Elucidating the folding problem of helical peptides using empirical parameters

Victor Muñoz and Luis Serrano

Using an empirical analysis of experimental data we have estimated a set of energy contributions which accounts for the stability of isolated α-helices. With this database and an algorithm based on statistical mechanics, we describe the average helical behaviour in solution of 323 peptides and the helicity per residue of those peptides analyzed by nuclear magnetic resonance. Moreover the algorithm successfully detects the α-helical tendency, in solution, of a peptide corresponding to a β-strand of ubiquitin.

EMBL, Meyerhofstrasse 1, Heidelberg D-69117, Germany.

Experimental analysis of protein folding has revealed that secondary structure is stabilized very early in the folding process, while tertiary structure is acquired later[1]. One of the earliest steps in the folding process is the formation of secondary structure elements, which may or may not be concomitant with a hydrophobic collapse[2]. It follows that if we want to understand protein folding, we need to know which factors determine the tendency of short amino acid sequences to populate various conformational states in solution.

Attempts to describe the energetics of systems formed by short polypeptide chains have concentrated on α-helix formation. The most commonly used model, the helix-coil transition, is based on statistical mechanics theory. In its simplest version[3] there were only two parameters, a nucleation factor and an elongation factor, responsible for the helical tendency of a particular sequence. These were postulated to be independent of the sequence environment of the amino acid and much work was devoted to the experimental determination of nucleation and elongation factors for every amino acid[4]. Later the theory was modified in order to incorporate data concerning side chain-side chain interactions[5–7]. These modifications considerably improved the predictive power of the helix-coil transition theory, but still fell short of accurately explaining the helical behaviour of peptides with non-repetitive sequences in solution[8–13].

In recent years a large body of information has accumulated regarding the factors which govern the stability of α-helices in proteins[14–20] and the helical behaviour of both isolated protein fragments[21–42] and designed helical sequences in solution[43–71]. We have analyzed and integrated all this information in an attempt to develop a method for predicting with great accuracy the solution behaviour of helical peptides.

Theory

Our model attempts to obtain an energetic description of a physical system, the peptide under study, that is considered closed and in equilibrium. For the definition of phase-space we used the helix-coil transition leading to a two-state model for each residue. The random coil state consists of a heterogeneous ensemble of conformations in which the amino acid is able to explore all the allowed conformational space. An important characteristic of this state, as defined here, is the lack of specific interactions between two or more residues stabilizing any particular random coil conformation. From this definition each residue in the random coil state could be considered as a quasi-closed sub-system and is therefore energetically independent of the rest of the peptide. In the α-helical state the amino acid being considered, and at least three contiguous residues, are restricted to α-helical angles. In this state there are specific interactions between this residue and the contiguous residues in α-helical angles, but not with those outside the helical segment. The only exception to this are the amino- and carboxy-caps (see below), thus the minimum helical length is four residues in helical angles plus two caps. The α-helical residue is energetically dependent on the system, but only on those residues comprising the α-helix, since the others are in random coil state[72].

The conformational energy of the peptide as a whole, may be separated into a sum of energies, one for each

residue. This is suficient to formulate a residue partition function[72]:

$$Z_X = U_{rc} + U_H \quad (1)$$

where Z_X is the partition function for a residue X in a polypeptide chain, U_{rc} is the statistical weight of the random coil and U_H of the helical state. Giving the random coil a statistical weight of one, we can express the statistical weight of the helical state referenced (K).

$$Z_X = 1 + K \quad (2)$$

The helical state of a residue is not unique but an ensemble of energetically different conformations consisting of all the possible helices with different lengths that include this residue. The term K, is consequently a summation of the statistical weights of each of these helical conformations with respect to the random coil. Since residues outside the helical segment, forming either another α-helical segment or a random coil, are not energetically coupled, the residue partition function may be defined in a simplified way (considering only those residues participating in each helical conformation).

$$Z_X = 1 + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} Kji \quad (3)$$
$$\forall\, x \,||\, i+j \leq x < i \rightarrow K_{ji} = 0$$

where x is the index (residue number) of the residue, K is the statistical weight of the helical conformation, i is the index of the first amino-acid within the particular helix, n is the length of the peptide, and j is the number of amino-acids forming the helix. The sub-index ($\forall\, x \,||\, i+j \leq x < i \rightarrow K_{ji} = 0$), assigns a statistical weight of zero to all the helical segments of the peptide which do not include the residue, x, being analyzed. The partition function for each residue is therefore different since it participates in different helical segments. Although the energy of a particular helical segment is constant, it contributes differently to the helical propensity of each of its amino acids. The helical propensity of a certain residue in a sequence (fraction of the time the residue is in the helical state), arises directly from eq. 4.

$$\langle \mathrm{Hel}\, X_{residue} \rangle = \frac{\sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} Kji}{1 + \sum_{j=6}^{n} \sum_{i=1}^{(n-j+1)} Kji} \quad (4)$$
$$\forall\, x \,||\, i+j \leq x < i \rightarrow K_{ji} = 0$$

To determine the average helical content of a peptide it is then only necessary to calculate the helical propensity of each residue from the individual partition functions and calculate their average value.

This theoretical approximation reduces enormously the conformational space to be explored, since it is only necessary to calculate the energy of each possible helical segment. Despite the simple formulation, the definition of the random coil state allows this model to account for more than one non-overlapping helical segment in the polypeptide chain, without the necessity of calculating them as different helical conformations.

The main assumption of this approximation is that the random coil state of each amino acid as energetically independent of the whole system. A particular sequence of amino acids may have specific interactions in certain conformations other than an α-helix which could produce an overprediction of α-helicity; although this is unlikely 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 is small, because the helical tendency will be small.

Definition of parameters

The calculation of the statistical weight, K, of any particular helical segment with respect to the random coil is carried out:

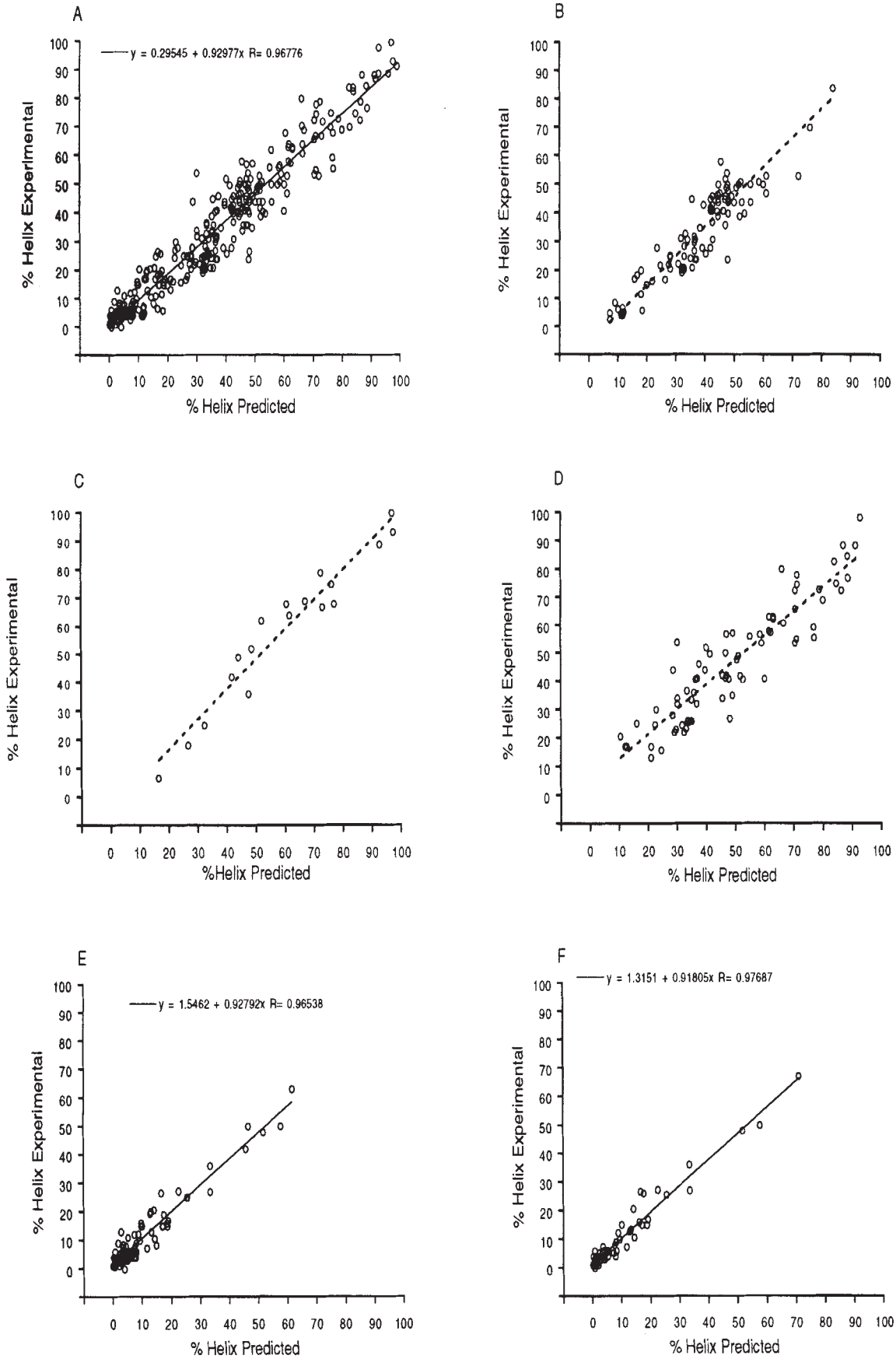
$$K = e^{-\Delta G_{Hel}/RT} \quad (5)$$

where ΔG_{Hel} is the difference in free energy between the random coil and helix states, and consists of a summation of contributions:

$$\Delta G_{Hel} = \Delta G_{Int} + \Delta G_{Hbond} + \Delta G_{SD} + \Delta G_{nonH} \quad (6)$$

ΔG_{Int} is the summation of the intrinsic tendency of the j residues to adopt helical dihedral angles. This term reflects the loss of conformational entropy and is expressed as the difference in free energy between the helix and ran-

Fig. 1 Correlation plots between the experimental CD average helical content (calculated using the method of Chen *et al.*[81]), and the predicted values using our method. All the peptides used in our database have been analyzed at 0–5 °C, pH 5–8 and less than 0.1 M NaCl, with the exception of some protected peptides without any i, i+1, or i, i+3 or i, i+4, electrostatic interactions[43,92]. In the case of the peptides containing His residues, we have used only measurements made below pH 6.0, where the His residue is protonated. To reproduce the same sigmoidal behaviour with peptide length found experimentally for polyalanine based peptides[59,92], we used the following parameters: A stabilizing, or destabilizing effect of the helix dipole of ±0.4 kcal mol^{-1}; a value of -0.92 kcal mol^{-1} per hydrogen bond, including the possible hydrogen bond made by the acetyl group at the N termini with a main-chain helical amide; an absolute helical intrinsic value for Ala, Glu and Lys of 0.8, 1.0 and 1.02 kcal mol^{-1} respectively, an electrostatic attraction between Glu and Lys at positions i, i+3 of ±0.15 kcal mol^{-1}. In the case of the peptides described by Chakrabartty *et al.*[60], we have introduced the correction indicated by the same authors[68]. *a*, Correlation with the total peptide database[21–71] (323 peptides). *b*, Correlation with those peptides designed to study capping interactions[43,63,67]. *c*, Correlation with those peptides designed to study electrostatic interactions[48,55,65]. *d*, Correlation with those peptides designed to study the intrinsic helical propensities of the twenty naturally occuring amino acids[46,50,52,53,56–58,60,69]. *e*, Correlation with 100 natural peptides corresponding to protein fragments[21–42]. The CD data at 4 °C for some of the peptides in the Fig. 1 were provided by M.A. Jimenez[29]. *f*, Correlation with 63 peptides not used in the refinement of the method[32–39,70,71]. 22 out of the 63 peptides are unpublished and they are described in Table 3.



dom coil states. There are several experimental data in proteins and peptides that assign relative free energies for helical stability to the twenty natural amino acids[19,20,52,73]. The hydrogen bond term is not included in these free energies, since they are relative to a reference amino acid and consequently reflect the relative intrinsic tendency of the different amino acids to be in helical angles. As a starting database we used the average values between the two different experimental scales that are most similar[19,73]. To calculate the absolute free energy values we have used the special characteristics of glycine. We considered that the energy contour of glycine in the Ramachandran plot is flat, implying two steps (one for the allowed and one for the forbidden region), and that the region corresponding to helical dihedral angles, (ϕ -130° to 50°; ψ -60° to 30°), is around 6% of the allowed area. Then, using the Boltzmann equation, we can estimate the amount of energy required to put a glycine in helical dihedral angles (ΔG= 1.62 kcal mol^{-1}). The absolute energy for the different amino acids would be the experimental ΔG with respect to glycine, plus 1.62 kcal mol^{-1}.

ΔG_{Hbond} is the sum of the net contribution of all the main chain–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 was obtained from calorimetric[74] and theoretical studies[75] (final value of 0.92 kcal mol^{-1}, Fig. 1). The typical pattern of the main chain–main chain hydrogen bonding network is between the CO group of residue i with the NH group of residue i+4, and consequently a helix of length n will have n-4 hydrogen bonds. The cooperativity of helix formation arises from this and from the entropic cost of fixing four residues in the helical conformation without a hydrogen bond (for four alanines this is 3.2 kcal mol^{-1}; Table 1). This could be considered similar to the classical nucleation factor σ, but in our model it is sequence dependent. The intrinsic value for each amino acid plus the hydrogen-bond contribution is similar to the classical elongation factor, s.

ΔG_{SD} is the sum of the net contribution of all the side chain–side chain interactions located at positions i,i+3 and i,i+4 within the helical region, with respect to the random-coil state, plus weakly attractive and repulsive coulombic interactions in the helical conformation between charged residues at positions i, i+1 (±0.05 kcal mol^{-1}). There are not accurate experimental values for all the i, i+3 and i, i+4 combinations of the twenty different amino acids, however it is possible to obtain a reasonable initial database and then refine it by looking at the experimental results for several different peptides[43–71]. We have used a statistical approach that has been developed and extensively utilised for protein-structure databases[76,77]. The mean force potential has been derived from a set of thirty protein fragments analyzed by NMR (Methods; V.M., M.A., Jimenez, M., Rico, & L.S., Serrano, manuscript in preparation). The ΔG_{SD} values obtained (Table 2) are derived from systems in real conformational equilibrium and not from a database of three dimensional structures.

ΔG_{nonH} is the sum of the net contribution to the stability of the helical region of all the residues that are

Table1 Free energies in kcal mol^{-1} for the intrinsic tendencies of the different amino-acids to be in helical dihedral angles, and for the N- and C-caps

	N-cap[a]	Capp.box[b]	Capp.box[c] (Pro N+1)	Asp+2[d]	C-cap[e]	Intrinsic[f]
A	0.00	0.08	0.00	-0.40	0.00	0.80
G	-0.75	-1.00	-1.00	-0.75	-0.40	1.62
S	-0.85	-1.75	-1.62	-0.85	-0.08	1.17
T	-0.62	-1.75	-1.62	-0.65	-0.08	1.27
N	-1.10	-1.45	-1.75	-1.10	-0.40	1.24
D	-1.10	-1.45	-1.75	-1.10	0.36	1.20
Q	0.15	0.00	0.00	-0.40	-0.08	1.01
E	-0.30	-0.45	-0.45	-0.40	0.08	1.02
H	0.08	0.08	0.08	-0.40	-0.40	1.34
K	0.00	0.08	0.00	-0.40	-0.10	1.00
R	0.00	0.08	0.00	-0.40	-0.25	0.89
F	0.08	0.08	0.08	-0.40	0.08	1.31
Y	0.08	0.08	0.08	-0.40	0.08	1.33
W	0.08	0.08	0.08	-0.40	0.08	1.33
L	0.08	0.08	0.08	-0.40	0.08	0.99
V	0.08	0.08	0.08	-0.40	0.08	1.40
I	0.08	0.08	0.08	-0.40	0.08	1.22
M	0.08	0.08	0.08	-0.40	0.08	1.11
C	0.08	0.08	0.08	-0.40	0.08	1.47
P	0.44	0.44	0.44	0.44	0.08	4.27

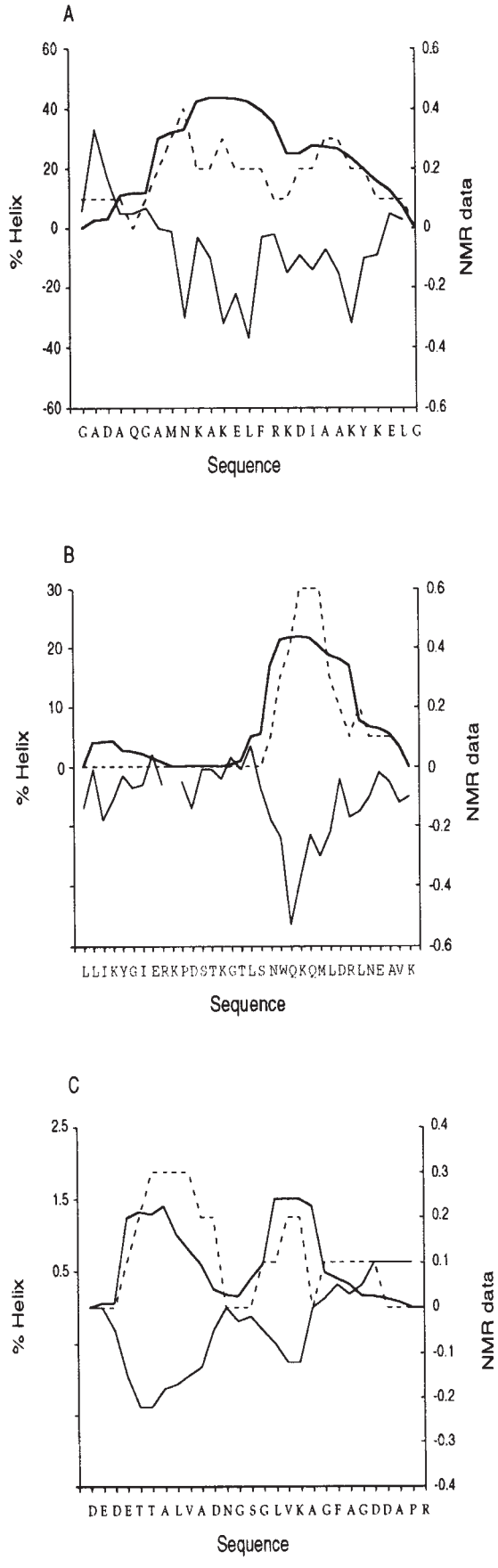
[a]N-cap values for the twenty different amino acids after refining the parameters.
[b]N-cap values of the different amino acid residues when there is a Glu, or Gln, residue at position N+3 and the possible formation of a capping box[79,80]. When an Asp was at position N+3, all the favourable values are multiplied by 0.75.
[c]N-cap values of the different amino acids when there is a Glu, or Gln at position N+3, and a Pro residue at position N+2. The result of having a Pro residue at position N+1 is a preferential stabilisation of the N-cap when there is an Asp or Asn residue[67].
[d]N-cap values when there is an Asp residue at position N+2. If there is not a good N-cap residue, an Asp residue at position N+2 can make a hydrogen-bond to the amide of the N-cap residue and stabilize the helix[17,80].
[e]C-cap values for the twenty amino-acids.
[f]Free energy required to put the twenty amino-acids in helical dihedral angles. This term does not include 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 1.32 kcal mol^{-1}. The correlation coefficients of these values with the data of O'Neil & DeGrado[73] and barnase[19] are 0.79 and 0.89, respectively.

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)[78]. The largest helix able to be formed in a peptide with n-residues consists of n-2 residues and the N- and C-caps. As a starting point we used the average data obtained from the experimental analysis of the the N- and C-caps of the two helices of barnase[14,15]. We have also considered as stabilizing interactions a special capping motif, the capping box[79,80], and the presence of an aspartate residue at position n+2 (ref. 79). Due to the complexity of the cap effect, which includes non-helical angles and hydrogen bond interactions, we could not refer to an absolute value. We have therefore used Ala as a reference value ($\Delta\Delta G = 0$), and rescaled the contribution of the capping effect during the refinement of the parameters (Methods). As a simplification, the interaction of the helix dipole with the charged amino acids was also included in the capping interaction of those residues.

Table 2 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)

	A	G	S	T	N	D	Q	E	H	K	R	Y	F	W	V	I	M	L
A	0.0	0.0	0.05	-0.05	-0.05	-0.05	-0.05	-0.15	-0.15	-0.15	-0.15	-0.13	-0.13	-0.13	-0.05	-0.15	-0.15	-0.25
	0.0	0.0	0.21	0.21	0.21	0.21	0.0	-0.1	-0.05	-0.1	-0.1	-0.1	-0.1	-0.1	-0.05	-0.1	-0.1	-0.2
G	0.2	0.4	0.6	0.6	0.6	0.6	0.0	0.0	0.02	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0	0.0
	0.3	0.6	0.6	0.6	0.6	0.6	-0.07	-0.07	0.0	0.1	0.1	0.17	0.17	0.17	0.4	0.4	0.4	0.4
S	0.15	0.4	0.4	0.4	0.6	0.6	0.0	0.0	0.02	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0	0.0
	0.1	0.6	0.6	0.6	0.6	0.6	-0.1	-0.1	-0.05	-0.1	-0.1	0.17	0.17	0.17	0.4	0.4	0.4	0.4
T	0.15	0.4	0.4	0.4	0.6	0.6	0.0	0.0	0.02	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0	0.0
	0.1	0.6	0.6	0.6	0.6	0.6	-0.1	-0.1	-0.05	-0.1	-0.1	0.17	0.17	0.17	0.4	0.4	0.4	0.4
N	0.15	0.4	0.6	0.6	0.6	0.6	0.0	0.0	0.02	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0	0.0
	0.1	0.6	0.6	0.6	0.6	0.6	-0.1	-0.1	-0.05	-0.1	-0.1	0.17	0.17	0.17	0.4	0.4	0.4	0.4
D	0.15	0.4	0.6	0.6	0.6	0.3	0.0	0.3	-0.1	-0.2	-0.2	0.05	0.05	0.05	0.0	0.0	0.0	0.0
	0.1	0.6	0.6	0.6	0.6	0.6	0.0	0.0	-0.2	-0.32	-0.4	0.17	0.17	0.17	0.4	0.4	0.4	0.4
Q	-0.0	-0.05	0.05	-0.05	-0.05	-0.05	0.0	-0.1	-0.05	-0.1	-0.1	-0.2	-0.2	-0.2	-0.15	-0.15	-0.15	-0.15
	0.0	0.0	0.0	0.0	0.0	-0.35	0.0	-0.1	-0.15	-0.25	-0.3	-0.26	-0.26	-0.26	-0.05	-0.05	-0.05	-0.05
E	-0.0	-0.05	0.05	-0.05	-0.05	0.3	-0.14	0.3	-0.1	-0.15	-0.15	-0.15	-0.15	-0.15	-0.1	-0.1	-0.1	-0.1
	0.0	0.02	0.0	0.0	0.0	0.0	0.0	0.0	-0.3	-0.42	-0.5	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.0
H	0.0	0.02	0.0	0.0	0.0	0.0	0.0	-0.2	0.1	0.15	0.15	-0.25	-0.2	-0.3	-0.11	-0.11	-0.11	-0.11
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.15	0.15	-0.25	-0.2	-0.3	-0.107	-0.107	-0.107	-0.107
K	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.2	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.02	0.0	0.0	0.0	0.15	0.0	0.1	0.2	0.3	0.3	-0.2	-0.2	-0.2	-0.15	-0.15	-0.15	-0.15
R	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.2	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.02	0.0	0.0	0.0	0.15	0.0	0.1	0.2	0.3	0.3	-0.2	-0.2	-0.2	-0.15	-0.15	-0.15	-0.15
F	0.0	0.13	0.13	0.13	0.13	0.13	0.0	-0.31	-0.25	-0.31	-0.31	-0.2	-0.2	-0.2	-0.25	-0.25	-0.25	-0.25
	0.0	0.32	0.32	0.32	0.32	0.32	0.0	0.0	-0.8	0.0	0.0	-0.5	-0.5	-0.5	-0.3	-0.3	-0.3	-0.3
Y	0.0	0.13	0.13	0.13	0.13	0.13	0.0	-0.31	-0.25	-0.31	-0.31	-0.2	-0.2	-0.2	-0.25	-0.25	-0.25	-0.25
	0.0	0.32	0.32	0.32	0.32	0.32	0.0	0.0	-0.85	0.0	0.0	-0.5	-0.5	-0.5	-0.3	-0.3	-0.3	-0.3
W	0.0	0.13	0.13	0.13	0.13	0.13	0.0	-0.31	-0.25	-0.31	-0.31	-0.2	-0.2	-0.2	-0.25	-0.25	-0.25	-0.25
	0.0	0.32	0.32	0.32	0.32	0.32	0.0	0.0	-0.9	0.0	0.0	-0.5	-0.5	-0.5	-0.3	-0.3	-0.3	-0.3
V	0.1	0.0	-0.0	0.0	0.0	0.0	-0.15	-0.15	-0.35	-0.15	-0.15	-0.45	-0.45	-0.45	-0.35	-0.35	-0.35	-0.35
	0.1	0.0	1.1	1.1	1.1	1.1	-0.08	-0.08	-0.1	-0.08	-0.08	0.0	0.0	0.0	-0.45	-0.45	-0.45	-0.45
I	0.10	0.0	-0.0	0.0	0.0	0.0	-0.15	-0.15	-0.35	-0.15	-0.15	-0.45	-0.45	-0.45	-0.35	-0.35	-0.35	-0.35
	0.1	0.0	1.1	1.1	1.1	1.1	-0.08	-0.08	-0.1	-0.08	-0.08	0.0	0.0	0.0	-0.45	-0.45	-0.45	-0.45
M	0.1	0.0	-0.0	0.0	0.0	0.0	-0.15	-0.15	-0.3	-0.15	-0.15	-0.45	-0.45	-0.45	-0.35	-0.35	-0.35	-0.35
	0.1	0.0	1.1	1.1	1.1	1.1	-0.08	-0.08	-0.1	-0.08	-0.08	0.0	0.0	0.0	-0.45	-0.45	-0.45	-0.45
L	0.1	0.0	-0.0	0.0	0.0	0.0	-0.15	-0.15	-0.35	-0.15	-0.15	-0.45	-0.45	-0.45	-0.35	-0.35	-0.35	-0.35
	0.1	0.0	1.1	1.1	1.1	1.1	-0.08	-0.08	-0.1	-0.08	-0.08	0.0	0.0	0.0	-0.45	-0.45	-0.45	-0.45

The lines in the table reflect the initial amino acid groups based on their physicochemical characteristics. The refinement of the side chain–side chain interactions resulted in slight modifications for some of the interactions, but the overall correlation between the initial and final values was good (r=0.9; data not shown). There is no data for the interactions between Pro, or Cis, with the other residues due to the small number of experimental cases.



Parameter refinement

The first problem we had to address concerned the differences in experimental conditions. Fortunately, the pH (5 to 7.0), salt concentration (10 to 100 mM) and temperature (0 to 5 °C) were similar for most cases, so only experiments performed under these conditions were used. More significant was the influence of chemical groups protecting both ends of the peptides which eliminates the charges at the ends and could act as capping residues providing a hydrogen bond donor or acceptor to the main-chain groups at the helix ends[43]. These factors were considered as explicit capping effects of the first and last amino acids of the peptide (Fig. 1).

It is clear that the contribution of individual residues, or interactions between particular residues, is subject to experimental and statistical error. To obtain statistically meaningful data for the side chain–side chain interactions, we initially grouped the amino acid residues (Methods). Subsequently the particular contributions of individual residues, or pairs of residues, were refined using the information present in synthetic peptides specifically designed to study defined contributions and interactions[21–31,43–69]. Correlation analysis of the parameters, without any refinement in the side chain–side chain interactions, with the peptide database resulted in a correlation coefficient of 0.7 (data not shown), thus indicating that our approximation and parameterization are accounting for α-helix stability and that the final correlation (Fig. 1*a*) is not an artifact of the refinement but only a fine tuning.

The final values for the intrinsic helical properties of the different residues, and the capping effects, are shown in Table 1. In Table 2 we show the final values for the i, i+3 and i, i+4 interactions for all the residues except Pro and Cis. All the values shown in Table 1 and 2 are in very good agreement with the experimental values determined for some specific interactions, such as ion-pairs, aromatic-histidine pairs, or hydrophobic interactions. There are some interactions in Table 2, such as those between an aliphatic residue and a small polar residue at position i+4, which are particularly unfavourable. This could be due to a bias in the database, but we found no reason to reduce it more.

Prediction of average helical populations

Fig. 1*a* illustrates the correlation between circular dichroism experimental results (the % helical population was calculated using the method of Chen *et al.*[81]), and the pre-

Fig. 2 Diagrams showing the agreement between the predicted helicity per residue (thick continuous lines) and that observed by NMR The upfield chemical shifts of the amide or Cα protons, with respect to random coil values[83] are shown as thin continuous lines. The total number of times a residue is found in a four-residue region having an i, i+3 NOE, divided by ten (in order to scale the right Y-axis for comparative purposes), is shown as a discontinuous line. *a*, helix H of myoglobin. *b*, a peptide corresponding to residues 532–565 of the adenylate cyclase from *B. antracis*. *c*, an actin fragment comprising residues 1–28, which forms an antiparallel β-hairpin and two coil regions in the protein structure.

Table 3 Sequence and circular dichroism estimation of the helical content in aqueous solution, of the unpublished peptides utilized in this work

Name	Sequence	Experimental conditons	Second. struct.	% helix
3FXC	NH_2-TYKVTELINEAEGINETIDCDD-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	β-hairpin	1
3LZM	NH_2-GFTNSLRMLQQKRWDEAVNLAKS-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	α–helix	10
3LZM-2	NH_2-GVAGFTNSLRMLQQKRWDEAAVNLAKS-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	α-helix	12
CIII	NH_2-ESLLERITRKLRDGWKRLIDIL-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	———	8
CIII-L	NH_2-ESLLERITRKL-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	———	15
CIII-R	NH_2-RDGWKRLIDIL-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	———	4
CIII-M	NH_2-RITRKLRDGWK-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	———	2
Sigma	NH_2-KVATTKAQRKLFFNLRKTKQRL-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	———	9
COMA1	NH_2-DHPAVMEGTKTILETDSNLS-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	4
COMA2	NH_2-EPSEQFIKQHDFSSY-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	3
COMA3	NH_2-VNGMELSKQILQENPH-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	6
COMA4	NH_2-EVEDYFEEAIRAGLH-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	20
COMA5	NH_2-KEKITQYIYHVLNGEIL-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	3
ARA1	NH_2-AVGKSNLLSRYARNEFSA-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	2
ARA2	NH_2-RFRAVTSAYYRGAVG-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	3
ARA3	NH_2-TRRTTFESVGRWLDELKIHSD-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	7.5
ARA4	NH_2-AVSVEEGKALAEEEGLF-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	4
ARA5	NH_2-STNVKTAFEMVILDIYNNV-COOH	pH7.0, 10mM Na-Phosphate, 4 °C	Helix	3
G1	NH_2-DTYKLILNGKTLKGETTTEA-COOH	pH3.0, water, 4 °C	β-hairpin	2
G2	NH_2-GDAATAEKVFKKIANDNGVD-COOH	pH5.1, 10mM Na-Acetate, 4 °C	Helix	4
G3	NH_2-GEWTYDDATKTFTVTE-COOH	pH5.1, 5mM Na-Phosphate, 4 °C	β-hairpin	2

The peptide 3FXC belongs to ferredoxin[88]. The peptide 3LZM-2 belongs to a β-hairpin in the protein[89]. The peptide 3LZM is a modified version of the previous peptide. The peptide CIII is a peptide corresponding to the bacteriophage λ cIII protein(McCarty *et al.*, manuscript in preparation). The peptides CIII-L, CIII-M and CIII-R are shorter versions of the same peptide. The peptide sigma is a peptide corresponding to the heat shock transcription factor σ (ref. 32) (McCarty *et al.*, manuscript in preparation). The peptides COMA1 to COMA5 correspond to the five α-helices of the COM-A protein, a member of the CheY family of proteins[85]. The peptides ARA1 to ARA5 correspond to the five α-helices of the ARA protein, a member of the H-ras family of proteins[89]. The peptides G1, G2 and G3 correspond to the first β-hairpin, α-helix and second β-hairpin, respectively, of the G-domain of the IgG binding protein[90].

dicted average helical content of protein fragments and designed peptides. A total of 323 peptides, corresponding to protein fragments[21–42] and designed sequences[43–71], were used in the analysis. The overall correlation coeficient is very good (r= 0.97), the slope is almost 1.0, (0.99 ± 0.01) and the line intersects the origin of the X,Y axes (0.51 ± 0.62). The correlation coefficient for individual groups of peptides (more than four peptides), is also higher in every case than 0.88 (with the exception of the peptides in ref. 59, r=0.75; data not shown). The linear regresion analysis of the group of peptides designed to study capping effects[43,63,67] shows a correlation of 0.95, a slope of 1.03 ± 0.03 and an intersection of the origin of -5.92 ± 1.5 (Fig. 1*b*), similar results are obtained for the group of peptides designed to study electrostatic interactions[48,55,65] (r=0.93, slope=1.05 ± 0.07 and intersection=-4.6 ± 4.4, Fig. 1*c*) as well as for that designed to study the intrinsic helical tendency of the commonly ocurring amino acids[46,50,52,53,56–58,60,69] (r = 0.91, slope=0.87 ± 0.04 and intersection = 3.82 ± 2.4, Fig. 1*d*). Fig. 1*e* shows the correlation between the predicted and observed helical content of protein fragments[21–42] (100 peptides). The correlation coefficient (r = 0.96),the slope (0.9 ± 0.03) and intersection of the origin (1.8 ± 0.39) are similar, indicating that we can succesfully predict the behaviour of peptides with more complicated sequences than those specifically designed to have a high α-helical content.

To show that the success of our method is not due to an overfitting of the initial parameters to the peptide database used to derive them, we set apart a number of peptides corresponding mainly to protein fragments[32–42] but also designed peptides[70,71] (63 sequences) to be used as a blind test. These peptides were not analyzed until the method was fully refined. They have no significant sequence homology with those analyzed before. The analysis of these peptides is shown in Fig. 1*f*. The data fit as well as for those peptides used to refine the method (r = 0.97). The slope is 1.09 ± 0.04 and the intersection of the origin is -2.1 ± 0.74.

Prediction of the helicity per residue

For protein folding, it is more important to correctly describe the helical behaviour of individual amino acids within a sequence than the overall behaviour of peptides in solution. Fortunately, NMR analysis can provide the experimental information, at a residue level, needed to test our results.

There is good correlation between the NMR analysis in water solution and the helicity per residue predicted by our method, for the monomeric helix H of myoglobin[40], and a peptide corresponding to residues 532–565 of the adenylate cyclase from *Bacilus antracis*[41] (Fig. 2). Our method is able to differenciate between helical and random coil regions within a peptide (Fig.2*b*), as well as to detect the existance of a non-uniform distribution of

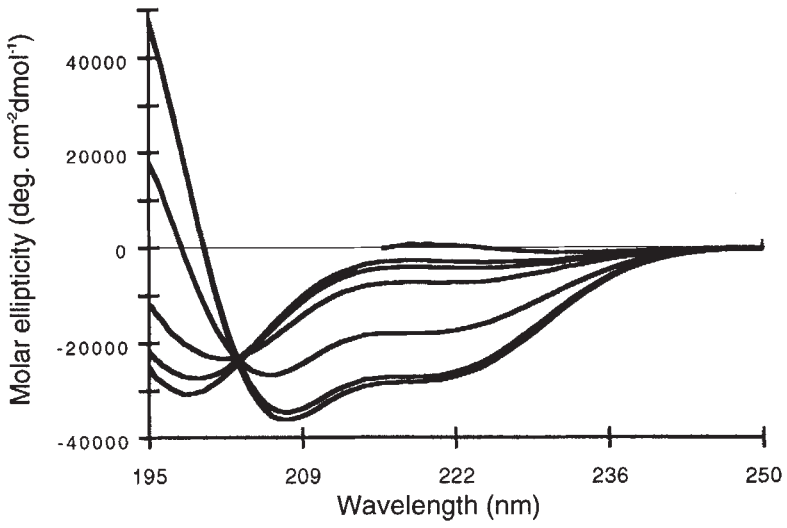


Fig. 3 Circular dichroism analysis of an unprotected peptide corresponding to the last β-strand of Ubiquitin. At pH5.0 there was no concentration dependence of the ellipticity of the peptide in the range 7 to 236 μM. The peptide was analyzed at 4 °C, in 5 mM sodium acetate pH 5.0. Baselines with the different solvents and in the same conditions were substracted from the peptide spectrum. In order of increasing negative ellipticity: the peptide in 6 M Guanidinium-HCl; with no addition of TFE; and in the presence of increasing concentrations of TFE, 7 % ,13 %, 20 %, 27.3 % and 34.4 %.

the helical population in a peptide (Fig.2*a*). We were also interested in determining the sensitivity of our method, when discriminating between regions with very small helical tendency and those with no tendency at all. Trifluoroethanol (TFE) apparently promotes helical formation only in those sequences that have a tendency to be helical in solution[25,26,27,82], and consequently can be used to detect low helical tendencies in unstructured peptides. An actin fragment comprising residues 1–28 which forms an antiparallel β-hairpin and two coil regions in the protein structure[42], is unstructured in water as we predicted (Fig. 2*c*). Addition of 80% TFE to the peptide results in the appearance of two helical regions as detected by the upfield chemical shifts for the Cα protons and the presence of i, i+3 NOEs. These two regions agree very well with the two regions in the peptide predicted to be more helical than the rest (around 1.5%), thus demonstrating the great sensitivity of our algorithm.

Helical prediction for non-helical regions

The experimental analysis, in aqueous solution, of peptides derived from protein fragments have shown either that they have no preferred conformation or that they populate a conformation similar to that in the folded protein[21–27,29–31,39–42]. To test the validity and applicability of our approach, we looked for β-strands in proteins that are predicted by our method to be helical in solution. One example is the sequence QKESTLHLVLRLRGG, which includes the last β-strand of ubiquitin[84] (residues STLHLVL), but is predicted by our method to have a 9% average helical content.

In Fig. 3 we show the CD analysis of this peptide in aqueous solution. The helical content of this peptide estimated by the method of Chen *et al.*[81] is around 8% at pH 5.0, in excellent agreement with our prediction. The negative ellipticity at 222 nm indicative of helical conformation was eliminated by guanidine-HCl, indicating that it was not due to spurious sequence effects in the spectrum. Moreover, addition of TFE significantly increased the helical content, and there is an isodicroic point indicating that we have a helix-coil two state transition. This result indicates that the secondary structure present in the protein does not necessarily reflect the intrinsic tendency of the amino acid sequence, and also explains why secondary structure prediction methods are not successful more than 60–70% of the time.

Concluding remarks

We have used a simplified approach that successfully predicts the helical behaviour of a large number of peptides in solution. The prediction is surprisingly good for peptides derived from protein sequences; this is probably because in these sequences a large number of different interactions are present, so that small errors in the quantification of the different parameters cancel each other out. In the case of polyalanine based peptides with a single interaction repeated several times, the errors would accumulate and the prediction may be worse. When more experimental data are available, further refinement will be needed, this can be accomplished in two ways: parameters which are the result of a combination of energy contributions (capping and dipole effects) could be split into the individual energy contributions, while differences could be introduced in the particular side chain–side chain interactions within the defined groups.

It is also possible to conclude that the major contributors to the helix stability already known (amino acid dependent entropic cost; n-4 hydrogen bonds between main chain groups; i+3 and i+4 side chain-side chain interactions; capping and dipole effects), are enough to describe in an appropriate framework the helical properties of the peptides experimentally analyzed to date.

Methods

Estimation of the side-chain side-chain interactions. In order to obtain an estimation of the free energy of interaction between different pairs of residues in the helical state, we have defined a potential of mean force similar to those previously derived from the protein structure database[76,77]. There are two main differences with the approaches mentioned above: the first is the higher simplicity of this potential since only two states (helix and coil) and therefore two distance intervals are considered. The second arises from the database used to

calculate the probability density functions that in this case is a set of thirty peptides in solution analyzed by NMR. We have looked at the residues whose NH or Cβ protons are involved in an NOE crosspeak with the Cα proton of a residue located at position i-3. If two residues, i and i+3, have a very favourable interaction in an α-helix, then the probability of their protons being close enough to produce an NOE crosspeak will be higher. If this is correct, then we should find that those interactions between side chains which are energetically favourable in a helical conformation will be more frequent, while those which result in unfavourable interactions will be less frequent. We have used the following peptides: Ribonuclease A and Angiogenin[27]. Thermolysin[29]; Plastocyanine peptides[26]; Haemoglobin[24], CheY, P21ras and Flavodoxin (V.M., M.A.J., M.R. & L.S., manuscript in preparation), and those analyzed by F. Blanco[86]. Peptides in solution are systems in equilibrium and it is possible to assume that their conformational behaviour follows a Bolzmann distribution. In other words the frequencies of the different conformations observed in the system are relative to their energies[76]. We assume that an i, i+3 NOE as defined above, is the result of these residues being in helical conformation[91]. The occurrence frequency between a certain pair of amino acids (i and i+3), in the peptide database, is related to their interaction energy in the helical conformation by eq. 7. In the case of the i, i+4 interactions, we assumed that the four residues are simultaneously in helical conformation when two NOEs are observed (i, i+3; i+1, i+4), and therefore the frequency of ocurrence of both NOEs simultaneously is related to the interaction energy between the two residues at i, i+4, by the same equation:

$$\Delta G_{int} = -RT\ln(\text{observed} / \text{possible}) \qquad (7)$$

where 'observed' are the detected i, i+3 NOEs and 'possible' are the number of times a specific pair of residues at i, i+3 positions are present. Assuming that the interaction between two alanine residues at positions i,i+3, or i, i+4 is null, then we consider the ΔG_{SD} value for this specific pair of residues as zero, and we refer the ΔG_{SD} values for the rest of the interaction pairs to it (V.M., M.A.J., M.R. & L.S, manuscript in preparation).

The problem for an analysis of this type is that a large database is needed in order to avoid statistical errors. One way to circumvent this problem is by grouping the amino acids according to their chemical and physical properties. We have grouped the amino acids according to their volume, polarity and hydrophobicity. The energy calculated with eq. 7 is consequently the average interaction energy between the defined amino acid groups. In the case of His, since we did not have enough cases in the database, we used as initial values those from Lys and Arg. Also, we have taken care that within each group there is no amino acid which exhibits a different behaviour from the other members of the group.

Received 17 December 1993; accepted 27 April 1994.

Parameter refinement. The variation of the average helical content of a peptide made of simple repeating units, $((AAXAA)_n$, where X is any amino acid), with the number of repeating units depends of the balance between the intrinsic helical tendency of the residues in the repeating unit, the contribution of the hydrogen bond, and, if the peptide is blocked, of the stabilizing effects of the N- and C-terminal blocking groups. If the peptide is not blocked the number of parameters is the same since the effect of the interaction of the charges at the N and C termini with the helix dipole needs to be taken into account. We have considered that the effect of an acetyl blocking group at the N terminus is equivalent to having an extra main-chain–main-chain hydrogen bond in the first turn of the helix[43,92]. The effect of a saccinyl blocking group is therefore that of the actyl plus -0.4 kcal mol^{-1} arising from the favourable interaction of its negative charge with the helix dipole (Fig. 1 legend). The effect of blocking the C terminus is much softer than when the N terminus is blocked[43], and so we have set the stabilising contribution at 50% of the N terminus. Thus we have only two parameters, the contribution of a hydrogen bond and the intrinsic helical tendency of the different amino acids (in the case of the set of peptides of different lengths used to refine these parameters, Ala and Lys)[92]. Once these two parameters were calibrated, we analyzed those peptides designed to study the intrinsic helical propensity of the different amino acids[46,50,52,53,56–58,60,69], to refine the values for their intrinsic helical propensities. Then those peptides with only one type of interaction between side-chains were considered[48,55,65]. Finally we analyzed those peptides specifically designed to study capping interactions[43,63,67], in this case, we found that the initial scale resulted in too strong an effect on the helical content of these peptides, and therefore we needed first to alter the relative scale derived from proteins, and second to refine the differences between the twenty amino acids as was done for the other parameters. With these refined parameters we analyzed a set of peptides with different lengths, based on repetitions of the following sequence $(AEAAKA)_n$ (ref. 59), to recheck the hydrogen bond value. Finally the more complex peptides with different types of interactions (that is, natural sequences from proteins), were analyzed[21–31,44,45,47,49,51,54,61,62,64,66,68]. Once this process was finished we reanalyzed the peptides in order to do a final tuning of the parameters. During all the refinement we used the experimental values as constraints. If the calorimetrically determined value for a hydrogen bond is between 0.9 and 1.1 kcal mol^{-1} (ref. 76), we did not exceeded those limits. The same can be said for the electrostatic interactions (0.3–0.5 kcal mol^{-1} for a Glu-Lys ion pair at i, i+4) and so on. The only exception being the capping interactions.

Note added in proof: The program AGADIR, containing the algorithm implemented in C-language, runs in a 486 based platform in around 20 sec for 100 residues, and is available on request .

Acknowledgements We are very grateful to T. Creighton for his useful remarks and suggestions that helped us to improve this work. We are also grateful to A.R. Ortiz for his comments on the manuscript.

1. Serrano, L., Matouschek. A.R. & Fersht, A.R. The folding of an enzyme VI: the folding pathway of Barnase. Comparison with theoretical models. *J. molec. Biol.* **224**, 847–859 (1992).
2. Matthews, C.R. Pathways of protein folding. *A. Rev. Biochem.* **62**, 653–683 (1993).
3. Zimm, B.H. & Brag, J.K. Theory of the phase transition between helix and random coil in polypeptide chains. *J. chem. Phys.* **31**, 526 (1959).
4. von Dreele, P.H, Lotan, N., Ananthanarayan, V.S, Andreatta, R.H., Poland, D. & Scheraga, H.A. Helix-coil stability parameters for the naturally occurring amino acids in water. II. Characterization of host polymers and application of host-guest technique to random poly[(hydroxypropyl)glutamine-co-(hydroxybuthyl)glutamine]. *Macromolecules* **4**, 408–417 (1971).
5. Finkelstein, A.V., Badretinov, A.Y. & Ptitsyn, O.B. Physical reasons for secondary structure stability: α-helices in short peptides. *Proteins Struc. Func. Genet.* **10**, 287 (1991)
6. Ptitsyn, O.B. Secondary structure formation and stability. *Curr. Opin. struct. Biol.* **2**, 13–20 (992).
7. Finkelstein, A.V. & Ptitsyn, O.B. A theory of protein molecule self organization. IV. Helical and irregular local structures of unfolded protein chains. *J. molec. Biol.* **103**, 15 (1976).
8. Gans, P.J., Lyu, P.C., Manning, M.C., Woody, R.W. & Kallenbach, N.R. The helix-coil transition in heterogeneous peptides with specific side-chain interactions: theory and comparison with CD spectral data. *Biopolymers* **31**, 1605–1614 (1991).
9. Chen,C.C., Zhu, Y., King, J.A. & Evans, L. A molecular thermodynamic approach to predict the secondary structure of homopolypeptides in aqueous systems. *Biopolymers* **32**, 1375–1392 (1992).
10. Qian, H. Single-residue substitution in homopolypeptides: perturbative helix-coil theory at a single site. *Biopolymers* **33**, 1605–1616 (1993).
11. Dyson, H.J. & Wright, P. Peptide conformation and protein folding. *Curr. Opin. struct. Biol.* **3**, 60–65 (1993).
12. Fukugita, M., Lancaster, D. & Mitchard, M.G. Kinematics and thermodynamics of a folding heteropolymer. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6365–6368 (1993).
13. Park, S.H., Shalongo, W. & Stellwagen, E. Residue helix parameters obtained from dichroic analysis of peptides of defined sequence.

Biochemistry **32**, 7048–7053 (1993).
14. Serrano, L., Sancho, J., Hirshberg, J. M. & Fersht., A.R. 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 glicine or serine at solvent -exposed surfaces. *J. molec. Biol.* **227**, 544–549 (1992).
15. Serrano, L. & Fersht, A.R. Capping and alpha-helix stability. *Nature* **342**, 296–299 (1989).
16. Nicholson, H., Anderson, D.E., Dao-Pin, S. & Matthews, B.W. Analysis of the interaction between charged side chains and the α-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* **30**, 981–9828 (1991).
17. Bell, J.A., Becktel, W.J., Sauer, U. Baase, W.A. & Matthews, B.W. Dissection of helix capping in T4 Lysozyme by structural and thermodynamical analysis of six aminoacid substitutions at Thr59. *Biochemistry* **31**, 3590–3596 (1992).
18. Sancho, S., Serrano, L. & Fersht, A.R. Histidine Residues at the N- and C-termini of α-Helices: Perturbed pKa_s and Protein Stability. *Biochemistry* **31**, 2253–2258 (1992).
19. Horovitz, A., Matthews, J. & Fersht, A.R. Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *J. molec. Biol.* **227**, 560–568 (1992).
20. Blaber, M., Zang, X. & Matthews, B. Structural basis of amino acid α helix propensity. *Science* **260**, 1637–1640 (1993).
21. Jimenez, M.A., Nieto, J.L., Herranz, J., Rico, M. & Santoro, J. ^{1}H NMR and CD evidence of the folding of the isolated ribonuclease 50-61 fragment. *FEBS Lett.* **221**, 320 (1987).
22. Jimenez, M.A., Rico, M, Herranz, J., Santoro, J. & Nieto, J.L. ^{1}H-NMR assignment and folding of the isolated ribonuclease 21-42 fragment. *Eur. J. Biochem.* **175**, 101–109 (1988).
23. Goodman, E.M.& Kim, P.S. Folding of a peptide corresponding to α-helix in bovine pancreatic typsin inhibitor. *Biochemistry* **28**, 4333–4337 (1989).
24. Peña, M.C., Rico, M., Jimenez, M.A., Herranz, J., Santoro, J. & Nieto, J.L. Conformational properties of the isolated 1-23 fragment of human hemoglobin α-chain. *Biochem. biophys. Acta.* **957**, 380–389 (1989).
25. Dyson, H.J., Merutka, G. Waltho, J.P., Lerner, R.A. & Wright, P.E. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. I Myohemerythrin. *J. molec. Biol.* **226**, 795–817 (1992).
26. Dyson, H.J., Merutka, G. Waltho, J.P., Lerner, R.A. & Wright, P.E. Folding of peptide fragments comprising the complete sequence of proteins. Models for initiation of protein folding. II Plastocyanin. *J. molec. Biol.* **226**, 817–835 (1992).
27. F.J. Blanco, Jimenez, M.A., Rico, M., Santoro, J., Herranz, J. & Nieto, J.L. The homologous angiogenin and ribonuclease N-terminal fragments fold into very similar helices when isolated. *Biochem. biophys. res. Commun.* **182**, 1491–1498 (1992).
28. Brown, J.E. & Klee, W.A. Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry* **10**, 470–476 (1971).
29. Jimenez, M.A. *et al.* CD and 1H-NMR studies on the conformational properties of peptide fragments from the C-terminal domain of thermolysin. *Euro. J. Biochem.* **211**, 569–581 (1993).
30. Shin, H.C., Merutka, G., Waltho, J. P. Wright, P.E. & Dyson, H.J. 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 (1993).
31. Shin, H.C., Merutka, G., Waltho, J.P, Tennant, L.L., Dyson, H.J. & Wright, P.E. Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry* **32**, 6356–6364 (1993).
32. Precheur, B., Siffert, O. Barzu, O. & Craescu, C.T. NMR and circular dichroic studies on the solution conformation of a synthetic peptide derived from the calmodulin-binding domain of *Bordetella pertussis* adenylate cyclase. *Euro. J. Biochem.* **196**, 67–72 (1991).
33. Mammi, S, Foffani, M.T., Improta, S., Tessari, M., Schierano, E. & Peggion, E. Conformation of uteroglobin fragments. *Biopolymers* **32**, 341 (1992).
34. Sancho, J., Neira, J.L. & Ferhst, A.R. An N-terminal fragment of Barnase has residual structure similar to that in a refolding intermediate. *J. molec. Biol.* **224**, 749–758. (1992).
35. Cushman, J.A., Mishra, P.K., Bothner, A.A. & Khosla, M.S. Conformations in solution of angiotensin II, and its 1-7 and 1-6 fragments. *Biopolymers* **32**, 1163–1171 (1992).
36. Kuroda, Y. Residual helical structure in the C-terminal fragment of cytochrome C. *Biochemistry* **32**, 1219–1224 (1993).
37. Yumoto, N., Murase, S., Hattori, T., Yamamoto, H., Tatsu, Y. & Yoshikawa, S. Stabilization of α-helix in C-terminal fragments of neuropeptide Y. *Biochem. biophys. res. Comm.* **196**, 1490–1495 (1993).
38. Doughty, M.B. & Hu, L. The contribution of helical potential to the in vitro receptor binding activity of a neuropeptide Y N-terminal deletion fragment. *Biopolymers* **33**, 1201–1206 (1993).
39. Kemmink, J. & Creighton, T.E. Conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor (BPTI) and their roles in folding. *J. molec. Biol.* (in the press).
40. Waltho, J.P., Feher, V.A., Merutka, G. Dyson, H.J. & Wright, P.E. 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 (1993).
41. Munier, H. *et al.* Characterisation of a Synthetic Calmodulin-binding Peptide derived. *J. biol. Chem.* **268**, 1695 (1993).
42. Sonnichsen, F.D., Van Eyk, J.E., Hodges. R.S. & Sykes, B.D. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry* **31**, 8790–8798 (1992).
43. Chakrabarty, B.D., Doig, A.J. & Baldwin, R.L. Helix N-cap propensities in peptides parallel those found in proteins. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
44. Mitchinson, C. & Baldwin, R.L. The design and production of semisynthetic ribonucleases with increased thermostability by incorporation of S-peptide analogs with enhanced helical stability. *Proteins Struct. Func. Genet.* **1**, 23–33 (1986).
45. Schoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M. & Baldwin, R.L. Tests of the helix dipole model for stabilisation of α-helices *Nature* **326**, 563–567 (1987).
46. Strehlow, K.G. & Baldwin, R.L. Effect of the substitution Ala-Gly at each of five residue positions in the C-peptide helix. *Biochemistry* **28**, 2130–2133 (1989).
47. Lyu, P.C., Marky, L. A. & Kallenbach, N.R. The role of ion pairs in α-helix stability: two new designed helical peptides. *J. Am. chem. Soc.* **111**, 2733–2734. (1989).
48. Marqusee, S., Robbins, V.H. & Baldwin, R.L. Unusually stable helix formation in short alanine-based peptides. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5286–5290 (1989).
49. Fairman, R., Shoemaker, K.R., York, E.J., Stewart, J.M. & Baldwin, R.L. Further studies of the helix dipole model: Effects of a free α-NH_3^+ or α-COO^- group on helix stability. *Proteins Struc. Func. Genet.* **5**, 1–7 (1989).
50. Merutka, G. & Stellwagen, E. Positional independence and additivity of amino acid replacements on helix stability in monomeric peptides. *Biochemistry* **29**, 894–898 (1990).
51. Merutka, G., Lipton, W., Shalongo, W., Park, S.H. & Stellwagen, E. Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry* **29**, 7511–7515 (1990).
52. Lyu, P.C., Liff, M.I., Marky, L.A. & Kallenbach, N.R. Side chain contributions to the stability of alpha-helical structures in peptides. *Science* **250**, 669–673 (1990).
53. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T.M. & Baldwin, R.L. Relative helix-forming tendencies of nonpolar amino acids. *Nature* **344**, 268–270 (1990).
54. Shoemaker, R.K. *et al.* Side Chain interactions in the C-peptide Helix Phe8...His12. *Biopolymers* **29**, 1 (1990).
55. Merutka, G. & Stellwagen, E. Effect of amino acid ion pairs on peptide helicity. *Biochemistry* **30**, 1591–1594 (1991).
56. Lyu, P.C., Wang, P.C., Liff, M.I. & Kallenbach, N.R. Local effect of glycine substitution in a model helical peptide. *J. Am. chem. Soc.* **113**, 3568–3572 (1991).
57. Padmanabhan, S. & Baldwin, R.L. Straight-chain non-polar amino acids are good helix-formers in water. *J. molec. Biol.* **219**, 135–137 (1991).
58. Strehlow, K.G., Robertson, A.D. & Baldwin, R.L. Proline for alanine substitutions in the C-peptide helix of ribonuclease A. *Biochemistry* **30**, 5810–5814 (1991).
59. Scholtz, M., Qian, H., York, E.J., Stewart, J.M. & Baldwin, R.L. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* **31**, 1463–1470 (1991).
60. Chakrabartty, A., Schellman, J.A. & Baldwin, R.L. Large differences in the helix propensities of alanine and glycine. *Nature* **351**, 586–588 (1991).
61. Lyu, P.C., Zhou, H.X., Jelveh, N., Wemmer, D.E. & Kallenbach, N.R. Position-dependent stabilizing effects in α-helices: N-terminal capping in synthetic model peptides *J. Am. chem. Soc.* **114**, 6560–6562 (1992).
62. Stellwagen, E., Park, S., Shalongo, W. & Jain, A. The contribution of residue ion pairs to the helical stability of a model peptide. *Biopolymers* **32**, 1193–1200 (1992).
63. Lyu, P.C., Wemmer, D.E., Hongxing, X.Z., Pinker, R.J. .& Kallenbach, N.R. Capping interactions in isolated α helices: Position-dependent substitution effects and structure of a Serine-capped peptide helix. *Biochemistry* **32**, 421–425. (1993).
64. Huygues-Despointes, B.M.P., Scholtz, J.M. & Baldwin, R.L. Effect of a single aspartate on helix stability at different positions in a neutral

alanine-based peptide. *Prot. Sci.* **2**, 1604–1611 (1993).
65. Huyghues-Despointes, B.M.P, Scholtz, J.M. & Baldwin, R.L. Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Prot. Sci.* **2**, 80–85 (1993).
66. Venkatachallapathi, Y.V., *et al.*. Effect of end group blockage on the properties of a class A amphipathic helical peptide. *Proteins Struct. Func. Genet.* **15**, 349–359 (1993).
67. Forood, B., Feliciano, E.J. & Nambiar, K.P. Stabilization of α-helical structures in short peptides via end capping. *Proc. natn. Acad. Sci. U.S.A.* **90**, 838–842 (1993).
68. Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R.L. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **32**, 5560–5565. (1993).
69. Zhou, N.E., Kay, C.M., Sykes, B.D., Hodges, R.S. 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 (1993).
70. Fairman, R., Armstrong, K.M., Shoemaker, K.R., York, E.J., Stewart, J.M. & Baldwin, R.L. Position effect on apparent helical propensities in the C-peptide helix. *J. molec. Biol.* **221**, 1395–1401 (1991).
71. Scholtz, J.M., York, E.J., Stewart, J.M. & Baldwin, R.L. A Neutral water-soluble, α-helical peptide: The effect of ionic strength on the helix-coil equilibrium. *J. Am. chem. Soc.* **113**, 5102 (1991).
72. Flory, P.J. *Statistical mechanics of chain molecules* (Hansen Publishers, 1989).
73. O'Neil, K.T. & DeGrado, W.F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **250**, 646–650 (1990).
74. Scholtz, J.M., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M. Bolen, D.W. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2854–2858 (1991).
75. Ooi, T. & Obatake, M. Prediction of the thermodynamics of protein unfolding: the helix-coil transition of poly(L-alanine). *Proc. natn. Acad. Sci. U.S.A.***88**, 2859–2863 (1991).
76. Sippl, M. Calculation of conformation ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. molec. Biol.* **213**, 859–883. (1990).
77. Rooman, M.J., Koger, J.P.A. & Wodak, S.J. Extracting information on folding from the aminoacid sequence: accurate predictions for protein regions with preferred conformations in the absence of tertiary interactions. *Biochemistry* **31**, 10226–10238 (1992).
78. Richardson, J.S. & Richardson, D.C. Amino acid preferences for specific locations at the ends of α helices. *Science* **240**, 1648–1652 (1988).
79. Dasgupta, S. & Bell, J.A. Design of helix ends. *Int. J. peptide protein Res.* **41**, 499–511 (1993).
80. Harper, E.T. & Rose,G.D. Helix stop signals in proteins and peptides: the capping box. *Biochemistry* **32**, 7605–7609 (1993).
81. Chen, Y.H. , Yang, J.T. & Chau, K.H. Determination of the helix and β-form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
82. Storrs, R.W., Truckses, D. & Wemmer, D.E. Helix propagation in trifluoroethanol solutions. *Biopolymers* **32**, 1695–1702 (1992)..
83. Jimenez, M.A., Blanco, F.J., Rico, M., Santoro, J., Herranz, J. & Nieto, J.L. Periodic properties of proton conformational shifts in isolated protein helices. *Eur. J. Biochem.* **207**, 39–71 (1992).
84. Vijay-Kumar, S., Bugg, C.E. & Cook, W.J. Structure of Ubiquitin refined at 1.8 Å resolution. *J. molec. Biol.* **194**, 531–544 (1987).
85. Stock, J.B., Stock, A.M. & Mottonen, J.M. Signal transduction in bacteria. *Nature* **344**, 395–400 (1990).
86. Blanco. F.J. "Papel de los giros β en el plegamiento de proteinas: estudio por resonancia magnetica nuclear". Universidad Complutense de Madrid. PhD dissertation (1992).
87. Ogawa, K. *et al.* X-ray analysis of fereedoxin from S. *platensis.* **84**, 1645–1652 (1978).
88. Weaber, L.H. & Matthews, B.W. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. molec. Biol.* **193**, 189 (1987).
89. Valencia, A., Chardin, P., Wittinghofer, A., & Sander, C. The ras protein family: evoultionary tree and role of conserved amino acids. *Biochemistry* **30**, 4637–4648 (1991).
90. Gronenborg *et al*. A novel, highly stable fold of the Immunoglobulin binding domain of Streptococcal protein G. *Science* **253**, 657–661 (1991).
91. Wutrich, K. *NMR of proteins and nucleic acids* (Wiley-Interscience, 1986).
92. Rohl, C.A, Scholtz, J.M., York, E.J. Stewart, J.M. & Baldwin, R.L. Kinetics of proton exchange in helical peptides of varying chain lengths: interpretation by the Lifson-Roig equation. *Biochemistry* **31**, 1263–1269 (1992).