ± 1	1.00
7	Leu	−87 ± 2	1.00	−46 ± 1	1.00
8	Ala	−49 ± 1	1.00	−36 ± 1	1.00
9	Lys	−58 ± 4	1.00	−38 ± 9	0.99
10	Leu	−65 ± 8	0.99	−40 ± 10	0.98
11	Gly	−52 ± 5	1.00	−24 ± 4	1.00
12	Lys	−62 ± 8	0.99	−42 ± 10	0.98
13	Val	−84 ± 20	0.94	−33 ± 13	0.97
14	Leu	−19 ± 51	0.65	3 ± 23	0.92
15	Ala	−65 ± 49	0.73	178 ± 79	0.59
16	His	−94 ± 71	0.50	180 ± 16	0.96
17	Val	−75 ± 46	0.77	−45 ± 38	0.80
18	Gly	−32 ± 75	0.45	−27 ± 65	0.50
19	Pro	−87 ± 8	1.00	25 ± 11	0.98
20	Val	−95 ± 46	0.799	36 ± 7	0.99
21	Asp	−88 ± 50	0.69	−172 ± 10	0.98
22	Ile	−71 ± 49	0.72	26 ± 101	0.26
23	Lys	−52 ± 55	0.62	1 ± 24	0.92
24	Ile	−104 ± 30	0.87	−42 ± 10	0.99
25	Met	−63 ± 5	1.00	−21 ± 3	0.99
26	Asp	−62 ± 9	0.99	−50 ± 9	0.99
27	Ile	−50 ± 5	1.00	−30 ± 3	1.00
28	Leu	−69 ± 2	1.00	−53 ± 6	0.99
29	Ala	−56 ± 5	1.00	−31 ± 1	1.00
30	Lys	−73 ± 3	1.00	−53 ± 4	1.00
31	Leu	−53 ± 3	1.00	−42 ± 3	1.00
32	Gly	−57 ± 4	1.00	−26. ± 5	1.00
33	Lys	−67 ± 10	0.99	−36 ± 2	1.00
34	Val	−94 ± 4	1.00	−40 ± 8	0.99
35	Leu	−7 ± 37	0.81	−2 ± 32	0.85
36	Ala	−65 ± 88	0.26	−44 ± 22	0.93
37	His	−109 ± 27	0.89	−56 ± 16	0.96
38	Val	−90 ± 54	0.65		

The interhelical connectivities involve protons of hydrophobic side chains and stress the importance of hydrophobic interactions for the stability of these structures. Unfortunately, overlap of NOESY cross-peaks involving side chain protons prevented the identification of additional interchain connectivities. The cis−trans isomerism in the loop sequence affects the local conformation but does not modify the relative orientation of the two helices. This is evident from the single set of resonances observed in the helical segments.

Two extremely important NOESY cross-peaks were found, that is, βH(Ile^3)−β2H(Pro^{19}) and βH(Val^{38})−αH(Gly^{18}). These connectivities are possible only between protons of different molecules and are consistent with the presence of dimeric or higher aggregates of **BIII-GPVD-BIII** molecules in a helix−loop−helix structure, oriented in an antiparallel fashion.

Structure Calculations. A. DG and SA Calculations. Integration of the NOESY spectra yielded 269 distance restraints (124 intraresidue, 93 sequential, 46 medium range, and 6 long range). These calculations were performed on the molecule as a monomer. Accordingly, the two long range interproton distance restraints βH(Ile^3)−β2H(Pro^{19}) and βH(Val^{38})−αH(Gly^{18}), indicative of dimer formation, were not included. From DG calculations and structural refinement using the SA protocol of the X-PLOR program previously described, 200 structures were generated, 29 of which fulfilled the experimentally determined interproton distances with violations lower than 0.55 Å. After energy minimization, 20 structures were obtained with energies lower than 500 kcal/mol. The average φ and ψ backbone dihedral angles with the relative standard deviations of the ensemble of low energy structures are reported in Table 1. The dispersion of the backbone dihedral angles around the average values can be quantified by the Havel order parameter,[26] also reported in Table 1. Order parameters very close to 1 are observed in the N-terminal and C-terminal sequences, indicating the convergence of low energy structures to the α-helical conformation. On the contrary, a large dispersion of dihedral angles, reflected by low order parameters, is observed in the sequence 14−23, indicative of a less defined structure.

The results of superimposition of heavy backbone atoms of the N-terminal 3−12 and C-terminal 25−34 sequences of the ensemble of low energy structures are shown in Figure 8a. The convergence of the ensemble toward the helix−loop−helix structural motif is clearly evident. In the central loop region, there is poor convergence indicating a higher structural flexibility. This is also evident from the figure obtained by superposition of the N-terminal 3−12 (Figure 8b) or C-terminal 25−34 segments (Figure 8c). For the sake of clarity, residues of the C-terminal 23−38 sequence were omitted in the first case. Similarly, in the superimposition of the C-terminal segments, the residues of the N-terminal 1−16 sequences were omitted. From this figure, the convergence of the N-terminal and C-terminal sequences toward the α-helical structure is clearly evident as well as the flexibility around the G-P-V-D tetrapeptide bridge.

The helix−loop−helix motif of one of the low energy structures is shown in Figure 8d, with the amino acids involved in long range NOE interactions indicated. In this, as well as in the other calculated structures, many of the hydrophobic residues of one helix are projecting toward the other. If one analyzes the disposition of the various amino acids on the helical surface in terms of the heptad repeat pattern usually utilized to describe a four helix bundle,[38,39] it is clear that **BIII** has fewer polar residues than ideal. Nevertheless, the calculated structures achieve the shielding of the hydrophobic residues from the solvent. This can be seen by placing Val^{13} in position "g" of the first helix and Ile^{24} and Ile^{27} in positions "d" and "g", respectively, of the second helix. Therefore, the observed interhelical NOEs are in line with the formation of the helix−loop−helix motif. In the four-helix bundle, many of the "a" positions would be occupied by Leu residues, the side chains of which are heavily overlapped in the NOESY spectrum and do not provide any information on the formation of such a dimeric structure.

B. Molecular Dynamics Simulations in Explicit Solvent. In all three simulations, the peptide samples mainly compact

(38) Bets, S. F.; Liebman, P. A.; DeGrado, W. F. *Biochemistry* **1997**, *36*, 2450−2458.
(39) Andersson, L. K.; Caspersson, M.; Baltzer, L. *Chem.—Eur. J.* **2002**, *8*, 3687−3697.

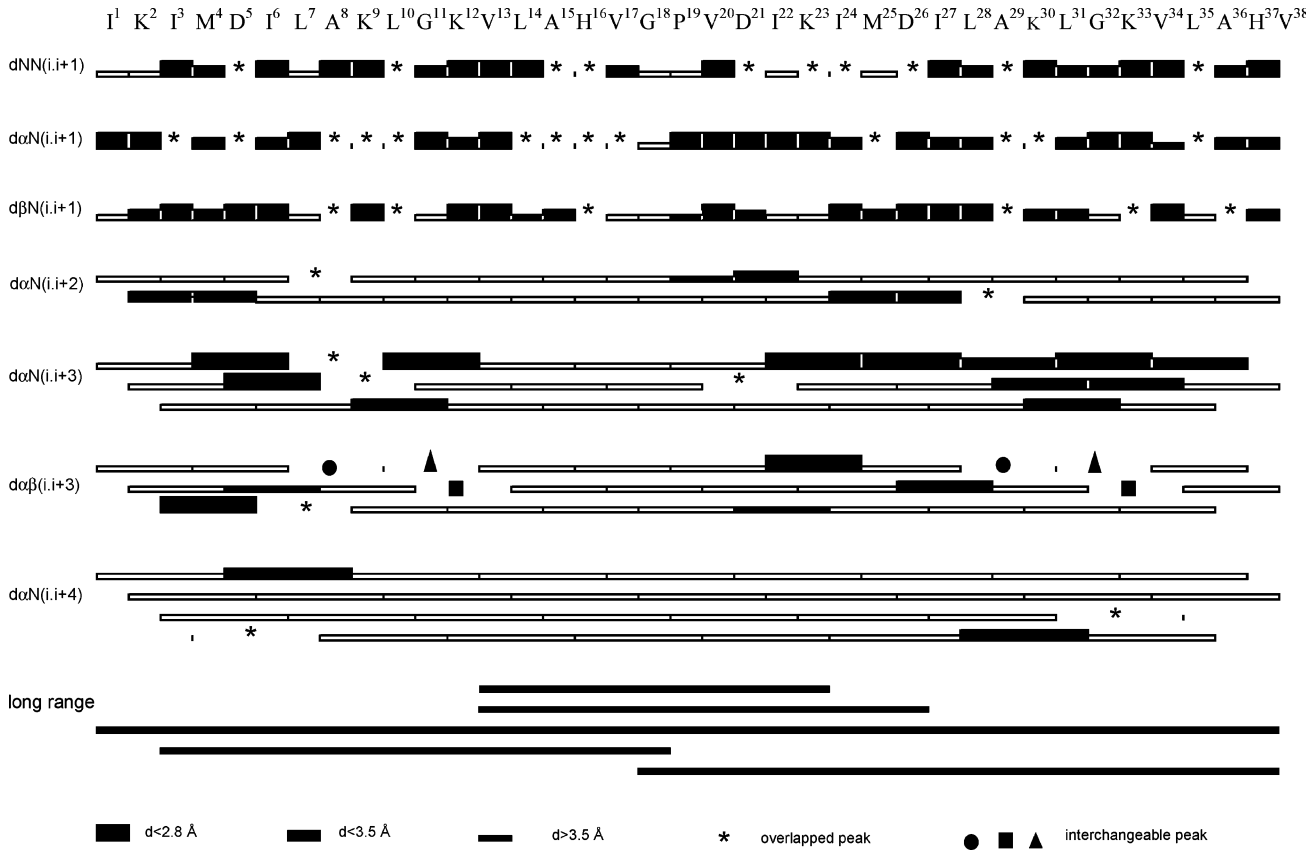

Figure 7. Summary of NOESY sequential, medium range, and long range connectivities. Empty bars indicate connectivities clearly missing. Peaks are grouped in three classes based upon their integrated volumes.

conformations and the helix−loop−helix structural motif is maintained throughout the trajectories. In simulation number 2, the average radius of gyration (R_g) decreases from 1.2 nm (at the beginning of the simulation) to about 1.13 nm after 40 ns (Figure 1 of the Supporting Information); in the last 60 ns of the MD run, it shows little oscillations about the value of 1.14 nm, indicating that the conformational change toward a more compact conformation takes about 40 ns. In simulation number 3, the initial value of R_g is much larger (approximately 1.8 nm), because the two peptide molecules are initially far apart, but reaches the value of 1.14 nm in about 12 ns; this is the amount of time required for the two units to get close in space. Once the two units are close to each other, there are still large conformational changes, but compact conformations are sampled for most of the simulation time and the dimer never breaks into two monomers.

Because the peptide is highly hydrophobic, a high percentage of its surface exposed to the solvent is hydrophobic. In the case of the monomer, it is about 72%, while, in the dimers, after an equilibration period, it is about 61% (Figure 2 of the Supporting Information). It is interesting to note that, in the case of simulation number 3, there is a substantial decrease of the solvent exposed surface area during the initial 1.4 ns; at the same time, the percentage of hydrophobic surface area accessible to the solvent decreases from about 70% (comparable to that of the monomer) to 61%. This should cause a remarkable increase in the entropy of the water surrounding the peptide, which may be the thermodynamic reason the dimer is much more stable than two monomers.

A conformational-clustering algorithm was used to characterize the conformational space sampled in the simulations. The number of clusters of conformations from the three trajectories as a function of time is shown in Figure 3 of the Supporting Information. A cluster radius of 0.1 nm was used. The number of clusters increases in all three simulations, and linear regression analysis gives a high correlation coefficient (>0.98), meaning that the number of clusters increases with time in a linear fashion. This is an indication that convergence, that is, a complete sampling of conformational space, is far from being reached; this is expected, because it is known that the time required for equilibration of all possible tertiary and quaternary structures is much longer than the total simulation time. Nevertheless, some of the most relevant conformations are likely to have been sampled within the 100 ns period.

It is important to remark that our simulations do not represent the equilibrium for either the monomer or the dimer systems; according to experimental data on the kinetics of folding of other α-helical proteins, the equilibration time for such large systems should be at least 2 or 3 order of magnitude longer than our simulation time, which makes equilibrium simulations not feasible with present day computers. This observation has important consequences in the interpretation of our results; in particular, values for the interproton distances and the percentages of secondary structure content have to be interpreted in a qualitative way.

The secondary structure content as a function of time is reported in Figure 9. The percentages of α-helix are 73%, 54%, and 57% in simulation number 1, 2, and 3, respectively. Table

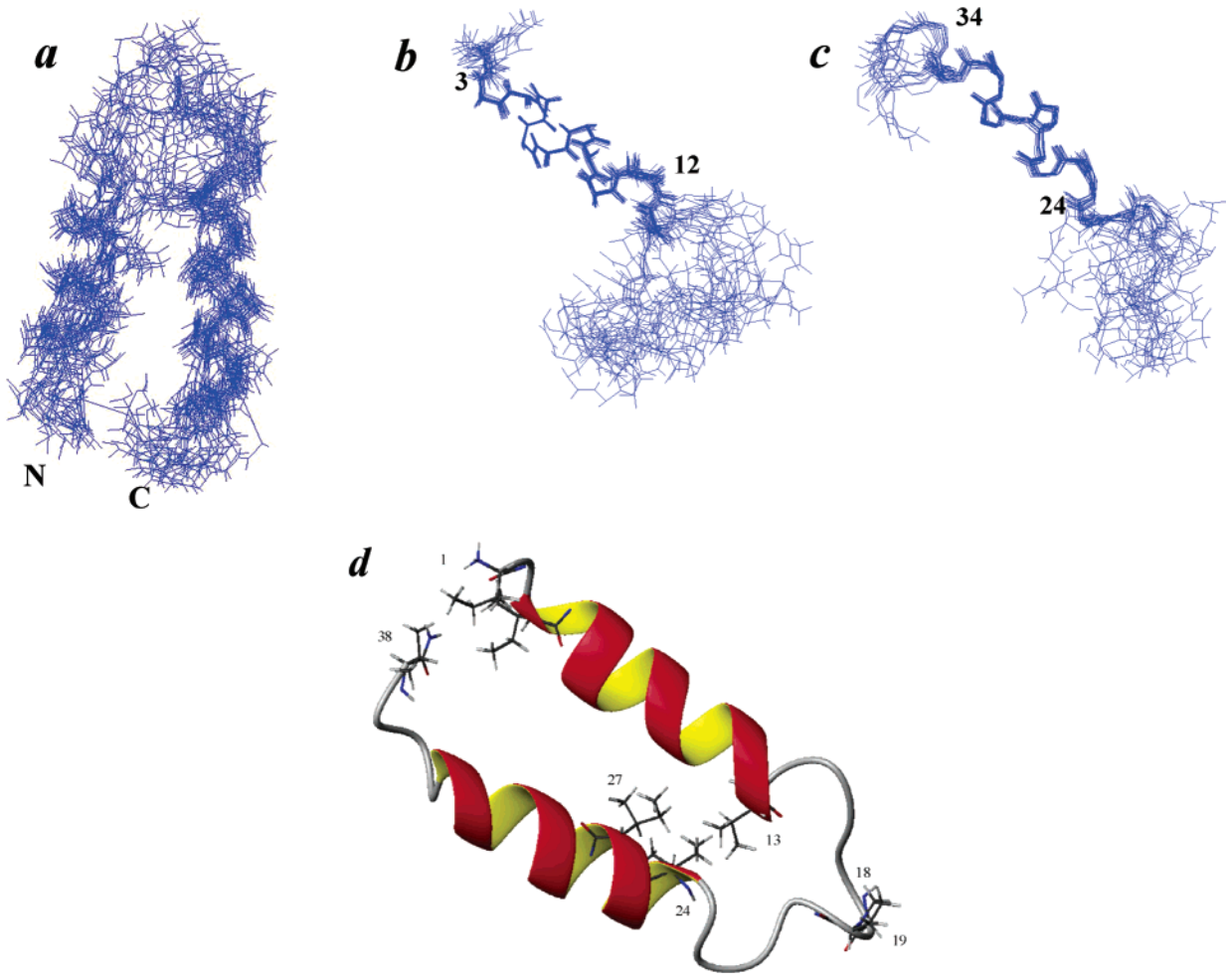

Figure 8. (a) Superimposition of the backbone heavy atoms of the N-terminal 3−12 and C-terminal 25−34 sequences of the ensemble of low energy structures; (b) superimposition of the backbone heavy atoms of the 3−12 sequence (residues of the C-terminal 23−38 sequence were omitted); (c) superimposition of the backbone heavy atoms of the 23−35 sequence (residues of the N-terminal 1−16 sequence were omitted); (d) one of the low energy structures.

Table 2. Summary of the Simulations Performed in the Present Work

number of simulation	% α-helix	localization of the helices	number of violations (>0.1 nm)
1	73	3−15 || 21−36	16
2	54	A: 3−12 || 23−33 B: 3−13 || 23−32	15
3	57	A: 3−12 || 23−33 B: 4−14 || 22−34	15

2 shows percentages and localization of the secondary structure elements that are in agreement with the experimental results.

A closer look at the trajectories reveals that several partial unfolding−refolding events are observed during all three simulations; this is consistent with the known time scale for the folding of α-helices, which is of the order of tens of nanoseconds. In the simulation, the dimerization to the putative four-helix bundle leads to a decrease of helix content from 75% to about 55%. This is due to the conformational rearrangements that are necessary to optimize the interaction between the side chains of the two peptide units. These rearrangements lead to significant distortions of the secondary structures in the time scale of the simulations but not on the time scale of NMR experiments.

Interproton distances derived from experimental NOEs were compared to the corresponding average distances in the simulations. In Table 2, the number of violations is reported for each of the three simulations. Using $<d(t)^{-6}>^{-1/6}$ distance weighting, important violations (higher than 0.1 nm) are observed for only 16 NOEs in the simulation of the monomer (corresponding to 5.9% of all measured NOEs) and 15 distances in both simulations of the dimer (corresponding to 5.5% of all measured NOEs). Again, the few large deviations from NMR data can be explained by the insufficient sampling of the conformational space, and the very low percentages of violations indicate good agreement between NMR data and MD calculations. A full description of the violations in each of the simulations is reported in Tables 2−4 of the Supporting Information. Large violations around P^{19} are found in all three simulations, and this may be an indication of the simultaneous presence of many conformers in solution, due to cis−trans isomerization of the G^{18}-P^{19} peptide bond. These violations were found also in the DG/SA calculations. The fact that long range NOEs are often violated in the MD calculations does not imply that the two helices in the peptide (nor the two peptide units) turn away from each other. In fact, many other long range contacts are found during the simulations, for example, between the side chains of V^{13}-I,[22]

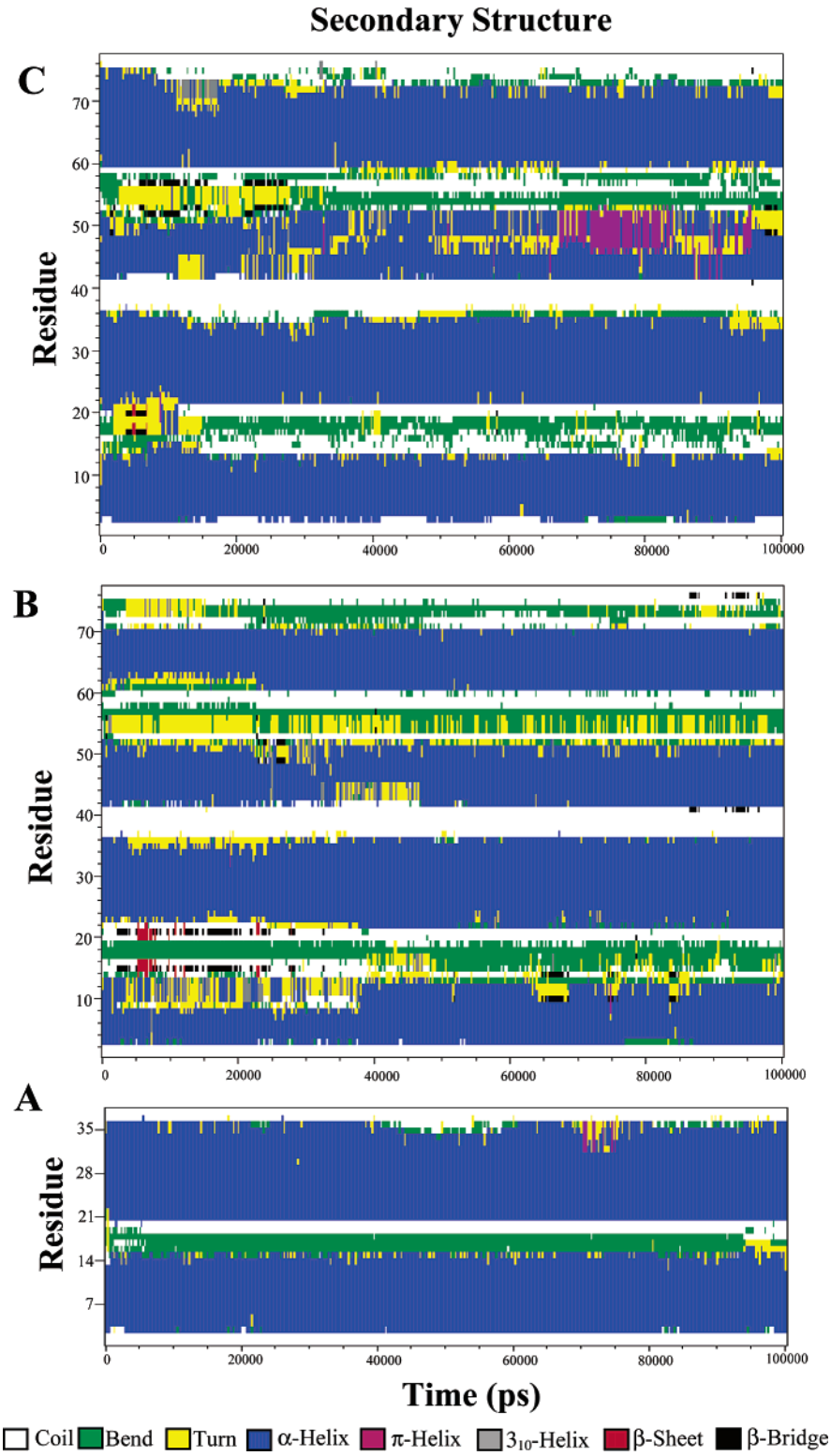

Figure 9. Secondary structures as a function of simulation time in each of the three simulations: (A) simulation 1 (monomer); (B) simulation 2 (dimer, with the two peptide molecules close to each other in the starting structure); (C) simulation 3 (dimer, with the two peptide molecules set at a distance of 1.8 nm in the starting structure).

V^{13}-D,^{26} K^{2}-H^{37} (all intramolecular), and I^{3}-A^{15} (intermolecular). Unfortunately, NOE cross-peaks supporting these contacts could not be experimentally detected because of extensive overlap.

We also performed cluster analysis with a cluster radius of 0.15 nm on the second and third simulations. One interesting result is that, in both cases, the four most populated clusters account for approximately 75% of all structures (see Figure 10 for simulation number 2).

No structure fulfils all the NOE restraints, and also the central structure of the most populated cluster has several violations; nevertheless, the ensemble of the central structures of the four most populated clusters shows very few violations. Therefore

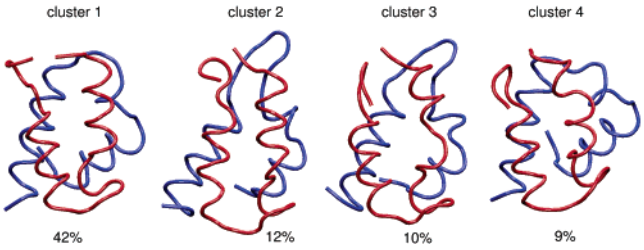

Figure 10. Backbone representation of the four most populated clusters in the second simulation. Cluster radius was 0.15 nm. The population of each cluster is indicated.

it represents the conformation of the peptide in solution much better than any single structure.

Conclusions

The CD results presented in this work gave a first, even if not unambiguous, indication of the possible folding of **BIII-GPVD-BIII** in the hairpin helix−loop−helix structure. However, the CD data could not provide detailed information about dimerization or further aggregation of the helix−loop−helix structure since these events should not enhance substantially the helix content.

At a 1 mM concentration, analysis of sequential and medium range NOESY connectivities and the results of structural calculations identify two helical segments in the sequences 3−14 and 23−37. The first helical segment is slightly shorter than that observed in aggregates of bombolitin III[3] where the helical sequence extends from Ile[3] to Val[17]. A possible reason for the slightly shorter N-terminal helix of **BIII-GPVD-BIII** can be the presence of the GPVD tetrapeptide bridge. In particular, the presence of cis−trans isomerism, which modifies the conformation of the Val[17]-Gly[18] residues immediately preceding Pro[19], could affect the structural stability of the C-terminal residues of the first helix. Medium range connectivities are clearly missing in the loop region, and it is quite possible that, in this segment, averaging of different structures in fast exchange occurs. Furthermore, secondary chemical shifts, NOESY connectivities, and structural calculations indicate that 27 residues, corresponding to 71% of the sequence, are comprised in the helical segments. This figure is slightly higher than that resulting from CD data and suggests that the helical segments in the hairpin fold are rather flexible, again indicating the presence of conformers with different helix content in fast exchange on the NMR time scale.

The results of NMR experiments and restrained molecular dynamics calculations unambiguously indicate the formation of the hairpin helix−loop−helix structure. Evidence for this structural motif arises from long-range NOESY connectivities between protons of amino acid residues distant along the helical sequences. Such NOEs are clearly due to interactions between helical segments oriented in an antiparallel way as in the helix−loop−helix motif. Furthermore, the two NOESY connectivities βH(Ile[3])−β2H(Pro[19]) and βH(Val[38])−αH(Gly[18]), which are possible only between protons of different molecules, are consistent with the presence of dimeric or higher aggregates of **BIII-GPVD-BIII** molecules with formation of helix bundles.

The tendency to dimerization with formation of the four-helix bundle is supported by the results of the extensive molecular dynamics simulation in explicit solvent. These data indicate a higher stability of the dimer in which a sizable portion of the hydrophobic surface is protected from the solvent.

The analysis of the location of the amino acid residues on the helical surfaces in terms of the heptad repeat pattern[38,39] showed that, in the hydrophobic core, the residues in "a" and "d" positions are Ile, Leu and Val. However, several hydrophobic residues occupy positions "b", "e", and "g", located at the helical interface. This suggests that these residues are solvent exposed in positions where they can interact with solvent exposed hydrophobic residues of neighboring helices. These interactions are similar to those observed in β-hairpins by Waters et al.[40,41] and could be one of the factors stabilizing the helix bundle motif. The presence of a single acidic residue per helix seems insufficient to provide enough stabilization of the helix bundle through interhelical salt bridges. The Asp residues occupy the "f" position in the heptad repeat in which they cannot contribute to interhelical interactions. The unique fold, stabilized by hydrophobic interactions between solvent exposed hydrophobic residues, could provide a new design strategy of four-helix bundles with high hydrophobic character.

From cluster analysis, it is clear that longer simulations are required to sample all conformations available to the peptide. Nevertheless, the length of the simulations carried out in the present work is sufficient to account for almost all experimental NOEs.

The presence of only two connectivities indicative of intermolecular aggregation cannot resolve the ambiguity of the formation of dimers or higher aggregates. Extensive overlap prevented the observation of additional interchain NOEs which would allow us to assess unambiguously dimerization and consequent formation of the four-helix bundle motif. To solve this problem, further work is needed. The design of a **BIII-GPVD-BIII** analogue with labeled side chains in the helical segments should allow the identification of a larger number of intermolecular NOE connectivities. This and the complete characterization in terms of molecular weight of the species in solution should allow us to assess unambiguously the formation of dimers reproducing the four-helix bundle motif. Work is in progress in this direction and will be reported elsewhere.

Acknowledgment. This work is supported by the National Research Council (CNR) of Italy.

Supporting Information Available: Table of proton chemical shifts (Table 1). Tables 2−4 with violations of NOE restraints in simulations number 1, 2, and 3. Figure 1: radius of gyration of the peptide as a function of the simulation time for each of the three simulations. Figure 2: solvent accessible surface area of the peptide as a function of the simulation time. Figure 3: Number of clusters as a function of the simulation time in each of the three simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0300970

(40) Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 9372−9373.
(41) Butterfiels, S. M.; Waters, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 9580−9581.