Exendin-4 and Glucagon-like-peptide-1: NMR Structural Comparisons in the Solution and Micelle-Associated States†

Jonathan W. Neidigh,‡ R. Matthew Fesinmeyer,‡ Kathryn S. Prickett,§ and Niels H. Andersen*,‡

Department of Chemistry, University of Washington, Seattle, Washington 98195, and Amylin Pharmaceuticals, Inc., 9373 Towne Centre Drive, San Diego, California 92121

Received May 2, 2001; Revised Manuscript Received August 24, 2001

ABSTRACT: Exendin-4, a 39 amino acid peptide originally isolated from the oral secretions of the lizard *Heloderma suspectum*, has been shown to share certain activities with glucagon-like-peptide-1 (GLP-1), a 30 amino acid peptide. We have determined the structuring preferences of exendin-4 and GLP-1 by NMR in both the solution and dodecylphosphocholine (DPC) micelle-associated states. Based on both chemical shift deviations and the pattern of intermediate range NOEs, both peptides display significant helicity from residue 7 to residue 28 with greater fraying at the N-terminus. Thornton and Gorenstein [(1994) *Biochemistry 33,* 3532−3539] reported that the presence of a flexible, helix-destabilizing, glycine at residue 16 in GLP-1 was an important feature for membrane and receptor binding. Exendin-4 has a helix-favoring glutamate as residue 16. In the micelle-associated state, NMR data indicate that GLP-1 is less helical than exendin-4 due to the presence of Gly^{16}; chemical shift deviations along the peptide sequence suggest that Gly^{16} serves as an N-cap for a second, more persistent, helix. In 30 vol-% trifluoroethanol (TFE), a single continuous helix is evident in a significant fraction of the GLP-1 conformers present. Exendin-4 has a more regular and less fluxional helix in both media and displays stable tertiary structure in the solution state. In the micelle-bound state of exendin-4, a single helix (residues 11−27) is observed with residues 31−39 completely disordered and undergoing rapid segmental motion. In aqueous fluoroalcohol or aqueous glycol, the Leu^{21}−Pro^{38} span of exendin-4 forms a compact tertiary fold (the Trp-cage) which shields the side chain of Trp^{25} from solvent exposure and produces ring current shifts as large as 3 ppm. This tertiary structure is partially populated in water and fully populated in aqueous TFE. The Leu^{21}−Pro^{38} segment of exendin-4 may be the smallest protein-like folding unit observed to date. When the Trp-cage forms, fraying of the exendin-4 helix occurs exclusively from the N-terminus; backbone NHs for the C-terminal residues of the helix display H/D exchange protection factors as large as 10^5 at 9 °C. In contrast, no tertiary structure is evident when exendin-4 binds to DPC micelles. An energetically favorable insertion of the tryptophan ring into the DPC micelle is suggested as the basis for this change. With the exception of exendin-4 in media containing fluoro alcohol cosolvents, NMR structure ensembles generated from the NOE data do not fully reflect the conformational averaging present in these systems. Secondary structure definition from chemical shift deviations may be the most appropriate treatment for peptides that lack tertiary structure.

Diabetes mellitus is a chronic disease characterized by multiple metabolic abnormalities arising primarily from an inadequate insulin effect. Glucagon-like-peptide-1 (GLP-1),[1] a 30 amino acid mammalian hormone, has been investigated as a potential therapeutic agent for the treatment of this disease because of its favorable spectrum of antidiabetic actions including a glucose-dependent insulinotropic action (*1, 2*), modulation of gastric emptying (*3, 4*), and a possible role in appetite control (*5*). Due to its short duration of action, GLP-1 may have limitations as a therapeutic agent.

Exendin-4 (*6*) is a 39 amino acid peptide that exhibits several of the antidiabetic actions of the mammalian hormone GLP-1. Exendin-4 displays approximately 50% sequence identity to GLP-1 and binds to and acts as an agonist at the GLP-1 receptor on RINm5f cell membranes (*7*). It has been

† The purchase of the 750 MHz NMR spectrometer was supported by grants from the M. J. Murdock Charitable Trust and the NSF Academic Research Infrastructure Program; the DRX console upgrade of the 500 MHz NMR spectrometer was supported by NSF Grant CHE-9710008. Studies at the University of Washington were supported by a grant from Amylin Pharmaceuticals, Inc.

* Corresponding author. E-mail: andersen@chem.washington.edu, Phone: 206 543-7099, FAX: 206 545-8665.

‡ University of Washington.

§ Amylin Pharmaceuticals, Inc.

[1] Abbreviations: the standard one- and three-letter symbols for amino acid residues are employed; ADC, anti-distance constraint; CD, circular dichroism; CNS, the Crystallography & NMR System software suite; CSD, chemical shift deviation from random coil reference values ($\delta_{obs} - \delta_{ref}$); DC, distance constraint; DPC, dodecylphosphocholine; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate; GLP-1, glucagon-like-peptide-1 (7−36)-amide; HFIP, hexafluoro-2-propanol; LBO, low-bound-only; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; se, standard error; TFE, trifluoroethanol; TOCSY, total correlation spectroscopy; $V(i,j)$, the volume or intensity of the NOE cross-peak between H_i and H_j; ncl, number of contour levels for a peak in a NOESY spectrum. NOESY peaks and derived distance constraints are designated as connectivities; $\alpha_i N_{i+1}$, for example, is a through-space interaction between Hα of residue i and the NH of residue $i+1$. All solvent compositions are given in vol-% terms; '%' should be understood to be 'vol-%'.

10.1021/bi010902s CCC: $20.00
Published on Web 10/13/2001

shown to have activities known to be important for improvement of glucose control and to be markedly more potent or long-lived in vivo than GLP-1 for these activities (*8*). In animal models, exendin-4 has been shown to stimulate secretion of insulin in the presence of elevated blood glucose concentrations, but not during hypoglycemic periods. Synthetic exendin-4 (AC2993) is currently in Phase 2 evaluation for the treatment of type 2 diabetes.

To better understand GLP-1 and exendin-4, we have compared them by circular dichroism (CD) and NMR in both the solution and dodecylphosphocholine (DPC) micelle-associated states. The CD studies are reported elsewhere (*9*). To our knowledge, there are only three previous structural studies related to GLP-1 and exendin-4. One study, using FT-IR and near-IR FT-Raman spectroscopies (*10*), investigated the effect of media and sample preparation on the peptide conformation and the resulting tendency of insulintropin to aggregate. Insulintropin differs from GLP-1 only at the C-terminus where Gly^{31} replaces the C-terminal amide. In the second study, GLP-1 was examined by NMR in aqueous media containing dodecylphosphocholine (DPC) micelles (*11*). That study suggested that GLP-1 is helical from residue 7 to residue 29 with a distortion in backbone geometry at residues 15−17. Thornton and Gorenstein suggested that the flexible link about Gly^{16} allows all of the aromatic residues to line up on one face of the helix facilitating membrane association. They further suggested that displaying these hydrophobic residues on a single face could be important for receptor binding affinity. Glucagon, which otherwise displays significant homology to GLP-1, lacks the flexible linker induced by Gly^{16} and displays greatly diminished binding to GLP-1 receptors (*12*). The sequences of the peptides are shown below:

	1	11	21	31
Exendin-4	HGEGTFTSDL	SKQMEEEAVR	LFIEWLKNGG	PSSGAPPPS-NH_2
GLP-1	HAEGTFTSDV	SSYLEGQAAK	EFIAWLVKGR-NH_2	
Insulintropin	HAEGTFTSDV	SSYLEGQAAK	EFIAWLVKGR G	
Glucagon	HSQGTFTSDY	SKYLDSRRAQ	DFVQWLMNT	

The central LEGQ segment of GLP-1, which contains the flexible glycine, is replaced by MEEE in the corresponding sequence of exendin-4. Thus, the ability of both GLP-1 and exendin-4 to bind and promote agonist activity at the GLP-1 receptor in vitro suggests that the hypothesized importance of the structuring effects of Gly^{16} in GLP-1 for receptor activation needs to be reexamined. In the third literature study, Parker et al. (*13*) examined a series of GLP-1 mutants, an exendin-4/GLP-1 chimera, and several cross-linked GLP-1 mutants by NMR in 50 vol-% aqueous TFE and concluded that most of the species examined display two helices (Asp^9−Glu^{15}, Ala^{19}−Val^{27}) in this highly helix-favoring medium. In 50% TFE, the more converged helical span was from Ala^{19} to Val^{27}. This C-terminal helix was viewed as an essential feature of the bioactive conformation of GLP-1 serving to orient key residues such as Phe^{22}, Ile^{23}, and Trp^{25} as required for interaction with the receptor.

MATERIALS AND METHODS

Peptide Samples. Exendin-4 and GLP-1 were prepared by standard solid-phase peptide synthesis protocols (Amylin lots #CS171297 and #P180597, respectively). Amino acid compositions and purity were confirmed by amino acid analyses (average error in residue count ×e1⟨6.1 ± 5.9%⟩ for Glx, Asx, and all residues other than Trp; ⟨1.7%⟩ for Asx/Glx/Pro/Gly/Met/Leu/Lys/His) and HPLC (>98% on Vydac C18). The average molecular weights observed by MS were within 0.6 amu of the expectation based on the sequence. The sequences were confirmed by the NMR studies reported herein.

NMR Experiments. Solution-state protic NMR samples (ca. 500 μL) were made by dissolving 1−4 mg of lyophilized peptide in 15 mM phosphate or phosphate/acetate buffer (pH 3.5 or 5.9) and immediately adding cosolvents as needed to the stated final compositions (vol-%, simply indicated as '%' from here forward): 30% d_3-TFE, 60% d_3-TFE, 36% d_2-HFIP (hexafluoro-2-propanol), 8.5% d_2-HFIP with 18.4% d_6-1,2-ethanediol (glycol), and 30% d_6-glycol. The micelle samples were prepared by mixing a solution of the peptide with a micelle-containing solution to give the desired peptide concentration and ratio of peptide to micelle. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used to calibrate the chemical shifts in all samples. NOESY and TOCSY spectra were collected for all media, typically at several temperatures to resolve partially overlapping peaks with different temperature gradients. The D_2O-exchanged sample of exendin-4 was prepared by repeated addition of 99.96% D_2O and lyophilization (3 times over 3 days) starting with an initial 700 μM peptide solution in 15 mM phosphate-buffered D_2O containing 15% TFE adjusted to pH* 5.35. TOCSY spectra employed the clean MLEV-17 spin-lock (*14*) with mixing periods of 10−55 ms. Solvent suppression was accomplished using the WATERGATE (*15*) pulse sequence. With the exception of the NH exchange studies, all the spectra were collected on a Bruker DMX operating at 750 MHz. NOESY spectra were also collected for the fully deuterium-exchanged state in each medium. Anti/gauche $\alpha/\beta\beta'$ assignments and $H_N/\alpha\alpha'$ coupling constants were confirmed by short mixing time (10−15 ms) TOCSY spectra. The NH exchange studies were acquired on a Bruker DRX spectrometer operating at 499.8 MHz. 1D spectra were collected over a period of 15 days at 297 ± 2 °C for both pD 3.5 and pD 5.9 solutions to estimate exchange half-lives. An NOESY spectrum of the pD 3.5 sample was collected 4 h after exchange was initiated to verify assignments of the 1D spectra for all slow-exchanging amide protons. More precise exchange rates were obtained at 7 °C (for the first 16 h, then at 9 °C for the remaining days) using 3:7 volume mixtures of pH* 6.03 and 7.20 aqueous D_2O buffer (25 mM phosphate) and TFE (corresponding to pD 6.0 and 6.6 due to the leveling effect of the added TFE). At pD 6.6, the exchange rates for the Ile^{23}−Lys^{27} segment ranged from 0.009 to 0.173 $days^{-1}$. The reference exchange rates employed for protection factor calculations for the Trp side chain and a backbone Ala-Ala unit at 7 °C were, respectively, at pD 6.0, 1.00 and 2.69 min^{-1}, and at pD 6.6, 2.94 and 10.72 min^{-1}, based on the closest literature analogies for strictly aqueous media (*16, 17*).

A precise conversion of the measured exchange rates to protection factors is problematic due to the undefined effects of TFE upon intrinsic exchange rates and Molday factors. For backbone NHs, we have assumed a 20% reduction in the intrinsic rate and that the Molday factors that are known for low-salt aqueous conditions (*16*) apply. Buck et al. (*18*) have reported that TFE addition at levels up to 50% changes

the intrinsic backbone NH exchange rates by less than a factor of 3. Our reported backbone NH exchange protection factors could be as much as 2−3 times the actual values. Control exchange studies (present work) for a Trp-containing tripeptide suggest that no TFE correction is required for the intrinsic exchange rate of the indole ring NH. The acidifying effect of TFE was taken in account in the calculations. At pH ≥5.5, it can safely be assumed that backbone NH exchange rates should increase 10-fold with a 1 unit increase in log[DO^-]. As a result, the leveling effect (and possible changes in the glass electrode response) could be calibrated from a semilog plot of pH meter reading versus k_{obs} over the pH meter reading range 5−8.

NMR Structure Ensemble Generation. NOESY data from all available experimental conditions were converted to distance constraints. Distance constraint tables of varying precision were developed from the NOE data; more precise constraint tables allowed for improved assignment error checking, while the less precise, "conservative" constraints reduce the possibility of data over-interpretation. The observed NOE intensities, recorded as the number of contour levels (ncl), were corrected for relative peak width in the directly detected dimension and for the H/D ratio for peaks involving NHs. These adjusted peak contour levels were converted to distances using the relation: $d_{exp} = d_{ref}\cdot[1.41^{(1/6)}]^{(ncl_{ref}-ncl_{exp})}$, reflecting the contour multiplier of 1.41 which was used for plotting the spectra. The reference distances for specific classes of NOEs were the following: smallest $\alpha_iN_{i+1} = 3.3-3.8$, largest $\alpha_iN_{i+1} = 2.0-2.4$, largest N_iN_{i+1} in a helical span = 2.35−2.9, and $\langle\alpha_iN_i\rangle = 2.8$ Å. As a result, relatively tight distance constraints, $r_{ij} = (d_{ij} - d_{-}) \rightarrow (d_{ij} + d_{+})$, were derived. For exendin-4, as an example, the average low and high distance bounds over all constraints with an upper bound in the range 3.3−3.9 Å were 2.81 and 3.63 Å ($n = 89$), considerably more restrictive than the 'medium' or 'small' category of a standard four-group NOE classification: L (large), M (medium, typically 2.0−3.5 Å), S (small), and VS (very small). Distance bounds for methyl and shift-coincident methylene groups derived from NOEs were increased by a factor of 1.2 and 1.04, respectively. These constraints were used in both XPLOR3.1 (*19*) and CNS0.9 (*20*) simulating annealing scripts. Further details of the XPLOR procedures, which included H-bond constraints and low-bound-only constraints, also known as ADCs (*21*), and of CNS runs using constraints of comparable and lesser precision appear in the Supporting Information. The exendin-4 ensemble appearing in Figures 5−7 was generated using a conservative CNS protocol with a universal 2.0 Å low-bound and four upper-bound categories [L (3.0), M (3.5), S (4.0), and VS (5.0 Å)]; for constraints involving methyl groups, the upper bounds for the four categories were extended to 3.15, 3.8, 4.6, and 6.0 Å. Similar upper limits have been used for both proteins and peptides; e.g., structures of a series of constrained GLP-1 analogues employed 3, 4, and 5 Å upper limits (*13*). The Supporting Information contains a complete listing of the distance constraints used to derive solution-state structure ensembles for exendin-4 with varying levels of constraint precision. The coordinates and NMR restraints for the "conservative" structure have been deposited in the Protein Data Bank under accession number 1jrj.

Stereospecific assignments were obtained for some of the Cα and Cβ methylenes. For the Cβ methylenes, the anti/gauche assignments were converted into alternative distance constraints ($d_{\alpha_i\beta_i} = 2.85-3.20$ and 2.30−2.75 Å, respectively, which were used in lieu of torsion constraints) when confirmed by $\alpha\beta/\alpha\beta'$ NOE ratios observed in the short mixing time NOESY; in the most conservative CNS protocol, these were viewed as M and S category NOEs, respectively. The two alternative assignments were examined for these and all $\beta_iN_{i+1}/\beta_iN_i/\beta_i\alpha_{i-3}$ constraints. If one assignment produced lower E_{NOE} values for a large majority of the structures in the ensemble, that stereospecific assignment version was used in the final ensemble generation. A similar procedure was used for Gly-$\alpha\alpha'$, Val-$\gamma\gamma'$, Ile-$\gamma_1\gamma_1'$, and Leu-$\delta\delta'$ assignments.

The default CNS simulated annealing script, which includes high-temperature dynamics and then a cooling cycle in torsion space (with $k_{NOE} = 150$) followed by further cooling in Cartesian space ($k_{NOE} = 50$) and minimization ($k_{NOE} = 75$), was modified in the following ways: (a) the high-temperature torsion dynamics and cooling phases were each doubled in length (2000 15-fs steps each, initially at 50 000 K with cooling during the last 2000 steps); (b) the Cartesian cooling cycle (starting at 2000 K, $k_{NOE} = 75$ or 50 kcal/Å^2) was also doubled in length (6000 5-fs steps) which allowed us to ramp the E_{repel} scale factor from 0.2 to 1.0 during the cycle; (c) chemical shift and H-bond constraints were not employed, and the dihedral angle energy function was disabled; (d) the E_{repel} function was replaced by a Lennard−Jones potential during the final Powell minimization—10 cycles of 100 steps each during which k_{NOE} is set at 75 (when loose, conservative constraints are employed) or 40 (for more precise constraints) kcal/Å^2. We have found that these changes greatly improve both the convergence and success rate of CNS annealing for exendin-4, and also for a helical peptide and a small single-domain protein. In the present study, 50 random structures were used, producing 36 accepted structures with no NOE distance restraint violations in excess of 0.2 Å. The other acceptance criteria were based on agreement with idealized covalent geometry (no single bond length, angle, or improper torsion rmsd exceeding 0.02 Å, 3°, and 2°, respectively, in the structure; in addition, the mean bond length, angle, and improper torsion rmsd had to be <0.01 Å, 0.5°, and 0.3°, respectively, for acceptance).

Evaluation of Convergence and Structural Deviations in NMR Ensembles. The structure (bond, angle, and improper torsion rmsd versus standard values) and violation statistics [fraction of constraints violated by >0.1 Å and rmsd (±se) from the distance constraints] are over the complete ensemble. Backbone and heavy atom rmsd measures of convergence are, unless otherwise noted, the mean ± se for pairwise comparisons over the entire accepted ensemble rather than rmsd versus the mean structure. The rmsd calculations and figures depicting structure ensembles were generated using MolMol (*22*).

RESULTS

Essentially complete chemical shift assignments and NOE intensity measurements were obtained for exendin-4 in six media: (1) aqueous DPC micelles; (2) 30% ethane-1,2-diol

(glycol) in water at 320 K; (3) aqueous buffer with added glycol (18%) and hexafluoro-2-propanol (HFIP, 8%) at 314 K; (4) 30% aqueous TFE; (5) aqueous 36% HFIP; and (6) aqueous 60% TFE. Chemical shift comparisons for GLP-1 in the DPC micelle-associated state were taken from the literature (*11*). Since Parker et al. (*13*) did not report the GLP-1 chemical shift data for their study in 50% TFE, we collected NMR data for GLP-1 in 30% TFE for comparison with the exendin-4 data. Preliminary CD comparisons (*9*) indicated that GLP-1 forms helix bundles (and within days larger aggregates) at NMR concentrations in the absence of added lipid micelles or fluoroalcohol. This precludes a meaningful NMR examination of GLP-1 in aqueous buffer. Under all of the conditions used in the present NMR investigations, CD spectra are identical at 20 μM and 1.2 mM, and the spectra are not time-dependent. For exendin-4 under conditions (2) and (3), line broadening at low temperatures suggests a modest level of aggregation; therefore, the data were collected at higher temperatures, and structures were not derived from the NOESY data. For the other media, line widths and relaxation rates at 280−320 K were consistent with expectations for a monomeric system of this molecular weight.

Resonance assignments were made by the usual TOCSY/NOESY strategy (*23*). Since there are residue differences between exendin-4 and GLP-1, all shift comparisons are expressed as chemical shift deviations (CSDs). The coil reference values have been previously reported along with the corrections for nearest neighbor, end, medium composition, and, in the case of NH's, temperature effects (*24*). Structure ensembles were generated using both conservative (loose bounds) and aggressive (more precise distance constraints) interpretations of the NOEs. In retrospect, only for exendin-4 in solutions containing some added fluoro alcohol, was the constraint density sufficient for the derivation of a high-resolution structure. Changes in structuring preferences with medium and the contrasts between GLP-1 and exendin-4 are based on CSD histograms and qualitative NOE comparisons. NMR ensembles for GLP-1 in aqueous TFE and exendin-4 in aqueous DPC micelles are presented in the Supporting Information.

Structuring Preferences of GLP-1 in Aqueous TFE and in the DPC Micelle-Associated State. The complete resonance assignment for GLP-1 in 30% TFE appears in Table 1S. NOESY spectra acquired for the solution state in 30% TFE/70% pH 5.9 aqueous buffer at 285 K (τ_m = 160 and 60 ms) and 300 K (τ_m = 60 ms) provided 119 structurally significant NOE distance constraints: 16 intraresidue and 104 short- to medium-range interresidue ($i/i\pm n$, n = 4). N_iN_{i+1} NOEs and a nearly continuous set of $i/i+3$ NOE connectivities (αN and/or $\alpha\beta$) were observed from Leu[7] through Ala[24]. This suggests significant helicity over this span. However, a more quantitative NOE indicator of helicity, α_iN_{i+1} to α_iN_i NOE ratios (*25*), provides further insight. For well-populated helices, α_iN_{i+1} NOE intensities should be equal to or less intense than α_iN_{i+3} NOEs for both α and 3_{10} helices. Local unfolding or end-fraying increases the intensity of the α_iN_{i+1} NOEs dramatically. The α_iN_{i+1} NOEs of GLP-1 are significantly smaller than the α_iN_i NOEs only for residues 10−14 and 19−24. Also, the intensities of $i/i+3$ and $i/i+4$ NOEs that spanned and immediately followed Gly[16] were reduced. Significant sampling of more extended conformations must be occurring for GLP-1 under these conditions; this is not fully reflected in the NMR structure ensemble (see Supporting Information). The conflict between medium-range and vicinal residue NOE constraints is more notable when relatively tight distance constraints are employed: the $i/i+1$ constraints will have less influence on the structure when wider, conservative, bounds are employed. Precise (tight) constraints yield relatively large ($\geq$20 kcal/mol) E_{NOE} values when conformational averaging is occurring producing either alternative conformer clusters or false convergence. The application of conservative (loose) bounds in peptide structure ensemble calculations is likely to emphasize the conformers with intermediate-range order; this is also a case of false convergence.

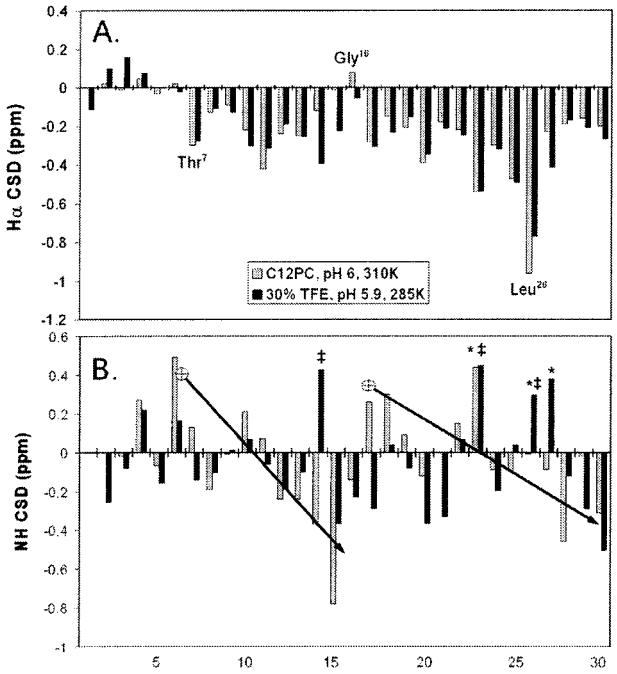

FIGURE 1: Chemical shift deviations for GLP-1 in aqueous 30% TFE and the micelle-associated state. Panels A and B show the Hα and NH CSD histograms, respectively. The NH random coil values were corrected to the experimental temperature using a slope of −7.6 ppb/°C for both the micelle-bound and aqueous 30% TFE states. In panel A, the positions of Thr[7], Gly[16], and Leu[26] are labeled. In panel B, the effects of two helix macrodipoles that may apply for the micelle-associated state and markers to indicate positions that correspond to residues that are exchange-protected (*) or at locations (‡) where a downfield ring current shift might be expected are included.

CSD histograms provide a semiquantitative basis for assessing differences in secondary structure along a peptide sequence. The interpretation of Hα CSDs is well established; significantly positive values indicate a β strand preference, and −0.39 ppm (*26*) is viewed as the value that represents 100% helicity. The interpretation of NH shift deviations is more controversial. It is essential to include the random coil temperature gradient of NH shifts in the calculation (*24, 27*). The H_N and Hα chemical shift deviation plots for GLP-1 in the micelle-associated and aqueous TFE solution states appear in Figure 1; the micelle values are derived from the shifts reported by Thornton and Gorenstein (*11*). The Hα CSD plot (panel A) indicates helicity from residues 7 to 30 for GLP-1 in both solvents. Both ends of this 'helix' are frayed, and the residues about Gly[16] sample ϕ/ψ conformation space outside the helical region. The only positive CSD

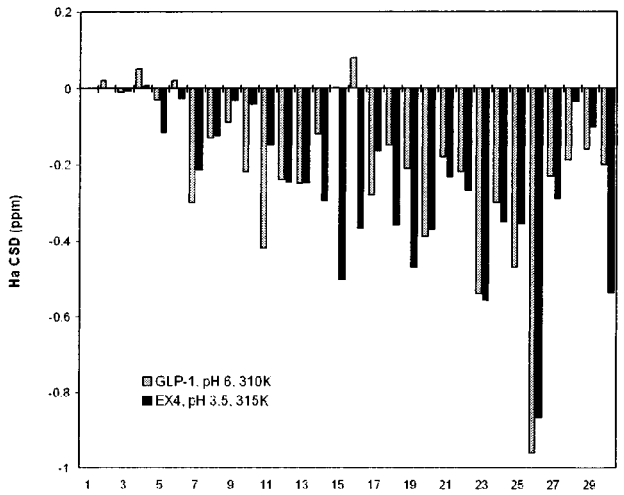

FIGURE 2: Sequence histogram Hα CSD comparison for exendin-4 and GLP-1 in the aqueous DPC micelle-bound state.

is for Gly[16] in the aqueous DPC micelle medium, and this residue displays a greatly diminished upfield shift in 30% TFE. In the DPC micelle data set, residues 14 and 15 also display greatly diminished upfield shifts. This is consistent with a more pronounced helix-interrupting, N-capping role for Gly[16] in the micelle-bound state.

Amide NH CSDs can also provide evidence for the medium-induced structural distinctions. It has been established that the helix macrodipole causes large position-dependent changes in H_N chemical shifts (*28*). On this basis, the increasingly upfield H_N shifts at residues 12−15 in the micelle data indicate the C-terminal end of a helix macrodipole (Figure 1B). After Gly[16], downfield H_N shifts are again observed reflecting the positive N-terminus of a new helix. The H_N shift data for 30% aqueous TFE do not display the anticipated effects of two helix macrodipoles. Apparently, the helix propagates through Gly[16] in most of the conformers present in 30% TFE. Four NHs (Leu[14,26], Ile[23], and Val[27]) are notably downfield in aqueous TFE; this probably reflects particularly strong hydrogen bonding on the hydrophobic face of the helix (*26*).

Structure of Exendin-4 in the DPC Micelle-Bound State. To compare the micelle-bound state of exendin-4 with that reported for GLP-1 (*11*), micellular solutions of ca. 200 mM aqueous DPC were used to solubilize exendin-4 (1.5 mM) at both pH 3.5 and pH 5.9. To minimize the line widths and obtain easily assignable 2D spectra, the definitive NMR data were collected using the pH 5.9 buffer at 314 K. The chemical shift assignments for exendin-4 in the DPC micelle-associated state appear in Table 2S. The Hα CSDs of exendin-4 and GLP-1 in the micelle-associated state are compared in Figure 2. The data for exendin-4 support a helical conformation for the residue 7−27 span with significant fraying in the N-terminal turn. The Hα CSD histogram indicates that both extreme termini are fully disordered (see Figure 3B, vide infra, for the CSDs at residue 30−39). Qualitative analysis of NOEs also supports this view. For example, the flexibility of the C-terminus is fully confirmed by NOESY/TOCSY peak intensity comparisons: interresidue NOESY peaks are present but of diminished intensity for residues 31−33 and absent any further toward the C-terminus. The C-terminal residues were assigned only from the sharp and intense TOCSY peaks. The lack of NOEs

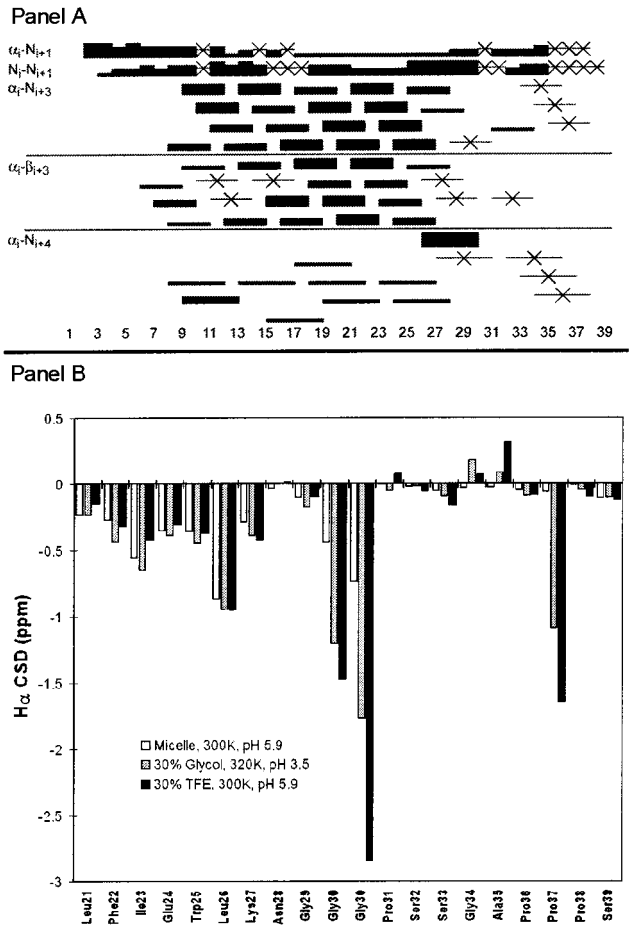

FIGURE 3: (A) Summary of NOE connectivities observed for exendin-4 in aqueous fluoro alcohol media: the NOE intensities are indicated by bar width; connectivities which are ambiguous or not applicable due to prolines in the sequence are indicated by —X—, where the bar width indicates the relative intensities of the NOEs within each connectivity class. (B) Hα CSD comparison for residues 21−39 of the exendin-4 in aqueous DPC micelles, 30% glycol, and 30% TFE; two entries appear for Gly[30]. A small decrease in helicity upon addition of 30% TFE is suggested by the decreased upfield shifts at the α methines of Leu[21]−Trp[25]. This is also observed for the remainder of the helical span; for residues 7−20, the average CSD change is $+0.07 \pm 0.08$ ppm upfield for transport from the micelle-bound state to 30% TFE medium. This effect is not observed for GLP-1 (see Figure 1A).

indicates that segmental motion reduces the effective correlation time to the point where NOEs cross over from large molecule cross-saturation to the 'positive' NOEs seen in small molecules. A similar analysis suggests that the side chains are fully disordered for residues 1, 3, and 31−39.

A more quantitative NOE analysis (based on the intensities in a 160 ms mixing time NOESY experiment at pH 3.5 and 300 K which provided 167 structurally significant NOE distance constraints—14 intraresidue, 149 short- to medium-range interresidue, and 4 longer range connectivities) provides a somewhat contrasting view. Given the low NOE density (ca. 5 constraints/residue over the 'structured' 7−31 residue span), the accuracy and usefulness of the resulting structure ensemble (Supporting Information) must be questioned. In this ensemble, the helix is limited to residues 11−27. This likely reflects the short, conformationally averaged $\alpha_i N_{i+1}$ distances that result from unfolded state contributions in the N-terminal portion of the helix.

Deriving a Solution-State Structure for Exendin-4. Solutions of exendin-4 in media lacking lipid micelles were thoroughly examined by NMR. The solubility of exendin-4 was sufficient even in aqueous buffers lacking cosolvents; however, most signals were too broad for clear definition in 1D spectra at 285 K (pH 5.9 buffer) presumably due to slow exchange between conformational states or with an aggregated state. Increasing the temperature to 300 K provided additional signals and also improved the intensity and line widths for all the observed resonances. But some of the Trp^{25} aromatic hydrogen signals, for example, were not observable. Further warming and the addition of glycol to 30 vol-% afforded TOCSY and NOESY spectra that provided a complete assignment. Aqueous fluoro alcohol media (36% HFIP, 30% and 60% TFE, and 8% HFIP/18% glycol) were also examined. The NOE contacts observed (Figure 3A, for 30% TFE) imply a helix spanning from residue 7 (or 8) to residue 28; the Hα chemical shift deviations afford the same conclusion for this medium and for aqueous 30% glycol.

Some very large Hα CSDs are observed in the solution state (30% aqueous glycol or 30% TFE) that are not observed in the micelle-associated state (Figure 3B). Resonances other than those of α methines also displayed unusual shifts; some diagnostic structuring shifts (as CSDs) for proline resonances illustrate this. The CSDs are listed in the order: micelle-state, 30% glycol (at 320 K), 30% TFE (at 300 K): Pro^{31}-δ3 (−0.29, −0.87, −0.76), Pro^{37}-β3 (+0.03, −0.56, −1.35), Pro^{37}-β2 (0.00, −0.26, −0.56), Pro^{37}-γ3 (−0.06, −0.16, −0.46), and Pro^{38}-δ3 (0.00, −0.26, −0.58 ppm). In aqueous glycol at 320 K, the 'structuring shifts' for all shifted proton resonances of $Pro^{37,38}$ are 47 ± 12% of those observed in aqueous TFE at 300 K. The chemical shift assignments for exendin-4 in 30% TFE at 300 K appear in Table 3S. The chemical shifts for the solution-state conformation(s) represent very large differences from the random coil values, particularly in the Trp^{25}−Pro^{37} span. This suggested tertiary structuring at the C-terminus in the solution state. The two most diagnostic NMR characteristics of protein tertiary structure, chemical shift dispersion and the protection of backbone NHs from H/D exchange by persistent H-bonds and/or reduced exposure to the bulk medium, are both observed for exendin-4 in the solution state. Figure 4 shows two panels of a NOESY recorded after extensive attempts (2+ days, pD > 5.6) to exchange deuterium into the peptide.

Figure 4 also serves to illustrate the existence of an extensive web of medium- and long-range NOE connectivities in the Phe^{22}−Pro^{38} span; for example, the tryptophan ring hydrogens have large- or medium-sized NOE connectivities to the α-protons of $Pro^{31,36,37}$, and the protons of Phe^{22} have connectivities with the $Pro^{37,38}$ unit. The average constraint density over the Leu^{21}−Pro^{38} segment was 11 distances/residue (the simulated annealing runs included 46 long-range constraints, $i/i\pm n$ with $n > 4$, which were all within this span); as a result, it was possible to obtain a well-converged structure ensemble based on the NOE constraints.

NMR structure ensembles were derived from NOESY distance constraints using a variety of protocols (see Supporting Information). The NOE intensities were taken largely from 750 MHz spectra recorded at 280−300 K for pH 5.9 buffer containing 30% TFE with mixing times between 60 and 200 ms. Ambiguities resulting from chemical shift overlaps were resolved by references to additional NOESY spectra: in 30% TFE (pH 3.5, 280 K, t_m = 60 ms; 300 K, t_m = 200 ms), in 36% HFIP (pH 3.5, 280 K, t_m = 60 and 150 ms). Here we present the results of the most conservative structure derivation; a nearly default CNS script (*20*) was employed with a universal low-bound of 2.0 Å. No torsion constraints were employed. In addition, even though we observed numerous exchange-protected NHs (vide infra), H-bond constraints were not employed. This was done to ensure that the structures generated were required by the NOEs observed and not the results of assumed H-bonding norms for secondary structure motifs. Statistics from the structure elucidation protocol appear in Table 1.

The accepted ensemble (36 structures from 50 random starts) appears in Figures 5 and 6; the primary acceptance criterion was no restraint violations greater than 0.2 Å. The pairwise backbone rmsd over residues 2−38 was 2.35 ± 0.76 Å; the poor fit was due largely to the significant disorder at the N-terminus (backbone rmsd 2.31 ± 0.72 Å over residues 2−10, versus 0.63 ± 0.22 Å for residues 11−38). The dihedral angle distribution within the ensemble appears in the Supporting Information (Table 4S). The convergence over the helical span is illustrated in Figure 5. Figure 5 also shows the essentially complete disorder in the first six residues of the peptide. The structure is best described as a helix spanning Thr^{7}−Asn^{28} with some fraying in the two N-terminal turns (pairwise backbone rmsd = 0.54 Å for residues 7−19 versus 0.21 Å for residues 16−28). The unusual phi value (135 ± 10°) for Gly^{29} signals the C-capping residue. A specific capping interaction is not evident at the N-terminus of this helix; however, NOEs between the Phe^{6} and Leu^{10} side chains suggest some local hydrophobic clustering. A shorter 3_{10} helical segment appears at Pro^{31}−Ser^{33}.

The convergence over the portion of the structure that displays tertiary structure (residues 21−38) was even tighter (0.76 Å rmsd for all heavy atoms). Figure 6 shows the structure of this fold; we designate the C-terminal structure of exendin-4 as the 'Trp-cage fold'. In the Trp-cage fold, the $Pro^{36,37,38}$ unit adopts a poly-Pro_{II} conformation and docks onto the aromatic rings of Phe^{22} and Trp^{25}. The helical portion of the motif (residues 21−28) is the C-terminal extension of the α helix previously observed in the micelle-associated state but ends with an $i-3$ H-bond (Gly^{29}-NH → O═C-Leu^{26}). In an ensemble generated with high precision constraints (see Supporting Information), the Trp^{25}-NH is situated equally well for H-bonding to the carbonyls of Leu^{21} and Phe^{22}. This ensemble also has (in all structures) Ser^{32}-NH → O═C-Gly^{29} and Ser^{33}-NH → O═C-Gly^{30} H-bonds reflecting a short 3_{10} helix; these are present in 35% of the structures in the most conservatively derived ensemble which is shown in Figure 6, panel A. A Trp^{25}-$N_{\epsilon 1}H$ → O═C−Ala^{35} H-bond is present in all structures of all ensembles *and* imparts significant exchange protection. Two additional H-bonds that appear in the NMR structure ensembles deserve note: Gly^{30}-NH → O═C-Trp^{25}, Ser^{33}-OH → O═C−Asn^{28} (or Ser^{33}-OH → O═C-Gly^{30} in some members of the ensemble). The former produces a modest level of NH protection, and there is experimental evidence to support the latter in some analogues. In two truncated mutants of the exendin-4 C-terminus, the Ser^{33}-OH appears in the spectrum and displays both intra- and interresidue NOEs (*29*). This suggests a degree of protection that is very unusual for an hydroxyl group in a peptide.

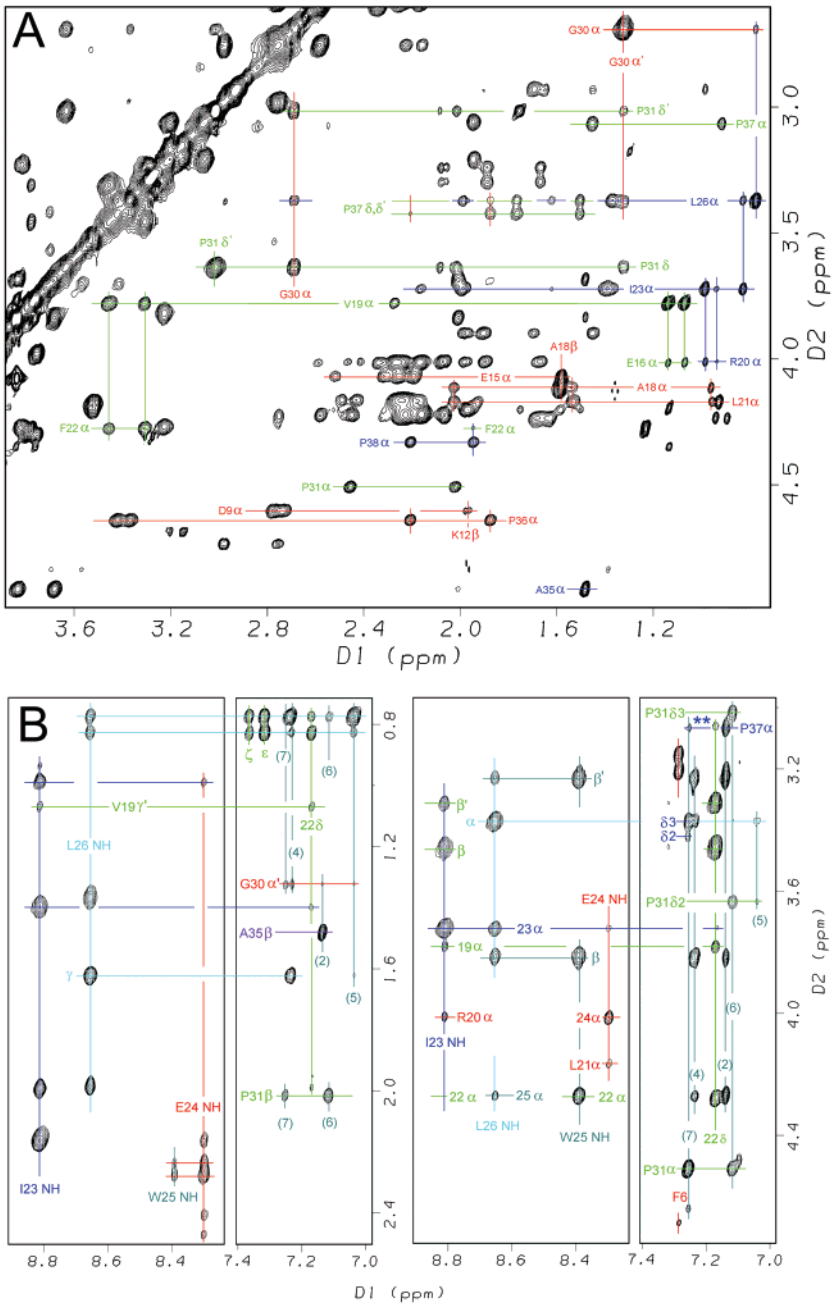

FIGURE 4: NOESY spectra (with a mixing time of 150 ms) of exendin-4 in 30% TFE (deuterated media), pH* 5.9, at 300 K. The sample was allowed to exchange in several serial batches of D_2O buffer (pH* adjusted to 5.35 for each batch) over a period of 3 days with lyophilization to dryness after each exchange. The final buffer was 99.95% deuterated. Panel A illustrates helical α_i/β_{i+3} connectivities and the assignments of Gly^{30} and the proline resonances. In panel B, NOESY connectivities that establish that the residual NHs are in a helical span and aryl-H/CH NOEs reflecting tertiary structure are labeled and color-coded by residue. R20-Hα and E24-Hα are shift-coincident (4.014 ppm), while F22-Hα (4.292 ppm) and W25-Hα (4.281 ppm) are nearly coincident. A weak 22Hα/23HN NOESY peak appears when the spectrum is plotted with a lower contour level cutoff. The Trp aryl-H lines are labeled using indole ring numbering (2−7). The upfield peak (**) on the 22δ line is due to Pro^{38}-δ2.

While the numerous H-bonds undoubtedly provide some fold stabilization, the primary feature of the Trp-cage is an extensive hydrophobic cluster. The close packing of the cluster is evident upon inspection of a space-filling model of the fold (panel B, Figure 6). The Leu^{26} side chain is fully ordered with the downfield methyl, $C_{\delta 1}H_3$, interacting exclusively with an edge of the phenyl ring. The δ2 methyl contacts with both aryl rings and the Pro^{31}-Cδ methylene (confirmed by NOEs). These interactions, Leu^{21}, Val^{19}, and backbone atoms, encase all but one surface of the phenyl ring. The $(CH_2)_3$ unit of Pro^{31}, Leu^{26}-$C_\gamma H$, the methylene of Gly^{30}, and Ala^{35} provide coverage of the periphery and the bottom surface of the indole ring. The association of two additional prolines completes the aryl ring cage: Pro^{37} is positioned immediately above the indole ring with the two ring planes nearly parallel, and the Pro^{38} δ-CH_2 fills the gap between the phenyl and indole ring. The remainder of the Pro^{38} ring provides additional coverage of the phenyl ring and possibly the Leu^{21} side chain. The close proximity and specific spatial relationships of numerous protons to the two

Table 1: Statistics from the Conservative CNS-Based NMR Structure Determination for Exendin-4: Distance Restraint Statistics and Distribution, Structural Statistics, Constraint Deviations, and Convergence Measures over 36 (from 50 Random Structure Starts) Simulated Annealing Structures

Distance Constraints by Peptide Segment and Type			
type of constraint	total	for residues 1−20	for residues 21−39
low-bounds only[a]	116	11	105
torsion constraints[a,b] (expressed as distances)	20	10	10
distance constraints	294	80	214
intraresidue	47	12	35
sequential	104	35	69
$i/i+n$, $n = 2-4$	97	33	64
$i/i+n$, $n \geq 5$	46	0	46

Distance Violations: Mean of Values for Accepted Structures (±SE)			
	total	for residues 1−20	for residues 21−39
rmsd violation (Å)	0.008 (0.002)	0.012 (0.005)	0.0061 (0.0008)
fraction of constraints violated by >0.1 Å		0.0014 (0.006)	none

Structure Statistics: Packing Measure and Deviations from Idealized Geometry	
E_{LJ} (kcal/mol)	−151.9 ± 5.7
bond violations (Å)	0.003 ± 0.000
angle violations (deg)	0.38 ± 0.008
improper torsions (deg)	0.18 ± 0.011

Ensemble Convergence: Atomic rms Deviations in Å, Pairwise over Ensemble (±SE)			
	residues 7−19[c]	residues 16−28	residues 21−38
backbone rmsd	0.54 ± 0.20	0.21 ± 0.08	0.35 ± 0.15
heavy atom rmsd	not determined	1.28 ± 0.21	0.76 ± 0.17

[a] These constraints were not employed in the conservative CNS protocol; they are used in alternative ensemble derivations that appear in the Supporting Information. [b] These are largely due to $\alpha\beta$ anti/gauche assignments coupled with βCH_2 prochiral assignments. [c] The C-terminal Ser and residues 1−6 are extensively disordered in the structure ensemble. Residues 7−15 and 16−28, respectively, represent the frayed-end and well-converged segments of the helix observed. We have compared them by backbone rmsds spanning 13 residues in both cases.

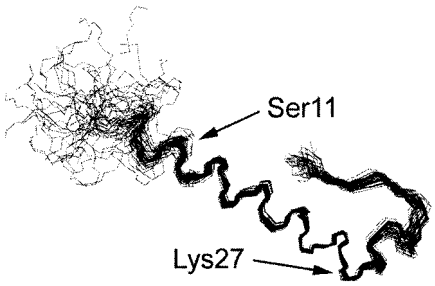

FIGURE 5: Conservative NMR ensemble for the solution-state structure of exendin-4. All 36 accepted structures are shown overlaid on the mean structure by least-squares-superimposing the backbone atoms of residues 11−29 (pairwise rmsd = 0.49 ± 0.21 Å). The locations of Cα of residues 11 and 27 are labeled for clarity. The coordinates and NMR restraints have been deposited in the Protein Data Bank under accession number 1jrj.

aromatic rings produce very large chemical shift deviations (Figure 3B and Table 2, vide infra) which are discussed in a later section.

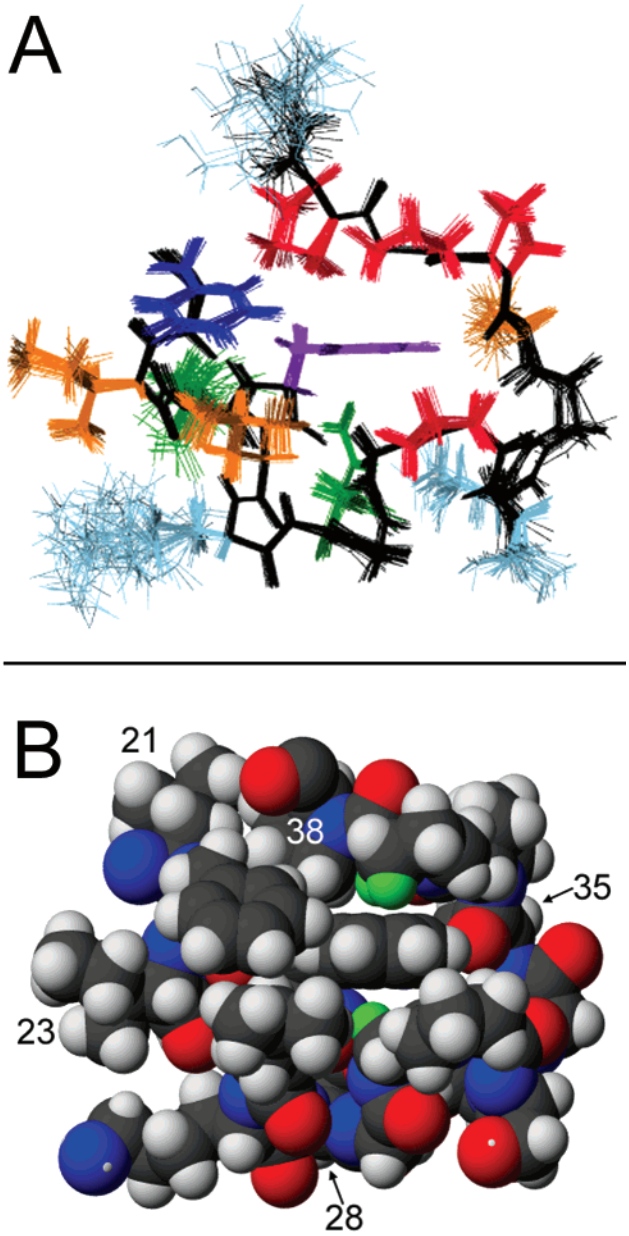

FIGURE 6: C-terminal fold of exendin-4 in aqueous 30% TFE. Panel A: all 36 structures in the 'conservative' CNS structure ensemble are shown, least-squares-fitted over the residue 22−38 backbone. All atoms of residues 22−39 are displayed. Glycines and the backbone are shown in black with the residue side chains coded by the following color scheme: Trp, purple; Phe, dark blue; Ser and Lys, light blue; Pro, red; Leu, Ile, and Ala, orange; Glu and Asn, green. Panel B is a CPK view of the 'Trp-cage' motif (residues 21−38) of a representative structure from the major cluster of the solution-state NMR structure ensemble of exendin-4. In panel B, residues 21, 23, 28, 35, and 38 are labeled, and the Gly^{30}-α2 and Pro^{37}-α,β3 hydrogens that display 1.5 ppm or greater upfield shifts are shown in green. In both panels, the edge-on indole ring of Trp^{25} is the central feature.

Amide NH Exchange Studies on Exendin-4 in Aqueous TFE. To probe the stability of the tertiary and secondary structural features of the NMR ensemble, NH exchange rates were measured. The key observations are given here: (a) at laboratory temperatures, residues 19−28 display half-lives of $0.6-30^{+}$ days at pD 3.6; (b) at pH 5.9, residues 23−27 had half-lives of 2−20 days at laboratory temperatures; (c)

Table 2: Ring Current Shifts in the Trp-Cage Motif[a]

		G30		P31		P37				P38	
	temp (K)	Hα2	Hα3	Hα	Hδ3	Hα	Hβ3	Hγ3	Hδ3	Hα	Hδ2
					Coil Reference Values						
aqueous	298	4.14		4.43	3.78	4.69	2.27	1.96	3.78	4.43	3.60
					CSD Values						
DPC micelle state	315	−0.70	−0.40	0.00	−0.29	−0.03	+0.03	−0.06	+0.01	−0.01	+0.03
					Solution-State						
30% glycol	320	−1.73	−1.16	−0.05	−0.87	−1.05	−0.56	−0.16	−0.28	−0.04	−0.24
18% glycol + 8% HFIP	314	−2.57	−1.45	0.00	−0.92	−1.23	−1.05	−0.45	−0.41	−0.09	−0.48
36% HFIP, pH 3.5	300	−2.83	−1.38	+0.10	−0.68	−1.55	−1.37	−0.41	−0.37	−0.10	−0.54
60% TFE	300	−2.77	−1.37	+0.11	−0.85	−1.54	−1.30	−0.39	−0.34	−0.08,	−0.53
30% TFE, pH 3.5	320	−2.40	−1.34	+0.05	−0.84	−1.24	−1.01	−0.35	−0.33	−0.08	−0.46
	310	−2.54	−1.39	+0.05	−0.82	−1.34	−1.11	−0.37	−0.35	−0.08	−0.48
	290	−2.75	−1.45	+0.06	−0.80	−1.53	−1.26	−0.43	−0.36	−0.09	−0.50
	280	−2.79	−1.47	+0.07	−0.81	−1.62	−1.31	−0.46	−0.35	−0.09	−0.50
30% TFE, pH 5.9	300	−2.80	−1.43	+0.08	−0.76	−1.62	−1.35	−0.46	−0.40	−0.10	−0.54
30% TFE, pD 6.6	280	−3.02	−1.44	+0.11	−0.66	−1.82	−1.52	−0.53	−0.46	−0.11	−0.56
ring current shifts expected based on NMR ensembles[b]	major	−2.56	−0.96	+0.33	−0.51	−2.94	−1.80	−0.23	−0.08	−0.17	−0.55
	minor	−2.32	−0.87	+0.38	−0.40	−2.82	−1.78	−0.20	−0.02	−0.17	−0.63
	precise	−2.82	−1.10	+0.24	−0.44	−1.41	−2.29	−0.54	−0.28	−0.09	−0.26

[a] The first row of the table gives the coil reference values employed; these include the +0.29 ppm shift effect for αH of Xaa when followed by a proline (Xaa-Pro, +0.17 for Gly-Pro) (*39*). All other entries are given as chemical shift deviations (CSD = $\delta_{obs} - \delta_{rc}$). Expected ring current contributions were calculated using the Shift 3.0 program (www.scripps.edu/case/). [b] Values for the mean structures are given in each case. There is considerable variation in predicted ring currents within each ensemble; for example, in the major cluster the ring current calculations for G30-Hα2 varied from −2.30 to −2.80 ppm. The alternate (precise) ensemble was generated with tighter NOE distance constraints (see Supporting Information); it differs (heavy atomic rmsd over residues 22, 25, 26, 30, 31, 37, 38) from the two clusters within the conservative ensemble by 0.50 Å; the two clusters from the conservative CNS ensemble differ by 0.24 Å based on the same measure. This shows how sensitive chemical shifts are to quite small displacements in a tight hydrophobic cluster including aromatic rings.

the ring-NH of Trp[25] was exchange-protected throughout the pH range 3.5−6.6. The indole NHε1 exchange rates were 0.23 and 0.79 h^{-1}, respectively, at pD 6.0 and 6.6 (7 °C). Assuming no large effects of the TFE cosolvent upon intrinsic exchange rates (see Materials and Methods), these indicate a protection factor of ca. 260 for Trp[25]-NHε1. The slowest exchange was observed for the backbone NHs of Ile[23] and Leu[26] (ca. 0.009 $days^{-1}$), which indicate protection factors 500-fold greater than that observed for the indole ring NH. The NH exchange data also provide a compelling argument for a helix that unravels exclusively from one end (the N-terminus): the C-terminus of the helix does not fray. Apparently, the Trp-cage serves as a very effective C-cap.

DISCUSSION

GLP-1 versus Exendin-4: Structural Differences. The present NMR data provide a basis for quantitating differences in the secondary structuring preference associated with the exendin-4 versus GLP-1 sequence. The uninterrupted series of upfield Hα shifts throughout the residue 12−16 span for exendin-4 (Figure 2) argues against the significant helix interruption or distortion that has been suggested to be a recognition determinant at the GLP-1 receptor (*11*, *13*). Residues 18−19 also display larger upfield shifts in exendin-4 (Ala-Val-Arg[20]) than in GLP-1 (Ala-Ala-Lys[20]); this may reflect either the increased helicity at residue 16 (Glu vs Gly) or the effect of the Arg/Lys change at position 20. The only other differences are the greater upfield shifts for the α methines of residues 9−11 of GLP-1 (these sites appear to have a diminished helix propensity in exendin-4), but potential ring current effects from Tyr[14] (present only in GLP-1) place this analysis in doubt.

In the GLP-1 ensemble previously reported for the DPC micelle-associated state (*11*), the distortions to the helix around residues 15−17 are not as substantial as those suggested by the Hα CSD histogram and differences in the NH exchange protection between the N- and C-terminal helices. The studies of Parker et al. (*13*) in 50% TFE defined the more converged helical span as Ala[19]−Val[27]. Our comparison of the TFE and micelle states clarifies the extent of helix disruption caused by Gly[16]. The Hα and NH CSD comparisons (Figure 1) indicate that the helix propagates through Gly[16] more often in the aqueous 30% TFE state. When bound to micelles, the major fraction of the GLP-1 conformations represents structures with two helices, in which Gly[16] serves as the N-cap for the C-terminal helix. The earlier reported structure for GLP-1 (*11*) appears to have underestimated the flexibility that results due to this glycine location in the 'helical' sequence.

Thornton and Gorenstein suggested that a helix phase shift (placing Phe[6], Tyr[13], Phe[22], and Trp[25] on one face of the 'helix') is important for membrane binding. They also suggested that the absence of Gly[16], purported to effect the helix phase shift in GLP-1, in glucagon explains the 1000-fold decrease in binding affinity to the GLP-1 receptor. Exendin-4, although it shares some biological actions of GLP-1, does not display similar helix distortions for the micelle-associated state. These results call into question the suggestion that the flexibility at Gly[16] is an important determinant for GLP-1 receptor binding activity and are consistent with the results of mutational studies in which the G16A mutation does not decrease either activity or binding affinity (*30*, *31*).

Relative Value of Chemical Shifts and NOEs in Peptide Structure Elucidation. The present study includes peptide 'states' that differ widely in the degree and nature of structuring, in order of increasing structure: GLP-1, micelle-associated exendin-4, and exendin-4 in aqueous fluoro

alcohol, with only the latter system displaying significant tertiary structure as reflected by an extensive web of long-range NOEs. In addition, each system included a helical span and a segment that was extensively disordered. The relative performance of chemical shift versus NOE-distance-based structural analysis for these systems deserves some comment. In the Supporting Information, we present the results obtained using XPLOR to derive structure ensembles for each system with the same algorithm used for deriving distances from NOE intensities and a comparably strong NOE forcing potentials throughout. Some key observations are the following: (1) when aggressive high-precision distance constraints are used to generate exendin-4 structures, $i/i+1$ NOEs are the major source of these penalties (due to the conflict between $i/i+1$ and $i/i+3$ NOEs); (2) violations are more pronounced in the nonconverged N-terminal portions of the structures presumably due to either extensive segmental motion or dynamic helix unfolding; and (3) over the structured portion of the sequence, the aggressive protocol affords the same structure as the conservative CNS protocol. We suggest that *for peptides* in the absence of multiple $i/i+n$ ($n > 4$) NOE connectivities (which will be present in β hairpins and other rare instances in which peptides display stable structuring motifs due to hydrophobic cluster formation), NOE-based structure ensemble derivations do not appear to be warranted as they will not define structural features with any greater precision or confidence than could be achieved by using chemical shift deviations to determine secondary structure preferences in conjunction with modeling. In the absence of tertiary structure and nearby aryl residues, chemical shift deviations accurately reflect secondary structure preferences and the extent of folding. A combined analysis of αH and NH CSDs along the sequence may represent the optimal level of structural analysis for most peptides. Cooperative structure formation can be detected based on the temperature dependence of CSDs (*24*). In our opinion, any meaningful structuring will be apparent in such an analysis of chemical shift data.

Tertiary Structure in a Modest-Sized Peptide. The most surprising finding of this study is the appearance of tertiary structure for exendin-4 in the solution state, particularly when fluoro alcohol cosolvents are present. Fluoro alcohols have been reported (*32*−*34*) to disrupt the tertiary (and quaternary) structure of many proteins, but there have been reports (*35, 36*) of increased structuring for designed β sheet structures in water/alcohol mixtures (including aqueous TFE), and the helix-stabilizing effect of fluoro alcohols is well-established. With exendin-4, fluoro alcohol addition stabilized (see Figure 3B) tertiary structure, the Trp-cage feature (Figures 6), within the C-terminal half of the sequence. For panel B of Figure 6, we selected a member of the major conformer cluster which displays a single side chain conformer for both Asn^{28} and Ser^{33} with a hydrogen bond from the side chain OH of Ser^{33} and the backbone carbonyl of Asn^{28} (vide supra). Chemical shift comparisons (Table 2, vide infra) indicate a fundamentally similar tertiary structure in 18−30% glycol. In the absence of added fluoroalcohol, however, the population of conformers with tertiary structure appears to be significantly diminished—$\chi_F = 0.47$ in 30% glycol at 320 K based on the magnitude of the diagnostic structural shifts. In the present case, essentially complete formation of a protein-like fold can be induced with the addition of as little as 8 vol-% of a fluoro alcohol. These results bolster other data that we have presented (*37, 38*) which indicate that fluoro alcohols can enhance the solvophobic effect and that this is probably a source of the structure-inducing effects that these cosolvents display with many peptides. The same structure is retained at high fluoro alcohol concentrations (36% HFIP, 60% TFE), that stands in contrast to the tertiary structure disruption observed for proteins under such conditions. However, the Trp-cage is completely absent in the micelle-associated state of exendin-4. The evidence for the structural details of the Trp-cage fold, observations that reveal the extent of fluxionality and the cooperativity of the folding process, and some hypotheses concerning the folding pathway follow. The intent is to establish whether the Trp-cage can serve as a paradigm for quantitative studies of the importance of the hydrophobic effect in polypeptide fold stabilization.

Since chemical shifts were not employed in deriving the NMR structure, they can provide a confirmation of the NOE-derived structural features of the Trp-cage. The close proximity and specific spatial relationships of numerous protons to the two aromatic rings produce very large chemical shift deviations. Table 2 shows the protons with the largest structuring shifts and how these chemical shifts vary with medium and temperature. These shifts are nearly as large as any reported for proline and glycine residues in well-folded non-heme proteins; shifts of this magnitude would not be observed if the folded state population was small. If the NMR structure ensemble correctly reflects the structure and the motions within the folded state energy well, it should predict the chemical shifts. Ring current shift calculations were performed for both clusters within the conservatively derived ensemble and an ensemble derived using tighter constraints. The structure ensemble fails to reproduce the exact relative shift deviations, but the direction and relative sizes of the shifts observed are in general accord with the model predictions. Notably, numerous large shielding effects and the unique deshielding effect at Pro^{31}-α/β3 are predicted by the ensemble. The NMR structure ensemble predicts larger ring current shifts than those observed for Pro^{37}-α/β3, but somewhat smaller shifts for the Gly^{30}-CH_2. The real structure probably has a more fluxional hydrophobic cluster than that reflected in the NMR ensembles. In addition, comparison of the calculated and observed ring current effects suggests that the time-averaged structure places Pro^{37} somewhat further away from and in a slightly different disposition relative to the indole ring. The structural implications of the experimental shifts are being tested using simulations of the dynamics of the folded state in a water box.

Notable stability is indicated for the Trp-cage folding motif by a number of NMR-derived parameters: chemical shifts, their temperature dependence, and NH exchange protection. Very large upfield shifts for Gly^{30}-α_2, Pro^{37}-β_3, and Pro^{37}-α are diagnostic of the Trp-cage fold (Table 2). Upon cooling to 280 K, these chemical shift deviations increase by 7−20%:Gly^{30}-α_2 appears at 1.12 ppm, 3.02 ppm upfield of its expected location. A more complete determination of the temperature dependence of these chemical shifts is available at pH 3.5; the structure loss on warming appears to be equal to or less than 5.5% for a 10 °C temperature increase in the 5−45 °C range. The folded state is significantly populated even at 50 °C. Some of the NHs of exendin-4 are remarkably

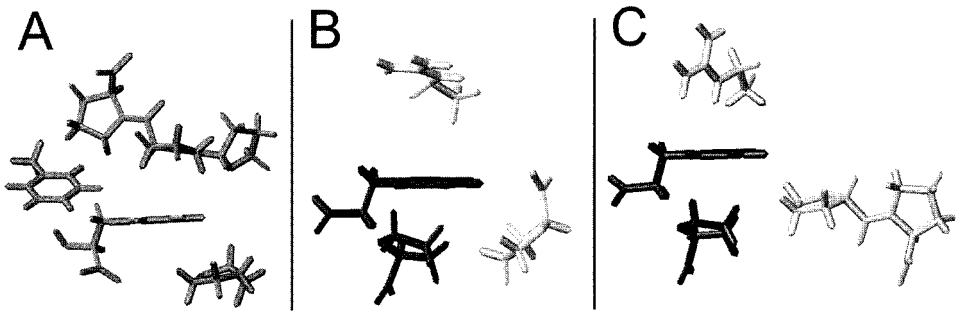

FIGURE 7: Comparison of the Trp/Pro interactions in the exendin-4 'Trp-cage' (panel A) and in SH3/peptide complexes. The proline-rich peptides illustrated are RALPPL**PR**Y (panel B) and VSLAR**R**PL**PP**LP (panel C). In the SH3/peptide complexes, the prolines in the peptide are shown in a lighter gray while the intramolecular proline is shown in dark gray. The $CH_2CH_2NHC(NH_2)_2$ portion of emboldened peptide arginine is also included in panels B and C.

resistant to exchange. The $10^{2.4}$ protection factor observed for the indole NH ($t_{1/2}$ = 3.0 h, at pD 6, 7 °C) argues for substantial tertiary structure stability and the lack of local unfolding events that would break the NϵH → O=C−Ala^{35} H-bond. The exchange half-lives of the backbone NHs in the portion of the helix that is capped within the Trp-cage are much longer than those we have observed for helical peptides in this medium. The average backbone NH protection factor over residues 23−27 of exendin-4 in 30% TFE appears to be $10^{5.1}$ at pD 6.6 (9 °C). Amide exchange protection data indicate facile unfolding in the N-terminal portion of the helix; protection factors fall off rapidly from 10^5 at the NH of Ile^{23} to $10^{2.0}$ at Ala^{18} and to less than $10^{0.8}$ from Glu^{16} to the N-terminus. Based on the data obtained thus far, particularly the disorder at the extreme N-terminus and the fraying of the N-terminal turns of the helix, it seems highly unlikely that the formation of the Trp-cage structure requires the first 14−17 residues of the exendin-4 sequence. The Leu^{21}−Pro^{38} segment of exendin-4 may be the smallest protein-like folding unit observed to date and the basis for designing even smaller "mini-proteins" based on this helix/Trp-cage motif.

Folding Requisites and Cooperativity. In all of our studies of the temperature dependence of chemical shifts, warming leads to similar fractional shifts toward random coil norms for essentially all[2] of the protons within the Phe^{22}−Pro^{38} span that display large structuring shifts. We view this as compelling evidence that Trp-cage formation is a two-state folding process[3] in 30% TFE. There is, however, evidence that supports the presence of helical structuring without complete Trp-cage formation. The smaller CSDs observed for α methines of the helix do not melt out as readily. In the micelle-associated state, the helix is fully formed with no trace of Trp-cage formation. Furthermore, over residues 18−27, excluding Ile^{23} and Leu^{26} (which are further upfield due to ring current effects), the average Hα CSDs are −0.308 ± 0.088 and −0.376 ± 0.076 ppm, respectively, in 30% TFE and 30% glycol. A shift of −0.39 ppm is generally equated with 100% helicity (*26*). This suggests that the helix is 85+% populated in aqueous glycol conditions under which the Trp-cage tertiary structure is present in only 47% of the conformational equilibrium mixture. The helical region appears to have a more regular α-conformation prior to tertiary structure formation; when prolines dock at the accessible solvent-exposed surface of aryl side chains (Phe^{22} and Trp^{25}) in the helix, the packing optimization associated with docking apparently causes some distortions of the α-helix. A specific indole−DPC interaction may be responsible for the loss of tertiary structure in the micelle-bound state. It is well established that indoles partition into membranes (*40*) and that tryptophan residues prefer to be located in membrane-spanning units (*41*), particularly near the positive PC headgroups (*42*).

Studies are underway to test these hypotheses by synthesizing a minimized version of the Trp-cage motif that is fully folded in aqueous medium. These constructs will be used to establish the folding pathway and to confirm the cooperativity of folding in water, the biologically significant medium.

Structural Analogies to the Trp-Cage Fold. The burial of tryptophan rings also appears to be an important feature in a wide variety of peptide structuring and binding phenomena; e.g., the hydrophobic clustering of a tryptophan ring and leucine side chain has been observed in a cyclic pentapeptide (*43*), the clustering of the side chains of Trp^{21} and Ile^{19} has been reported to be an important structural determinant in endothelins (*25, 44*−*46*), and Trp residues appear at numerous binding sites (vide infra). However, we have found no close analogy to this specific Trp-cage in protein structures, although some similar component features can be noted: one of the tryptophans in a WW domain is encaged with a leucine side chain located immediately on the face of the indole ring (*47*); in the archetypal SH3 domain (*48, 49*), both Trp residues show some hydrophobic interactions—Trp^{43} has isoleucines both above and below the $C_{\epsilon 3}$−$C_{\zeta 3}$ edge of the

[2] The exceptions are 31-δ2,δ3 and 30α3 resonance shifts. The 31δ3 resonance moves further upfield from the random coil value on warming and under other conditions where the complete Trp-cage is less populated or more fluxional. The fractional loss of upfield shift of 30α3 (but not the more shifted diastereotopic 30α2) on warming is significantly smaller than that of the other ring current shifted resonances. These observations can be rationalized by assuming that the unfolded state has a residual hydrophobic cluster (Trp^{25}/Gly^{30}/Pro^{31}) similar to that observed in the DPC micelle-associated state.

[3] All of our data indicate that conformers of the *all-trans*-Xaa-Pro isomer are in sufficiently rapid equilibrium to result in population-averaged chemical shifts. As a result, the temperature dependence of chemical shifts reveals the changes in the folded population. If all CSDs are due to a single folded state, the fractional loss of the CSD at each position should reflect the loss of structure at each locus. Thus, the observation of identical fractional losses[2] of shift deviations upon warming indicates the same degree of structure loss throughout the residue 21−38 span. The simplest rationale for this is a fully cooperative two-state equilibrium with only fully folded and fully disordered states present. We suggest that, for systems in a rapid folding equilibrium, the response of chemical shifts throughout the sequence to perturbations of the folding equilibrium provides the ultimate test of cooperativity.

ring; Trp[42] has a parallel proline ring (Pro[57]) in close proximity (see Figure 7). The SH3 domain has one additional aryl/proline ring-stacking interaction; Pro[63] is parallel to Phe[10] and in close spatial proximity. Both of the Pro/Trp stacking interactions in SH3 produce upfield shifts (0.20−1.66 ppm) for the proline-α,β3,γ3 resonances, but the shifts for the proline perched on Trp[42] are not as large as those observed for Pro[37] in the Trp-cage of exendin-4. In some analogy to Pro[37] of exendin-4, Pro[63] of SH3 displays an α methine resonance that is well-upfield of the δ methylene signals.

Exposed tryptophan residues are the key feature of cellulose binding proteins (*50*−*52*), and the aryl rings of Trp (and Tyr) residues frequently appear as the side or bottom walls of receptors for saccharides (*53*−*58*) and proline-rich peptides (for example, SH3 and WW domains). In this regard, we note that the axial hydrogens on one face of a sugar ring present a similar surface to that presented by the face of proline that is cis to the α methine.

The closest analogies to the Trp-cage structure of exendin-4 that we have been able to find are in structures reported for complexes of SH3 domains and proline-rich peptides (*59*, *60*). The Trp and Pro residues (and Phe[22]) of the Trp-cage and two views of SH3/peptide complexes are shown in Figure 7. The intramolecular proline of SH3 (Pro[57]) clearly occupies a position comparable to Pro[37] in the Trp-cage (but on the *opposing* face of the indole ring). The actual position of Pro[37] is consistently occupied by a peptide arginine residue in the SH3/peptide complexes, reflecting a recognition determinant associated with a salt-bridging interaction with Asp[23] of SH3. The prolines in the peptides bound to SH3 appear in positions that bear some analogy to the locations of Pro[31] and Pro[36] in the Trp-cage. These analogies suggest that quantitative studies of the energetics of proline association in the Trp-cage may provide insights into recognition phenomena involved in signaling by common protein motifs.

Implications in Defining the Bioactive Conformation at the GLP-1 Receptor. For peptides with conformational versatility, there is always reason to doubt that any particular structural feature (or structuring preference) observed in any state other than the receptor-bound state can be used to derive the requisites for binding. In the present case, however, one significant structural preference, a well-converged helix over residues 18−27, has now been observed for both GLP-1 and exendin-4 in the solution-state *and* when associated with DPC micelles. The latter may mimic the context of the peptide just prior to binding to the receptor. Exendin-4 displays a greater intrinsic propensity toward the formation of a monomeric helix than GLP-1. This is attributed, at least in part, to the absence in exendin-4 of the helix-interrupting glycine residue at position 16 of GLP-1. CD data (*9*) indicate that GLP-1 is significantly helical only in helix-favoring media (aqueous TFE or DPC micelles) or at high concentrations when helix bundle formation occurs. In exendin-4, the helix goes much further toward the N-terminus; this feature and the efficient C-capping of the helix provided by the Trp-cage may reduce the entropic cost of binding at a receptor that requires the C-terminal helix state (*13*). The pronounced helical tendency of exendin-4 in conjunction with its binding at the GLP-1 receptor argues against the importance of Gly[16] of GLP-1 as a flexible linker and biorecognition feature (*11*).

ACKNOWLEDGMENT

We are indebted to John Tomaszewski for providing his shift assignments for SH3 and the views of the SH3/peptide complex appearing in Figure 7. We thank Ved P. Srivastava (Amylin Pharmaceuticals) for revision suggestions following a critical reading of a draft of the manuscript.

SUPPORTING INFORMATION AVAILABLE

All ^{1}H chemical shift assignments (3 tables). Also included are a complete list of NOE constraints, torsion angles in the resulting ensembles and other statistics, and three structural figures (29 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Fehmann, H. C., and Haebener, J. F. (1992) *Trends Endocrinol. Metab. 3*, 158−163.
2. Creutzfeldt, W. O., Kleine, N., Willms, B., Orskov, C., Holst, J. J., and Nauck, M. A. (1996) *Diabetes Care 19*, 580−586.
3. Wettergren, A., Schjoldager, B., Mortensen, P. E., Myhre, J., Christiansen, J., and Holst, J. J. (1993) *Dig. Dis. Sci. 38*, 665−673.
4. Dupre, J., Behme, M. T., Hramiak, I. M., McFarlane, P., Williamson, M. P., Zabel, P., and McDonald, T. J. (1995) *Diabetes 44*, 626−630.
5. Turton, M. D., O'Shea, D., Gunn, I., Beak, S. A., Edwards, C. M., Meeran, K., Choi, S. J., Taylor, G. M., Heath, M. M., Lambert, P. D., Wilding, J. P., Smith, D. M., Ghatei, M. A., Herbert, J., and Bloom, S. R. (1996) *Nature 379*, 69−72.
6. Eng, J., Kleinman, W. A., Singh, L., Singh, G., and Raufman, J. P. (1992) *J. Biol. Chem. 267*, 7402−7405.
7. Göke, R., Fehmann, H. C., Linn, T., Schmidt, H., Krause, M., Eng, J., and Göke, B. (1993) *J. Biol. Chem. 268*, 19650−19655.
8. Young, A. A., Gedulin, B. R., Bhavsar, S., Bodkin, N., Jodka, C., Hansen, B., and Denaro, M. (1999) *Diabetes 48*, 1026−1034.
9. Andersen, N. H., Brodsky, Y., Neidigh, J. W., and Prickett, K. S. (2001) *Bioorg. Med. Chem.* (in press).
10. Kim, Y., Rose, C. A., Liu, Y., Ozaki, Y., Datta, G., and Tu, A. T. (1994) *J. Pharm. Sci. 83*, 1175−1180.
11. Thornton, K., and Gorenstein, D. G. (1994) *Biochemistry 33*, 3532−3539.
12. Thorens, B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8641−8645.
13. Parker, J. C., Andrews, K. M., Rescek, D. M., Massefski, W., Jr., Andrews, G. C., Contillo, L. G., Stevenson, R. W., Singleton, D. H., and Suleske, R. T. (1998) *J. Pept. Res. 52*, 398−409.
14. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
15. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
16. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
17. Rohl, C. A., and Baldwin, R. L. (1997) *Biochemistry 36*, 8435−8442.
18. Buck, M., Radford, S. E., and Dobson, C. M. (1993) *Biochemistry 32*, 669−678.
19. Brünger, A. T. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
20. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D.: Biol. Crystallogr. 54*, 905−921.
21. Brüschweiler, R., Blackledge, M., and Ernst, R. R. (1991) *J. Biomol. NMR 1*, 3−11.

22. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
23. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley, New York.
24. Andersen, N. H., Neidigh, J. W., Harris, S. M., Lee, G. M., Liu, Z. H., and Tong, H. (1997) *J. Am. Chem. Soc. 119*, 8547−8561.
25. Lee, G. M., Chen, C., Marschner, T. M., and Andersen, N. H. (1994) *FEBS Lett. 355*, 140−146.
26. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
27. Merutka, G., Dyson, H. J., and Wright, P. E. (1995) *J. Biomol. NMR 5*, 14−24.
28. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
29. Neidigh, J. W. (1999) Ph.D. Thesis, University of Washington, Seattle, WA.
30. Adelhorst, K., Hedegaard, B. B., Knudsen, L. B., and Kirk, O. (1994) *J. Biol. Chem. 269*, 6275−6278.
31. Gallwitz, B., Witt, M., Paetzold, G., Morys-Wortmann, C., Zimmermann, B., Eckart, K., Folsch, U. R., and Schmidt, W. E. (1994) *Eur. J. Biochem. 225*, 1151−1156.
32. Sönnichsen, F. D., Van Eyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) *Biochemistry 31*, 8790−8798.
33. Slupsky, C. M., Kay, C. M., Reinach, F. C., Smillie, L. B., and Sykes, B. D. (1995) *Biochemistry 34*, 7365−7375.
34. Cort, J. R., and Andersen, N. H. (1997) *Biochem. Biophys. Res. Commun. 233*, 687−691.
35. Sharman, G. J., and Searle, M. S. (1997) *Chem. Commun.*, 1955−1956.
36. Kortemme, T., Ramírez-Alvarado, M., and Serrano, L. (1998) *Science 281*, 253−256.
37. Andersen, N. H., Cort, J. R., Liu, Z., Sjoberg, S. J., and Tong, H. (1996) *J. Am. Chem. Soc. 118*, 10309−10310.
38. Andersen, N. H., Dyer, R. B., Fesinmeyer, R. M., Gai, F., Liu, Z. H., Neidigh, J. W., and Tong, H. (1999) *J. Am. Chem. Soc. 121*, 9879−9880.
39. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
40. Wimley, W. C., and White, S. H. (1993) *Biochemistry 32*, 6307−6312.
41. Jacobs, R. E., and White, S. H. (1989) *Biochemistry 28*, 3421−3437.
42. Schiffer, M., Chang, C. H., and Stevens, F. J. (1992) *Protein Eng. 5*, 213−214.
43. Krystek, S. R., Jr., Bassolino, D. A., Bruccoleri, R. E., Hunt, J. T., Porubcan, M. A., Wandler, C. F., and Andersen, N. H. (1992) *FEBS Lett. 299*, 255−261.
44. Andersen, N. H., Chen, C., and Lee, G. M. (1995) *Protein Pept. Lett. 1*, 215−222.
45. Mills, R. G., O'Donoghue, S. I., Smith, R., and King, G. F. (1992) *Biochemistry 31*, 5640−5644.
46. Ragg, E., Mondelli, R., Penco, S., Bolis, G., Baumer, L., and Guaragna, A. (1994) *J. Chem. Soc., Perkin Trans. 2*, 1317−1326.
47. Macias, M. J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996) *Nature 382*, 646−649.
48. Yu, H., Rosen, M. K., and Schreiber, S. L. (1993) *FEBS Lett. 324*, 87−92.
49. Grantcharova, V. P., and Baker, D. (1997) *Biochemistry 36*, 15685−15692.
50. Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J. K., and Teeri, T. T. (1992) *Proteins: Struct., Funct., Genet. 14*, 475−482.
51. Poole, D. M., Hazlewood, G. P., Huskisson, N. S., Virden, R., and Gilbert, H. J. (1993) *FEMS Microbiol. Lett. 80*, 77−83.
52. Tomme, P., Warren, R. A. J., Miller, R. C. J., and Gilkes, N. R. (1995) in *Enzymatic degradation of insoluble carbohydrates* (Saddler, J. N., and Penner, M. H., Eds.) pp 142−163, American Chemical Society, Washington, DC.
53. Vyas, N. K., Vyas, M. N., and Quiocho, F. A. (1988) *Science 242*, 1290−1295.
54. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) *J. Biol. Chem. 266*, 5202−5219.
55. Sharff, A. J., Rodseth, L. E., and Quiocho, F. A. (1993) *Biochemistry 32*, 10553−10559.
56. Vyas, N. K. (1991) *Curr. Opin. Struct. Biol. 1*, 732−740.
57. Andersen, N. H., Cao, B., Rodriguez-Romero, A., and Arreguin, B. (1993) *Biochemistry 32*, 1407−1422.
58. Asensio, J. L., Cañada, F. J., Bruix, M., Rodriguez-Romero, A., and Jimenez-Barbero, J. (1995) *Eur. J. Biochem. 230*, 621−633.
59. Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994) *Science 266*, 1241−1247.
60. Feng, S., Kasahara, C., Rickles, R. J., and Schreiber, S. L. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 12408−12415.

BI010902S