J. Mol. Biol. (1995) **245**, 275–296

JMB

Elucidating the Folding Problem of Helical Peptides using Empirical Parameters. II†. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides

Victor Muñoz and Luis Serrano

EMBL, Meyerhofstrasse 1
Heidelberg D-69117
Germany

Explaining the helical behaviour of amino acid sequences in solution could be one of the first steps in solving the protein folding problem in a rational way. The information about the conformational behaviour of helical peptides in solution, as well as the α-helix stability in proteins, has been utilised to derive a database with the energy contributions for various interaction taking place in an α-helix: intrinsic helical propensities, side-chain to side-chain interactions, main-chain to main-chain hydrogen bonds, and capping effects. This database was implemented in a algorithm based on the helix-coil transition theory (AGADIR). Here, the effects on helix stability due to interactions between charged groups and the helix macrodipole are described, quantified and implemented in AGADIR. The algorithm correctly calculates the average helical behaviour in solution of 423 peptides analysed by circular dichroism and it describes the helicity at a residue level, as found when comparing the prediction for each amino acid residue with the data derived from nuclear magnetic resonance studies. Using AGADIR we have done a rational modification of peptides corresponding to protein secondary structure elements in order to increase their helical content. The circular dichroism analysis of the mutant peptides showed a very good agreement between the experimental and calculated helical content. Moreover, in certain specific cases in which strong tertiary contacts in folded proteins do not exist, the algorithm successfully predicts the length of mutagenised α-helices. It is interesting to note that the final values of the parameters used do not significantly differ in absolute terms from those extracted from mutagenesis studies in proteins. This indicates that the same physico-chemical principles stand for both systems.

Keywords: α-helix stability; protein folding; secondary structure prediction; biotechnology

Introduction

Protein folding is one of the most important problems in modern biochemistry remaining to be solved. It has been found that in the majority of cases proteins refold spontaneously *in vitro* to the native conformation, thus indicating that the three-dimensional information is contained in the linear sequence of amino acid residues. It also seems clear that the refolding polypeptide chain cannot explore all the available conformational space and, consequently, that one or more folding pathways should exist that will limit the conformational search (Anfinsen *et al*., 1961). Experimental analysis of protein folding has revealed that secondary structure is formed very early in the folding process, while tertiary structure is acquired later (Matthews, 1993). Moreover, it has been found in many cases that isolated peptides have a strong tendency to populate significantly the same conformation that they have in the native state (Dyson & Wright, 1993). It is then possible to suggest that one of the earliest steps in the folding process will be the formation of secondary structure elements, concomitant or not with a hydrophobic collapse (Serrano *et al*., 1992*a*). It follows that if we want to understand protein folding in a rational way, we need to know what the factors are that determine the tendency of short amino acid sequences to populate different conformational states in solution.

Abbreviations used: TFA, trifluoroacetic acid; EDT, ethane dithiol; TFE, trifluoroethanol.

† Papo I in this series is Muñoz & Serano (1994).

0022–2836/95/030275–22 $08.00/0

Currently, the attempts to describe the energetics of systems formed by short polypeptide chains have been addressed to α-helices. The model most often used for that purpose, the helix-coil transition, is based on the statistical mechanics theory and needs the assumption of certain simplifications. In its simplest version, as was postulated by Zimm & Bragg (1959), there were only two parameters, an elongation factor and a nucleation factor corresponding to equilibrium constants that are characteristic of each amino acid type responsible for the helical tendency of a particular sequence. Both parameters are independent of the sequence environment of the amino acid residue. That means side-chain to side-chain interactions do not participate in the α-helix stability of a given peptide. A great deal of work has been devoted to the experimental determination of nucleation and elongation factors for every type of amino acid (von Dreele *et al*., 1971), hoping that once parameterised, a complete picture of the helix-coil transition would arise. Later on the theory was modified in order to incorporate data about side-chain to side-chain interactions (Finkelstein *et al*., 1991) as well as capping effects (Chakrabarty *et al*., 1993). These modifications, although they improve considerably the predictive power of the helix-coil transition theory, still fell short of explaining accurately the helical behaviour of complex heteropolypeptides in solution (Gans *et al*., 1991; Chen *et al*., 1992; Quian, 1993; Fukugita *et al*., 1993; Park *et al*., 1993).

In a previous work we described an algorithm (AGADIR) that, based on the helix-coil theory and using experimentally derived parameters, was able to predict the average helical behaviour of 323 peptides in aqueous solution at low temperatures (0 to 4°C), ionic strengths (0 to 100 mM), and in the pH range 6 to 7. This algorithm was able to describe the helical behaviour of peptides in solution at a residue level (Muñoz & Serrano, 1994, 1995). Here, we provide a more detailed explanation of the method and parameters used, as well as a more realistic description of the interactions between charged residues and the helix macrodipole, previously included in the capping effect. This results in a slight modification of the values for some of the parameters previously described (Muñoz & Serrano, 1994, 1995), and improves the predictive power of the algorithm.

Results

Variation of the helical content with the length of the peptide

In Figure 1, we show the far-UV CD spectra of three peptides having one (Figure 1C), two (Figure 1B) and four repeats (Figure 1A) of the unit $(AAQAA)_n$. In all cases two different peptides with the N terminus, acetylated or not, were analysed. The helical content of both five-residue peptides is zero with respect to the ellipticity at 222 nm, although if we look at the N-blocked peptide there is a slight difference involving a displacement of the minimum

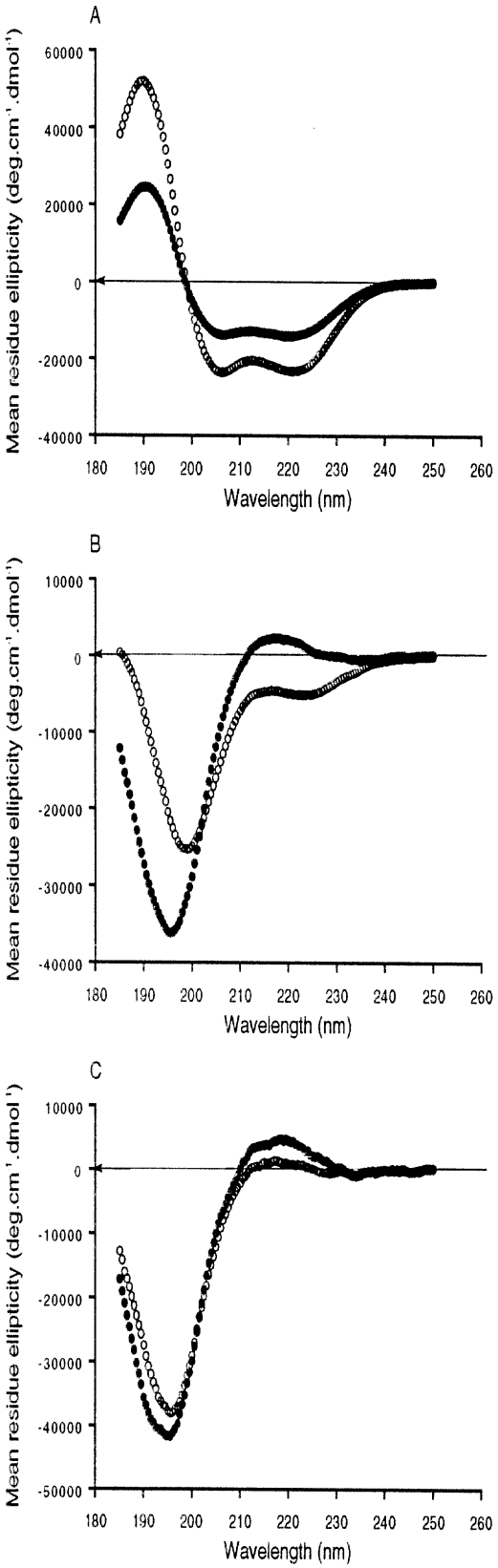

Figure 1. Far-UV CD analysis of peptides of different length based on different repeats of the unit AAQAA. The spectra were acquired at 5° C, in 2.5 mM phosphate buffer (pH 7.0), peptide concentration was 25 mM. A, Four repeats; B, 2 repeats; C, 1 repeat. (○) N terminus acetylated and C terminus amidated; (●) N terminus free and C terminus amidated.

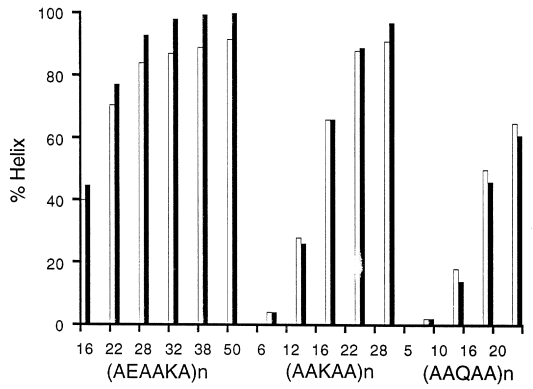

Figure 2. Variation of the helical content of polyalanine-based peptides with the length of the peptides. The open columns represent the experimentally determined helical content calculated using the method of Chen *et al.* (1974), and the filled columns represent the calculated numbers with AGADIR. The type of repeating unit is shown on the *x*-axis.

towards higher wavelengths (Figure 1C). The ten-residue peptide has a significant helical content when the N terminus is blocked ($\sim$18%), and it looks like a random coil when it is unprotected (Figure 1B). The CD spectrum of the N terminus unblocked ten-residue peptide is almost identical with that of the N-blocked five-residue peptide, suggesting that in both cases a very small helical content could be present. Finally, the 20-residue peptide presented a significant degree of helical content in both the acetylated and N-terminal unblocked peptides (35% and 65%, respectively: Figure 1A).

In Figure 2 we show the length-dependence of the helical content for different N and C-terminal blocked polyalanine-based peptides; $(AAKAA)_n$ (Rohl *et al*., 1992), $(AEAAKA)_n$ (Scholtz *et al*., 1991*c*) and $(AAQAA)_n$ (analysed here with the exception of the 16-residue peptide, which has been analysed by Scholtz *et al*., 1991*b*). The first two sets of peptides were used for the parameter refinement, while the last set was kept apart to finally test the refined algorithm. The predicted values are shown in the same Figure. It is clear that within the experimental error, the calculated values are very close to the experimental values, indicating that the basic parameters and assumptions are correct. The main differences between the predicted and experimental values are with the peptides based on the AEAAKA repeat. In this case, for all of the different lengths we are predicting more helix that is found experimentally. However, we must take into account that there is a Tyr residue at the beginning of the sequence, and a Phe residue at the end. It has been demonstrated that in some polyalanine-based peptides the presence of an aromatic residue, not separated by glycine residues from the rest of the peptide, results in the appearance of a positive band around 222 nm (Chakrabartty *et al*., 1993*a*). This band provokes an underestimation of the helical content. Consequently, it is possible that this difference could be due to the presence of this band. The $(AAQAA)_n$ peptides were synthesised without a Tyr residue (with the exception of that with 16 residues, Scholtz *et al*., 1991*b*). In the case of the $(AEAAKA)_n$ peptides, the helical content was calculated directly from the hydrogen exchange-rates and consequently they are not affected by the presence of a Tyr residue (Rohl *et al*., 1992). In any case there is also a slight overprediction of the helical content for the longest peptides, which could be due to the limits of the helix-coil assumption and the simplicity of the phase-space definition as explained in Materials and Methods, Theoretical procedures, Theory.

Separation of the capping effect from the helix dipole

The interaction of charged residues with the helix macrodipole has been separated from the capping effect and described in a more realistic fashion. In Figure 3, we show the experimental variation of the average helical content of polyalanine-based peptides, at pH 7.0 (2.5 in the case of the Ala to His substitutions), when differently charged residues, Asp^- (Figure 3A), Glu^- (Figure 3C) and His^+ (Figure 3E), are located at different positions in a peptide (Armstrong & Baldwin, 1993; Huyghues-Despointes *et al*., 1993*b*; Scholtz *et al*., 1993). Comparison of the calculated values for these peptides using the previous version of our algorithm, in which the interaction of charged residues with the helix dipole was included in the capping effect (Muñoz & Serrano, 1994, 1995), results in an overprediction for the peptides having a Glu residue close to the C terminus or His close to the N terminus (data not shown). Introduction of a more realistic model for the helix dipole interactions with charged residues (see Materials and Methods, Theoretical procedures, Theory), results in a very good prediction for the helical behaviour of these peptides when the side-chain groups are charged (Figure 3A, C and E), as well as when they are neutral (Figure 3B, D and F). The introduction of this refined model results in the modification of some of the values of the other parameters (hydrogen bond, capping effects, etc.), mainly because they were calibrated with polyalanine-based peptides containing several charged residues (Muñoz & Serrano, 1994, 1995). The parameters are now likely to be closer to the real ones and the intrinsic helical tendencies and capping interactions are purer since they do not include part of the energy term arising from the interaction with the helix macrodipole.

Average helical prediction

A total of 423 peptides, including those described in this work, were used in the analysis. From these, 135 peptides correspond to wild-type and modified protein fragments (Blanco *et al*., 1992; Bruch *et al*., 1991; Burke *et al*., 1991; Cushman *et al*., 1992; Doughty & Hu, 1993; Dyson *et al*., 1993*a*,*b*; Goodman & Kim,

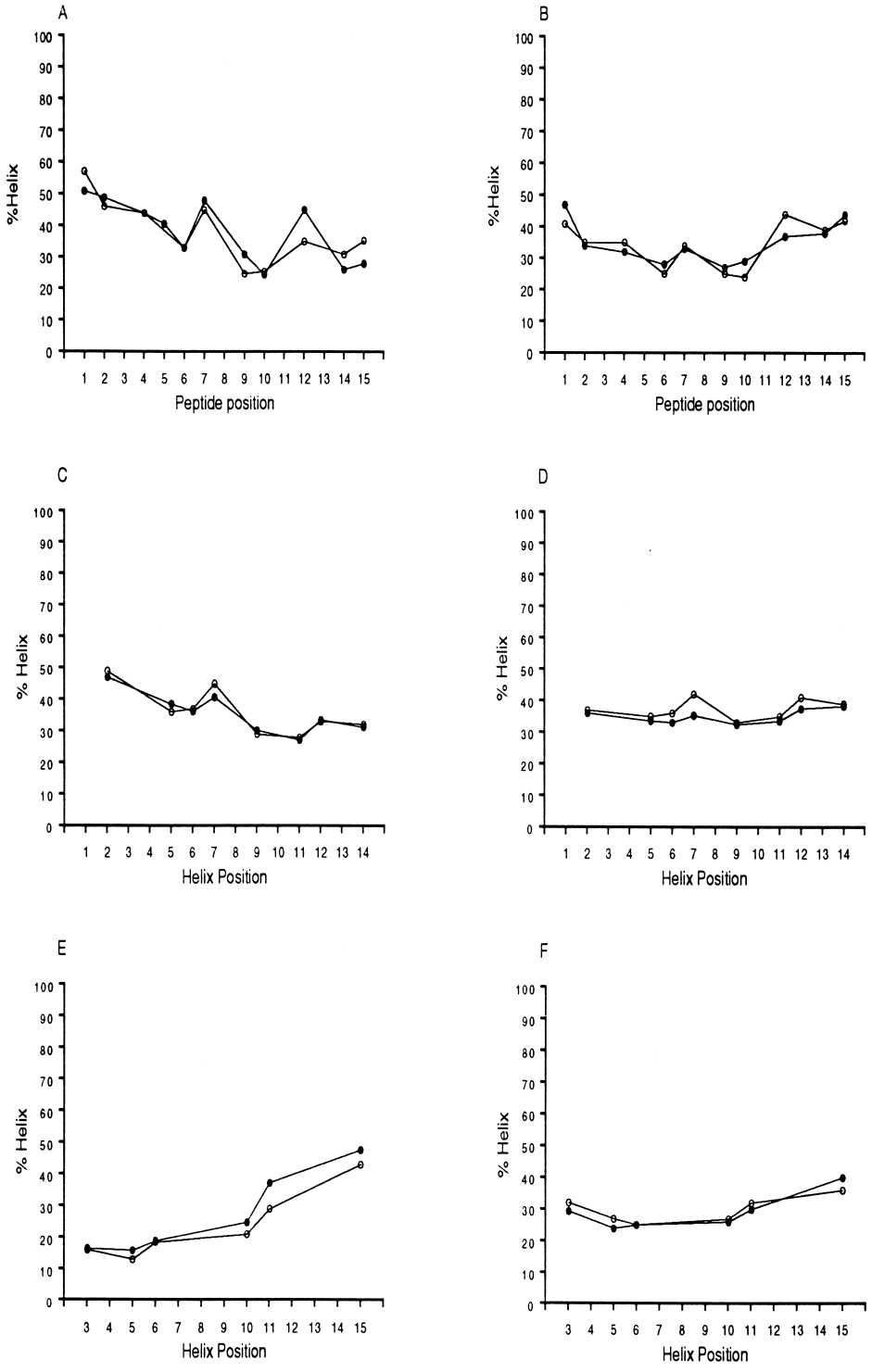

Figure 3. Interaction of charged groups with the helix macrodipole. A, Variation of the helical content of a polyalanine-based peptide when negatively charged Asp residues (pH 7.0) are placed at different peptide positions. (○) Calculated values with AGADIR; (●) experimental values from the work of Huyghues-Despointes *et al.* (1993*b*). B, The same as A but at pH 2.5. C, Variation of the helical content of a polyalanine-based peptide when negatively charged Glu residues (pH 7.0) are placed at different peptide positions. (○) Calculated values with AGADIR; (●) experimental values from the work of Scholtz *et al.* (1993). D, The same as C but at pH 2.5. E, Variation of the helical content of a polyalanine-based peptide when positively charged His residues (pH 2.5) are placed at different peptide positions in a peptide based on the repeat AAQAA. (○) Calculated values with AGADIR; (●) experimental values from the work of Armstrong & Baldwin (1993). F, The same as at E but pH 10.

1989; Greenfield *et al*., 1994; Jarvis *et al*., 1994; Jimenez *et al*., 1987, 1988, 1992, 1993, 1994; Kemmink & Creighton, 1993; Kuroda, 1993; Mammi *et al*., 1992; Maulet & Cox, 1983; McDowell *et al*., 1985; Morii *et al*., 1994; Morton *et al*., 1994; Munier *et al*., 1993; Muñoz & Serrano, 1994, 1995; Musco *et al*., unpublished results; Peña *et al*., 1989; Pintar *et al*., 1994; Precheur *et al*., 1991; Reed *et al*., 1992; Sancho *et al*., 1992; Shin *et al*., 1993*a*,*b*; Siligard *et al*., 1991; Terzi *et al*., 1994; Waltho *et al*., 1993; Yumoto *et al*., 1993; Zhang *et al*., 1993; this work), while the other 288 peptides correspond to designed sequences (Armstrong & Baldwin, 1993; Armstrong *et al*., 1993; Bradley *et al*., 1990; Chakrabartty *et al*., 1993*b*, 1994; DeGrado *et al*., 1985; Fairman *et al*., 1989, 1991; Forood *et al*., 1993; Ghadiri & Chong, 1990; Huyghues-Despointes *et al*., 1993*a*,*b*; Li & Deber, 1993; Lyu *et al*., 1989, 1990, 1991, 1992, 1993; Marqusee *et al*., 1989; Merutka & Stellwagen, 1990, 1991; Merutka *et al*., 1990, 1994; Padmanabhan *et al*., 1990; Padmanabhan & Baldwin, 1991; Rohl *et al*., 1992; Scholtz *et al*., 1991*a*,*b*,*c*, 1993; Shoemaker *et al*., 1987, 1990; Stellwagen *et al*., 1992; Strehlow *et al*., 1991; Venkatachallapathi *et al*., 1993; Zhou *et al*., 1993, 1994*a*,*b*; this work).

Figure 4A illustrates the correlation between the helical population for the 423 different peptides estimated from CD (the percentage helical population was calculated using the method of Chen *et al*., 1974), and the calculated average helical content. The overall correlation coefficient is extremely good ($r = 0.97$), especially if we consider the inherent error in any CD measurement due to concentration determination problems, the presence of aromatic residues, or the presence of alternative secondary structure conformations with different spectral properties. The slope is almost 1.0 (0.96 ± 0.01) and the line intersects the origin of the x and y axes (0.8 ± 0.49).

In Figure 4B, we show the correlation between the calculated values for those peptides having $i, i + 3$ and $i, i + 4$ electrostatic interactions, and the experimental values (Huyghues-Despointes *et al*., 1993*a*,*b*; Marqusee *et al*., 1989; Merutka & Stellwagen, 1991; Scholtz *et al*., 1993). Figure 4C shows the correlation analysis between the calculated values for those peptides specifically designed to analyse the contribution of capping interactions, and the experimental results (Forood *et al*., 1993; Chakrabartty *et al*., 1993*a*; Lyu *et al*., 1993). The correlation between the calculated values for those peptides specifically designed to analyse the intrinsic contribution of specific residues and the experimental data are shown in Figure 4D (Chakrabartty *et al*., 1991, 1994; Lyu *et al*., 1990, 1991; Merutka & Stellwagen, 1990; Padmanahban & Baldwin, 1991; Strehlow *et al*., 1991; Zhou *et al*., 1993). In all cases, the slopes, intersection of the origin and correlation coefficients are very similar to the values found in the general correlation.

Figure 4E shows the correlation between the calculated and observed helical content of peptides corresponding to protein fragments (135 peptides). The correlation coefficient ($r = 0.93$), the slope (0.88 ± 0.04) and intersection of the origin (1.9 ± 0.4) are similar to the designed peptides. This is remarkable since it indicates that the behaviour of peptides with more complicated sequences than those specifically designed to have a high α-helical content might also be determined with this algorithm and parameters. The dispersion of the points is similar to the overall correlation, and logically it is higher for helical populations lower than 15%, due to higher relative errors in the determination of the helical content. It is clear that the method we have developed with the refined parameters is able to calculate from the sequence the helical behaviour of all these peptides, within experimental error.

To show that the success of our method is not due to an overfitting of the initial parameters to the peptide database used to refine them, it is necessary to carry out a blind test. For this purpose we set apart a number of peptides, which has been extended from that presented by Muñoz & Serrano (1994, 1995), corresponding mainly to protein fragments (57 peptides: Burke *et al*., 1991; Bruch *et al*., 1991; Doughty & Hu, 1993; Greenfield *et al*., 1991; Kuroda, 1993; Mammi *et al*., 1992; Maulet & Cox, 1983; McDowell *et al*., 1985; Morii *et al*., 1994; Munier *et al*., 1993; Muñoz & Serrano, 1994, 1995; Pintar *et al*., 1994; Precheur *et al*., 1991; Reed *et al*., 1992; Sancho *et al*., 1992; Siligard *et al*., 1991; Terzi *et al*., 1994; Yumoto *et al*., 1993; Zhang *et al*., 1993; Musco *et al*., unpublished results) but also designed peptides (56 peptides: DeGrado *et al*., 1985; Fairman *et al*., 1989; Ghadiri & Chong, 1990; Chakrabartty *et al*., 1994; Li & Deber, 1993; Scholtz *et al*., 1991*b*; Zhou *et al*., 1994*b*; the $(AAQAA)_n$ peptides described above). These peptides were obviously not analysed until the method was finally adjusted. The analysis of these peptides is shown in Figure 4F. The correlation between the calculated and experimental values for the sequences derived from protein fragments is as good as that from the designed peptides as well as that in which we included all the peptides. This indicates that the refinement of the parameters did not result in an overfitting of some of the parameters.

Precision of the calculated values

From a practical point of view it is interesting to know what is the precision of the calculated helical values for a given sequence. Before discussing this point, it is worth mentioning that in the case of CD, the presence of aromatic residues (Chakrabartty *et al*., 1993*a*), the determination of the peptide concentration (Gill & von Hippel, 1989) and the possible coexistence of other secondary structure elements in the peptides, introduces an error, which in the case of the peptide concentration could be at least 10%. Consequently, it is impossible to attain a precision in the calculation of the helical content of a peptide, better than ~10%, being optimistic. Statistical analysis of the differences between the calculated and experimental helical values for the

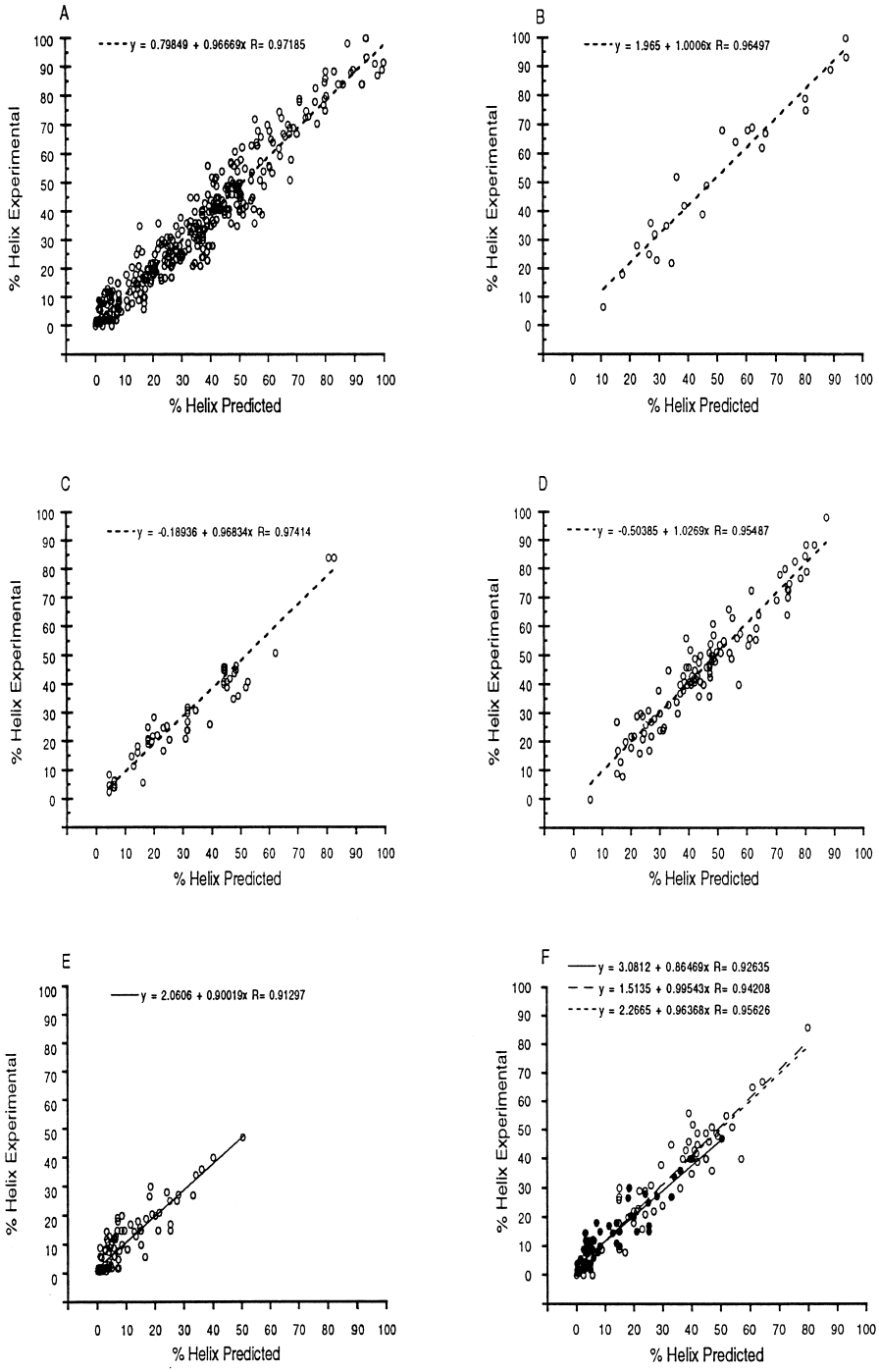

Figure 4. Correlation plots between the experimental CD average helical content (determined using the method of Chen *et al.*, 1974), and the calculated values using our algorithm. All of the peptides used in our database have been analysed at less than 0.1 M ionic strength, with the exception of some protected peptides without any *i*, *i* + 1, or *i*, *i* + 3 or *i*, *i* + 4 electrostatic interactions (Chakrabartty *et al.*, 1991, 1994). Some of the peptides have been analysed at temperatures higher than 4°C, and at very low or high values; see accompanying paper (Muñoz & Serrano, 1995). A, Correlation with the total peptide database. B, Correlation with those peptides designed to analyse electrostatic interactions. C, Analysis of those peptides designed to study capping interactions. D, Correlation with the peptides designed to study the intrinsic helical contribution of the different amino acids. E, Analysis of those peptides corresponding to wild-type, or modified protein fragments. F, Correlation analysis of those peptides not used in the refinement of the method (10 peptides corresponding to the α-helices of the Ara and Com proteins (Stock *et al.*, 1990), V. Muñoz, F. J. Blanco & L. Serrano (unpublished results); 3 peptides corresponding to the G-domain of protein G (Gronenborg *et al.*, 1991), F. J. Blanco & L. Serrano (unpublished results); and 4 peptides corresponding to the peptide CIII of 1 bacteriophage. J. McCarthy (personal communication)). (●) Wild-type and mutant peptides; (○) all the peptides.

423 peptides of the database, gives an overall standard deviation value of 5.86. The standard deviation value is lower for those peptides having less than 20% helical content (3 for 0 to 10% helical content; 5 for those having between 10 and 20%). This means that in the majority of the cases, we can calculate the helical content of a peptide with an error of $\pm 6\%$, and in 99% of the cases with an error of $\pm 12\%$.

Prediction of the helicity per residue and determination of helix limits

Our model also attempts to calculate the helical behaviour of individual amino acid residues in a polypeptide chain, since the calculation of the relative tendencies of the residues in a sequence, arising from short-range interactions, may be viewed as a first step in the solution of the protein folding problem. Nuclear magnetic resonance (NMR) analysis can provide the experimental information, at a residue level, needed to test our results. A good indicator of the presence of helix structure in peptides is provided by the upfield conformational $C^{\alpha}H$ $\Delta\delta$ values, which plotted in front of the sequence often showed a 3 to 4 periodicity in their minima for those regions populating the α-helical conformation (Jimenez *et al*., 1992). The $C^{\alpha}H$ protons belonging to N-terminal, C-terminal and Pro-preceding residues have specific position $\Delta\delta$ values, unrelated to the secondary structure, and are not included in the plots. We have used these vaues to compare them with the prediction of helical content at a residue level. In Figure 5A we show the good correlation between the NMR analysis in water solution and the helicity per residue calculated with our method, for the C-terminal 19 residues of murine interleukin-6 (Morton *et al*., 1994) (Figure 5C) a peptide corresponding to the amino acid region 71 to 93 of the plasma protein transthyretin (Jarvis *et al*., 1994), (Figure 5B) a peptide corresponding to the α-helical region 63 to 73 of *Rhodospirillum* cytochrome c_2 (Pintar *et al*., 1994) (Figure 5D) for single amino acid type region of titin (residues 1977 to 2014; Musco *et al*., unpublished results), and (Figure 5E) for a peptide corresponding to the 532 to 565 region of adenylate cyclase from *Baiallus antracis* (Munier *et al*., 1993). Similar good results were obtained for the single amino acid type helix H of myoglobin (Shin *et al*., 1993*a*,*b*; Muñoz & Serrano, 1994).

Our method is able to determine correctly the limits of the helical conformations present in complex peptides (Figure 5A to C), which is itself a remarkable result, and the helical distribution per residue when compared with the upfield chemical shifts. Furthermore, it successfully detects the existence of more than one helical segment (Figure 5D), or a non-uniform distribution of the helical population (Figure 5E). This means that AGADIR may calculate correctly both the average helical population of a peptide, and the distribution of the helical conformation along the peptide sequence. This works well for peptides that contain regions with a significant helical population per residue (>10%). In peptides that have a low helical tendency, AGADIR detects the regions that have some local helical tendency that is revealed upon addition of trifluoroethanol (TFE; Blanco *et al*., 1992; Dyson *et al*., 1993*a*,*b*; Storrs *et al*., 1992), as can be seen in Figure 5F. This Figure shows the CD and NMR analysis in 80% TFE of an actin fragment comprising residues 1 to 28, that does not have any detectable helical population in water solution (Sonnichsen *et al*., 1992).

Rational modification of the helical content of peptides

In a previous work we demonstrated that using this algorithm it was possible to localise non-helical regions in proteins that have a high intrinsic tendency to be helical in solution (Muñoz & Serrano, 1994). A stringent way of testing the validity and applicability of our approach is to design the minimum number of specific mutations on peptide sequences so the α-helical content increases to the level desired. Three short polypeptide sequences have been mutated to increase their α-helical population: the α-helices 2 (Figure 6A) and 5 (Figure 6B) of the chemotactic protein from *Escherichia coli*, CheY (Stock *et al*., 1990; Volz & Matsumura, 1991; Belsollel *et al*., 1994), and a peptide corresponding to the long loop in the SH3-spectrin domain (Musacchio *et al*., 1992; and see Figure 6C). In Figure 6 we show the CD spectra in water solution of the wild-type and designed peptides. In the case of the SH3-loop, since the increase in helical content is moderate (2% *versus* 10%), we have added increasing concentrations of TFE, to show that the appearance of a minimum at 222 nm reflects a real helical population. Comparison of the two titration curves (wild-type and mutant) shows that while the wild-type peptide is not affected very much by TFE, the mutant peptide suffers a dramatic conformational change in the direction of helix formation (Figure 6C and D). The TFE induced-transition has an isodichroic point around 202 nm, indicating that there is a two-state transition. The average helical population, together with the calculated vaues, for the three peptides are shown in Table 1. In the case of the peptide corresponding to the α-helix 2 of CheY (Figure 5A), the CD spectrum of the wild-type peptide indicates that it is not completely unstructured, adopting a secondary structure conformation that is not helicoidal (V. Muñoz, M. A. Jimenez, M. Rico & L. Serrano, unpublished results). After modification of the peptide, the helical content increases to a record 40% for a short peptide (15 residues) that does not have blocked ends and is not polyalanine-based. These results indicate that the algorithm is successfully calculating the effect of point mutations on the helical content of peptides.

Applicability of the algorithm to proteins

In principle, this algorithm is designed to calculate the helical content of homopolymeric peptides in

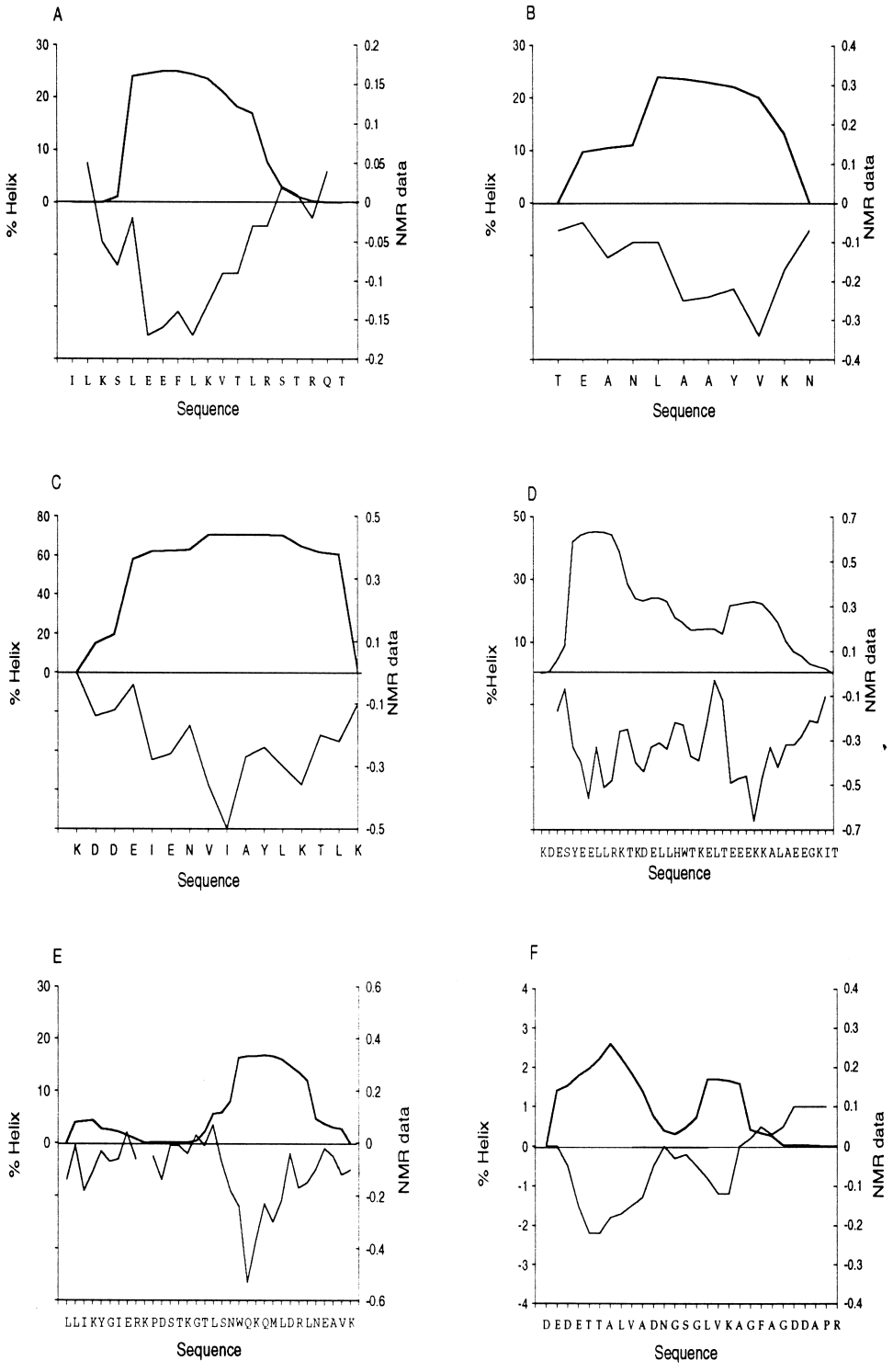

Figure 5. Diagrams showing the agreement between the predicted helicity per residue (thick continuous lines) and that observed by NMR. The upfield chemical shifts of the C^{α} protons, with respect to random-coil values, are shown as thin continuous lines. A, The C-terminal 19 residues of murine interleukin-6. B, A peptide corresponding to the α-helical region 63 to 73 of *Rhodospirillum* cytochrome *c*2. C, A peptide corresponding to the region 71 to 93 of the plasma protein transthyretin. D, A fragment of titin (residues 1977 to 2014). E, A peptide corresponding to residues 532 to 565 of the adenylate cyclase from *B. antracis*. F, An actin fragment, in 80% TFE, comprising residues 1 to 28, that forms an antiparallel β-hairpin and 2 coil regions in the protein structure. The algorithm was run using the same conditions of pH and temperature as were indicated by the respective authors (see Results).

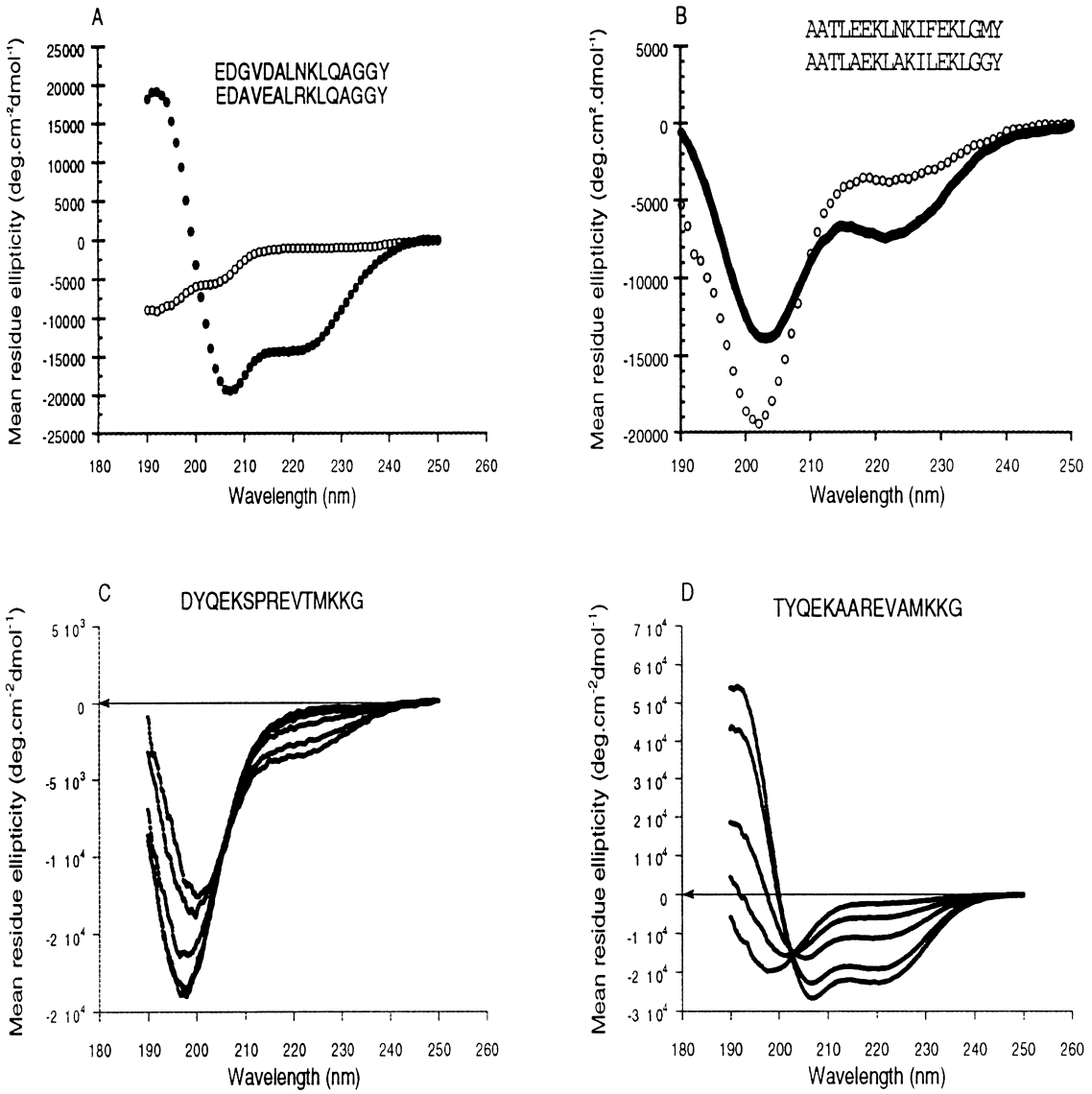

Figure 6. Circular dichroism spectra of the wild-type and mutant peptides corresponding to A the α-helix 2 of CheY. (○) Far-UV CD spectra of the wild-type peptide (top sequence); (●) far-UV CD spectra of the mutant peptide (bottom sequence). B, The α-helix 5 of CheY. (○) Far-UV CD spectra of the wild-type peptide (top sequence); (●) far-UV CD spectra of the mutant peptide (bottom sequence). C, The long loop of the spectrin SH3 domain in the presence of increasing TFE concentrations (0, 9.1, 16.7, 24.5 and 32.2%). D, The mutant form of the long loop of the spectrin SH3 domain in the presence of increasing TFE concentrations (0, 9.1, 16.7, 24.5 and 32.2%).

solution with no tertiary interactions. However, it is possible, as we have seen above, to use it in protein design or to increase the stability of isolated α-helices by choosing the right mutations. A nice example of the putative applicability of this method to protein design projects is found in the rational modification of the triosephosphate isomerase, to render an active monomer (Borchet *et al*., 1993). In the dimer of this protein, a long loop protrudes and packs against one of the monomers. The authors deleted this loop in order to dissociate the dimer, and the way they connected the β-strand with the α-helix was by designing an extra helical turn on the helix. This extra turn was present in the crystal structure, and it did have very few contacts with the rest of the protein. Running AGADIR on the wild-type sequence indicated a very poor helical tendency in solution for the α-helical region of the wild-type protein (Figure 7A). On the other hand for the mutant protein a stronger α-helical tendency is predicted for the designed α-helix, having Asn69 as the N-cap, in perfect agreement with the crystal structure (Figure 6B).

Another interesting case is found in the design of a nucleotide-binding site on CheY (P. Cronet & L. Serrano, unpublished results). Modification of the amino acid sequence of the first α-helix and loop of CheY in order to introduce a phosphate-binding loop was supposed to shorten the α-helix by one turn (P. Cronet, L. Belsollel, M. Col & L. Serrano, unpublished results). However, in the crystal structure of the mutant it was found that the helix was shorter by almost two turns, with Thr20 acting as the N-cap

Table 1

Amino acid sequences of peptides corresponding to wild-type regions of CheY and of the SH3 domain of spectrin, as well as of the mutant sequences design to increase the α-helical content

Peptide	Sequence	% Helix[a]	% Predicted
CheY2	EDGVDALNKLQAGGY	2	4
CheY2-Mo	EDAVEALRKLQAGGY	39	40
CheY5	AATLEEKLNKIFEKLGMY	13	7
CheY5-Mo	AATLAEKLAKILEKLGGY	20	20
SH3Lo	DYQEKSPREVAMKKG	2	2
Sh3Lo-Mo	TYQEKAAREVAMKKG	10	15

[a]The experimental helical percentage was obtained by using the method of Chen *et al*. (1974). All the peptides were analysed at pH 7.0 in 5 mM phosphate buffer at 4°C.

residue. Again that is what AGADIR calculates for this sequence in solution (Figure 7C and D).

Discussion

Theoretical development

The work presented here and in a previous paper (Muñoz & Serrano, 1994) is basically an attempt to put the large amount of experimental information available to date into an appropriate theoretical framework. The ultimate idea is to explain the helical behaviour of all linear, homopolymeric peptides having no tertiary interaction at a residue level. The theoretical framework utilised in this work is based on the classical helix-coil transition as first postulated by Zimm & Bragg (1959) and by Lifson & Roig (1961). However, certain characteristics make our approximation different from those modifications, recently developed, of the original helix-coil transition theory (Chen *et al*., 1992; Doig *et al*., 1994; Finkelstein *et al*., 1991; Gans *et al*., 1991; Quian, 1993). Our model, rather than describing the helix in terms of the two normal parameters, nucleation and elongation, attempts to differentiate the different energy contributions to the stability of the helical conformation and thus a more extensive parameterisation is required. This makes it difficult to compare our energy contributions directly with the nucleation and elongation parameters extracted by several groups from their experimental data. Under our theoretical development it is possible to combine in one parameter the intrinsic tendency of a particular amino acid residue to populate helical dihedral angles with the enthalpy of hydrogen-bond formation, to obtain the classical elongation factor *s*. The nucleation factor σ is more difficult to compare. In the classical helix-coil transition theory, σ does not arise from the entropic cost of fixing four amino acid residues in helical angles without forming main-chain to main-chain hydrogen bonds, as has frequently been considered the case. Rather, it accounts for the four main-chain to main-chain possible hydrogen bonds that are not formed in a particular helix, but are included in the elongation parameter, *s*, of each participating residue. This makes it possible to assume it to be independent of the sequence and almost invariant upon changes in temperature (Scholtz *et al*., 1991*c*). Using this definition we could calculate a sort of nucleation factor obtained from the sum of four hydrogen bonds (3.12 kcal mol^{-1}), $\sigma = 0.005$, or 0.002 (if we consider a value of 0.92 kcal mol^{-1} per hydrogen bond). In our approach the contributions of the main-chain to main-chain hydrogen bonds and of the intrinsic helical propensities of the amino acid residues are clearly separated. We have assumed that the intrinsic helical propensities reflect the entropy loss upon fixing the different residues in α-helical dihedral angles. Then, the entropic cost of having the first four residues in helical angles, without formation of a hydrogen bond, is easily determined and is obviously strongly dependent on the particular sequence of amino acids and on the temperature. Our formulation, although equally involving two parameters, could be more precise because, in principle, it is closer to the real energy contributions.

The partition function utilised here merits some comment. From a strict point of view this partition function might be considered a one-sequence approximation. The calculations performed involve only the different individual possible helical segments and do not analyse all the possible molecular conformations with more than one helical segment. However, since we are calculating the partition function for each amino acid residue within the polypeptide chain, and not for the whole peptide, more than one helical segment may be present at the same time in a single molecule. In other words, the one-sequence approximation for the calculation of helicity of each residue renders an approximation short of multiple sequence for the whole polypeptide chain. This is due to the total helicity of the polypeptide chain arising from the average of the helicities of the individual residues. One piece of evidence for this statement is that our formulation is able to describe, in an appropriate way, the breaking of helices by bad helix-former residues when placed in the middle of the polypeptide chain and so it can detect the presence of more than one helical segment in a polypeptide chain. The approach used in this work is therefore as simple as the one-sequence approximation but it provides a more realistic model for helix formation and stability, without involving so many calculations.

Refinement of the parameters

It is necessary to assess whether the set of parameters being used accounts for the phenomenon to be described or is the consequence of overfitting to the database utilised. In our case, all of the parameters are derived from experimental data. This is somehow a warranty of their physical significance. However, there are important uncertainties that make the parameter refinement necessary.

Experimental error results in a range of values rather than in a discrete value for each parameter (i.e.

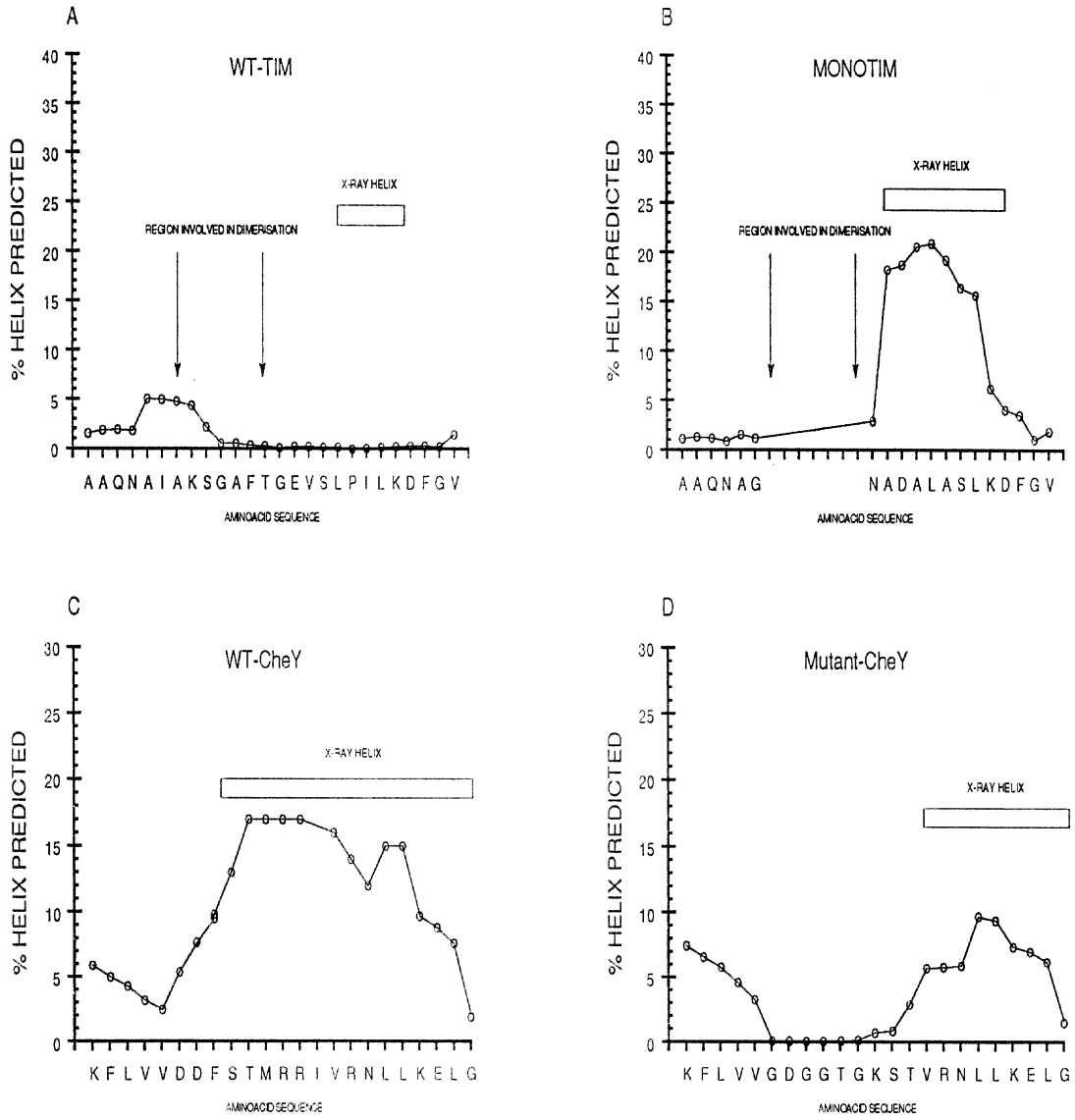

Figure 7. Comparison between the crystal structure secondary structures of wild-type TIM, mono TIM, wild-type CheY and a mutant form of CheY, and the helical content per residue calculated by AGADIR. A, Wild-type TIM; B, mono Tim; C, wild-type CheY; D, mutant CheY. The continuous line represents the helical content per residue calculated by AGADIR. The box represents the extent of the α-helix in the 3-dimensional structure.

intrinsic helical propensities). Besides, most of the experimental parameters are not derived from single energy contributions, but from a combination of them. For the simplest polyalanine-based peptide, because it contains some polar residues to make it soluble, at least four parameters need to be considered (hydrogen bond, intrinsic helical tendency, capping properties and the charges or blocking groups at the end of the peptide). This results in an infinite number of solutions. A typical case is the quantification of the enthalpic contribution of a hydrogen bond (Scholtz *et al*., 1991*a*). This was determined from a 50-residue polyalanine peptide with $i, i+3$ favourable electrostatic interactions $(AEAAKA)_n$. It is clear that there is a large experimental error due to the broad calorimetric transition observed and the interpolation of the baselines (heat capacities). Moreover, the enthalpic contribution of this peptide is the result of the main-chain to main-chain hydrogen bonds and of the possible hydrogen bonds made by the blocking groups at the end of the peptide, the van der Waals interactions between the side-chains of Glu and Lys residues, their interaction with the alanine residues and of the interactions between the C^{β} atom of residue $i+4$ with the C^{α} atom of residue i. This results in a large degree of uncertainty regarding the determination of the real value of the enthalpic contribution of a hydrogen bond, and consequently in a set of possible solutions, all of them quasi-equivalent. The refinement is then addressed to select the optimal solution, which by using values

within the experimental margins calculates correctly the average helical behaviour of all the peptides analysed to date.

The refinement process needs to be carried out sequentially instead of doing a global fitting. In this way it is possible to fix the parameters one by one or by small sets. The refined parameters presented here show some differences with those presented in the previous work (Muñoz & Serrano, 1994). These differences arise from the fact that the original parameters were adjusted in the framework of a more simplistic model whereby interactions of charged residues with the helix macrodipole were included in the capping effects. This simple approximation was found to work very well in predicting the average helical behaviour of 323 peptides. However, when the algorithm was applied to some polyalanine-based peptides specifically designed to analyse the helical dipole (Armstrong & Baldwin, 1993; Huyghues-Despointes *et al*., 1993*b*; Scholtz *et al*., 1993), we found that the two interactions (capping and helix dipole) needed to be separated. Since most of the peptides used for the previous refinement were polyalanine-based with charged residues, the separation of the capping effect from the interaction with the helix dipole of charged groups done here results in slight changes in the values of the previous basic parameters. The more significant change is the reduction of the main-chain to main-chain hydrogen-bond contribution (0.78 kcal mol^{-1}), below the experimental margins determined by calorimetry for a polyalanine-based peptide (Scholtz *et al*., 1991*a*: 0.8 kcal mol^{-1} to 1.2 kcal mol^{-1}). This is an absolute requirement in order to obtain the right cooperativity for the experimentally observed length-dependent helix-coil transition of different polyalanine-based peptides (Rohl *et al*., 1992; Scholtz *et al*., 1991*c*; this work). Higher values, closer to the experimental data, always produced over-cooperative transitions, as is explained in Results. This discrepancy could be due to our theoretical approximation, or to the experimental error of the calorimetric studies. In the calorimetric analysis, and due to the broad transition, it was not possible to determine the base-lines of the unfolded and folded states. Then, to calculate the base-lines and the thermodynamic parameters it was assumed that ΔC_p was zero and that the complete transition was a pure Gaussian curve (Scholtz *et al*., 1991*a*). This could result in an overestimation of the area of the curve (ΔH_{cal}), and consequently of the enthalpic contribution of a hydrogen bond. Moreover, in this peptide there could be enthalpic contributions different from the hydrogen bond (see above). In support of this is the fact that ΔH_{cal} for polyglutamic and polylysine peptides is different (–1.1 *versus* –0.9 kcal mol^{-1} per residue: Hermas, 1966), showing that factors apart from hydrogen-bond formation contribute to ΔH_{cal}.

Remarkably, the modification of the helix-coil theory and the parameters described here predict very well the helical content of more than 423 peptides analysed by CD. They also predict with a high degree of success how the helical population is distributed along the polypeptide chain for those peptides analysed by NMR. The helix-coil transition theory using the appropriate parameters, it is able to calculate correctly the average helical content of short polypeptide chains without tertiary interactions in solution, as well as the individual behaviour of the different component residues.

Comparison with the parameters determined by other authors

There are several scales of intrinsic helical tendencies of the different amino acids, determined from the thermodynamic analysis of proteins (Horovitz *et al*., 1992; Blaber *et al*., 1993) or peptides (von Dreele *et al*., 1971; Lyu *et al*., 1990; O'Neil & DeGrado, 1990; Horovitz *et al*., 1992; Chakrabartty *et al*., 1994). Each scale has its own problems derived from the system being used, as well as the different experimental conditions. Since every system is completely different from the others, but the results are related (with the exception perhaps of that of von Dreele *et al*., 1971), it means that the average value of the different experimental scales (Horovitz *et al*., 1992; Blaber *et al*., 1993; Lyu *et al*., 1990; O'Neil & DeGrado, 1990; Horovitz *et al*., 1992; Chakrabartty *et al*., 1994), could diminish the context effects.

After refinement, the intrinsic helical propensities were slightly different from the initial values (slope = 0.84 and r = 0.89; data not shown). Comparison of the refined values with the different experimental scales (Figure 8A to E), excluding Pro, indicates that the refined parameters correlate approximately equally well with all of them (the correlation with the scale used by Lyu *et al*. (1990) is better but we must take into account that the number of data is much smaller, 10 *versus* 19). The slopes are similar, varying from 0.7 to 0.9, with the exception of that of Baldwin and co-workers, which is higher, 1.48 (Chakrabartty *et al*., 1994). The slopes of the correlations between the data of Baldwin and co-workers and the other scales (including AGADIR), are very similar, varying between 0.4 and 0.6 (data not shown). This means that the differences in energy between Ala and the different residues are much higher in this case than in all the other scales. Interestingly, our algorithm is able to calculate correctly the experimental data used by Baldwin and co-workers to derive their helix scale (Chakrabartty *et al*., 1994: slope = 1 $\pm$ 0.07; r = 0.89; intersection = –1.6 $\pm$ 3; Figure 5F), although it was not used for the refinement of the parameters, without having such large energy differences between the intrinsic helical energies of the amino acid residues.

Most importantly, the best correlation was found with the average data of the five experimental scales (slope = 0.94 and r = 0.97), indicating that we have probably reached values that are very close to the real ones. Also, the fact that by using values similar

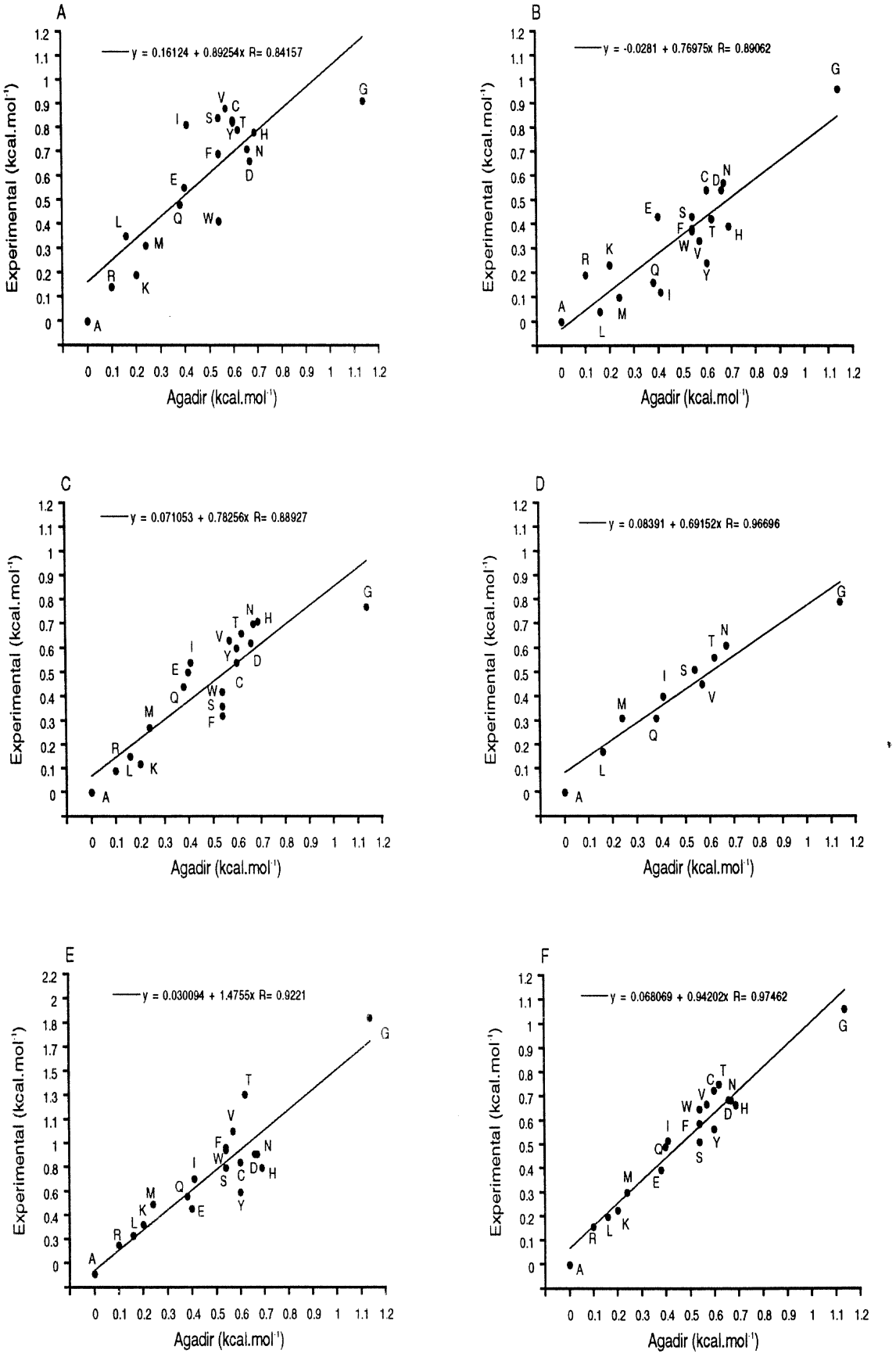

Figure 8. Correlation analysis between the different experimental helical intrinsic scales and that implemented in AGADIR. A, Fersht and co-workers (Horovitz *et al.*, 1992). B, Matthews and co-workers (Blaber *et al.*, 1993). C, O'Neil & De Grado (1990). D, Kallenbach and co-workers (Lyu *et al.*, 1990). E, Baldwin and co-workers (Chakrabartty *et al.*, 1994). F, Average scale obtained from the 5 previous scales.

to those found in proteins we are able to reproduce the helical behaviour of 423 peptides in solution, indicates that the physico-chemical parameters in proteins and peptides are the same.

Conclusions

Here, we have described a more complete approach to the folding of helical peptides in solution, in which we considered five different parameters; hydrogen bond, intrinsic helical tendencies, capping interactions (including the effect of blocking the ends of the peptides), charge-dipole interactions and side-chain to side-chain interactions. Using these five parameters we predicted the helical behaviour of 423 homopolymeric peptides in solution, as well as the effect of designed mutations on the helical content of different peptides. More important, as we have seen for several different peptides, AGADIR also calculates correctly the helical population at a residue level. This we believe is an important step in the direction of the elucidation of the folding problem of helical peptides, but we must be cautious since there are many energy parameters involved and we are making several assumptions. The analysis of more peptides by CD and NMR will allow a better refinement of the parameters, and will indicate if the assumptions made here are reasonable. In any case, it is clear that the parameters used in this work describe α-helix stability. When more peptides have been studied the model might be improved and therefore further refinement could be needed. This should be in only two directions: splitting of pair-interactions groups into their components and inclusion in the theoretical framework of newly discovered factors.

Materials and Methods

Experimental procedures

Peptide synthesis

The solid-phase synthesis of the peptides was performed on an Abimed AMS422 multiple peptide synthesiser using Fmoc chemistry and PyBOP activation at a 0.025 mmol scale. After synthesis was completed, protecting groups were removed and the peptide chains were cleaved from the resin with a mixture of 10 ml of TFA, 0.75 g of phenol, 0.2 ml of EDT, 0.5 ml of thioanisole and 0.5 ml of water for three hours. The peptides were purified on a Vydac C-18 reverse phase column (20 mm × 250 mm, 0.01 mm particle) at a flow-rate of 10 ml/min. Solvent A was water containing 0.1% TFA and solvent B was 70% acetonitrile, 0.1% TFA in water. Peptide homogeneity (>98%) was determined by HPLC using an acetonitrile gradient of 0.7% per minute. The peptide composition was confirmed by amino acid analysis and the molecular mass was checked by matrix-assisted laser desorption time-of-flight mass spectrometry.

Peptide concentration

The concentration of the different peptides was determined by amino acid analysis, or UV absorbance using the method of Gill & von Hippel (1989). For peptides that do not contain Tyr or Trp residues, the concentration was determined by amino acid analysis. The error in both cases is around 10%.

Circular dichroism analysis

Circular dichroism (CD) spectra were recorded on a Jasco-710 instrument at a temperature of 5°C. The peptides (roughly 1.5 mg) were dissolved in 1 ml of 2.5 mM sodium phosphate buffer (pH 7.0), unless otherwise indicated. To check for concentration dependence of the CD spectra, different dilutions of the peptides (10 to 750 μM), using cuvettes with different pathlengths (0.1 mm to 0.5 cm), were scanned. CD spectra in the range 190 to 250 nm were obtained using the continuous scan option (100 nm/min scan speed), with a one second response time and taking points every 0.1 nm. For every sample we took 30 scans and the experiment was repeated three times on different days. The ellipticity was calibrated using D-10-camphorsulphonic acid.

Determination of the helical percentage from the CD spectra

In order to estimate the helical population of the different peptides we used the mean residue ellipticity at 222 nm, taking into account the peptide length (Chen *et al*., 1974).

Theoretical procedures

Theory

The development of an appropriate statistical mechanics framework to model the conformational behaviour of peptides in aqueous solution needs, as a first step, a precise definition of the physical system under study. In this case the physical system is obviously a polypeptide chain in aqueous solution. The polypeptide chain as defined here comprises $n - 1$ peptide bonds (for a peptide of n residues) of fixed length at fixed bond angles. The different conformations of each residue are defined by the angles between its alpha carbon atom and the atoms of the flanking peptide bonds. Following this definition, the phase space for each residue is restricted by means of simplicity to the allowed region of the Ramachandran plot.

As in the classical helix-coil transition formulations (Zimm & Brag, 1959), the phase space of each residue is simplified by considering only two states. The first is the random-coil state, which consists of a heterogeneous ensemble of conformations where each amino acid residue is able to explore all the allowed conformational space. The second is the helical state, which is defined as the state where the amino acid residue being considered, and at least three contiguous residues, are restricted to α-helical angles plus the two capping residues (N-cap and C-cap).

The calculation of the helical propensity of each residue on a polypeptide chain may be addressed upon calculation of the molecular partition function. The molecular partition function would be the summation of the statistical weights of all the possible helical conformations of the polypeptide chain, plus the statistical weight of the molecular random coil. A helical conformation would be any containing at least one helical segment (at least four residues in helical angles and the two caps) in the

Table 2

Free energies in kcal mol^{-1} for the intrinsic tendencies of the different amino acids to be in helical dihedral angles, or to be at the N or C-cap positions, for the contribution of a hydrogen bond, for the contribution of an N-terminal acetyl or succinyl blocking groups, and for an amide C-terminal blocking group

	N-cap[a]	Capp. box[b]	(Pro N + 1)[c]	Asp + 2[d]	C-cap[e]	Intrinsic[f]	Acetyl[g]	Succinyl[h]	Amide[i]	Hbond[j]
A	0.40	0.40		0.00	0.00	0.61	–1.375	–1.875	–0.91	–0.775
G	–0.60	–0.60		–0.65	–0.40	1.71	–1.275	–1.775	–0.81	–0.775
S	–0.75	–1.00	S = –0'3	–0.75	–0.08	1.13	–1.275	–1.775	–0.81	–0.775
T	–0.65	–0.80	T = –0'3	–0.50	–0.08	1.18	–1.275	–1.775	–0.81	–0.775
N	–1.10	–1.00	N = –0'3	–1.10	–0.20	1.21	–1.275	–1.775	–0.81	–0.775
D	0.00	–1.00	D = –0'3	0.00	0.08	1.20	–1.275	–1.775	–0.81	–0.775
Q	0.50	0.25		–0.40	0.00	0.93	–1.275	–1.775	–0.81	–0.775
E	–0.08	–0.10		–0.40	0.08	0.95	–1.275	–1.775	–0.81	–0.775
H	0.25	0.25		–0.40	–0.35	1.23	–1.275	–1.775	–0.81	–0.775
K	0.00	0.00		–0.40	0.00	0.76	–1.275	–1.775	–0.81	–0.775
R	0.00	0.00		–0.40	–0.10	0.67	–1.275	–1.775	–0.81	–0.775
F	0.40	0.40		–0.40	0.08	1.08	–1.275	–1.775	–0.81	–0.775
Y	0.40	0.40		–0.40	0.08	1.08	–1.275	–1.775	–0.81	–0.775
W	0.08	0.08		–0.40	0.08	1.08	–1.275	–1.775	–0.81	–0.775
L	0.40	0.40		–0.40	0.08	0.80	–1.275	–1.775	–0.81	–0.775
V	0.40	0.40		–0.40	0.08	1.12	–1.275	–1.775	–0.81	–0.775
I	0.40	0.40		–0.40	0.08	0.96	–1.275	–1.775	–0.81	–0.775
M	0.40	0.40		–0.40	0.08	0.82	–1.275	–1.775	–0.81	–0.775
C	0.08	0.08		–0.40	0.08	1.21	–1.275	–1.775	–0.81	–0.775
P*	0.44	0.44		0.44	0.08	3.33	–1.275	–1.775	–0.81	–0.775

[a]N-cap values for 20 different amino acids after refining the parameters.

[b]N-cap values of the different amino acid residues when there is Glu at position N + 3 and the possible formation of a capping box (Harper & Rose, 1993; Dasgupta & Bell, 1993). In the case of having Gln or Asp at position N + 3, all the favourable values are multiplied by 0.625.

[c]Favourable free energy values added to that of the capping box when there is Glu at position N + 3, and Pro at position N + 1. In the case of having Gln or Asp at position N + 3, all the favourable values are multiplied by 0.625.

[d]N-cap values when there is Asp at position N + 2. If there is not a good N-cap residue, Asp at position N + 2 can make a hydrogen bond to the amide group of the N-cap residue and stabilise the helix (Bell *et al*., 1992; Dasgurpta & Bell, 1993).

[e]C-cap values for the 20 amino acids.

[f]Free energy required to put the 20 amino acids in helical dihedral angles. This term does not incude the contribution of the hydrogen bond, and in our model it reflects the free energy cost of putting the different amino acid residues in helical dihedral angles. In the case of Pro, it includes the result of breaking a hydrogen bond. When Pro is at position N + 1, its intrinsic value is 0.66 kcal mol^{-1}.

[g]Net contribution of an acetyl group at the N terminus. It is the result of the formation of a hydrogen bond (–0.775 kcal mol^{-1}), plus the disappearance of the electrostatic repulsion with the helix macrodipole (0.5 kcal mol^{-1}). In the case of Ala, there is an extra positive contribution of –0.1 kcal mol^{-1}.

[h]Net contribution of a succinyl group at the N terminus. It is the result of the formation of a hydrogen bond (–0.775 kcal mol^{-1}), plus the disappearance of the electrostatic repulsion with the helix macrodipole (0.5 kcal mol^{-1}), plus the introduction of a favourable electrostatic interaction with the helix macrodipole (–0.5 kcal mol^{-1}). In the case of Ala, there is an extra positive contribution of –0.1 kcal mol^{-1}.

[i]Net contribution of an amide group at the C terminus. It is the result of the formation of a weak hydrogen bond (–0.31 kcal mol^{-1}), plus the disappearance of the electrostatic repulsion with the helix macrodipole (0.5 kcal mol^{-1}). In the case of Ala, there is an extra positive contribution of –0.1 kcal mol^{-1}.

[j]Hydrogen bond contribution.

polypeptide chain. The random coil comprises the ensemble of molecular conformations that do not fulfil these requirements. Using the random coil as the reference state (statistical weight 1), the statistical weight of a certain helical conformation arises from:

$$K_{\text{Confor}} = \mathrm{e}^{-\Delta G_{\text{Confor}}/RT} \tag{1}$$

$$\Delta G_{\text{Confor}} = \Delta G_{\text{helical-segment}} + \Delta G_{\text{rest}}, \tag{2}$$

where K_{Confor} is the statistical weight of the helical conformation and ΔG_{Confor} is the difference in free energy between the helical conformation and the random coil state. This ΔG can be divided into two components: one corresponding to the particular helical segment we are analysing ($\Delta G_{\text{helical-segment}}$) and the other to the rest of the polypeptide chain including other possible helical segments (ΔG_{rest}).

In order to simplify the formulation of the partition function it is possible to define K_{Hel} as the summation of the statistical weights of all the possible conformations containing a specific helical segment (equation (3)). This equation might be directly transformed in equation (4), and since the $\Delta G_{\text{helical-segment}}$ term is constant it may be extracted from the summation as indicated in equation (5):

$$K_{\text{Hel}} = \sum \mathrm{e}^{-(\Delta G_{\text{helical-segment}} + \Delta G_{\text{rest}})/RT} \tag{3}$$

$$K_{\text{Hel}} = \sum (\mathrm{e}^{-\Delta G_{\text{helical-segment}}/RT} \mathrm{e}^{-\Delta G_{\text{rest}}/RT}) \tag{4}$$

$$K_{\text{Hel}} = \mathrm{e}^{-\Delta G_{\text{helical-segment}}/RT} \sum \mathrm{e}^{-\Delta G_{\text{rest}}/RT}. \tag{5}$$

The molecular partition function might therefore be expressed, with respect to a specific residue, as the summation of all of the statistical weights of the molecular conformations containing the helical segments that comprise this residue, plus the summation of all of the statistical weights of the molecular conformations in which this residue is in the random-coil state (equation (6)).

$$Z_{\text{X}} = 1 + (Z_{\text{N}} + Z_{\text{C}}) + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\text{Hel}}ji$$

$$\text{“}\forall x \| i + j \leqslant x < i \Rightarrow K_{\text{Hel}}ji = 0\text{”}, \tag{6}$$

Table 3

Free energy of interaction, in kcal mol^{-1}, between different charged residues and the helix macrodipole as a function of the distance to the N or C-cap residue

Res.	N-cap	$N+1$	$N+2$	$N+3$	$N+4$	$N+5$	$N+6$	$N+7$	$N+8$	$N+9$
Asp	–0.34	–0.51	–0.53	–0.42	–0.18	–0.15	–0.13	–0.12	–0.11	–0.09
Glu	–0.26	–0.39	–0.22	–0.19	–0.18	–0.15	–0.13	–0.12	–0.09	–0.08
His^+	0.33	1.40	1.40	0.52	0.39	0.36	0.34	0.32	0.26	0.23
Lys	0.43	0.38	0.64	0.58	0.48	0.27	0.24	0.20	0.18	0.11
Arg	0.29	0.33	0.44	0.40	0.34	0.21	0.19	0.16	0.15	0.09
Res.	C-cap	C –1	C –2	C –3	C –4	C –5	C –6	C –7	C –8	C –9
Asp	0.58	0.90	0.53	0.26	0.22	0.18	0.17	0.13	0.10	0.08
Glu	0.40	0.46	0.38	0.20	0.16	0.15	0.13	0.10	0.09	0.07
His^+	–0.44	–0.34	–0.23	–0.19	–0.13	–0.11	–0.09	–0.09	–0.09	–0.08
Lys	–0.51	–0.36	–0.34	–0.32	–0.26	–0.13	–0.12	–0.10	–0.09	–0.08
Arg	–0.36	–0.27	–0.26	–0.24	–0.25	–0.11	–0.09	–0.09	–0.07	–0.07

The nomenclature is that of Richardson & Richardson (1988).

where x is the index of the residue with respect to the peptide sequence, i is the index with respect to the peptide sequence of the first residue within the particular helix, n is the length of the peptide, j is the number of residues forming the helix, and $K_{\mathrm{Hel}}ji$ is the summation of all the statistical weights of conformations with the segment defined by i and j in the helical conformation (see equation (5)). A value of 1 corresponds to the statistical weight of the random-coil conformation of the polypeptide chain, Z_{N} is the partition function of the segment from the first residue of the polypeptide chain until residue $x-1$, and Z_{C} is the partition function of the segment from residue $x+1$ until the last residue of the polypeptide chain. The sub-index of equation (6), (''$\forall x \| i + j \leqslant x < i \Rightarrow K_{ji} = 0$''), assigns a statistical weight of zero to all the helical conformations of the peptide being analysed that do not include the residue X.

From this formulation of the molecular partition function it is straightforward to define the helical propensity of a specific residue of a polypeptide chain in the following terms:

$$\langle \text{Hel Xresidue} \rangle = \left(\sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\mathrm{Hel}}ji \right) \Bigg/ \left(1 + (Z_{\mathrm{N}} + Z_{\mathrm{C}}) + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\mathrm{Hel}}ji \right)$$

$$\text{''}\forall x \| i + j \leqslant x < i \Rightarrow K_{\mathrm{Hel}}ji = 0\text{''}. \quad (7)$$

The determination of the helical propensity of a residue in this way depends upon the conformational state of every residue in the polypeptide chain, requiring a large number of calculations. This approximation might be simplified if we consider that the random-coil state of a residue comprises all of its conformational space, including the helical dihedral angles, which it explores freely. Under this definition, it becomes clear that the random-coil state of a residue is not energetically coupled to other residues in the polypeptide chain. On the other hand, a residue in the helical state is obviously energetically coupled to those participating in the same helical segment. Then the polypeptide chain might be viewed for a sufficiently small fraction of the time, as several quasi-closed systems (helical segments) separated by decoupling residues (residues in the random-coil state). Under this assumption, the term $K_{\mathrm{Hel}}ji$ might be simplified, since it is not necessary to consider the energy contributions from residues not participating in the helical segment. Equation (5) is subsequently simplified to equation (8). The partition functions Z_{N} and Z_{C} might also be eliminated from the calculation, rendering a much simplified equation:

$$K_{\mathrm{Hel}} = \mathrm{e}^{-\Delta G_{\text{helical-segment}}/RT} \quad (8)$$

$$\langle \text{Hel Xresidue} \rangle = \left(\sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\mathrm{Hel}}ji \right) \Bigg/ \left(1 + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} K_{\mathrm{Hel}}ji \right)$$

$$\text{''}\forall x \| i + j \leqslant x < i \Rightarrow K_{\mathrm{Hel}}ji = 0\text{''}. \quad (9)$$

To determine the average helical content of a polypeptide chain, it is then necessary only to calculate the helical propensity of each residue from the individual partition functions and then the average value for all of them.

This formulation, despite the simple formulation involved, accounts for the possibility of having more than one non-overlapping helical segment simultaneously, without the necessity of calculating all the possible conformations for the polypeptide chain. The main restriction for this kind of analysis lies in the proper reduction of the phase space to only two states, random coil and α-helix. It is very likely that a complex sequence of residues will have specific interactions in certain conformations other than an α-helix. Their inclusion in the random-coil state could therefore be too severe an assumption, producing an overprediction of α-helicity. However, it is unlikely to happen in a peptide with a very high helical tendency. For those peptides with low helical tendency, the likelihood of finding another stable conformation increases, but the absolute error introduced into the calculation is small since the helical tendency is small. Another source of error could arise from the use of this approximation for very long polypeptide chains. A linear increase in the polypeptide chain produces a geometrical increase in the possible conformations being considered as random coil. The result is that longer polypeptide chains are likely to be overpredicted.

Definition of the parameters

The calculation of the statistical weight K_{Hel} of any particular helical segment with respect to the random coil is carried out using equation (8), where $\Delta G_{\text{helical-segment}}$ is the difference in free energy between the random coil and helical states, and consists of a summation of contributions:

$$\Delta G_{\text{helical-segment}} = \Delta G_{\mathrm{Int}} + \Delta G_{\mathrm{Hbond}} + \Delta G_{\mathrm{SD}} + \Delta G_{\mathrm{nonH}} + \Delta G_{\mathrm{dipole}}. \quad (10)$$

Table 4

Energy contributions in kcal mol^{-1} of the interactions between the different amino acids at positions *i*, *i* + 3 (first line) and *i*, *i* + 4 (second line)

	P	C	A	G	S	T	N	D	Q	E	H	K	R	Y	F	W	L	M	V	I
P	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.30	−0.30	−0.30	0.00	0.00	0.00	0.00
C	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.20	0.20	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.10	0.20	0.20	0.20	0.20
A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.10	0.00	−0.05	−0.10
G	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.10	0.10	0.15	0.15	0.15	0.30	0.30	0.30	0.30
S	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.10	−0.10	−0.10	0.00	0.00	0.00	0.00
	0.00	0.30	0.00	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.40	0.00	0.00	0.10	0.10	0.10	0.20	0.20	0.20	0.20
T	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.30	0.00	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10	0.10	0.00	0.00	0.00	0.00
	0.00	0.30	0.00	0.00	0.20	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.20	−0.20	−0.20	−0.30	0.15	0.15	0.15	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.20	0.00	0.10	−0.20	−0.25	−0.34	0.20	0.20	0.20	0.00	0.00	0.00	0.00
Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.15	−0.10	−0.10	−0.20	0.00	−0.10	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	−0.30	−0.10	−0.30	−0.60	0.00	−0.25	−0.05	−0.10	−0.10	−0.20	−0.20	−0.20	0.00	0.00	0.00	0.00
E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	−0.30	−0.10	−0.20	−0.10	0.00	0.00	0.00	−0.15	0.00	0.00
	0.00	0.00	0.00	0.00	−0.05	−0.05	−0.30	0.10	0.00	0.20	−0.40	−0.33	−0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.20	0.50	0.50	0.50	−0.25	−0.20	−0.30	−0.10	−0.10	−0.10	−0.10
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.20	−0.15	−0.20	0.15	0.15	0.15	−0.40	−0.40	−0.40	−0.10	−0.10	0.10	−0.10
K	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	−0.10	0.25	0.25	0.25	−0.10	0.00	0.00	−0.20	0.00	−0.05	−0.05
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.30	0.00	−0.30	0.20	0.20	0.25	−0.40	−0.20	−0.20	−0.15	−0.10	−0.15	−0.15
R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	−0.10	0.25	0.25	0.25	−0.10	0.00	0.00	−0.20	0.00	−0.05	−0.05
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.30	−0.10	−0.30	0.20	0.25	0.25	−0.40	−0.20	−0.20	−0.15	−0.10	−0.15	−0.15
Y	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.35	−0.40	−0.30	0.00	0.00	−0.15	−0.15	−0.15	−0.25	−0.25	−0.25	−0.25
	0.00	0.00	0.00	0.00	0.10	−0.10	0.10	0.10	−0.30	−0.20	−1.10	−0.20	−0.20	−0.35	−0.35	−0.45	−0.35	−0.40	−0.25	−0.35
F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.30	−0.40	−0.30	0.00	0.00	−0.15	−0.15	−0.15	−0.25	−0.25	−0.25	−0.25
	0.00	0.00	0.00	0.00	0.10	−0.10	0.10	0.10	−0.30	−0.20	−0.90	0.00	0.00	−0.25	−0.25	−0.35	−0.35	−0.40	−0.25	−0.35
W	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.35	−0.40	−0.30	0.00	0.00	−0.15	−0.15	−0.15	−0.25	−0.25	−0.25	−0.25
	0.00	0.00	0.00	0.00	0.10	−0.10	0.10	0.10	−0.30	−0.20	−1.30	0.00	0.00	−0.45	−0.45	−0.45	−0.30	−0.40	−0.25	−0.30
L	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	−0.15	0.00	−0.10	−0.30	−0.30	−0.30	−0.20	−0.20	−0.30	−0.25
	0.00	0.00	0.00	0.00	0.40	0.10	0.40	0.50	0.00	0.15	0.20	0.20	0.20	−0.10	−0.10	−0.10	−0.20	−0.20	−0.30	−0.25
M	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	−0.20	−0.20	−0.10	−0.30	−0.30	−0.30	−0.25	−0.20	−0.20	−0.20
	0.00	0.00	0.00	0.00	0.20	0.10	0.20	0.50	0.00	−0.10	0.20	0.20	0.20	−0.20	−0.20	−0.30	−0.20	−0.20	−0.40	−0.30
V	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.10	0.00	−0.15	0.00	0.00	−0.30	−0.30	−0.30	−0.30	−0.20	−0.20	−0.20
	0.00	0.00	0.00	0.00	0.30	0.10	0.30	0.50	0.00	0.00	0.00	−0.15	−0.15	−0.15	−0.15	−0.15	−0.25	−0.20	−0.25	−0.25
I	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	−0.10	0.00	−0.15	0.00	0.00	−0.30	−0.30	−0.30	−0.30	−0.25	−0.20	−0.30
	0.00	0.00	0.00	0.00	0.40	0.10	0.40	0.50	0.00	0.10	0.20	0.20	0.20	−0.10	−0.10	−0.10	−0.25	−0.25	−0.30	−0.30

(1) ΔG_{Int} is the summation of the intrinsic tendency of the *j* residues to adopt the helical dihedral angles, and has been defined (Muñoz & Serrano, 1994, 1995). This term reflects the loss of conformational entropy and it is expressed as the difference in free energy between the helix and random-coil states (see Table 2). There is a very good correlation ($r = 0.92$, data not shown), between the intrinsic values empirically determined here and the values determined by Monte Carlo simulations of the hydrophobic side-chains (Ala, Val, Leu, Ile, Phe, Tyr and Trp; Creamer & Rose, 1994). In the Monte Carlo simulations the hypothesis was that the difference in intrinsic propensities of these residues was mainly due to the loss of conformational entropy, thus supporting our assumption that the intrinsic propensities of the different amino acids reflect their loss of conformational entropy upon adopting α-helical angles.

(2) ΔG_{Hbond} is the sum of the net contribution of all the main-chain hydrogen bonds within the helical region, and reflects the difference in energy between a main-chain hydrogen bond made in the peptide and the hydrogen bonds made by the same groups with water molecules. This term is the main contributor to helix stability and the initial value was obtained from calorimetric (Scholtz *et al*., 1991*b*) and theoretical studies (Ooi & Obatake, 1991).

(3) ΔG_{SD} is the sum of the net contribution, with respect to the random-coil state, of all the side-chain to side-chain interactions located at positions $i, i + 3$ and $i, i + 4$ within the helical region. It also includes a weakly attractive, as well as repulsive, coulombic interaction in the helical conformation between charged residues at positions $i, i + 1$ (± 0.05 kcal mol^{-1}; Muñoz & Serrano, 1994, 1995; and see Table 2).

(4) ΔG_{nonH} is the sum of the net contribution to the stability of the helical region of all the residues that are not in the helical conformation. This term is different from zero only for the interactions of the helix with the first residues before (N-cap), and after, the helical conformation (C-cap: Richardson & Richardson, 1988). Consequently, the largest helix that can be formed in a peptide with *n* residues consists of $n - 2$ residues and the N and C-caps. In a previous work the interaction of charged groups with the helix dipole was included here (Muñoz & Serrano, 1994, 1995). This was clearly an oversimplification, since it is clear that charged groups located at different positions within the helix also interact with the helix dipole. In this work, we have separated the capping effect of charged groups from their interaction with the helix dipole (see Table 2).

(5) Dipole interactions. This represents the interaction of charged groups with the helix macrodipole (Shoemaker *et al*., 1987). Experimental analysis on proteins (Serrano & Fersht, 1989; 1992*b*,*c*; Nicholson *et al*., 1991; Sancho *et al*., 1992) and peptides (Armstrong & Baldwin, 1993; Huyghues-Despointes *et al*., 1993*b*; Scholtz *et al*., 1993), as well as statistical analysis of the protein database (Richardson & Richardson, 1988; Dasgupta & Bell, 1993), have shown that negatively charged residues are favourable at the N terminus of the helix, while positively charged residues are favourable at the C terminus. As has been experimentally demonstrated, those effects are found when the N or C-cap residue is charged and, as expected from an electrostatic interaction, they take place at any position inside the helical conformation (Armstrong & Baldwin, 1993; Huyghues-Despointes *et al*., 1993*b*; Scholtz *et al*., 1993). To estimate the interaction of a charged group in the helix macrodipole, we have placed in a polyalanine model helix different charged groups at all positions (using their more favourable rotamers), and then measured and averaged the distance between the charged groups and the first four amide groups or the last four carbonyl groups. Positions that are more than nine residues distant from the N or C-cap are assumed to have zero contribution. The average free energy value for the interaction of a charged histidine residue with the helix macrodipole (–0.6 kcal mol^{-1} with the C terminus; Sancho *et al*., 1992), and the average distance of the two NH side-chain groups of the histidine residue to the four main-chain carbonyl groups of the last turn of the first α-helix of barnase (4.9 Å; Mauguen *et al*., 1982; Baudet & Janin, 1991), have been used as reference values to obtain the final energy values for the interaction with the dipole of the different charged amino acids at different positions:

$$\Delta G = \pm 0.6(4.9^2)/r^2 \text{ kcal mol}^{-1}, \quad (11)$$

where *r* is the average distance between the charged atom of the side-chain to the four main-chain carbonyl groups of the last turn of the helix or the four amide groups of the first turn of the helix. We have considered only the residues that are totally or partly charged at pH 7.0 (Asp, Glu, His, Lys and Arg), since there are experimental data for them. The final energy contributions of the intercation of these residues with the helix dipole are presented in Table 4.

In non-blocked peptides there is a repulsion energy between the main-chain ionisable end groups and the helix macrodipole. We have then given a similar value to the repulsive interaction when the charged main-chain group is in the N or C-cap residue, as for the experimental average repulsion of a charged histidine residue with the N terminus of the helices in barnase (0.6 kcal mol^{-1}; Sancho *et al*., 1992). Outside the N and C-cap positions the main-chain charged groups, as well as charged side-chain groups, could interact with the helix macrodipole. The residues at these positions are in a random-coil conformation (very mobile) and we have considered the interaction of their side-chain charged groups with the helix dipole as negligible. On the other hand, the main-chain groups are more rigid, since they have a lesser degree of freedom and they could have a more significant interaction with the helix macrodipole. We have then arbitrarily assigned a value of 0.3 kcal mol^{-1} when the main-chain charged end-group belongs to the N – 1 or C + 1 residue, and a value of 0.15 kcal mol^{-1} when it belongs to the residue N – 2 or C + 2. At the rest of the positions we have assumed it to be zero. These values have been tested against those peptides analysed by CD with blocked or unblocked ends, and we found them to be satisfactory (data not shown). These numbers also agree quite well with those published by Doig *et al*. (1994).

Parameter refinement

One of the critical points of this kind of analysis lies in the acquisition of the energy contributions to parameterise the algorithm. In our approach, we have used data from studies of helical peptides as well as of proteins, which is better than utilising arbitrary parameters derived from theoretical assumptions. However, the empirical parameters used are subject to experimental and statistical error (Muñoz & Serrano, 1994, 1995). It is highly noteworthy that with the initial parameters, without any refinement, a correlation of 0.7 is obtained (data not shown). This indicates that the theoretical approximation and the parameters are really describing the α-helix tendency of peptides in solution. Despite this, some refinement is needed to check how far it is possible to go with this kind of approximation and to improve the quality of the parameters. We have carried out this refinement

sequentially as described (Muñoz & Serrano, 1994, 1995), to avoid possible overfitting when trying to refine large sets of parameters, using the large body of experimental information about helical peptides available. The refinement basically consisted of calibrating the energy contributions between themselves, fixing those parameters with an interval of experimental uncertainty to one discrete value and slightly modifying the side-chain to side-chain interactions to account for differences among amino acids within the defined groups. It is important to note that the changes made during the refinement have been generally supported by physicochemical reasoning. This is the main reason why we opted for a manual refinement procedure rather than an automatic fitting. The main difference between this refinement and the previous one is the separation of the capping effect and the interaction of charged residues with the helix macrodipole. The introduction of this extra parameter, as well as of the temperature and pH dependence (see the accompanying paper, Muñoz & Serrano, 1995), results in a slight modification of some of the previous values (Muñoz & Serrano, 1994). The new values for the different refined parameters are shown in Tables 2 to 4.

Acknowledgements

We are very grateful to Dr Francisco Blanco for communicating the results of the CD analysis of the protein G fragments. We are very grateful to Dr Giovanna Musco and Dr Annalisa Pastore for communicating the results of the CD and NMR analysis of the peptide corresponding to the region 1977 to 2014 of titin.

References

Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. J. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Nat. Acad. Sci., U.S.A.* **47**, 1309–1314.

Armstrong, K. M. & Baldwin, R. L. (1993). Charged histidine affects α-helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 11337–11340.

Armstrong, K. M., Fairman, R. & Baldwin, R. L. (1993). The (*i*, *i* + 4) Phe-His interaction studied in an alanine-based α-helix. *J. Mol. Biol.* **230**, 284–291.

Baudet, S. & Janin, J. (1991). Crystal structure of a barnase d(GpC) complex at 1.9 Å resolution. *J. Mol. Biol.* **214**, 123–132.

Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A. & Matthews, B. W. (1992). Dissection of helix capping in T4 lysozyme by structural and thermodynamical analysis of six aminoacid substitutions at Thr59. *Biochemistry*, **31**, 3590–3596.

Bellsolel, L., Prieto, J., Serrano, L. & Coll, M. (1994). Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface. *J. Mol. Biol.* **238**, 489–495.

Blaber, M., Zhang, X. & Matthews, B. (1993). Structural basis of amino acid a helix propensity. *Science*, **260**, 1637–1640.

Blanco, F. J., Jimenez, M. A., Rico, M., Santoro, J., Herranz, J. & Nieto, J. L. (1992). The homologous angiogenin and ribonuclease N-terminal fragments fold into very similar helices when isolated. *Biochem. Biophys. Res. Commun.* **182**, 1491–1498.

Borchet, T. V., Abagyan, R., Radhα-Kishan, K. V., Zeelen, J. P. H. & Wierenga, R. K. (1993). The crystal structure of an engineered monomeric triosephosphate isomerase, monσTIM: the correct modelling of an eight residue loop. *Curr. Opin. Struct. Biol.* **1**, 205–213.

Bradley, E. K., Thomason, J. F., Cohen, F. E., Kosen, P. A. K. & Kuntz, I. D. (1990). Studies of synthetic peptides using circular dichroism and nuclear magnetic resonance. *J. Mol. Biol.* **215**, 607–622.

Bruch, M. D., Dhingra, M. M. & Gierasch, L. M. (1991). Side chain-backbone hydrogen bonding contributes to helix stability in peptides derived from an α-helical region of carboxypeptidase A. *Proteins: Struct. Funct. Genet.* **10**, 130–139.

Burke, C., Mayo, K. H., Skubitz, A. P. N. & Furcht, L. T. (1991). ^{1}H-NMR and CD secondary structure analysis of cell adhesion promoting peptide F-9 from laminin. *J. Biol. Chem.* **266**, 19407–19412.

Chakrabartty, A., Schellman, J. A. & Baldwin, R. L. (1991). Large differences in the helix propensities of alanine and glycine. *Nature (London)*, **351**, 586–588.

Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R. L. (1993*a*). Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry*, **32**, 5560–5565.

Chakrabartty, A., Doig, A. J. & Baldwin, R. L. (1993*b*). Helix N-cap propensities in peptides parallel those found in proteins. *Proc. Nat. Acad. Sci., U.S.A.* In the press.

Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino-acids measured in alanine-based peptides without helix-stabilizing sidechain interactions. *Protein Sci.* **3**, 843–852.

Chen, C. C., Zhu, Y., King, J. A. & Evans, L. A. (1992). A molecular thermodynamic approach to predict the secondary structure of homopolypeptides in aqueous systems. *Biopolymers*, **32**, 1375–1392.

Chen, Y. H., Yang, J. T. & Chau, K. H. (1974). Determination of the helix and β-form of proteins in aqueous solution by circular dichroism. *Biochemistry*, **13**, 3350–3359.

Creamer, T. P. & Rose, G. D. (1994). α-Helix-forming properties in peptides and proteins. *Proteins: Struct. Funct. Genet.* **19**, 85–97.

Cushman, J. A., Mishra, P. K., Bothner, A. A. & Khosla, M. S. (1992). Conformations in solution of angiotensin II, and its 1-7 and 1-6 fragments. *Biopolymers*, **32**, 1163–1171.

Dasgupta, S. & Bell, J. A. (1993). Design of helix ends. *Int. J. Pept. Protein Res.* **41**, 499–511.

DeGrado, W. F., Prendergast, F. G., Wolfe, H. R., Jr & Cox, A. C. (1985). The design, synthesis, and characterisation of tight-binding inhibitors of calmodulin. *J. Cell. Biochem.* **29**, 83–93.

Doig, A. J., Chakrabartty, A., Klinger, T. M. & Baldwin, R. L. (1994). Determination of the free energies of N-capping in α-helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. *Biochemistry*, **33**, 3396–3403.

Doughty, M. B. & Hu, L. (1993). The contribution of helical potential to the in vitro receptor binding activity of a neuropeptide Y N-terminal deletion fragment. *Biopolymers*, **33**, 1201–1206.

Dyson, H. J. & Wright, P. (1993). Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* **3**, 60–65.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A. & Wright, P. E. (1993*a*). 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., Merutka, G., Waltho, J. P., Lerner, R. A. & Wright, P. E. (1993*b*). Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II. Plastocyanin. *J. Mol. Biol.* **226**, 818–835.

Fairman, R., Shoemaker, K. R., York, E. J., Stewart, J. M. & Baldwin, R. L. (1989). Further studies of the helix dipole model: effects of a free α-NH_3^+ or α-COO^- group on helix stability. *Proteins: Struct. Funct. Genet.* **5**, 1–7.

Fairman, R., Armstrong, K. M., Shoemaker, K. R., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991). Position effect on apparent helical propensities in the C-peptide helix. *J. Mol. Biol.* **221**, 1395–1401.

Finkelstein, A. V., Badretinov, A. Y. & Ptitsyn, O. B. (1991). Physical reasons for secondary structure stability: α-helices in short peptides. *Proteins: Struct. Funct. Genet.* **10**, 287–290.

Forood, B., Feliciano, E. J. & Nambiar, K. P. (1993). Stabilization of α-helical structures in short peptides via end capping. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 838–842.

Fukugita, M., Lancaster, D. & Mitchard, M. G. (1993). Kinematics and thermodynamics of a folding heteropolymer. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 6365–6368.

Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W. & Kallenbach, N. R. (1991). The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data. *Biopolymers*, **31**, 1605–1614.

Ghadiri, M. R. & Choi, C. (1990). Secondary structure nucleation in peptides. Transition metal ion stabilised α-helices. *J. Amer. Chem. Soc.* **112**, 1630–1632.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.

Goodman, E. M. & Kim, P. S. (1989). Folding of a peptide corresponding to α-helix in bovine pancreatic trypsin inhibitor. *Biochemistry*, **28**, 4333–4337.

Greenfield, N. J., Stafford, W. F. & Hitchcock-DeGregori, S. E. (1994). The effect of N-terminal tropomyosin peptide and α-tropomyosin. *Protein Sci.* **3**, 402–410.

Gronenborg, A. M., Filpula, D. R., Essig, N. Z., Achani, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science*, **253**, 657–661.

Harper, E. T. & Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry*, **32**, 7605–7609.

Horovitz, A., Matthews, J. & Fersht, A. R. (1992). α-Helix stability in proteins. II. Factors that influence stability at an internal position. *J. Mol. Biol.* **227**, 560–568.

Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1993*a*). Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci.* **2**, 80–85.

Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1993*b*). Effect of a single aspartate on helix stability at different positions in a neutral alanine-base peptide. *Protein Sci.* **2**, 1604–1611.

Jarvis, J. A., Munro, S. L. A. & Craik, D. J. (1994). Structural analysis of peptide fragment 71-93 of transthyretin by NMR spectroscopy and electron microscopy: insight into amyloid fibril formation. *Biochemistry*, **33**, 33–41.

Jimenez, M. A., Nieto, J. L., Herranz, J., Rico, M. & Santoro, J. (1987). ^{1}H NMR and CD evidence of the folding of the isolated ribonuclease 50-61 fragment. *FEBS Letters*, **221**, 320.

Jimenez, M. A., Rico, M., Herranz, J., Santoro, J. & Nieto, J. L. (1988). ^{1}H-NMR assignment and folding of the isolated ribonuclease 21-42 fragment. *Eur. J. Biochem.* **175**, 101–109.

Jimenez, M. A., Blanco, F. J., Rico, M., Santoro, J., Herranz, J. & Nieto, J. L. (1992). Periodic properties of proton conformational shifts in isolated protein helices. *Eur. J. Biochem.* **207**, 39–71.

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

Jimenez, M. A., Mũnoz, V., Rico, M. & Serrano, L. (1994). Helix stop and start signals in peptides and proteins: the capping box does not necessarily prevent helix elongation. *J. Mol. Biol.* **242**, 487–496.

Kemmink, J. & Creighton, T. E. (1993). Conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor (BPTI) and their roles in folding. *J. Mol. Biol.* **234**, 861–878.

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

Li, S. & Deber, C. M. (1993). Peptide environment specifies conformation. *J. Biol. Chem.* **268**, 22975–22978.

Lifson, R. & Roig, A. (1961). On the theory of helix-coil transitions in biopolymers. *J. Chem. Phys.* **34**, 1963–1974.

Lockhart, D. J. & Kim, P. S. (1993). Electrostatic screening of charge and dipole interactions with the helix backbone. *Science*, **260**, 198–202.

Lyu, P. C., Marky, L. A. & Kallenbach, N. R. (1989). The role of ion pairs in α-helix stability: two new designed helical peptides. *J. Amer. Chem. Soc.* **111**, 2733–2734.

Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. R. (1990). Side chain contributions to the stability of alphα-helical structures in peptides. *Science*, **250**, 669–673.

Lyu, P. C., Wang, P. C., Liff, M. I. & Kallenbach, N. R. (1991). Local effect of glycine substitution in a model helical peptide. *J. Amer. Chem. Soc.* **113**, 3568–3572.

Lyu, P. C., Zhou, H. X., Jelveh, N., Wemmer, D. E. & Kallenbach, N. R. (1992). Position-dependent stabilizing effects in α-helices: N-terminal capping in synthetic model peptides. *J. Amer. Chem. Soc.* **114**, 6560–6562.

Lyu, P. C., Wemmer, D. E., Hongxing, X. Z., Pinker, R. J. & Kallenbach, N. R. (1993). Capping interactions in isolated α helices: position-dependent substitution effects and structure of a serine-capped peptide helix. *Biochemistry*, **32**, 421–425.

Mammi, S., Foffani, M. T., Improta, S., Tessari, M., Schierano, E. & Peggion, E. (1992). Conformation of uteroglobin fragments. *Biopolymers*, **32**, 341.

Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989). Unusually stable helix formation in short alanine-based peptides. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5286–5290.

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

Mauguen, Y., Hartley, R. W., Dodson, E. J., Dodson, G. G., Bricogne, G., Chothia, C. & Jack, A. (1982). Molecular structure of a new family of ribonucleases. *Nature (London)*, **297**, 162–164.

Maulet, Y. & Cox, J. A. (1983). Structural changes in melittin and calmodulin upon complex formation and their modulation by calcium. *Biochemistry*, **22**, 5680–5686.

McDowell, L., Sanayal, G. & Prendergast, F. G. (1985). Probable role of amphiphilicity in the binding of

mastoparan to camodulin. *Biochemistry*, **24**, 2979–2984.

Merutka, G. & Stellwagen, E. (1990). Positional independence and additivity of amino acid replacements on helix stability in monomeric peptides. *Biochemistry*, **29**, 894–898.

Merutka, G. & Stellwagen, E. (1991). Effect of amino acid ion pairs on peptide helicity. *Biochemistry*, **30**, 1591–1594.

Merutka, G., Lipton, W., Shalongo, W., Park, S. H. & Stellwagen, E. (1990). Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry*, **29**, 7511–7515.

Merutka, G., Morikis, D., Bruschweiler, R. & Wright, P. E. (1994). NMR evidence for multiple conformations in a highly helical model peptide. *Biochemistry*, **32**, 13089–13097.

Morii, H., Honda, S., Ohashi, S. & Uedaira, H. (1994). α-Helical assembly of biologically active peptides and designed helix bundle protein. *Biopolymers*, **34**, 481–488.

Morton, C. J., Simpson, R. J. & Norton, R. S. (1994). Solution structure of synthetic peptides corresponding to the C-terminal helix of interleukin-6. *Eur. J. Biochem.* **219**, 97–107.

Munier, H., Blanco, F. J., Precheur, B., Dieisis, E., Nieto, J. L., Craescu, C. T. & Barzu, O. (1993). Characterisation of a synthetic calmodulin-binding peptide derived ? *J. Biol. Chem.* **268**, 1695–1701.

Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature: Struct. Biol.* **1**, 399–409.

Muñoz, V. & Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. III: Temperature and pH dependence. *J. Mol. Biol.* **245**, 297–308.

Musacchio, A., Noble, M. E. M., Pautit, R., Wierenga, R. & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature (London)*, **359**, 851.

Nicholson, H., Anderson, D. E., Dao-Pin, S. & Matthews, B. W. (1991). Analysis of the interaction between charged side chains and the α-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry*, **30**, 9816.

O'Neil, K. T. & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–650.

Ooi, T. & Obatake, M. (1991). Prediction of the thermodynamics of protein unfolding: the helix-coil transition of poly(L-alanine). *Proc. Nat. Acad. Sci., U.S.A.* **88**, 2859–2863.

Padmanabhan, S. & Baldwin, R. L. (1991). Straight-chain non-polar amino acids are good helix-formers in water. *J. Mol. Biol.* **219**, 135–137.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990). Relative helix-forming tendencies of nonpolar amino acids. *Nature (London)*, **344**, 268–270.

Park, S. H., Shalongo, W. & Stellwagen, E. (1993). Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. *Biochemistry*, **32**, 7048–7053.

Peña, M. C., Rico, M., Jimenez, M. A., Herranz, J., Santoro, J. & Nieto, J. L. (1989). Conformational properties of the isolated 1-23 fragment of human hemoglobin α-chain. *Biochim. Biophys. Acta*, **957**, 380–389.

Pintar, A., Rooman, M. J., Hallenga, K. & Wodak, S. J. (1994). Conformational properties of four peptides corresponding to α-helical regions of *Rhodospirillum* cytochrome c_2 and bovine calcium binding protein. *Biochemistry*, **33**, 11158–11173.

Precheur, B., Siffert, O., Barzu, O. & Craescu, C. T. (1991). NMR and circular dichroic studies on the solution conformation of a synthetic peptide derived from the calmodulin-binding domain of *Bordetella pertussis* adenylate cyclase. *Eur. J. Biochem.* **196**, 67–72.

Qian, H. (1993). Single-residue substitution in homopolypeptides: perturbative helix-coil theory at a single site. *Biopolymers*, **33**, 1605–1616.

Reed, J., Hull, W. E., Ponstingl, H. & Himes, R. H. (1992). Conformational properties of the b(400-436) and b(400-445) C-terminal peptides of porcine tubulin. *Biochemistry*, **31**, 11888–11895.

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

Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M. & Baldwin, R. L. (1992). Kinetics of proton exchange in helical peptides of varying chain lengths: interpretation by the Lifson-Roig equation. *Biochemistry*, **31**, 1263–1269.

Sancho, J., Neira, J. L. & Fersht, A. R. (1992). An N-terminal fragment of barnase has residual structure similar to that in a refolding intermediate. *J. Mol. Biol.* **224**, 749–758.

Sancho, S., Serrano, L. & Fersht, A. R. (1992). Histidine residues at the N- and C-termini of α-helices: perturbed p*K*a$_s$ and protein stability. *Biochemistry*, **31**, 2253.

Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. & Bolen, D. W. (1991*a*). Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 2854–2858.

Scholtz, J. M., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991*b*). A neutral water-soluble, α-helical peptide: the effect of ionic strength on the helix-coil equilibrium. *J. Amer. Chem. Soc.* **113**, 5102–5104.

Scholtz, M., Qian, H., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991*c*). Parameters of helix-coil theory for alanine-based peptides of varying chain lengths in water. *Biopolymers*, **31**, 1463–1470.

Scholtz, J. M., Quian, H., Robbins, V. H. & Baldwin, R. L. (1993). The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry*, **32**, 9668–9676.

Serrano, L. & Fersht, A. R. (1989). Capping and α-helix stability. *Nature (London)*, **342**, 296–299.

Serrano, L., Matouschek, A. R. & Fersht, A. R. (1992*a*). The folding of an enzyme VI: the folding pathway of barnase. Comparison with theoretical models. *J. Mol. Biol.* **224**, 847–859.

Serrano, L., Neira, J. L., Sancho, J. & Fersht, A. R. (1992*b*). Effect of alanine *versus* glycine in α-helices on protein stability. *Nature (London)*, **356**, 453–455.

Serrano, L., Sancho, J., Hirshberg, J. M. & Fersht, A. R. (1992*c*). α-Helix stability in proteins. I. Empirical correlations concerning substitution of side chains at the N- and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J. Mol. Biol.* **227**, 544–549.

Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E. & Dyson, H. J. (1993*a*). Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin act as a helix stop signal. *Biochemistry*, **32**, 6348–6355.

Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993*b*). Peptide models of protein

folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry*, **32**, 6356–6364.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1987). Tests of the helix dipole model for stabilisation of α-helices. *Nature (London)*, **326**, 563–567.
Shoemaker, R. K., Fairman, R., Schultz, D. A., Robertson, A. D., York, E. J., Stewart, J. M. & Baldwin, R. L. (1990). Side chain interactions in the C-peptide helix Phe8 . . . His12. *Biopolymers*, **29**, 1–11.
Siligard, G., Drake, A. F., Mascagni, P., Rowlands, D. J., Brown, F. & Gibbons, W. A. (1991). A CD strategy for the study of polypeptide folding / unfolding. *Int. J. Pept. Protein Res.* **38**, 519–527.
Sonnichsen, 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.
Stellwagen, E., Park, S., Shalongo, W. & Jain, A. (1992). The contribution of residue ion pairs to the helical stability of a model peptide. *Biopolymers*, **32**, 1193–1200.
Stock, J. B., Stock, A. M. & Mottonen, J. M. (1990). Signal transduction in bacteria. *Nature (London)*, **344**, 395–400.
Storrs, R. W., Truckses, D. & Wemmer, D. E. (1992). Helix propagation in trifluoroethanol solutions. *Biopolymers*, **32**, 1695–1702.
Strehlow, K. G. & Baldwin, R. L. (1989). Effect of the substitution Ala-Gly at each of five residue positions in the C-peptide helix. *Biochemistry*, **28**, 2130–2133.
Strehlow, K. G., Robertson, A. D. & Baldwin, R. L. (1991). Proline for alanine substitutions in the C-peptide helix of ribonuclease A. *Biochemistry*, **30**, 5810–5814.
Terzi, E., Holzemann, G. & Seelig, J. (1994). Reversible coil-β-sheet transition of the Alzheimer β-amyloid fragment (25-35). *Biochemistry*, **33**, 1345–1350.
Venkatachallapathi, Y. V., Phillips, M. C., Epand, R. M., Epand, R. F., Tytler, E. M., Segrest, J. P. & Anatharamaiah, G. M. (1993). Effect of end group blockage on the properties of a class A amphipathic helical peptide. *Proteins: Struct. Funct. Genet.* **15**, 349–359.
Volz, K. & Matsumura, P. (1991). Crystal structure of *Escherichia coli* CheY refined at 1.7 Å resolution. *J. Biol. Chem.* **266**, 15511–15519.
von Dreele, P. H., Lotan, N., Ananthanarayan, V. S., Andreatta, R. H., Poland, D. & Scheraga, H. A. (1971). Helix-coil stability parameters for the naturally occurring amino acids in water. II. Characterization of host polymers and application of host-gues technique to random poly[(hydroxypropyl)glutamine-co-(hydroxybuthyl)glutamine]. *Macromolecules*, **4**, 408–417.
Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry*, **32**, 6337–6347.
Yumoto, N., Murase, S., Hattori, T., Yamamoto, H., Tatsu, Y. & Yoshikawa, S. (1993). Stabilization of α-helix in C-terminal fragments of neuropeptide Y. *Biochem. Biophys. Res. Commun.* **196**, 1490–1495.
Zhang, M., Yuan, T. & Vogel, H. J. (1993). A peptide analog of the calmodulin-binding domain of myosin light chain kinase adopts an α-helical structure in aqueous trifluoroethanol. *Protein Sci.* **2**, 1931–1937.
Zhou, N. E., Kay, C. M., Sykes, B. D. & Hodges, R. S. (1993). A single-stranded amphipatic α-helix in aqueous solution: design, structural characterisation, and its application for determining α-helical propensities of amino acids. *Biochemistry*, **32**, 6190–6197.
Zhou, H. X., Lyu, P., Wemmer, D. E. & Kallenbach, N. R. (1994*a*). α-Helix capping in synthetic model peptides by reciprocal side chain-main chain interactions: evidence for an N-terminal ''capping box''. *Proteins: Struct. Funct. Genet.* **18**, 1–7.
Zhou, H. X., Kay, C. M. & Hodges, R. S. (1994*b*). The role of interhelical ionic interactions in controlling protein folding and stability. *J. Mol. Biol.* **237**, 500–512.
Zimm, B. H. & Brag, J. K. (1959). Theory of the phase transition between helix and random coil. *J. Chem. Phys.* **31**, 526.

Edited by A. R. Fersht

(Received 28 June 1994; accepted 15 September 1994)