# Conformational Properties of Four Peptides Corresponding to α-Helical Regions of *Rhodospirillum* Cytochrome $c_2$ and Bovine Calcium Binding Protein†

Alessandro Pintar,‡,§ André Chollet,∥ Charles Bradshaw,∥ Alain Chaffotte,⊥ Céline Cadieux,⊥ Marianne J. Rooman,‡,∞ Klaas Hallenga,‡ Jonathan Knowles,∥ Michel Goldberg,⊥ and Shoshana J. Wodak*,‡

*Unité de Conformation de Macromolécules Biologiques, Université Libre de Bruxelles, CP 160/16, Avenue P. Héger, 1050 Brussels, Belgium, Glaxo Institute for Molecular Biology, Chemin des Aulx 14, 1228 Plan-les-Ouates, Geneva, Switzerland, and Unité de Biochimie Cellulaire, Institut Pasteur, Rue du Docteur Roux 28, 75724 Paris Cedex 15, France*

*Received March 17, 1994; Revised Manuscript Received June 27, 1994*®

ABSTRACT: Four peptides corresponding to α-helical regions delimited by residues 63–73 and 97–112 of cytochrome $c_2$ (*Rhodospirillum*) and residues 24–36 and 45–55 of bovine calcium binding protein are predicted to be α-helical by a recently developed method [Rooman, M., Kocher, J. P., & Wodak, S. J. (1991) *J. Mol. Biol. 221*, 961–979], synthesized by solid phase methods, and purified by HPLC, and their solution conformations are determined by NMR and CD. The observed conformational properties of these peptides in solution confirmed prediction results: in water/TFE (60/40, v/v) at room temperature, these peptides adopt an α-helical conformation, as shown by an extended pattern of strong, sequential $d_{NN}(i,i+1)$ NOE cross-peaks, $d_{\alpha N}(i,i+1)$ NOEs of reduced intensity, several medium-range [$d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{\alpha\beta}(i,i+3)$] NOE connectivities, small $^3J_{H\alpha N}$ values, and more upfield α-proton chemical shifts. CD studies at different TFE concentrations and at room temperature provide further evidence of the propensity of these peptides to adopt an α-helical conformation in solution, as determined by the ellipticity values at 222 nm, and by deconvolution of the CD spectra. According to the method used, helicities in the range 34–50% and 55–75% are found for the 63–73 and 97–112 fragments of cytochrome $c_2$, respectively, and in the range 53–80% and 42–65% for the fragments 24–36 and 45–55 of calcium binding protein in water/TFE (60/40, v/v) at 298 K. In addition, the experiments and predictions agree for those residues that are more flexible. Finally, the relevance of our results for the protein folding pathways is discussed.

The last few years have brought significant new insights into the protein folding process. Both experimental (for a review see, e.g., Matthews 1993) and theoretical (for a review see, e.g., Rose & Wolfenden, 1993) analyses have contributed to this progress. Although different models for protein folding have been proposed (Baldwin, 1990), clear evidence is emerging from experimental studies of refolding kinetics (Radford et al., 1992; Roder et al., 1988; Udgaonkar & Baldwin, 1988; Udgaonkar & Baldwin, 1990; Matouschek et al., 1989; Bycroft et al., 1990), from studies on stable intermediates (Buck et al., 1993; Alexandrescu et al., 1993; Jeng & Englander, 1991; Hughson et al., 1990; Jeng et al., 1990), and from peptide models (Oas & Kim, 1988; Staley & Kim, 1990) that certain regions of the polypeptide chain can adopt their native conformation at the early stages of folding before the rest of the protein structure is formed. In particular, it has been shown that short peptides corresponding to secondary structure elements in their parent proteins can adopt a well-defined conformation in solution (Dyson et al., 1988b,c, 1992a,b). These results suggest that interactions between neighboring residues along the chain (local interactions) may play an important role at the early stages of protein folding, thereby "driving" the folding process along a preferred pathway (Wright et al., 1988).

Conformational studies of protein fragments in solution and the evaluation of their propensity to adopt a nativelike structure have proven to be useful in outlining the folding pathways of several proteins and in improving our understanding of how tertiary interactions, those involving residues far apart along the sequence but close in space, influence the folding process.

This study reports the conformational properties in solution of four chemically synthesized peptides corresponding to the α-helical regions delimited by residues 63–73 and 97–112 of cytochrome $c_2$ (*Rhodospirillum rubrum*) and residues 24–36 and 45–55 of calcium binding protein (bovine intestine), as determined by high-resolution NMR[1] and CD spectroscopy. These peptides correspond to protein segments predicted to adopt a well-defined α-helical conformation in absence of interactions with the rest of the protein and were selected by applying the recently developed computational procedure of

† A.P. was supported by a grant of the Commission of the European Communities (BRIDGE programme) and by the Brachet Foundation. The Belgian programme on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture, is gratefully acknowledged for support.

‡ Université Libre de Bruxelles.

§ Present address: Department of Biochemistry, University of Oxford, Oxford OX1 3QU, U.K.

∥ Glaxo Institute for Molecular Biology.

⊥ Institut Pasteur.

∞ Research Associate at the National Fund for Scientific Research (Belgium).

® Abstract published in *Advance ACS Abstracts,* August 15, 1994.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; Fmoc, 9-fluorenylmethoxycarbonyl; HOBt, 1-hydroxybenzotriazole; DCC, *N,N'*-dicyclohexylcarbodiimide; DMSO, dimethyl sulfoxide; THF, tetrahydrofuran; NMP, *N*-methyl-2-pyrrolidone; TFA, trifluoroacetic acid; HPLC, high-performance liquid chromatography; FAB-MS, fast atom bombardment mass spectrometry; TFE, 2,2,2-trifluoroethanol; DQF-COSY, two-dimensional double-quantum-filtered correlation spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; ROESY, two-dimensional rotating frame nuclear Overhauser effect spectroscopy; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; UV, ultraviolet.

0006-2960/94/0433-11158$04.50/0 © 1994 American Chemical Society

Rooman et al. (1991, 1992) to the amino acid sequences of 69 proteins of known three-dimensional (3D) structure. This procedure uses a knowledge-based force field that incorporates only local interactions between residues along the sequence. Segments with a well-defined preferred conformation are identified as those whose lowest energy structure displays a sizable energy gap relative to other computed structures. Such segments often correspond to $\alpha$-helices and $\beta$-strands well buried in the protein core, and it has been suggested that they could serve as initiation sites for folding (Rooman et al., 1990, 1991, 1992; Rooman & Wodak, 1991, 1992).

As means of checking the validity of the predictions, the aforementioned peptides were chosen from among the helical fragments identified with the largest energy gap by the prediction procedure. The choice of helical peptides, as opposed to peptide adopting an extended or turn conformation, was dictated primarily by the fact that they are easier to characterize experimentally. Further criteria for peptide selection were their solubility, the ease of chemical synthesis, and finally, availability of experimental data on the folding kinetics of the parent protein. Two of the studied peptides are segments of cytochrome $c_2$. This member of the cytochrome family is structurally homologous to cytochrome $c$, the folding kinetics of which have been extensively analyzed by both NMR and CD spectroscopy (Kuwajima et al., 1987a; Roder et al., 1988).

This paper starts by giving the details of the theoretical prediction, the peptide selection procedure, and the experimental protocols. Then, the theoretical predictions for all four peptides are described and compared to the experimentally determined properties and to the experimental data, when available, on the folding of the intact protein to which they belong.

## MATERIALS AND METHODS

### *1. Theoretical Prediction Procedure and Peptide Selection Criteria*

The algorithm for predicting protein segments that adopt a well-defined conformation in the absence of tertiary interactions has been described previously (Rooman et al., 1991, 1992). Only a brief summary is therefore given here.

In this procedure, implemented in the computer programs PRELUDE and FUGUE (Rooman et al., 1992), the protein backbone conformation is described as a combination of seven structural states (A, C, B, P, G, E, O), adopted by individual residues. These states represent domains in the Ramachandran map, corresponding to allowed conformations of the isolated peptide. A and C represent an $\alpha$- and $3_{10}$-helix, respectively; B and P correspond to extended structures, with B representing $\beta$-strand conformations; G and E correspond to positive values of $\phi$ and occur essentially in glycine residues; O merges all $\phi$ and $\psi$ values corresponding to cis peptide conformations that occur almost exclusively in prolines. On the basis of this representation, probabilities for a given residue $i$ to adopt each of the seven $(\phi, \psi, \omega)$ states are computed from the database of known protein 3D structures. These probabilities take into account influences of single residues and residue pairs (including gaps), removed from residue $i$ by at most eight positions along the sequence. The structural preferences of individual residues being considered as independent, the overall probability of a sequence to adopt a conformation is expressed as the product of individual residue preferences. After appropriate normalization, and assuming a Boltzmann distribution, derived probabilities are translated into potentials of mean force.

Using these potentials, PRELUDE computes a series of lowest energy conformations from the amino acid sequence of a peptide or protein segment, without requiring a conformational search. These are ranked by order of increasing energies. The extent to which a particular conformation is preferred over others can be evaluated by analyzing the ranked series. A predicted conformation is considered as "preferred", when its energy is lower by a certain margin—the energy gap ($E$-gap)—than that of the next predicted structure in the energy ranking, whose conformation is sufficiently different. To determine structural similarity, backbones are constructed from the computed structural assignments using standard bond lengths and bond angles, and a single $(\phi, \psi, \omega)$ value representing the conformational domain of the structural assignment (Rooman et al., 1991). Structures are considered as significantly different if the root mean square deviation of their backbone atoms (N, C$\alpha$, C) after coordinate superposition (Kabsch, 1978) is larger than $\varphi = (0.5 + 0.05L)$ Å, where $L$ is the number of residues in the considered fragment.

An elaboration of the above described prediction program, which applies to full protein sequences, is implemented in the program FUGUE. In this procedure, windows of fixed length (5–15 residues) were slid along the protein sequence, one residue at a time, and the lowest energy conformations of the protein segment delimited by the window was computed. Segments whose lowest energy structure displayed an $E$-gap $\geq 0.5$ kcal/mol (internal energy units) were retained. Regions where several such segments overlap and predict the same conformation were defined as likely to have a well-defined structure. This procedure can be used in two modes: one in which only the residues in the window are included in computing the energies and one in which residues along the polypeptide flanking each window are also considered.

To select the peptides for the present study, FUGUE was applied to the amino acid sequence of 69 highly resolved and refined protein structures. Only protein regions defined as having well-defined conformations with the highest score in both prediction modes were considered for final evaluation. PRELUDE was then applied to compute the lowest energy conformations of the isolated peptides corresponding to these regions, and among them, only those whose lowest energy conformation have an $E$-gap $\geq 0.5$ kcal/mol were retained. The jackknife cross validation procedure (Efron, 1982) was applied throughout: the protein on which the predictions were performed, as well as any other related protein with more than 20% sequence identity, were excluded from the dataset when deriving the force field parameters.

Candidates obtained by this protocol were subjected to further selection using the following criteria: (1) only those segments predicted to adopt a well-defined $\alpha$-helical conformation were retained, since the latter is more readily amenable to experimental analysis by NMR and CD spectroscopy, (2) among the predicted helical peptides, only those observed to be helical in the native parent protein were considered at this stage. Segments predicted to be helical but observed in a different conformation in the native protein are also of interest, but will be considered subsequently, (3) to ensure solubility, peptides with too few polar and/or charged residues were not considered.

Here we report results on four peptides from among the candidates that complied with the above criteria. Two of them correspond to the $\alpha$-helices 63–73 and 98–108 of *Rhodospirillum* cytochrome $c_2$ (3C2C) (Salemme et al., 1973) (Figure 1a), which is structurally homologous to horse cytochrome $c$. The first peptide corresponds to one of the middle helices and the second to the C-terminal helix. Both helices interact with the heme. The two other peptides

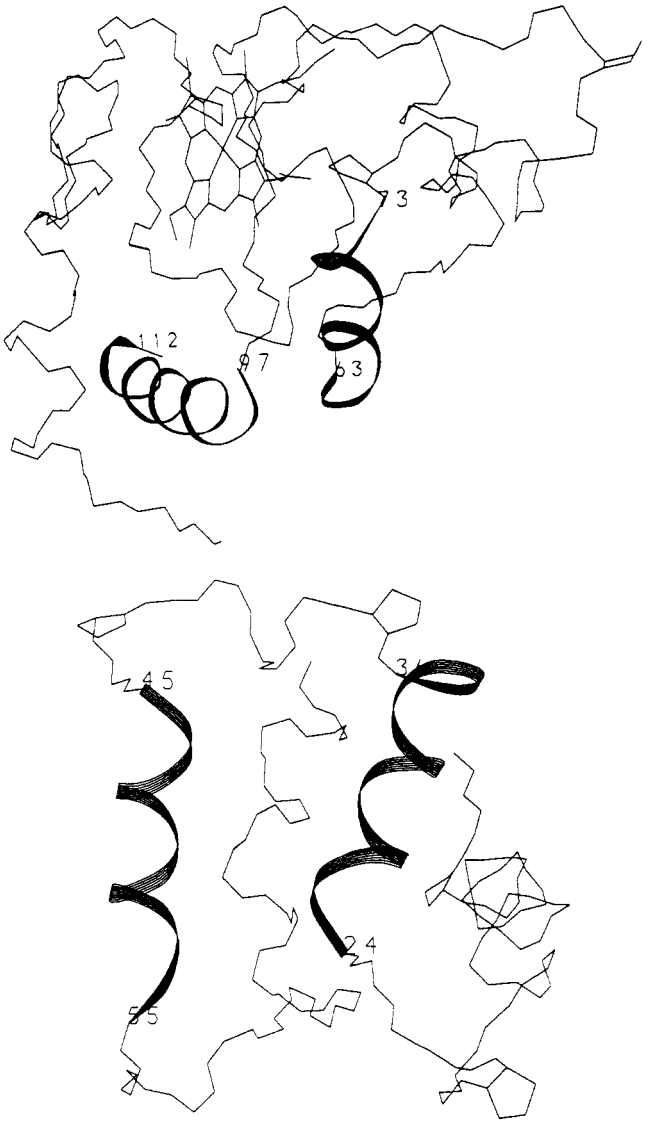



FIGURE 1: Three-dimensional backbone structure of cytochrome $c_2$ (a, top) and calcium binding protein (b, bottom). The studied segments are depicted as ribbons using the molecular modelling package BRUGEL (Delhaise et al., 1985).

correspond to $\alpha$-helices 24–36 and 45–54 of bovine calcium binding protein (3ICB) (Szebenyi & Moffat, 1986). This protein consists of two helix–loop–helix calcium-binding domains forming two EF-hands (Figure 1b). The first peptide corresponds to the second helix of the first calcium-binding domain and the second to the first helix of the second calcium-binding domain of the native protein.

To increase the chance of observing the expected $\alpha$-helical conformation in the isolated peptide in solution, the limits of the peptides were extended beyond the regions identified by FUGUE, to take into account possible stabilizing interactions with polar groups nearby along the sequence. In the case of the C-terminal cytochrome $c_2$ helix (98–109), the initial region predicted to be $\alpha$-helical (residues 100–112) was extended in the N-terminal direction to include residues Lys 97, Asp 98, and Asp 99, seen to form "capping" interactions with the helix in the native 3D structure. Likewise, the 45–54 helix of calcium binding protein, being on the short side, was extended at its C-terminus to include Lys 55, which appeared to interact favorably with the helix C-terminus in the native structure. No extensions seemed necessary for the other two peptides. All peptides were N-terminally succinylated and C-terminally amidated.

The four peptides that were finally synthesized are denoted respectively as C2C[63–73], C2C[97–112], ICB[24–36], and ICB[45–55], with the numbers in brackets indicating their limits along the native protein sequence. The sequences are reported in Figure 2.

### 2. Peptide Synthesis

The peptides were synthesized on "Dod-resin" (Knolle et al., 1987), with an Applied Biosystems 431A automated peptide synthesizer (batch reactor) using one or more of the following procedures: (a) Fmoc HOBt/DCC coupling protocols supplied by Applied Biosystems; (b) similar protocols with modified solvent systems, i.e, DMSO/THF/NMP (2:3:5 v/v/v) with or without the addition of 2 M ethylene carbonate as chaotropic agent; (c) FastMoc chemistry protocol (Knorr et al., 1989) as supplied by Applied Biosystems. The resin and the Fmoc-protected amino acids were purchased from Bachem Feinchemikalien AG, Basel. Solid phase N-terminal acylation was effected with succinic anhydride. The peptide was cleaved from the resin and amino acids side chains deprotected with TFA in the presence of byproduct scavengers (King et al., 1990) to yield the C-terminally amidated peptides. Peptides were recovered by precipitation with ice-cold *tert*-butyl methyl ether and purified by size-exclusion chromatography (Sephadex G10, $CH_3COOH/CH_3CN/H_2O$ 6/20/74) followed by reversed-phase HPLC (Nucleosil C18–7mm, 10 mm internal diameter, 250 mm length, 120 Å pore size; buffer A, 0.1% TFA in $H_2O$; buffer B, 0.08% TFA in $CH_3CN/H_2O$ 80/20 using a gradient of 25% to 60% B in 90 min at a flow rate of 6 mL/min). The resulting material was analyzed by positive-ion FAB-MS or by electrospray-MS (Gaertner et al., 1992) and found to have the expected relative molecular mass and to be >90% pure by analytical HPLC (Nucleosil C8–5 mm, 4 mm internal diameter, 250 mm length, pore size 300 Å; buffers A and B as above, using a gradient from 0% to 75% B in 60 min at a flow rate of 1 mL/min). Typically, 4–5 mg of lyophylized material was obtained, for each HPLC purification, for each peptide.

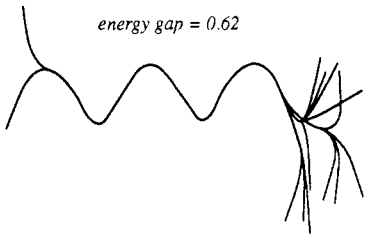
### 3. Nuclear Magnetic Resonance Spectroscopy

Peptide solutions for NMR spectroscopy were prepared by dissolving 1–2 mg of the peptide in a mixture of $^1H_2O$/trifluoroethanol-$d_3$ (Euriso-top, Gif-sur-Yvette, France) (0.6 mL; 60/40, v/v). TFE concentrations will usually be expressed as percentages in volume, not taking into account the nonideal nature of water/TFE mixtures. A water/TFE 60/40 (v/v) mixture will be indicated as 40% TFE, which corresponds to an 86/14 (mol/mol) solution. The pH was adjusted by adding 1 $\mu$L aliquots of 0.1 N HCl or 0.1 N NaOH and the solution was filtered through a 0.45 $\mu$m microfilter (Gelman Sciences). TSP (3-(trimethylsilyl)propionic-$d_4$ acid, sodium salt, purchased from Aldrich Chem. Co.) was used as the internal standard. $^1$H NMR spectra were recorded on a Varian Unity-600 spectrometer operating at 600.068 MHz. DQF-COSY (Rance et al., 1983) and TOCSY (Bax & Davis, 1985) were used to assign the spin systems. ROESY (Bothner-By et al., 1984), NOESY (Jeener et al., 1979), and NOESY-11 (Sklenár & Bax, 1987) were used to obtain the complete sequential assignments (Wüthrich, 1986). Spectra were acquired without sample spinning, using 2K complex points, 32 transients for each FID, and 256-600 $t_1$ increments, with mixing times of 60 ms for the TOCSY, 200 ms for the ROESY, and 100-400 ms for the NOESY spectra. The transmitter frequency was placed on the solvent resonance. Solvent suppression was achieved by presaturation during the relaxation delay (0.7 s)

C2C[63–73]

T E A N L A A Y V K N

*energy gap = 0.62*

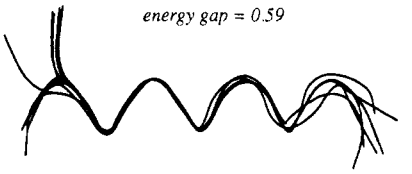


| | r.m.s. | structure | E | E-gap |
|---|---|---|---|---|
| 1 | | AAAAAAAAAAC | -4.59 | |
| 2 | 0.12 | AAAAAAAAAAA | -4.52 | 0.08 |
| 3 | 0.45 | AAAAAAAAAAG | -4.51 | 0.09 |
| 4 | 0.11 | AAAAAAAAAAB | -4.47 | 0.12 |
| 5 | 0.94 | AAAAAAAAABC | -4.14 | 0.45 |
| 6 | 0.82 | AAAAAAAAAPC | -4.10 | 0.49 |
| 7 | 0.09 | AAAAAAAAAAP | -4.07 | 0.53 |
| 8 | 0.94 | AAAAAAAAABA | -4.07 | 0.53 |
| 9 | 0.67 | AAAAAAAAABG | -4.06 | 0.54 |
| 10 | 0.92 | AAAAAAAAABB | -4.02 | 0.57 |
| 11 | 0.76 | AAAAAAAAAPA | -4.02 | 0.57 |
| 12 | 0.58 | AAAAAAAAAPG | -4.01 | 0.58 |
| 13 | 0.85 | AAAAAAAAAPB | -3.98 | 0.61 |
| 14 | 0.45 | BAAAAAAAAAC | -3.97 | 0.62 |
| 15 | 2.99 | AAAAAAABAAC | -3.97 | 0.62 |

ICB[24–36]

S K E E L K L L L Q T E F

*energy gap = 0.59*

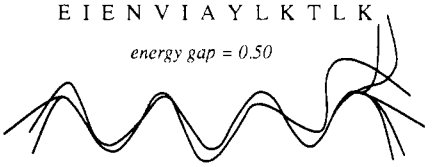


| | r.m.s. | structure | E | E-gap |
|---|---|---|---|---|
| 1 | | AAAAAAAAAAAAA | -6.28 | |
| 2 | 0.42 | PAAAAAAAAAAAA | -6.28 | 0.00 |
| 3 | 0.11 | AAAAAAAAAAAAC | -5.94 | 0.34 |
| 4 | 0.43 | PAAAAAAAAAAAC | -5.93 | 0.35 |
| 5 | 0.55 | AAAAAAAAAACAA | -5.89 | 0.40 |
| 6 | 0.72 | PAAAAAAAAACAA | -5.88 | 0.40 |
| 7 | 0.82 | AAAAAAAACAAAA | -5.87 | 0.41 |
| 8 | 0.88 | PAAAAAAACAAAA | -5.87 | 0.41 |
| 9 | 0.23 | AAAAAAAAAAAAB | -5.84 | 0.44 |
| 10 | 0.47 | PAAAAAAAAAAAB | -5.84 | 0.45 |
| 11 | 0.48 | ACAAAAAAAAAAA | -5.82 | 0.46 |
| 12 | 0.45 | PCAAAAAAAAAAA | -5.81 | 0.47 |
| 13 | 0.16 | AAAAAAAAAACA | -5.81 | 0.47 |
| 14 | 0.45 | PAAAAAAAAAACA | -5.81 | 0.48 |
| 15 | 0.66 | AAAAAAAAACAAA | -5.70 | 0.58 |
| 16 | 0.73 | PAAAAAAAACAAA | -5.70 | 0.58 |
| 17 | 3.09 | AAAAAAAAPAAAA | -5.69 | 0.59 |

C2C[97–112]

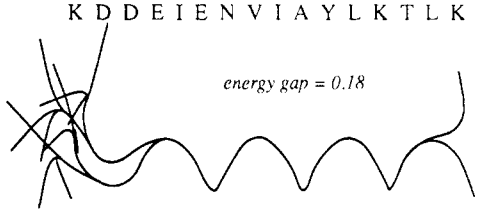
E I E N V I A Y L K T L K

*energy gap = 0.50*



K D D E I E N V I A Y L K T L K

*energy gap = 0.18*

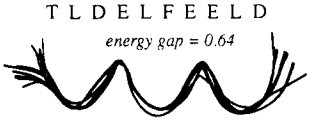


| | r.m.s. | structure | E | E-gap |
|---|---|---|---|---|
| 1 | | CACCAAAAAAAAAAAA | -5.45 | |
| 2 | 0.72 | CACAAAAAAAAAAAA | -5.44 | 0.01 |
| 3 | 0.36 | PACCAAAAAAAAAAAA | -5.44 | 0.01 |
| 4 | 0.73 | PACAAAAAAAAAAAAA | -5.43 | 0.02 |
| 5 | 0.11 | AACCAAAAAAAAAAAA | -5.42 | 0.03 |
| 6 | 0.73 | AACAAAAAAAAAAAAA | -5.41 | 0.04 |
| 7 | 0.55 | CCCCAAAAAAAAAAAA | -5.36 | 0.09 |
| 8 | 1.06 | CCCAAAAAAAAAAAAA | -5.35 | 0.10 |
| 9 | 0.46 | PCCCAAAAAAAAAAAA | -5.35 | 0.10 |
| 10 | 0.97 | PCCAAAAAAAAAAAAA | -5.34 | 0.11 |
| 11 | 0.53 | ACCCAAAAAAAAAAAA | -5.33 | 0.12 |
| 12 | 1.05 | ACCAAAAAAAAAAAAA | -5.32 | 0.13 |
| 13 | 0.60 | CACCAAAAAAAAAAPA | -5.29 | 0.17 |
| 14 | 0.90 | CACAAAAAAAAAAPA | -5.28 | 0.18 |
| 15 | 0.72 | PACCAAAAAAAAAAPA | -5.28 | 0.18 |
| 16 | 1.44 | CPCCAAAAAAAAAAAA | -5.28 | 0.18 |

ICB[45–55]

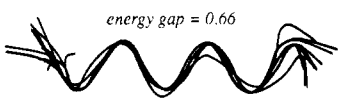
T L D E L F E E L D

*energy gap = 0.64*



T L D E L F E E L D K

*energy gap = 0.66*



| | r.m.s. | structure | E | E-gap |
|---|---|---|---|---|
| 1 | | BAAAAAAAAAA | -5.32 | |
| 2 | 0.45 | AAAAAAAAAAA | -5.31 | 0.01 |
| 3 | 0.08 | BAAAAAAAAAC | -5.16 | 0.16 |
| 4 | 0.43 | CAAAAAAAAAA | -5.15 | 0.17 |
| 5 | 0.45 | AAAAAAAAAAC | -5.15 | 0.17 |
| 6 | 0.01 | PAAAAAAAAAA | -5.01 | 0.31 |
| 7 | 0.44 | CAAAAAAAAAC | -4.99 | 0.33 |
| 8 | 0.76 | BAACAAAAAAA | -4.93 | 0.39 |
| 9 | 0.74 | AAACAAAAAAA | -4.92 | 0.40 |
| 10 | 0.08 | PAAAAAAAAAC | -4.85 | 0.47 |
| 11 | 0.23 | BAAAAAAAACA | -4.85 | 0.47 |
| 12 | 0.51 | AAAAAAAAACA | -4.84 | 0.48 |
| 13 | 0.77 | BAACAAAAAAC | -4.77 | 0.55 |
| 14 | 0.72 | CAACAAAAAAA | -4.76 | 0.55 |
| 15 | 0.76 | AAACAAAAAAC | -4.76 | 0.56 |
| 16 | 0.77 | BAAAAACAAAA | -4.74 | 0.57 |
| 17 | 0.88 | AAAAAACAAAA | -4.73 | 0.58 |
| 18 | 0.67 | BAAAAACAAA | -4.70 | 0.62 |
| 19 | 0.20 | BAAAAAAAACC | -4.69 | 0.63 |
| 20 | 0.75 | AAAAAAACAAA | -4.69 | 0.63 |
| 21 | 0.49 | CAAAAAAACA | -4.68 | 0.63 |
| 22 | 0.49 | AAAAAAAAACC | -4.68 | 0.64 |
| 23 | 1.66 | BBAAAAAAAAA | -4.66 | 0.66 |

FIGURE 2: Predicted lowest energy conformations for (a, upper left) C2C[63–73], (b, lower left) C2C[97–112], (c, upper right) ICB[24–36], and (d, lower right) ICB[45–55]. The lower part of each panel shows the predicted lowest energy conformations up to the first one which diplays a substantial disruption of the helical structure, as estimated by the rms value with respect to the lowest energy conformation, which must be larger than $(0.5 + 0.05L)$ (Å), where $L$ is the peptide length. These conformations are described in terms of the structural states A ($\alpha$-helix), C ($3_{10}$-helix), B and P (extended), and G (positive $\phi$ values). For each listed conformation, its rms deviation (Å) with respect to the lowest energy structure, calculated after superposition of the N, C$\alpha$, and C backbone atoms (Kabsch, 1978), is given; the coordinates are obtained using representative values for each structural state and average bond lengths and angles. The energy ($E$) of each conformation (kcal/mol, internal energy units) is also indicated, as well as the energy gap ($E$-gap, kcal/mol) with respect to the lowest energy conformation. The upper part of each panel shows the amino acid sequence of the peptides, the superposition of the N, C$\alpha$, and C backbone atoms of the lowest energy conformations reported in the lower panel, except for the last one, and the $E$-gap. In b and d, the amino acid sequence, the backbone superposition, and the corresponding $E$-gap of the fragment which was originally predicted to be $\alpha$-helical are also reported.

and also during the mixing time in the NOESY experiments. All spectra were recorded at 298 K. Data were processed on a Sun Sparc workstation using the Varian VNMR software.

Data were Fourier transformed using a shifted Gaussian weighting function in both $t_2$ and $t_1$ domains, with zero-filling of the data to 1K complex points in the $t_1$ domain and application of a first-order baseline correction in the $t_2$ domain. NH–CH coupling constants ($^3J_{H\alpha N}$) were estimated from high signal to noise ratio 1D spectra taken with 8K complex points, zero-filled to 16K, and Fourier transformed using a Henning-like weighting function to enhance resolution. In the case of amide peak overlap, the pH was slightly changed (±0.1 pH units) to separate the amide resonances. The intensity of NOE cross-peaks was evaluated by counting the number of contours. The possible occurrence of aggregation was tested by measuring chemical shifts and linewidths in 1D spectra obtained for different peptide concentrations.

The percentage of helix was calculated from $^3J_{H\alpha N}$ using the expression $1/n\sum[(J_{RC} - J_{obs})/(J_{RC} - J_h)]$, where $n$ is the number of residues in the peptide, $J_{RC}$ is the random coil value of $^3J_{H\alpha N}$ for the given amino acid type, as reported by Wüthrich (Bundi & Wüthrich, 1979), $J_{obs}$ is the experimental value of $^3J_{H\alpha N}$, and $J_h$ is the reference value of $^3J_{H\alpha N}$ for an $\alpha$-helix (3.9 Hz) (Wüthrich, 1986); when $J_{obs} > J_{RC}$, the contribution of that residue to the $\alpha$-helical content was set to zero. No corrections were made for $J_{obs}$ values smaller than 3.9 Hz or for solvent effects. The percentage of helix was also calculated on the basis of $\alpha$-protons chemical shift deviations from the random coil values (Jimenez et al. 1993) using the following expression: $1/n\sum[(\delta_{obs} - \delta_{RC})/(\delta_h - \delta_{RC})]$ where $n$ is the number of residues in the peptide, $\delta_{obs}$ is the experimental value of the $\alpha$-proton chemical shift for that residue, $\delta_{RC}$ is the random coil value of the $\alpha$-proton chemical shift for that amino acid type, as reported by Wüthrich (1986), and $\delta_h$ is the $\alpha$-proton chemical shift for that amino acid type in an $\alpha$-helical conformation, as reported by Wishart et al. (1991). No corrections were made for solvent effects.

### *4. Circular Dichroism Spectroscopy*

Peptide stock solutions for CD were prepared by dissolution of 2 mg of the peptide in 1 mL of 5 mM acetate buffer (pH 3.3 for C2C[62–73], 4.5 for C2C[97–112], 4.0 for ICB[24–36], and ICB[45–55]), centrifugation, and filtration on a 0.22 $\mu$m microfilter (Gelman Sciences). The resulting concentration was 0.342, 0.873, 0.073, 1.232 mM for C2C[63–73], C2C[97–112], ICB[24–36], and ICB[45–55], respectively. The stock solution was appropriately diluted either in acetate buffer or in a mixture of acetate buffer (5 mM) and trifluoroethanol (Aldrich Chem. Co.) to give 0, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60% TFE solutions. The resulting peptide concentration was 86, 87, 29, 308 $\mu$M for the four peptides, respectively. The peptide concentration was determined by amino acid analysis of the stock solutions, with an estimated error of about 5%. CD spectra were recorded on a Yvon-Jobin dichograph equipped with a thermostatic bath. A quartz cell of 0.2 mm pathlength was generally used. In the near-UV region, three spectra were usually acquired between 250 and 250 nm, using a wavelength step of 1 nm and an acquisition time of 5 s. In the far-UV region, five spectra were usually acquired between 185 and 255 nm using a wavelength step of 1 nm and an acquisition time of 5 s.

The absence of a contribution from the aromatic residues (a tyrosine for C2C[63–73] and C2C[97–112], a phenylalanine for ICB[24–36] and ICB[45–55]) to the CD spectra of the peptides was tested on the buffer and the 40% TFE peptide solutions before the calculations of the secondary structure content by deconvolution of the CD bands. The eventuality of peptide aggregation was tested by studying the concentration dependence of the intensity of CD bands. Deconvolution of the far-UV CD spectra and analysis of the secondary structure content was calculated by the method of Yang et al. (1986) using the polypeptide database of Brahms and Brahms (1980) or the 18-protein database of Yang et al. (1986). The helix length was considered as infinite when using the database of Brahms and Brahms (1980), whereas the actual peptide length was considered when using the 18-protein database. The percentage of helix in 40% TFE was also calculated from the mean residue ellipticity at 222 nm, using the expression, % helicity = $(\Theta/39\,500) \times 100$).

The pH was measured with an Ingold combined microelectrode, previously calibrated in buffer and buffer/TFE mixtures, and pH readings were corrected for the TFE effect.

## RESULTS

### *1. Predicted Conformational Properties*

The predicted lowest energy conformations of the four peptides are reported in Figure 2. These conformations are expressed in terms of the seven structural states accessible to each residue. Each state being represented by single values of the ($\phi$, $\psi$, $\omega$) angles, the predicted backbone structures of each peptide were reconstructed using ideal bond distances and angles and compared, using as a criterion the root mean square (rms) deviations of the atomic coordinates after superposition (Kabsch, 1978). The superimposed backbones of the lowest energy conformations of the regions originally identified as $\alpha$-helical, and of the extended segments, are also depicted in Figure 2.

We can remark that using PRELUDE to compute the energy of the elongated peptides may yield energy gaps that are lower than the 0.5 kcal/mol threshold imposed on the original peptides. This is the case with C2C[97–112] and is a consequence of the predicted higher flexibility of the N-terminal elongated region of this peptide.

*Cytochrome $c_2$ Peptides.* The lowest energy conformations of C2C[63–73] were predicted to be $\alpha$-helical, with some variability in the conformation of the last two residues at the C-terminus. A substantial disruption of the $\alpha$-helical conformation (rms = 2.99 Å) is observed in the 15th structure in the energy ranking ($E$-gap = 0.62 kcal/mol), which features an extended conformation (B) at the eighth position (Y70) in the sequence.

For C2C[97–112], a whole set of lowest energy conformers differing only slightly in energy ($E$-gap ~0.1 kcal/mol) was predicted. All of them display a common, fully $\alpha$-helical part, extending from the fifth (I101) to the last residue in the segment, and a more variable part, spanning over the first four residues. The variability in this part is however mostly limited to switches between the closely related A and C ($3_{10}$-helix) states in individual residues, entailing only small rms deviations between the different conformations. Descending in the energy ranking, some variability is found also at L111, whose conformation switches from A to P. The 16th conformation in the ranking is the first to show a significant rms deviation (1.44 Å).

*Calcium Binding Protein Peptides.* For ICB[24–36], two predicted $\alpha$-helical structures rank first by their energy. They differ only by the conformation of the N-terminal residue (A or P). These structures display an $E$-gap of 0.34 kcal/mol relative to the third conformation in the ranking. A substantial

disruption of the α-helix (rms = 3.09 Å), accompanied by an *E*-gap of 0.59 kcal/mol, occurs only in the 17th lowest energy structure, when the 9th residue in the segment is predicted in the extended conformation (P).

Two α-helical structures of lowest energy are also predicted for ICB[45–55]. Here too, these structures differ only in the conformation of the N-terminal residue (A or B). The first structure in the list which differs significantly from these structures (rms = 1.66 Å) has the first two residues at the N-terminus in the B conformation and displays an energy gap of 0.66 kcal/mol.

### 2. Experimentally Determined Properties

*2a. Solubility and Aggregation.* The sometimes limited solubility of short peptides and the eventual aggregation occurring already at millimolar concentrations are problems often encountered when studying protein fragments in solution. Finding the correct experimental conditions such that a monomeric species is present in solution, at the suitable concentration and pH, is thus of crucial importance in spectroscopic studies of peptides.

Here, to monitor aggregation, we measured the concentration dependence of chemical shifts, of NMR linewidths, and of the intensity of far-UV CD bands and checked for the presence of chiroptical activity in the near-UV. These physical properties are respectively sensitive to the chemical environment, to the size of the molecule, to the secondary structure content, and to the presence of aromatic residues immobilized in a chiral environment and can thus provide complementary information on the formation of aggregates.

We found no evidence for aggregation in 40% TFE for the four peptides in the concentration range used. In water (or buffer), we found ambiguous results for C2C[63–73], strong aggregation occurring for C2C[97–112], and no evidence for aggregation for ICB[24–36] and ICB[45–55]. Hereafter, we report the results in detail.

*Cytochrome $c_2$ Peptides.* C2C[63–73] displayed low solubility in water and dissolved with some difficulty in 40% TFE at pH ~3.5, to precipitate at pH > 4.5. Upon dilution 1:10 of the NMR sample in 40% TFE, small $\Delta\delta_{NH}$ (0.05 ppm) were found for the amide protons of A68, N73, and A65, but no difference larger than 0.01 ppm was found for the other amide protons or for the aliphatic or aromatic protons. Also linewidths were independent of the peptide concentration.

A negative CD band of very small intensity could be detected around 290 nm in buffer, but the CD spectra were completely flat in the same region in 40% TFE. A small aromatic contribution to the CD spectrum can possibly arise from aggregation and explain the somewhat anomalous ellipticity value at 222 nm in buffer. Nevertheless no differences were detected by superposition of the CD spectra taken at different concentrations and no aggregation was found by CD in water/TFE mixtures.

The solubility of C2C[97–112] was high both in water and in water/TFE mixtures. Very broad lines were observed in NMR spectra in water over a broad range of pH in the millimolar concentration, but neither chemical shift changes larger than 0.01 ppm nor linewidth differences were measured for the samples in 40% TFE at different concentrations. In the case of C2C[97–112] in buffer, a decrease of the ellipticity by 20% at 222 nm was observed by 1:10 dilution of the sample. A systematic study of the concentration dependence of CD spectra was then carried out at different TFE concentrations, and no difference in the CD spectra of the concentrated and the diluted sample was found in 15% TFE or at higher TFE percentage. A negative band of very low intensity could be detected around 280 nm in buffer, but the CD spectra are completely flat in the same region in 40% TFE. A small contribution from aromatic residues can be envisaged in the case of C2C[97–112] and can be correlated with the extensive aggregation that occurs in buffer (and water). These results suggest that C2C[97–112] aggregates in buffer (and water) but not at high TFE concentrations.

*Calcium Binding Protein Peptides.* The solubility of ICB[24–36] was low in water, but the peptide was readily dissolved in 40% TFE. The chemical shifts of the amide protons of the charged residues E26, E27, and E35 slightly changed upon dilution (~0.05 ppm), but they were not accompanied by changes in any of the other protons. No concentration dependence of linewidths was observed. The normalized CD spectra of the parent and the diluted solutions, either in buffer or in 40% TFE, were superimposable within experimental error and we can conclude that no aggregation occurs, the small changes of chemical shift being caused probably by very slight pH changes. The ellipticity in the region 255–350 nm was zero both in buffer and in 40% TFE, thus showing no contribution to the CD spectra from the phenylalanine.

For ICB[45–55] no concentration dependence was observed for the chemical shifts and linewidths in 40% TFE. Likewise, no concentration dependence was observed for CD spectra, either in buffer or in 40% TFE. The ellipticity in the region 255–350 nm was zero both in buffer and in 40% TFE, thus showing no contribution to the CD spectra from the phenylalanine.

*2b. Conformational Properties.* Both NMR and CD spectra of α-helices present some unique, easily recognizable features. We used NOESY spectra, $^3J_{H\alpha N}$ values, and α-proton chemical shifts to characterize in detail the preferred conformation in 40% TFE for each peptide and to define the limits of the helix ($^3J_{H\alpha N}$ values of ~4 Hz and negative α-proton $\Delta\delta_{H\alpha}$ being associated with the α-helical conformation), $^3J_{H\alpha N}$ values and α-proton chemical shifts to calculate the overall α-helical content, the ellipticity at 222 nm as a function of TFE concentration to estimate the helical propensity, and the decomposition of far-UV CD spectra in secondary structure elements to calculate the α-helical content by CD. In this section, we report the experimental results obtained from NMR and CD studies for each peptide. These results are also summarized in Tables 1 and 2 and in Figures 3–7.

*Cytochrome $c_2$ Peptides. a. C2C[63–73].* The $^1$H chemical shifts are reported in Table 1. The scheme of the observed NOE connectivities, the $^3J_{H\alpha N}$ values, and the chemical shift differences for the α-protons are plotted in Figure 3a,b.

A complete pattern of strong, sequential $d_{NN}(i,i+1)$ NOE connectivities and $d_{\alpha N}(i,i+1)$ NOEs of reduced intensities was observed. At least one unambiguous $d_{\alpha N}(i,i+3)$ connectivity in the central portion of the sequence and a nearly complete pattern of $d_{\alpha\beta}(i,i+3)$ NOEs involving residues T63 through K72 were identified, indicating the presence of an α-helical conformation for nearly the entire length of the peptide, which is also confirmed by the negative values of the α-proton $\Delta\delta_{H\alpha}$. No $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ NOEs could be unambiguously assigned to the N-terminal partly due to peak overlap. The average value of $^3J_{H\alpha N}$ was of 5.5 Hz, with a remarkably higher value for the last residue (7.4 Hz), which thus displays a larger degree of flexibility than the rest of the chain. However, the presence of a strong $d_{NN}(72,73)$ NOE cross-peak, together with a weak $d_{\alpha N}(72,73)$ NOE connectivity and a negative value for the $\Delta\delta_{H\alpha}$, suggests that also the last

Table 1: Chemical Shifts of the Four Analyzed Peptides[a]

| | proton chemical shifts, ppm | | | | | |
|---|---|---|---|---|---|---|
| | N | α | β | γ | δ | others |
| | | | | C2C[63–73] | | |
| T63 | 8.17 | 4.28 | 4.31 | 1.30 | | |
| E64 | 8.40 | 4.23 | 2.20, 2.13 | 2.53, 2.53 | | |
| A65 | 7.90 | 4.21 | 1.45 | | | |
| N66 | 8.01 | 4.64 | 2.89, 2.85 | | | $\gamma NH_2$, 7.52, 6.81 |
| L67 | 8.01 | 4.27 | 1.76, 1.46 | 1.76 | 0.97, 0.92 | |
| A68 | 8.10 | 4.10 | 1.48 | | | |
| A69 | 7.74 | 4.11 | 1.43 | | | |
| Y70 | 7.86 | 4.38 | 3.23, 3.15 | | | 2,6H, 7.15; 3,5H, 6.84 |
| V71 | 8.10 | 3.84 | 2.20 | 1.10, 1.00 | | |
| K72 | 8.08 | 4.19 | 1.91 | 1.54, 1.51 | 1.72, 1.72 | ε, 3.02, 3.02 |
| N73 | 7.96 | 4.68 | 2.88, 2.81 | | | $\gamma NH_2$, 7.57, 6.82; C-ter $NH_2$, 7.28, 7.02 |
| | | | | C2C[97–112] | | |
| K97 | 8.26 | 4.36 | 1.93, 1.84 | 1.52, 1.52 | 1.75, 1.75 | ε, 3.05, 3.05 |
| D98 | 8.42 | 4.62 | 2.85, 2.77 | | | |
| D99 | 8.19 | 4.63 | 2.81, 2.81 | | | |
| E100 | 8.21 | 4.26 | 2.21, 2.21 | 2.51, 2.44 | | |
| I101 | 8.13 | 3.96 | 1.99 | 1.63, 1.34 | 0.94 | γMe, 1.00 |
| E102 | 8.22 | 4.03 | 2.17, 2.17 | 2.56, 2.49 | | |
| N103 | 7.90 | 4.57 | 3.04, 2.82 | | | $\gamma NH_2$, 6.86, 7.47 |
| V104 | 7.97 | 3.81 | 2.34 | 1.12, 1.04 | | |
| I105 | 8.37 | 3.72 | 1.99 | 1.78, 1.29 | 0.86 | γMe, 0.97 |
| A106 | 8.06 | 4.07 | 1.60 | | | |
| Y107 | 8.01 | 4.34 | 3.36, 3.29 | | | 2,6H, 7.11; 3,5H, 6.80 |
| L108 | 8.88 | 3.92 | 2.08, 1.46 | 2.08 | 0.92, 0.92 | |
| K109 | 8.59 | 4.01 | 2.02, 1.94 | 1.47, 1.47 | 1.75, 1.71 | ε, 2.96, 2.96 |
| T110 | 7.86 | 4.14 | 4.39 | 1.36 | | |
| L111 | 7.74 | 4.15 | 1.73, 1.50 | 1.50 | 0.80, 0.76 | |
| K112 | 7.91 | 4.25 | 1.94, 1.86 | 1.55, 1.49 | 1.72, 1.72 | ε, 3.01.3.01; C-ter $NH_2$, 7.21, 6.99 |
| | | | | ICB[24–36] | | |
| S24 | 8.22 | 4.47 | 4.09, 4.01 | | | |
| K25 | 8.55 | 4.21 | 1.96, 1.96 | 1.59, 1.52 | 1.80, 1.80 | ε, 3.07, 3.07 |
| E26 | 8.14 | 4.14 | 2.12, 2.12 | 2.45, 2.45 | | |
| E27 | 8.13 | 4.07 | 2.18, 2.18 | 2.49, 2.49 | | |
| L28 | 7.91 | 4.18 | 1.84, 1.72 | 1.72 | 1.00, 0.95 | |
| K29 | 7.84 | 3.94 | 1.98, 1.98 | 1.66, 1.45 | 1.76, 1.76 | ε, 2.98, 2.98 |
| L30 | 7.73 | 4.18 | 1.87, 1.77 | 1.76 | 0.95, 0.92 | |
| L31 | 8.27 | 4.10 | 1.91, 1.85 | 1.75 | 0.97, 0.93 | |
| L32 | 8.65 | 4.18 | 1.97, 1.56 | 1.97 | 0.94, 0.91 | |
| Q33 | 8.31 | 4.13 | 2.33, 2.23 | 2.63, 2.45 | | $\delta NH_2$, 7.18, 6.66 |
| T34 | 8.05 | 4.14 | 4.39 | 1.32 | | |
| E35 | 8.25 | 4.16 | 1.90, 1.74 | 2.15, 2.15 | | |
| F36 | 8.13 | 4.71 | 3.37, 3.02 | | | 3,5H, 7.38; 2,6H, 7.32; 4H, 7.26; C-ter $NH_2$, 7.24, 7.01 |
| | | | | ICB[45–55] | | |
| T45 | 7.80 | 4.38 | 4.38 | | | |
| L46 | 8.26 | 4.14 | 1.66, 1.66 | 1.63 | 0.94, 0.89 | |
| D47 | 7.98 | 4.46 | 2.79, 2.79 | | | |
| E48 | 7.80 | 4.12 | 2.25, 2.18 | 2.55, 2.49 | | |
| L49 | 7.90 | 4.14 | 1.80, 1.67 | 1.56 | 0.89, 0.82 | |
| F50 | 8.23 | 4.34 | 3.27, 3.23 | | | 3,5H, 7.34; 2,6H, 7.31; 4H, 7.27 |
| E51 | 8.10 | 4.14 | 2.31, 2.31 | 2.68, 2.59 | | |
| E52 | 8.03 | 4.20 | 2.30, 2.25 | 2.59, 2.59 | | |
| L53 | 8.28 | 4.20 | 1.87, 1.57 | 1.86 | 0.87, 0.87 | |
| D54 | 8.16 | 4.60 | 2.84, 2.84 | | | |
| K55 | 7.77 | 4.27 | 1.99, 1.93 | 1.63, 1.55 | 1.75, 1.75 | ε, 3.05; 3.05; C-ter $NH_2$, 7.34, 6.98 |

[a] The peptides are denoted respectively as C2C[63–73], C2C[97–112], ICB[24–36], ICB[45–55], with the number in parentheses indicating the limits of the peptide along the native protein sequence. C2C and ICB are the codes used here for *Rhodospirillum* cytochrome $c_2$ and for bovine calcium binding protein, respectively. The residue names (one letter code) and numbers are given in the left most column. Other columns give (from left to right) the chemical shifts for the amide, α, β, γ, δ, and "other" protons, measured relative to TSP. The nature of the "other" protons, is indicated explicitly using the common convention. All results were obtained for measures in water/TFE (60/40; v/v) at 298 K.

residue adopts a helical or turnlike conformation. The α-helical content of the peptide was evaluated to be 49% and 44%, from coupling constant and chemical shift values, respectively, as described in Materials and Methods.

Typical far-UV CD spectra at different TFE concentrations are reported in Figure 3b and the ellipticity values at 222 nm in Figure 3. We found a somewhat anomalous value of the ellipticity in buffer, possibly due to a small degree of aggregation, as discussed in a previous section. The ellipticity increased between 5% and 25% TFE and levelled off above 25% TFE (Figure 3c). The α-helical content of the peptide in 40% TFE was 43 and 50%, obtained using the protein database of Yang et al. (1986) and the polypeptide database of Brahms and Brahms (1980), respectively (Table 2). An α-helical content of 34% was calculated from the mean residue ellipticity at 222 nm.

*b. C2C[97–112].* NMR results are reported in Table 1 and in Figure 4a,b. The HN–HN, αN, and αβ regions of NOESY spectra are reproduced in Figure 5.

A complete pattern of strong, sequential $d_{NN}(i,i+1)$ NOE connectivities and $d_{\alpha N}(i,i+1)$ NOEs of reduced intensities was observed (Figure 4a). We furthermore identified a nearly

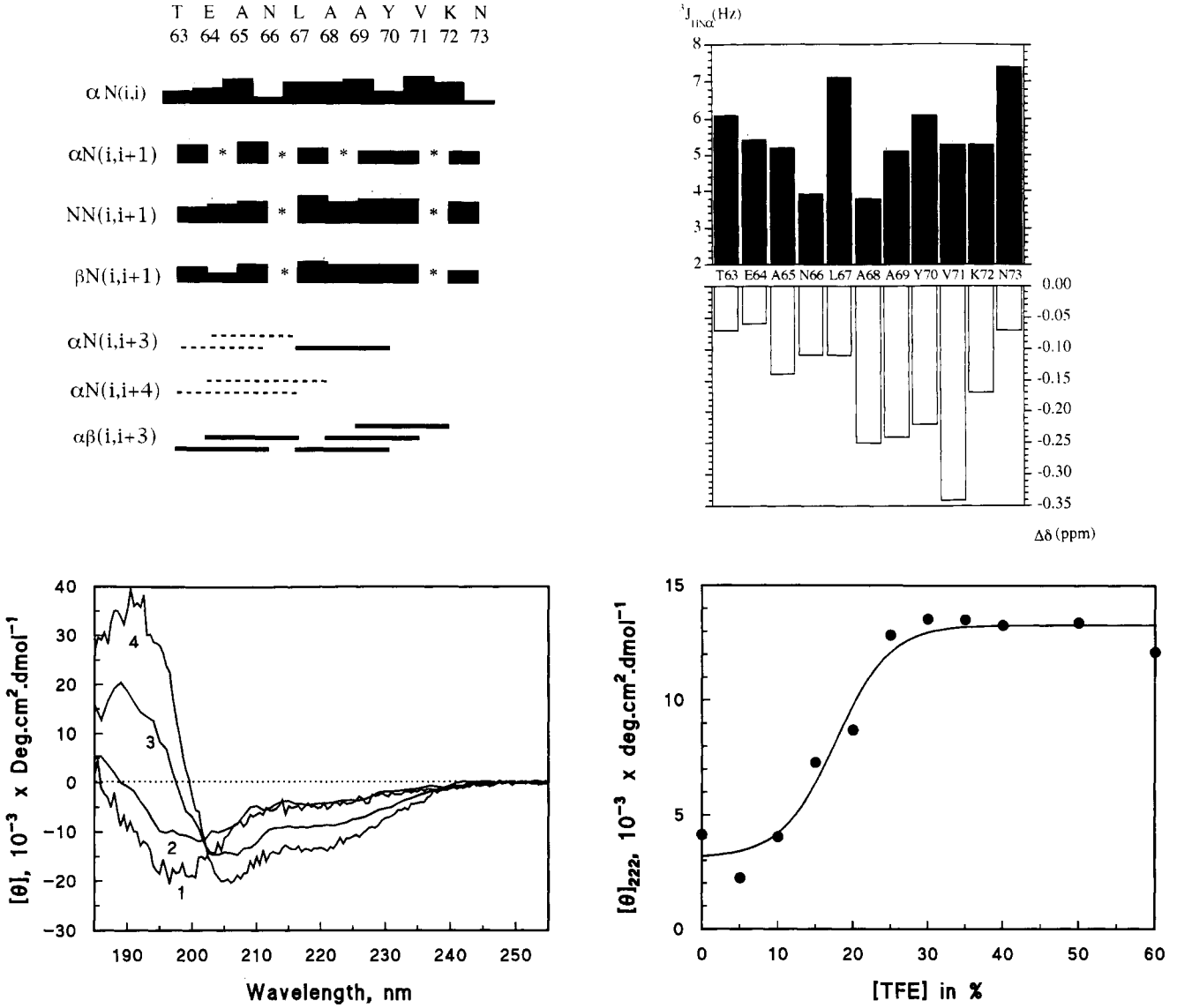



FIGURE 3: NMR and CD results for fragment C2C[63–73]. (a, upper left) Schematic representation of NOE connectivities. The thickness of the bar is proportional to the intensity of the NOE cross-peak; *, not observed because of peak overlap; —, observed medium-range NOE cross-peak; - - -, medium-range NOE cross-peak overlapping with intraresidue or sequential cross-peaks. (b, upper right) Schematic representation of $^3J_{H\alpha N}$ values (filled bars) and α-proton chemical shift deviations from random coil values $\Delta\delta_{H\alpha}$ (empty bars). Experimental conditions for NMR: ~2 mM peptide in TFE 40%, pH 3.3, 298 K, 300 ms mixing time for NOESY spectra. (c, lower left) Representative far-UV CD spectra at different TFE concentrations. The spectra labeled as 1, 2, 3, and 4 correspond respectively to 0%, 10%, 20%, and 40% TFE peptide solutions. (d, lower right) Mean residue ellipticity values at 222 nm for different TFE concentrations. The experimental conditions for CD are 86 μM peptide, pH 3.3, 298 K.

Table 2: Percent Helicity in the Four Peptides Derived from CD and NMR Data[a]

| | CD | | | NMR | |
|---|---|---|---|---|---|
| | Yang | B&B | $\Theta_{222}$ | $^3J_{H\alpha N}$ | $\Delta\delta_{H\alpha}$ |
| C2C[63–73] | 43 | 50 | 34 | 49 | 44 |
| C2C[97–112] | 65 | 75 | 64 | 65 | 55 |
| ICB[24–36] | 75 | 80 | 79 | 66 | 53 |
| ICB[45–55] | 55 | 65 | 42 | 53 | 51 |

[a] The left most column gives the peptide identity using the convention described in the text and in the legend of Table 1. Percent helicity computed from CD data is obtained using, respectively, the databases of Yang et al., and Brahms and Brahms (B&B), and from the mean residue ellipticity at 222 nm ($\Theta_{222}$). Percent helicity derived from NMR data is computed from the amide $J$ coupling constants ($^3J_{H\alpha N}$) and from the changes in the α-proton chemical shift ($\Delta\delta_{H\alpha}$). All results were obtained from measures in water/TFE (60/40, v/v) at 298 K.

complete pattern of $d_{\alpha\beta}(i,i+3)$ and $d_{\alpha N}(i,i+3)$ NOEs between residues 100 and 112, and several $d_{\alpha N}(i,i+4)$ NOEs in the central portion of the chain. Fewer medium-range connectivities, somewhat reduced intensity of the $d_{NN}(i,i+1)$ NOEs, and larger coupling constants (Figure 4b) were found for the first three residues of the peptide. These results indicate that residues 100–112 are more predominantly α-helical than residues 97–99, which appear to be more flexible. Although we find that the $^3J_{H\alpha N}$ coupling constants of residues 110–112 are above the random coil value, the presence of several medium-range NOEs [$d_{\alpha N}(108,111)$, $d_{\alpha N}(109,112)$, $d_{\alpha\beta}(107,110)$, $d_{\alpha\beta}(108,111)$] confirms the presence of a threshold population of conformers in which the helix is extending to the peptide C-terminus. This view is substantially confirmed by the plot of $\Delta\delta_{H\alpha}$, wherein more negative values are found in the central region of the peptide (residues 101–109) than in the N- and C-terminal regions.

The α-helical content of this peptide was estimated to be 65% and 55% from $^3J_{H\alpha N}$ and α-proton chemical shifts, respectively (Table 2).

The behavior of C2C[97–112] at low TFE concentrations was affected by strong aggregation even in the concentration range used for CD measurements. Selected far-UV CD

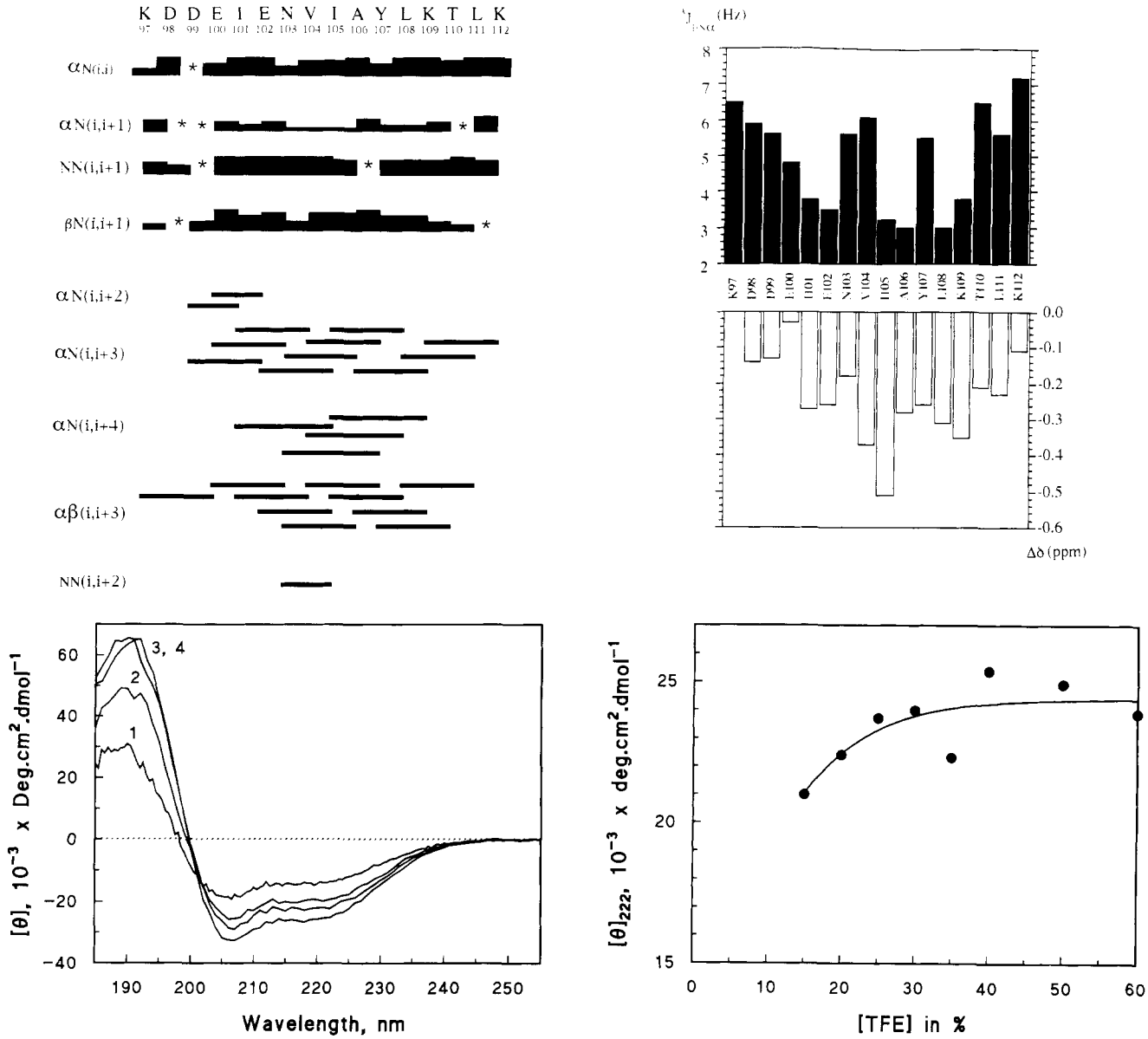



FIGURE 4: NMR and CD results for fragment C2C[97–112]. (a, upper left) Schematic representation of NOE connectivities. The thickness of the bar is proportional to the intensity of the NOE cross-peak; *, not observed because of peak overlap; —, observed medium-range NOE cross-peak; - - -, medium-range NOE cross-peak overlapping with intraresidue or sequential cross-peaks. (b, upper right) Schematic representation of $^3J_{H\alpha N}$ values (filled bars) and $\alpha$-proton chemical shift deviations from random coil values (empty bars). Experimental conditions: ~2 mM peptide in TFE 40%, pH 4.5, 298 K, 300 ms mixing time for NOESY spectra. (c, lower left) Representative far-UV CD spectra at different TFE concentrations. The spectra labeled as 1, 2, 3, and 4 correspond respectively to 0%, 10%, 20%, and 40% TFE peptide solutions. Aggregation is occurring below 15% TFE. (d, lower right) Mean residue ellipticity values at 222 nm for different TFE concentrations. Experimental conditions 87 $\mu$M peptide, pH 4.5, 298 K.

spectra and ellipticity values at 222 nm at different TFE concentration have been reported in Figure 5, parts c and d, respectively. Between 15% and 30% TFE the ellipticity increased linearly, to decrease slightly at higher TFE concentrations. From the secondary structure analysis of the 40% TFE peptide solution, an $\alpha$-helical content of 65% and 75% was calculated, using the databases of Yang et al. (1986) and Brahms and Brahms (1980), respectively (Table 2). The $\alpha$-helical content determined from the mean residue ellipticity at 222 nm was 64%.

*Calcium Binding Protein Peptides. a. ICB[24–36].* The full assignment of the NMR spectra of this peptide was considerably more difficult than for C2C[63–73] and C2C-[97–112], because of the presence of four leucines in the sequence (three of which are consecutive) and the clustering of eight $\alpha$-protons in ~0.1 ppm. The $^1$H chemical shifts are reported in Table 1. A complete pattern of strong, sequential $d_{NN}(i,i+1)$ NOE connectivities and $d_{\alpha N}(i,i+1)$ NOEs of reduced intensities was observed (Figure 6a). We also identified a $d_{\alpha\beta}(i,i+3)$ connectivity pattern that extends from the N-terminus (S24) to the 11th residue (T34) of the peptide and $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ NOEs connecting residues E27 through Q33. As estimated from medium-range NOEs, the helical region thus starts at S24 and extends to T34. In the C-terminal part of the peptide, we observed two strong $d_{NN}$-(34,35) and $d_{NN}$(35,36) NOEs. However, the reduced $^3J_{H\alpha N}$ value (6.0 Hz) and the negative $\Delta\delta_{H\alpha}$ of E35 are accompanied by the large $^3J_{H\alpha N}$ (8.3 Hz) and the positive $\Delta\delta_{H\alpha}$ of F36 (Figure 6b), thus suggesting that the C-terminal residue is probably no longer in a helical conformation.

We found an average value for the $^3J_{H\alpha N}$ coupling constants of 4.6 Hz and an $\alpha$-helical content of 66% from $^3J_{H\alpha N}$ values and of 53% from chemical shift values (Table 2).

Circular dichroism spectra (Figure 6c) showed a linear increase in the ellipticity between 0 and 20% TFE with levelling off at higher TFE concentrations (Figure 6d). The $\alpha$-helical

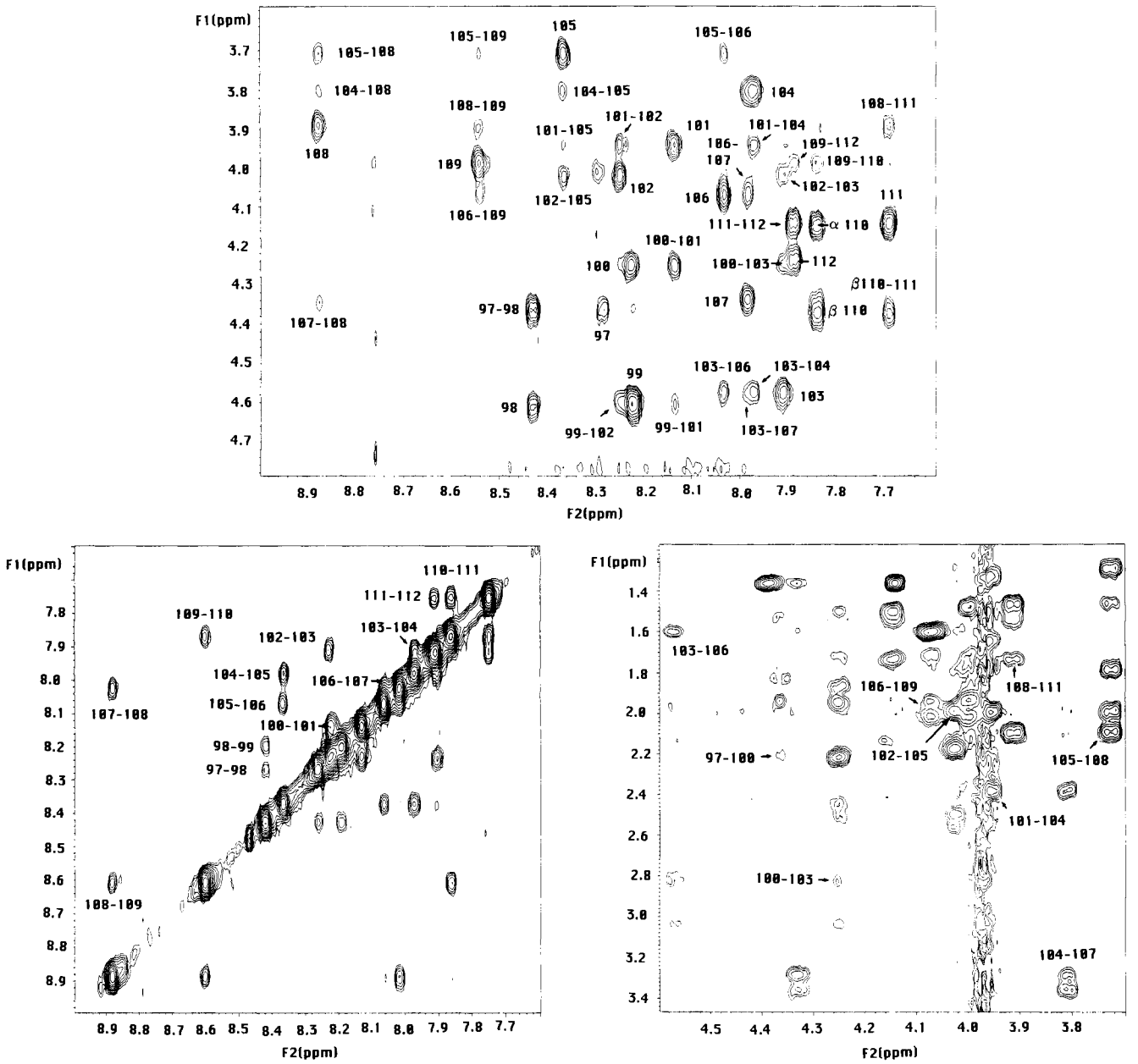



FIGURE 5: Contour plots of 600 MHz $^1$H-NMR NOESY spectra of fragment C2C[97–112]: a (top), NH–$\alpha$H region; b (bottom left), NH–NH region; c (bottom right), $\alpha$H–$\beta$H region; $d_{NN}(i,i+1)$, $d_{\alpha N}(i,i)$, $d_{\alpha N}(i,i+1)$, $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, and $d_{\alpha\beta}(i,i+3)$ NOE connectivities are labeled with the residue numbers. Intraresidue NOE cross-peaks are labeled with the residue number. Experimental conditions: ~2 mM peptide in TFE 40%, pH 4.5, 298 K, 300 ms mixing time, with the exception of a, which was recorded using a NOESY 11-pulse sequence and 200 ms mixing time.

content, as calculated from the deconvolution of CD spectra, was 75 and 80%, with the databases of Yang et al. (1986) and Brahms and Brahms (1980), respectively (Table 2), and 79% from the mean residue ellipticity at 222 nm.

*b. ICB[45–55].* As in the case of ICB[24–36], the complete assignment of NMR resonances (Table 1) was made difficult by the presence of three leucines and of six $\alpha$-protons clustered in ~0.1 ppm. This leads to very crowded spectra and in some occasions prevented us from assigning NOESY cross-peaks in an unambiguous way because of peak overlap.

The NOE connectivities are sketched in Figure 7a. The complete pattern of strong, sequential $d_{NN}(i,i+1)$ NOE connectivities is accompanied by $d_{\alpha N}(i,i+1)$ NOEs of equal intensities. We identified a series of medium-range NOEs between the first residue (T45) through the last (K55), although many of them could not be unambiguously assigned because of extensive peak overlap. With the exception of the N-terminal residue (T45), all the others displayed negative $\Delta\delta_{H\alpha}$ values.

The value of the $^3J_{H\alpha N}$ coupling constants (Figure 7b), averaged over the entire length of the peptide, is 5.2 Hz and the helix content, determined from $^3J_{H\alpha N}$ coupling constants, is 53%. A very similar helical content (51%) was found by analysis of chemical shift values (Table 2).

Selected far-UV CD spectra at different TFE concentrations and the ellipticity values at 222 nm are shown in Figure 7, parts c and d, respectively. ICB[45–55] displays a very similar behavior to that of ICB[24–36]: its ellipticity value increases linearly between 0 and 20% TFE, but levels off at higher TFE concentrations. The $\alpha$-helical content of 55% and 65% was determined from the deconvolution of the CD spectra using the databases of Yang et al. (1986) and Brahms and Brahms (1980), respectively (Table 2). Using the mean residue ellipticity at 222 nm, an $\alpha$-helical content of 42% was determined.

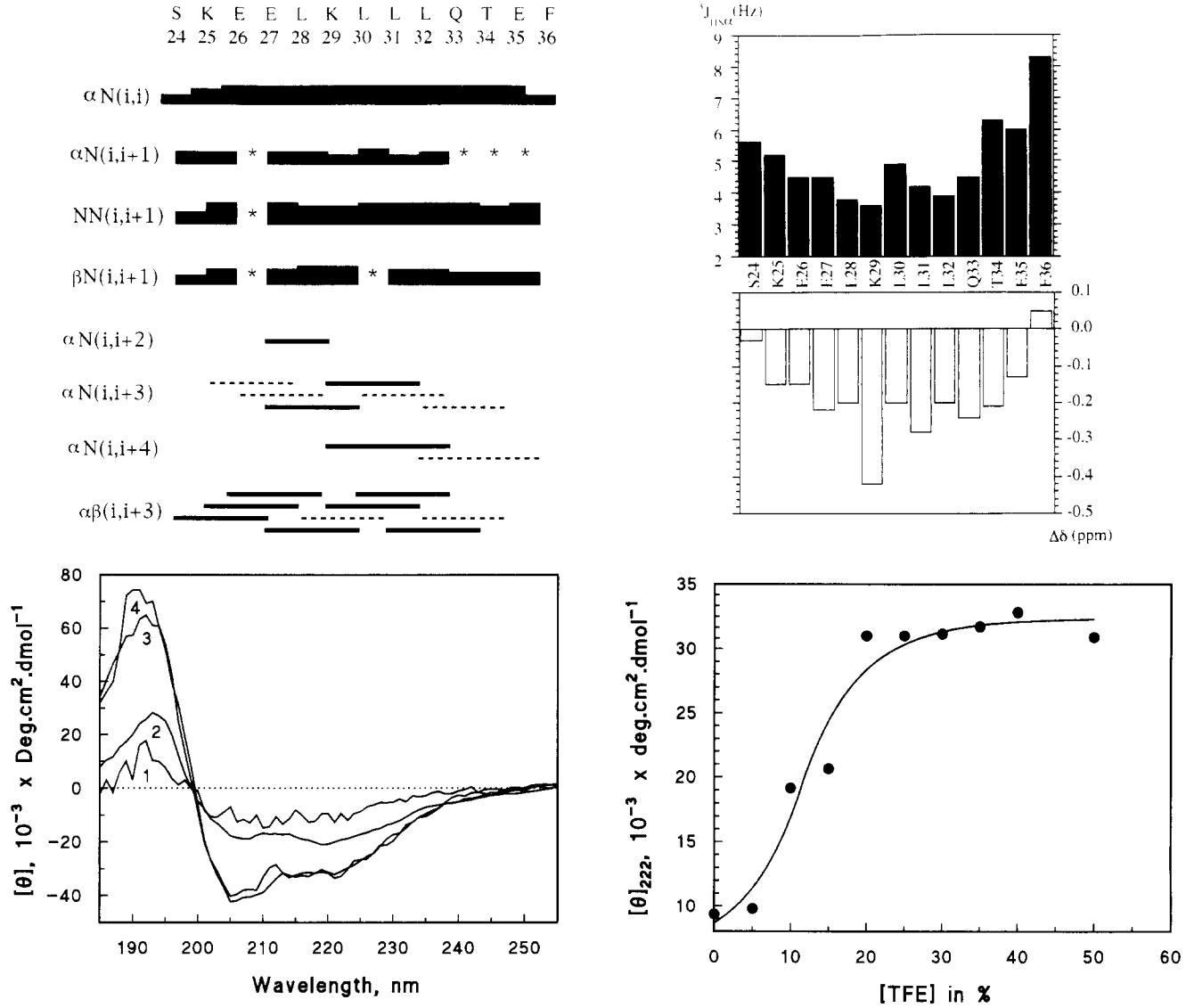



FIGURE 6: NMR and CD results for fragment ICB[24–36]. (a, upper left) Schematic representation of NOE connectivities. The thickness of the bar is proportional to the intensity of the NOE cross-peak; *, not observed because of peak overlap; —, observed medium-range NOE cross-peak; - - -, medium-range NOE cross-peak overlapping with intraresidue or sequential cross-peaks. (b, upper right) Schematic representation of $^3J_{H\alpha N}$ values (filled bars) and α-proton chemical shift deviations from random coil values (empty bars). Experimental conditions: ~2 mM peptide in TFE 40%, pH 4.0, 298 K, 300 ms mixing time for NOESY spectra. (c, lower left) Representative far-UV CD spectra at different TFE concentrations. The spectra labeled as 1, 2, 3, and 4 correspond respectively to 0%, 10%, 20%, and 40% TFE peptide solutions. (d, lower right) Mean residue ellipticity values at 222 nm for different TFE concentrations. Experimental conditions: 29 μM peptide, pH 4.0, 298 K.

## DISCUSSION

### *1. Level of Agreement between Experiments and Predictions*

The computational method used here has the advantage of predicting not only the most probable conformation but a whole set of low energy structures, ranked in order of increasing energy. It is thus possible to compare not only the predicted lowest energy structure with that observed in solution but also a range of low-energy conformations predicted for each individual residue with the ($\phi$, $\psi$) space sampled by the same residue in solution.

For C2C[63–73], NMR data show that the helix extends from the first residue (T63) to K72, with the last residue (N73) having a somewhat larger $^3J_{H\alpha N}$. This is in agreement with the predictions, and with threonine and lysine being good helix N- and C-caps, respectively (Richardson & Richardson, 1988).

For C2C[97–112], some variability in the predicted conformation of the first four residues (K97, D98, D99, and E100) and of L111 is seen in the lowest energy structures. In the N-terminal region of the peptide, this variability is mostly limited to a switch between the A (α-helix) and C ($3_{10}$-helix) conformations. NMR data confirm this view, suggesting that the N- and C-terminal parts of the peptide are more flexible, as deduced from the larger $^3J_{H\alpha N}$ values, the $\Delta\delta_{H\alpha}$, and the relative intensities of the $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOE cross-peaks. This is consistent with the fact that D98 or D99 could act as good N-caps for C2C[97–112], whereas K109 or K112 could act as the C-cap.

In ICB[24–26], as in C2C[63–73], the prediction results show that the N-terminal residue (S24) is free to adopt either a helical or extended conformation and that a large disruption of the helical structure occurs when the residues in the central portion of the fragment adopt an extended conformation. The slightly negative value of $\Delta\delta_{H\alpha}$ observed for S24 in the N-terminal region, the slightly positive $\Delta\delta_{H\alpha}$, and the large $^3J_{H\alpha N}$ values observed for F36 in the C-terminal region, are consistent with this view.

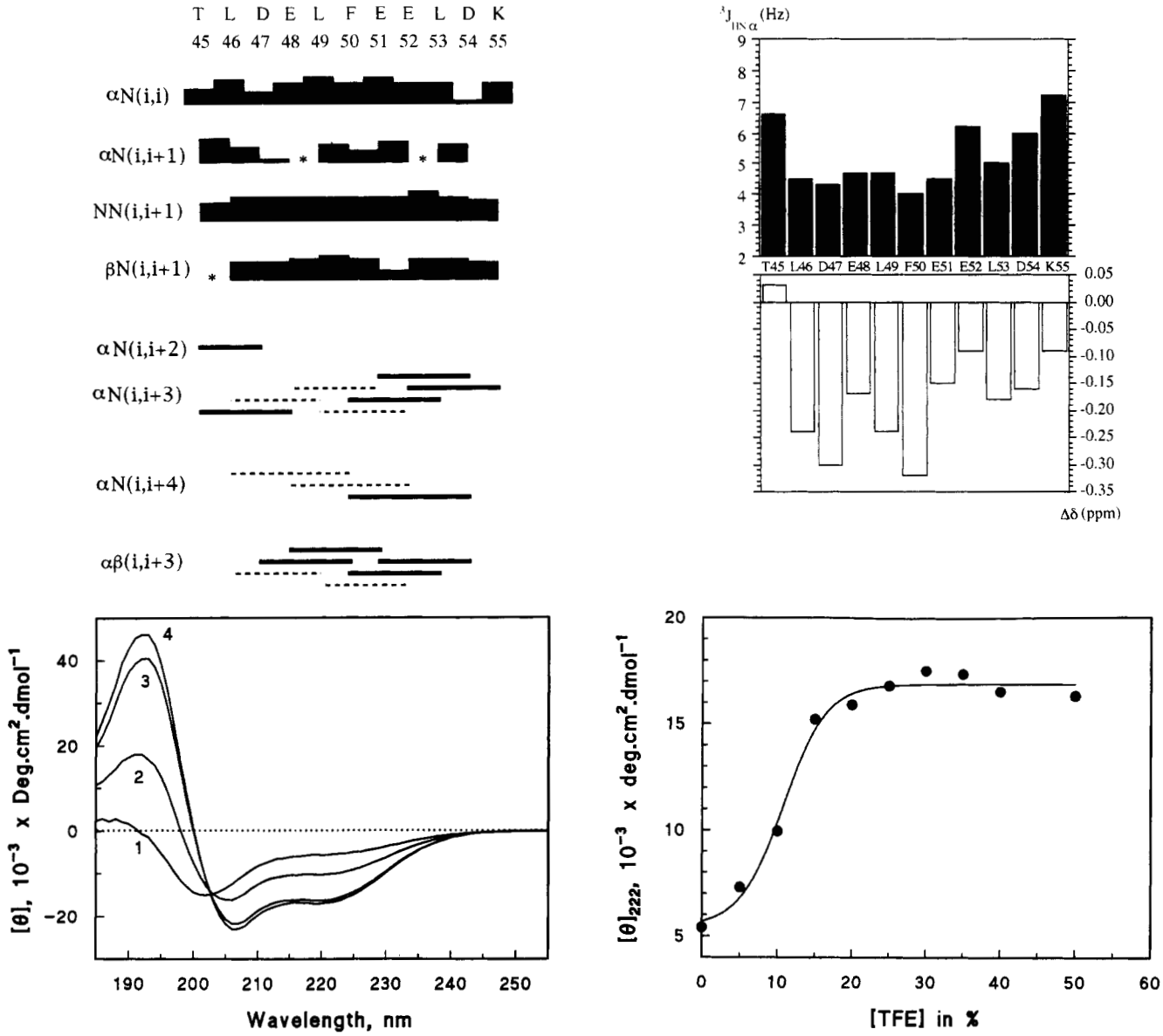



FIGURE 7: NMR and CD results for fragment ICB[45–55]. (a, upper left) Schematic representation of NOE connectivities. The thickness of the bar is proportional to the intensity of the NOE cross-peak; *, not observed because of peak overlap; —, observed medium-range NOE cross-peak; - - -, medium-range NOE cross-peak overlapping with intraresidue or sequential cross-peaks. (b, upper right) Schematic representation of $^3J_{H\alpha N}$ values (filled bars) and $\alpha$-proton chemical shift deviations from random coil values (empty bars). Experimental conditions: ~2 mM peptide in TFE 40%, pH 4.0, 298 K, 300 ms mixing time for NOESY spectra. (c, lower left) Representative far-UV CD spectra at different TFE concentrations. The spectra labeled as 1, 2, 3, and 4 correspond respectively to 0%, 10%, 20%, and 40% TFE peptide solutions. (d, lower right) mean residue ellipticity values at 222 nm for different TFE concentrations. Experimental conditions: 308 $\mu$M peptide, pH 4.0, 298 K.

In ICB[45–55], the two lowest energy structures have their N-terminal residue (T45) in either helical or extended conformation. A significant rms deviation from the fully helical structure is computed only when the first two residues adopt an extended conformation. NMR data partially confirm this view. The $\Delta\delta_{H\alpha}$ of T45 is slightly positive, thus reflecting a higher conformational flexibility, whereas the second residue in the sequence (L46) seems to adopt a fully helical conformation. Both T45 and D47 could act as good N-caps, whereas K55, which was added to the original fragment to provide a good C-cap, is seen to display a higher value of $^3J_{H\alpha N}$.

We can thus conclude that the prediction method proved to be useful not only in identifying short protein segments that are likely to adopt a preferred conformation but also, to a good extent, in delimiting within a fragment the regions with well-defined conformation and those which are more flexible.

As far as the helix limits are concerned, we observed a good correspondence between the limits predicted by our method, those observed in the crystal structure of the native proteins [by the DSSP criteria (Kabsch & Sander, 1983)], and the helix limits determined in the corresponding peptides in solution, as defined on the basis of medium-range NOE connectivities, $^3J_{H\alpha N}$ values, and $\Delta\delta_{H\alpha}$ (Figure 8). In the case of C2C[97–112] and ICB[45–55], the experimentally determined conformation in solution was closer to the predicted conformation than to that observed in the crystal structure. This is not surprising, as the prediction method takes into account only local interactions along the sequence. Small differences between the helix limits in the crystal structure and those predicted or found in the peptides in solution could arise either from specific polar interactions (H-bonds or salt bridges) of the helix with other parts of the protein or from nonspecific packing interactions that are absent in solution. As no specific interactions involving these fragments were

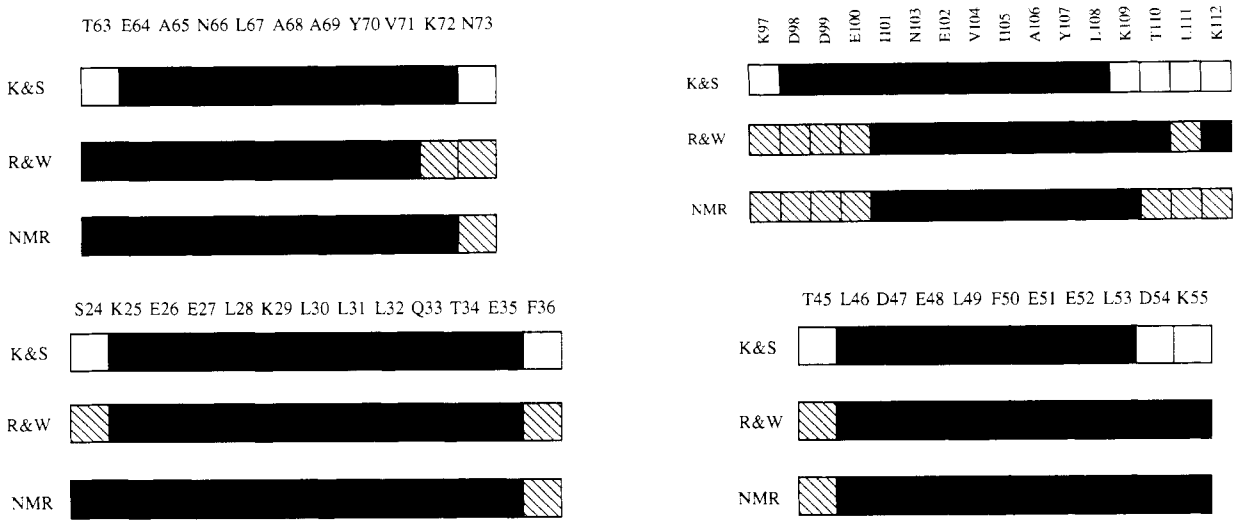



FIGURE 8: Helix length as determined from the crystal structure (RX) using the DSSP algorithm (Kabsch & Sander, 1983), as predicted (PRED.), and as observed experimentally in TFE 40%, 298 K, by NMR (NMR). Empty boxes, not helical; black boxes, helical; hatched boxes, helical, but displaying some flexibility. The relative intensity of the $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ NOE cross-peaks, the presence or absence of medium-range NOEs, the $^3J_{H\alpha N}$ values, and the $\alpha$-proton chemical shift deviations from random coil values were used to define the $\alpha$-helical regions and the more flexible ones.

observed in the native protein structures, the second hypothesis seems to be more plausible, if abstraction is made of factors influencing the experimental measures (see below).

In addition to the very good agreement between the experiments and predictions described above, we find that there is a linear correlation between the minimum energy value computed for each peptide (Figure 2) and its $\alpha$-helical content in 40% TFE, as determined from the CD and NMR data (Table 2). The corresponding correlation coefficients are 0.97, 0.94, and 0.93 when the helix content is determined from CD data using the databases of Yang et al. (1986), and Brahms and Brahms (1980) and the mean residue ellipticities, respectively. Somewhat poorer correlation coefficients of 0.85 and 0.77 were obtained when the helix content is determined from the $J$ coupling constants and chemical shift changes, respectively. This is an intriguing result because, in principle, energy values computed by our prediction methods for peptides with different amino acid sequences cannot be compared, since the expressions used to derive them lack the partition function term necessary for normalization (Rooman et al., 1991; Kocher et al., 1994). Furthermore, the experimental conditions in which the helical content of the peptides was measured are rather arbitrary. Analysis on a larger number of peptides is in progress to verify the present finding.

Finally, it should be added that the same prediction method has previously been applied rather successfully to identify protein segments that have a preferred extended or helical conformation, either in early folding intermediates or when excised from the chain (Rooman & Wodak, 1992) or to peptides that do not have a well-defined conformation in solution (Rooman et al., 1991), predicting for the latter many different conformations with similar energies.

### *2. Factors Influencing the Experiments and Their Interpretation*

One of the major aims of this study has been to provide a direct experimental assessment of the performance of theoretical structure prediction methods. For such an assessment to be objective, factors which may affect the experimental studies and the interpretation of their results must be evaluated. One is the use of TFE, a cosolvent whose influence on peptide conformation is not well understood. In some cases the peptide conformation observed in the presence of TFE was found to differ from that in the parent native protein structure (Zagorski & Barrow, 1992; Sönnichsen et al., 1992). However, a growing number of studies on protein fragments from all-$\alpha$, all-$\beta$ and $\alpha,\beta$ proteins suggest that TFE tends to stabilize the $\alpha$-helical conformation only when there is an intrinsic helical propensity in the amino acid sequence (Lehrman et al., 1990; Dyson et al., 1992b). Furthermore the length of the helical portion of peptides in water/TFE was found to correspond well with the helix length in the native protein structure (Nelson & Kallenbach, 1986). It cannot be ruled out, however, that TFE at the concentrations used here may somewhat affect helix lengths (Nelson & Kallenbach, 1989).

A second non-natural factor that may influence the helical propensity of our peptides is the presence of the blocking groups at the N- and C-termini. The succinyl group used to block the N-terminus was shown to enhance the helical content of short linear peptides in water (Shoemaker et al., 1987). This group can indeed mimic the N-capping action of an aspartic acid side chain, as its carboxyl moiety can form an H-bond with one of the free amide protons in the last turn of the helix. We see that the $^3J_{H\alpha N}$ value of the N-terminal residue is relatively small (5.6–6.6 Hz) in all four peptides. In particular, they are significantly smaller than those of the last residue of each peptide (7.2–8.3 Hz). These results are consistent with a reduced mobility of the amide group of the N-terminal residue.

A third factor may be the use of different pH conditions, some of which surround the p$K$ of carboxylate groups, which have been shown to have a significant sequence specific effect on helix formation in short peptides (Shoemaker et al., 1985, 1987). In the case of C2C[63–73], a particularly low pH was used for solubility reasons, most probably decreasing the measured helical content as a result of protonating E64. The effect of pH on C2C[97–112] and ICB[24–36] is more difficult to assess as these peptides contain several ionizable side chains. This also applies to ICB[45–55], except that here one would expect a salt bridge to form between E51 and K55 at higher pH's, thereby stabilizing the helix.

Aside from the factors that may influence the experiments, one must be aware of the limitations in our ability to extract quantitative information from CD or NMR data on the helix content of short peptides. Commonly, several methods are used for this purpose. We preferred to use the deconvolution

of the far-UV CD spectra in terms of contributions from individual secondary structures (Table 2) rather than the ellipticity value at 222 nm, which is probably more affected by the presence of nonhelical conformations and by the contribution of aromatic residues. The choice of reference values for these contributions may however be a source of error. Indeed, Table 2 shows that the $\alpha$-helical contents calculated using the polypeptide database of Brahms and Brahms (1980) are generally higher than those calculated using the 18 protein database of Yang et al. (1986). In addition we used the $^3J_{H\alpha N}$ coupling constants and $\alpha$-proton chemical shifts to evaluate the average $\alpha$-helical content in 40% TFE. The latter two approaches can be considered rather crude, due to the lack of reliable reference values and to the uncertainty associated with the measurement of coupling constants. Nevertheless a fairly good agreement was found between the helix content determined from CD spectra decomposition [using the protein database of Yang et al. (1986) with parametrization of the actual peptide length] and from $^3J_{H\alpha N}$ values. However, using the chemical shift values, we found an agreement between CD and NMR results only in two of the four peptides studied. It is worth noting here that the $\delta_{H\alpha}$'s are influenced not only by the neighboring NH and CO groups that determine the observed average shift of $\alpha$-protons in an $\alpha$-helix but also by the presence of aromatic rings, carboxyl groups, and charges on side chains of other residues.

With the limitations cited above, both the NMR and the CD data support the conclusion that the peptides studied here have a significant $\alpha$-helical propensity, which is displayed at room temperature and at relatively low TFE concentrations (20–30%). These results are in very good general agreement with the theoretical predictions.

### 3. *Relevance for Protein Folding*

Identifying protein segments with a well-defined conformational preference in the absence of interactions with the rest of the chain has interesting applications. We have shown here that these segments, when excised from their native protein, could yield peptides that adopt a well-defined conformation in solution. Such peptides could be used as immunogens to raise protein-reactive antibodies and be instrumental in the design of synthetic vaccines (Dyson et al., 1988a). In addition, due to their local conformational robustness, such segments are likely to adopt a well-defined structure early during the folding process, thereby reducing the degrees of freedom of the polypeptide backbone (Rooman et al., 1991, 1992). They could furthermore act as initiation sites for the folding (Wright et al., 1988).

While the folding pathway of *Rhodospirillum* cytochrome $c_2$ has not been analyzed as yet, a large body of experimental data is available on structurally related horse cytochrome *c*. Its central helices were shown to fold on a slower time scale than the N- and C-terminal helices by quenched-flow NMR experiments (Roder et al., 1988). CD measurements in water at 5 °C yielded an $\alpha$-helical content of 27%, 7%, and 3% for the peptides corresponding to the C-terminal, middle, and N-terminal helices of cytochrome *c*, respectively (Kuroda, 1993). In another study, an $\alpha$-helical content of only 15% was determined by CD for a synthetic peptide closely homologous to the C-terminal helix (residues 87–104), but the helical content was shown to increase upon complex formation of this peptide with a heme-containing N-terminal fragment of cytochrome *c* (Wu et al., 1993). Notwithstanding these small discrepancies, the experiments on cytochrome *c* all reach the conclusion that the C-terminal helix is relatively stable in solution even in absence of native tertiary interactions, whereas the middle and N-terminal fragments are only marginally helical in absence of these interactions.

Theoretical analyses made previously using our prediction method (Rooman & Wodak, 1992) independently reached very similar conclusions, suggesting furthermore that they could be extended to other members of the cytochrome *c* family. The C-terminal fragment was predicted as adopting a preferred $\alpha$-helical conformation in the large majority of the cytochrome *c*'s family members, the region corresponding to one of the middle helices (residue 71–75 in rice cytochrome *c* numbering, corresponding to helix 63–73 in *Rhodospirillum* cytochrome $c_2$) was predicted to have a preferred helical conformation in a smaller, yet appreciable number of members, whereas the N-terminal helix and the other central helices were never predicted as locally structured with a high enough score.

Our results show that both [63–73] and [97–112] cytochrome $c_2$ fragments have appreciable propensity for helical structure, with that of the C-terminal fragment being somewhat higher. This could mean that cytochrome $c_2$ may have a somewhat different folding pathway than cytochrome *c*, consistent with our previous proposal that variations in the main folding pathway could exist in more distant members of the same protein family (Rooman & Wodak, 1992).

The other two peptides analyzed correspond to $\alpha$-helical regions of bovine calcium binding protein. Despite the proximity of these helices to the calcium ions and the presence of several complex interactions between the helices and other parts of the protein, these two helices were found to have a significant stability as isolated fragments in water/TFE solutions at room temperature. Also, taking into account that among the four helices of calcium binding protein, only the two corresponding to the fragments analyzed here were predicted to have a preferred $\alpha$-helical conformation, it can be speculated that these particular helical regions play a role in the early stage of the folding of the intact protein even in the absence of calcium ions. The corresponding helices pack against each other in the native crystal structure. Their folding and association could thus produce a core around which the remainder of the polypeptide chain would fold. The $\alpha$-helical content calculated both by CD and NMR (from $^3J_{H\alpha N}$ values) is higher for ICB[24–36] than for ICB[45–55] and, in analogy with what is observed for C2C[63–73] and C2C[97–112], a differential role of the two segments in the folding process might be expected.

Experimental data on the folding of calcium binding protein are not available. However, folding and unfolding kinetics of carp parvalbumin, which is homologous in sequence to calcium binding protein, were studied by CD (Kuwajima et al., 1987b). A folding intermediate was found to accumulate in this protein on an 18 ms time scale. This intermediate contains 60–80% of the $\alpha$-helix content of the native protein, but lacks the structural requirements for strong calcium binding. Though caution should be taken when comparing the folding pathways of different proteins, these results are consistent with our hypothesis that the peptides ICB[24–36] and ICB[45–55] could fold and pack together in the early stages of folding. Stopped-flow NMR experiments would be necessary to obtain more detailed information about the folding kinetics of the individual helices in these proteins.

## ACKNOWLEDGMENT

We are grateful to K. Rose (University of Geneva) for mass spectrometry analysis.

## REFERENCES


Alexandrescu, A. T., Evans, P. A., Pitkeathly, M., Baum, J., & Dobson, C. M. (1993) Structure and dynamics of the acid-denatured molten globule state of α-lactalbumin: A two-dimensional NMR study, *Biochemistry 32*, 1707–1718.

Baldwin, R. L. (1990) Pieces of the folding puzzle, *Nature 346*, 409–410.

Bax, A. & Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, *J. Magn. Res. 65*, 355–360.

Bothner-By, A. A., Stephens, R. L., Lee, J. M., Warren, C. D., & Jeanloz, R. W. (1984) Structure determination of a tetrasaccharide:transient nuclear Overhauser effects in the rotating frame, *J. Am. Chem. Soc. 106*, 811–813.

Brahms, S., & Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism, *J. Mol. Biol. 138*, 149–178.

Buck, M., Radford, S. E., & Dobson, C. M. (1993) A partially folded state of hen egg white lysozyme in trifluoroethanol: structural characterization and implications for protein folding, *Biochemistry 32*, 669–678.

Bundi, A., & Wüthrich, K. (1979) $^1$H-NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH, *Biopolymers 18*, 285–297.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) Detection and characterization of a folding intermediate in barnase by NMR, *Nature 346*, 488–490.

Delhaise, P., Van Belle, D., Bardiaux, M., & Wodak, S. J. (1985) Analysis of data from computer simulations on macromolecules using the CERAM package, *J. Mol. Graph. 3*, 116–119.

Dyson, H. J., Lerner, R. A., & Wright, P. E. (1988a) The physical basis for induction of protein-reactive antipeptide antibodies, *Annu. Rev. Biophys. Chem. 17*, 305–324.

Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., & Lerner, R. A. (1988b) Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn, *J. Mol. Biol. 201*, 161–200.

Dyson, H. J., Rance, M., Houghten, R. A., Wright, P. E., & Lerner, R. A. (1988c) Folding of immunogenic peptide fragments of proteins in water solution. II. The nascent helix, *J. Mol. Biol. 201*, 201–217.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992a) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I. Myohemerythrin, *J. Mol. Biol. 226*, 795–817.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H., Lerner, R. A., & Wright, P. E. (1992b) Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin, *J. Mol. Biol. 226*, 819–835.

Efron, B. (1982) *The Jack Knife, the Bootstrap, and other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia.

Gaertner, H. F., Rose, K., Cotton, R., Timms, D., Camble, R., & Offord, R. E. (1992) Construction of protein analogues by site-specific condensation of unprotected fragments, *Bioconjugate Chem. 3*, 262–268.

Hughson, M. F., Wright, P. E., & Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate, *Science 249*, 1544–1548.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys. 71*, 4546–4553.

Jeng, M., & Englander, S. W. (1991) Stable submolecular folding units in a non-compact form of cytochrome c, *J. Mol. Biol. 221*, 1045–1061.

Jeng, M., Englander, S. W., Elöve, G., Wand, A. J., & Roder, H. (1990) Structural description of acid-denatured cytochrome c by hydrogen exchange and 2D NMR, *Biochemistry 29*, 10433–10437.

Jimenez, M. A., Bruix, M., Gonzalez, C., Blanco, F. J., Nieto, J. H., Herranz, J., & Rico, M. (1993) CD and $^1$H-NMR studies on the conformational properties of peptide fragments from the C-terminal domain of thermolysin, *Eur. J. Biochem. 211*, 569–581.

Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors, *Acta Crystallogr. Sect. A, 34*, 827–828.

Kabsch, W., & Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577–2637.

King, D., Fields, C. G., & Fields, G. B. (1990) A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis, *Int. J. Pept. Protein Res. 36*, 255–266.

Knolle, J., Breipol, Y., & Stüber, W. (1987) Preparation and application of new acid labile anchor groups for the synthesis of peptide amides by Fmoc solid phase synthesis, *Tetrahedron Lett. 28*, 5651–5654.

Knorr, R., Trezeciak, A., Bannwarth, W., & Gillesen, D. (1989) New coupling reagents in peptide chemistry, *Tetrahedron Lett. 30*, 1927–1930.

Kocher, J. P., Rooman, M. J., & Wodak, S. J. (1994) Factors influencing the ability of knowledge based potentials to identify native sequence-structure matches, *J. Mol. Biol. 235*, 1598–1613.

Kuroda, Y. (1993) *Residual* helical structure in the C-terminal fragment of cytochrome c, *Biochemistry 32*, 1219–1224.

Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., & Nagamura, T. (1987a) Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism, *FEBS Lett. 221*, 115–118.

Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M., & Sugai, S. (1987b) Folding of carp parvalbumin studied by equilibrium and kinetic circular dichroism spectra, *Biochemistry 27*, 7419–7428.

Lehrman, S. R., Tuls, J. L., & Lund, M. (1990) Peptide α-helicity in aqueous trifluoroethanol: correlations with predicted α-helicity and the secondary structure of the corresponding regions of bovine growth hormone, *Biochemistry 29*, 5590–5596.

Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering, *Nature 340*, 122–126.

Matthews, C. R. (1993) Pathways of protein folding, *Annu. Rev. Biochem. 62*, 653–683.

Nelson, J. W., & Kallenbach, N. R. (1986) Stabilization of the ribonuclease S-peptide α-helix by trifluoroethanol, *Proteins 1*, 211–217.

Nelson, J. W., & Kallenbach, N. R. (1989) Persistence of the α-helix stop signal in the S-peptide in trifluoroethanol solutions, *Biochemistry 28*, 5256–5261.

Oas, T. G., & Kim, P. S. (1988) A peptide model of a protein folding intermediate, *Nature 336*, 42–48.

Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature 358*, 302–307.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) Improved spectral resolution in COSY $^1$H-NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Commun. 117*, 479–485.

Richardson, J. S., & Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of α-helices, *Science 240*, 1648–1652.

Roder, H., Elöve, G. A., & Englander, W. S. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR, *Nature 335*, 700–704.

Rooman, M. J., & Wodak, S. J., (1991) Weak correlation between predictive power of individual sequence patterns and overall prediction accuracy in proteins, *Proteins 9*, 69–78.

Rooman, M. J., & Wodak, S. J. (1992) Extracting information on folding from the amino acid sequence: Consensus regions with preferred conformation in homologous proteins, *Biochemistry 31*, 10239–10249.

Rooman, M. J., Rodriguez, J., & Wodak, S. J. (1990) Relations between protein sequence and structure and their significance, *J. Mol. Biol. 213*, 337–350.

Rooman, M. J., Kocher, J. P., & Wodak, S. J. (1991) Prediction of protein backbone conformation based on seven structure assignements. Influence of local interactions, *J. Mol. Biol., 221*, 961–979.

Rooman, M. J., Kocher, J. P., & Wodak, S. J. (1992) Extracting information on folding from the amino acid sequence: Accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions, *Biochemistry 31*, 10226–10238.

Rose, G. D., & Wolfenden, R. (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding, *Annu. Rev. Biophys. Biomol. Struct. 22*, 381–415.

Salemme, F. R., Freer, S. T., Xuong, N. H., Alden, R. A., & Kraut, J. (1973) The structure of oxidized cytochrome $c_2$ of Rhodospirillum rubrum, *J. Biol. Chem. 248*, 3910–3921.

Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., & Baldwin, R. L. (1985) Nature of the charged group effect on the stability of the C-peptide helix, *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.

Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) Tests of the helix dipole model for stabilization of $\alpha$-helices, *Nature 326*, 563–567.

Sklenár, V., & Bax, A. (1987) Spin–echo water suppression for the generation of pure-phase two-dimensional NMR spectra, *J. Magn. Res. 74*, 469–479.

Sönnichsen, F. D., Van Eyk, J. E., Hodges, R. S., & Sykes, B. D. (1992) Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide, *Biochemistry 31*, 8790–8798.

Staley, J. P., & Kim, P. S. (1990) Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor, *Nature 344*, 685–688.

Szebenyi, D. M. E. & Moffat, K. (1986) The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine, *J. Biol. Chem. 261*, 8761–8777.

Udgaonkar, J. B., & Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A, *Nature 335*, 694–699.

Udgaonkar, J. B., & Baldwin, R. L. (1990) Early folding intermediate of ribonuclease A, *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197–8201.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *J. Mol. Biol. 222*, 311–333.

Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding, *Biochemistry 27*, 7167–7175.

Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J., & Roder, H. (1993) A noncovalent peptide complex as a model for an early folding intermediate of cytochrome c, *Biochemistry 32*, 10271–10276.

Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley & Sons, Inc., ed.

Yang, J. T., Wu, C., & Martinez, H. M. (1986) Calculation of protein conformation from circular dichroism, *Methods Enzymol. 130*, 208–269.

Zagorski, M. G., & Barrow, C. J. (1992) NMR studies of amyloid $\beta$-peptides: Proton assignments, secondary structure, and mechanism of an $\alpha$-helix$\rightarrow\beta$-sheet conversion for a homologous, 28-residue, N-terminal fragment, *Biochemistry 31*, 5621–5631.