The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding

Victor Muñoz, Francisco J. Blanco and Luis Serrano

A recurrent local structural motif is described at the amino terminus of α-helices, that consists of a specific hydrophobic interaction between a residue located before the N-cap, with a residue within the helix (i,i+5 interaction). NMR and CD analysis of designed peptides demonstrate its presence in aqueous solution, its contribution to α-helix stability and its role in defining the α-helix N terminus limit. Comparison between the N-terminal structures of the peptide and those in proteins with the same fingerprint sequence, shows striking similarities. The change in the polypeptide chain direction produced by the motif suggests an important role in protein folding for residues located in polypeptide segments between secondary structure elements.

EMBL, Postfach 102209, Meyerhofstrasse 1, 69012 Heidelberg, Germany

Correspondence should be addressed to L.S.

It is known that the folded conformation acquired by a protein is determined by its primary sequence[1], however, we still cannot predict the three-dimensional structure of a protein from its sequence alone. What makes the solution of this problem so difficult is that the folded state is unique within a vast number of possible conformations, but proteins find their folded conformation by following preferred folding pathways and not through a random search of the conformational space[2,3]. The solution of the problem could therefore be simplified by deciphering the rules underlying the folding pathways. In fact, current protein folding models atribute to local interactions a critical role in defining the protein folding pathways[4]. The identification of local structure motifs that would be sufficiently stable to reduce the conformational flexibility at certain points of the polypeptide chain is thus quite important.

The search for local stable motifs has been focused mainly on secondary structure elements rather than the region between two of these elements[5–7]. Statistical analyses of protein databases have indicated that certain hydrophobic amino acids are especially frequent in the sequence segment preceding an α-helix (position N' before the N-cap)[6,7], as well as at position N4 within the α-helix[6,7] (we adopt here the same nomenclature for α-helices adopted by Richardson and Richardson[6]: N"-N'-Ncap-N1-N2-N3-N4-.....-C4-C3-C2-C1-Ccap-C'-C"; where N1 to C1 belong to the helix, Ncap and Ccap are the boundary residues, (the first and last residues whose Cα lays on the helix axis, and N', N", C' and C" are non-helical residues). Very recently it has been reported that these two hydrophobic residues may be associated with the presence of a capping-box motif[5]: the expanded-box motif[8].

In this work we further investigate the role of two hydrophobic residues at positions N' and N4. For this purpose we have extended the statistical analysis previously published by other groups and carried out NMR and CD studies in aqueous solution of designed peptides.

Statistical analysis

We have looked at the frequency of the simultaneous appearance of a hydrophobic residue (Leu, Ile, Val, Met and Phe) at positions N' and N4 of α-helices in a structure database of 279 proteins. Our results show that two hydrophobic residues at these positions are more abundant than expected from a random distribution, with or without the simultaneous presence of the capping-box, suggesting that they could make a favourable interaction[9] (Table 1). The fact that the frequency of appearance of hydrophobic residues in these positions without the capping-box is higher than expected (2.9 times), indicates that rather than an expansion of the capping-box motif, as has been previously described[8], the interaction between these hydrophobic residues may be a motif by itself. The distribution of the possible hydrophobic pairs is biased towards some specific residue pairs, indicating that there may be some stereospecificity in the interaction between the two hydrophobic residues (Fig. 1). We designate this putative local supersecondary interaction between two hydrophobic residues at positions N' and N4 as the hydrophobic-staple motif. The motif is very frequent occurring in one out of six α-helices in the protein database.

Table 1 The hydrophobic staple contributes to the determination of the helix N terminus limits

Fingerprint[1]	Database[2]	α_9[3]	α_{9expec}[4]	$freq_{\alpha 9}$[5]	$c_2\alpha_7$[6]	$c_2\alpha_{7expec}$[7]	$freq_{c2\alpha 7}$[8]
XCXXXXXXX	14091	1295	1219	1.06	636	303	2.10
HXXXXHXXX	3834	457	603	0.76	177	90	1.97
HCXXXHXXX	979	79	109	0.72	108	23	4.70
XCXXEXXXX	1354	135	160	0.84	168	40	4.20
HCXXZHXXX	862	76	95	0.80	58	20	2.90
HCXXEHXXX	117	3	14	0.21	50	3	17.00

[1]Sequence fingerprint: X is any amino acid; H is Phe, Leu, Ile, Val or Met; C is Ser, Thr, Asn or Asp; Z is any amino-acid except Glu or Gln; and E is Glu or Gln.
[2]Number of nine-residue segments which have each sequence fingerprint in the protein structure database.
[3]Number of nine-residue segments in helical conformation which have each sequence fingerprint.
[4]Number of expected nine-residue segments if the residue distribution were random.
[5]Normalized frequency of occurrence for the conformation.
[6]Number of nine-residue segments with the two first residues in a non-helical conformation and the other seven in helical conformation, which have each sequence fingerprint.
[7]Number of expected nine-residue segments in column 6, if the residue distribution were random.
[8]Normalized frequency of occurrence for the conformation. The search was done in a protein database of 279 proteins with less than 50 % homology filtered for the quality of data[24,25], with the program WHATIF[26].

Comparison of the normalized frequency of occurrence of the two hydrophobic residues at positions N' and N4 with that of appearance within the α-helix, reveals that the hydrophobic staple might help to define the α-helix amino terminus limit (Table 1). The presence of the hydrophobic-staple sequence fingerprint (Leu, Ile, Val, Met or Phe at positions i and i+5) can define the α-helix starting point, as can the presence of a good capping residue (Asp, Asn, Ser or Thr), but both motifs together have an additive effect defining the α-helix starting point even better than a capping-box motif[5,7]. The combination of a capping-box and a hydrophobic-staple motif is clearly greater than either alone, showing a cooperative effect in defining the α-helix starting point. These results show that different motifs can cooperate to define the helix N terminus limit.

Solution structure

To experimentally asses the local nature of the hydrophobic staple motif we have designed a peptide with the following sequence:

GFSKAELAKARAAKRGGY

This peptide has a hydrophobic-staple motif (residues Phe 2 and Leu 7) and a capping-box motif (Ser 3 and Glu 6). A capping box has been included to increase the effect of the hydrophobic staple as illustrated by the statistical analysis. It has been studied by NMR in aqueous solution without any organic solvent (Fig.2, Table 2). Although there is some signal overlapping at the carboxy terminus of the peptide, the presence of several i, i+3 NOEs, as well as the upfield conformational shifts, show the existence of a highly populated α-helical conformation starting at Lys 4, which gradually decreases up to residue Arg 15. The abrupt change in the CαH Δδ values between Ser 3 and Lys 4 and the dNN(i,i+1) NOE intensities indicates that the α-helix starts at Lys 4 and there is no fraying at the N terminus of the α-helix. The CαH resonance of Phe 2 is significantly downfield shifted with respect to random coil values, indicating that it adopts dihedral angles in the β-strand region of the Ramachandran plot[10,11]. There are several NOE crosspeaks between the side chains and

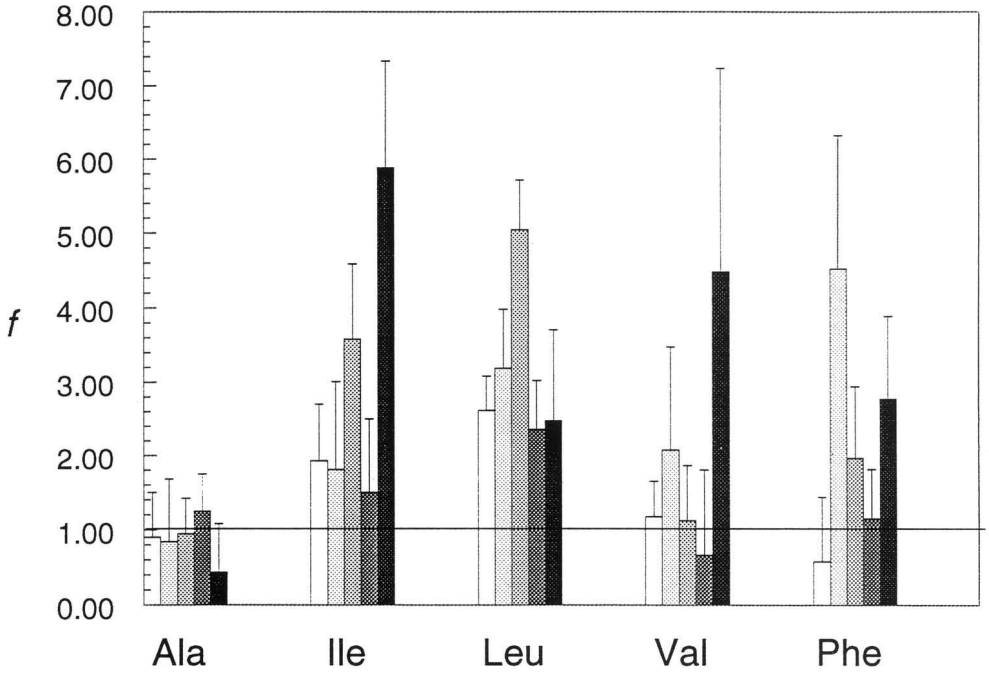

Fig. 1 Histogram of normalized pairwise residue frequencies for positions N' and N4 in the hydrophobic staple motif, when the N-cap residue is a Ser, Thr, Asn or Asp. The five hydrophobic-residues at N' are shown in the X-axis. For each, five histogram bars are shown that correspond to the same residues at position N4. In increasing order of greyness, the bars corresponding to Ala, Ile, Val, Leu and Phe. The respective *f* value of each pair is shown on the y-axis. All values greater than 1 imply that these pairs occur at a greater frequency at positions N' and N4 of helices than expected by chance.

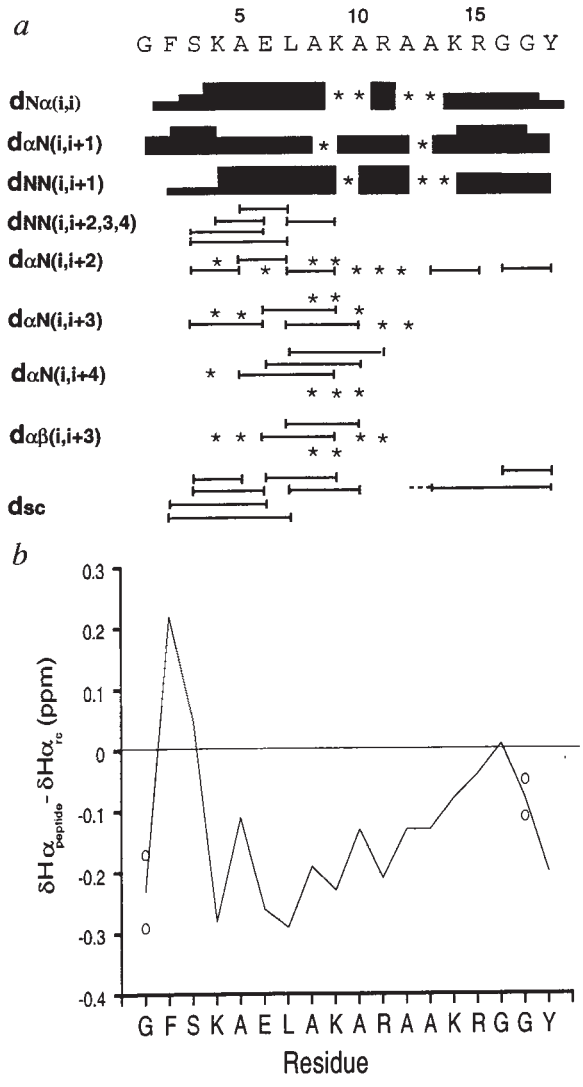

Fig. 2 *a*, summary of the NOEs connectivities found for the F-L peptide at 278 K, pH 6.1 in aqueous solution. Asterisks refer to NOEs that cannot be detected due to signal overlap. NOEs involving side chains are grouped as dsc. The dashed line indicates that it was not possible to differentiate between the methyl protons of Ala 12 and Ala 13. *b*, Plot of the CαH protons conformational shifts against the peptide sequence. The F-L peptide was designed to have a moderate helical content (~25 %) and high solubility. To achieve this we used the algorithm AGADIR, developed for predicting the helical content of monomeric peptides in aqueous solution[21–23]. Several Lys and Arg residues were placed along the sequence to prevent peptide aggregation. A capping-box motif[5,7], was included to strongly nucleate the α-helix at Ser 3 (N-cap)[12]. The peptide has free ends to prevent spurious effects of the blocking groups on the helical content of the peptide[21,22]. The Phe-Leu pair was choosen for NMR analysis because any possible contact between the side-chain protons of these two residues should be easily identified in a clean area of the NOESY spectrum, free of other resonances. Leu was used at position N4 because it is the hydrophobic residue with one of the highest intrinsic helical propensities (excluding Ala)[14–19].

the backbone amide protons of Ser 3 and Glu 6, which are typical of the formation of a capping box motif[12], but most remarkably we find several NOEs between the side chain protons of Phe 2 and those of Glu 6 and Leu 7 (a weak NOE is also observed with the methyl protons of Ala 10).

Calculation of an NMR structure of a peptide in aqueous solution is difficult because of conformational averaging affecting the NMR parameters. In our peptide, however, the number of non-sequential NOEs at the N terminus of the peptide (residues 2–7), the large differences of the CαH protons with respect to random-coil values (Fig. 2) and the small coupling constants $3J_{HN\alpha}$ for some of the residues (Table 3), are similar to those found in proteins[11,13], thus indicating that this region of the peptide is quite rigid. This permits the calculation of the structure in the N-terminal half of the peptide. The calculated structures (Fig. 3*a*), fullfil all the NMR constraints and show the formation of a well defined helix with a capping box involving Ser 3 (N-cap) and Glu 6 (N3). Phe 2 (N') adopts dihedral angles in the β-region of the Ramachandran plot and its side chain makes a small hydrophobic pocket with the side chains of Glu 6 and Leu 7 (N4).

Table 2 Spin-Spin couplings $^3J_{HN\alpha}$

	$^3J_{HN\alpha}$
Gly1	—
Phe2	5.9
Ser3	6.4
Lys4	3.5
Ala5	5.1
Glu6	5.7
Leu7	5.1
Ala8	3.9
Lys9	5.5
Ala10	4.3
Arg11	5.1
Ala12	4.9
Ala13	5.6
Lys14	6.0
Arg15	6.3
Gly16	5.3, 4.4
Gly17	6.3, 6.3
Tyr18	8.1

The Spin-Spin couplings $^3J_{HN\alpha}$ were carefully measured in a proton 1D spectrum with high digital resolution and signal to noise ratio. Resolution enhancement functions were used prior to Fourier transformation, thus minimising the distorting effect of the linewidth in some of the amide signals.

Effects on helix stability

Once experimentally established that the hydrophobic staple motif is present in the template peptide we can analyze its relative contribution to α-helix stability through the circular dichroism (CD) analysis of the helical content of a series of mutant peptides in which the residues at positions N' and N4 are varied (Fig. 4; Table 3). In these peptides we have substituted the Phe residue at position N', by Gly, Ala, Val, Ile, Met and Leu, and the Leu at position N4 by Ala. The peptides with a Gly at position N' and either Ala or Leu at position N4 (G-A and G-L, respectively), are used as the reference peptides. The peptide with Ala at positions N' and N4 (A-A), shows the same helical content as the peptides G-A and G-L. This suggests that in these peptides the α-helix does not bypass Ser 3 (N-cap), as was observed by NMR in the F-L peptide, since Ala has a much higher helical propensity than Gly[14–19]. Introduction of Ala, Val, Ile, Leu, Met or Phe at position N' increases the helical content when there is a Leu at position N4, showing that the interaction between the hydrophobic residues at these positions stabilizes the helical conformation (an Ala-Leu i,i+5 interaction is geometrically feasible as found in the protein structure database; data not shown). As expected from the statistical analysis of the preferences for different hydrophobic pairs at the hydrophobic staple motif, the helical contents of the peptides vary widely depending on the N'-N4 pair. Ala (A-L), Val (V-L) or Ile (I-L) increase slightly the helical content with respect to the reference peptides (G-A, G-L and A-A) while Leu (L-L), Met (M-L) and Phe (F-L), at position N' more than double it (Table 3).

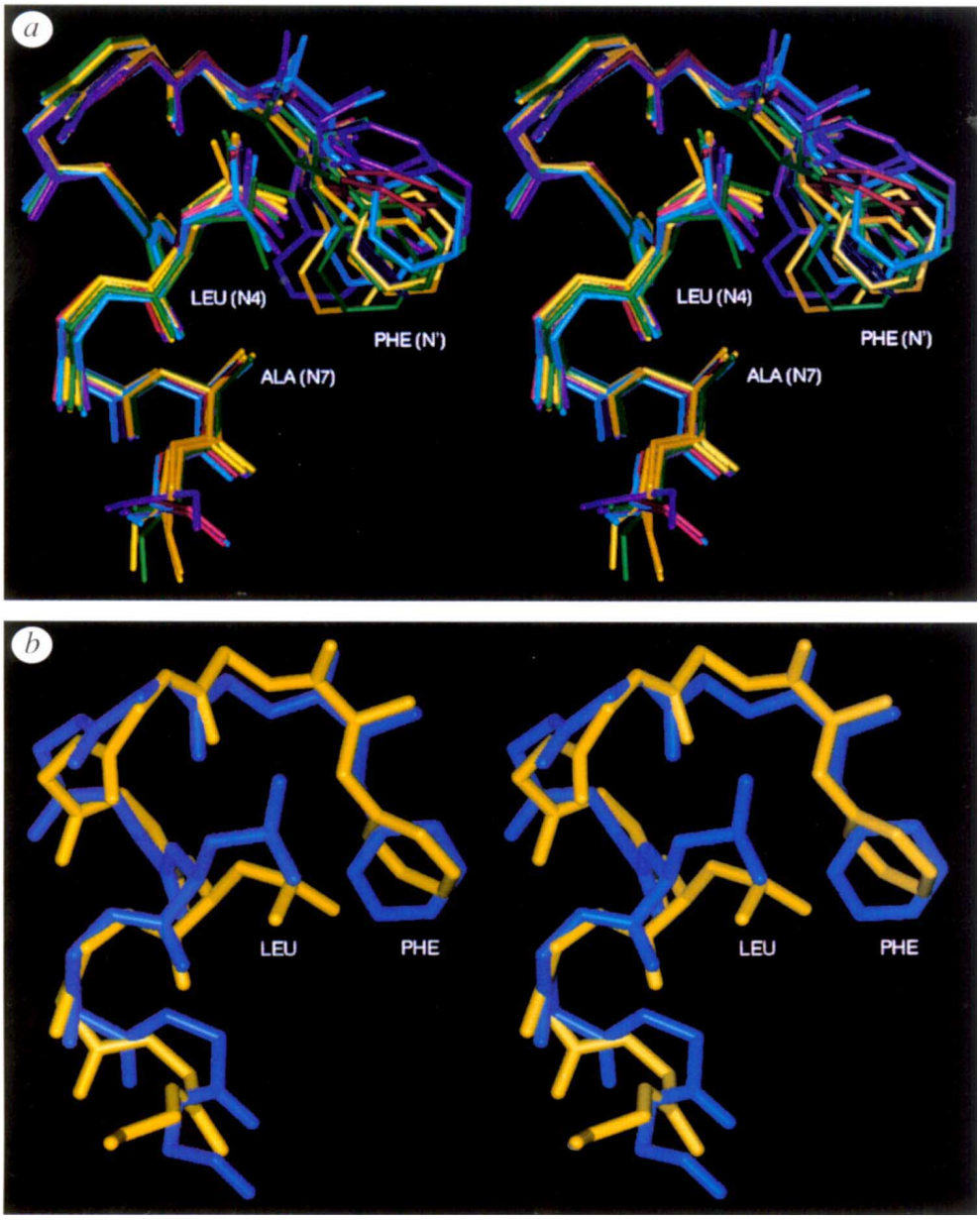

Fig. 3 *a*, Superposition of the 14 best structures of the peptide F-L in water solution. The backbone for residues Phe 2 to Arg 11 and the side chains of Phe 2 (N') and Leu 7 (N7) are shown. *b*, Superposition of one of the conformations in peptide F-L and the beginning of the α-helix of 1alc (residues 3–12).

The hydrophobic-staple, which appears to be a recurrent motif in protein α-helices involving only local specific interactions, is an autonomous structural unit as it is found in an isolated peptide. The participation together in the same local motif of amino acids in the α-helix and in the preceding segment is especially noteworthy. In this work we have found that two hydrophobic residues at position N' and N4 of an α-helix produce two important effects.

Firstly, it helps to define the N terminus of an α-helix and it stabilizes the α-helix through a hydrophobic interaction between the side-chains of a loop-residue and an α-helical residue. The strength of the interaction is very dependent on the specific stereochemistry of the two residues. Our statistical analysis indicates that the hydrophobic-staple is independent of other motifs and previous work has shown its formation in the absence of a capping-box[20]. Its presence in phase with a capping-box motif enhances this effect (Table 1), as indicated by the high frequency in which both motifs are found in the same α-helix (almost one out of three α-helices containing at least one of them). The combination of the hydrophobic-staple and the capping-box motif provides a large stabilization (only three helices out of 53, found in the protein structure database, bypass the motifs), so that it is tempting to define them as a stereochemical rule for the determination of the N-termini of α-helices[8]. It is very likely that certain protein sequences may provide other local interactions, however, which can counteract the energy input of these motifs, invalidating the rule[20]. Quantification of the energetics of all of the different local interactions involved is then necessary to predict local structures[21,22]. (The energetic determination of the i,i+5 hydrophobic interaction carried out by the CD analysis of the different peptides is now being incorporated in the program AGADIR[21–23].)

Secondly, this motif together with a capping-box motif determine the polypeptide chain direction as it comes into the α-helix. In the peptide F-L a reversal of the direction of the polypeptide main-chain is produced by the residues at positions N' and N-cap adopting dihedral angles in the β-region of the Ramachandran plot (average values, $\phi = -104 \pm 36$, $\psi = 131 \pm 33$; $\phi = -92 \pm 34$, $\psi = 163, \pm 9$; calculated from the 14 best structures). The reversal of the polypeptide chain is also observed in the database of protein structures and has been reported[8]. The 50 cases found containing a capping box plus a hydrophobic-staple, show that the N' and N-cap residues have dihedral angles very similar to those calculated in the peptide F-L (average values, $\phi = -106 \pm 22$, $\psi = 141 \pm 17$ and $\phi = -88 \pm 20$, $y = 164 \pm 16$, respectively; data not shown). The superposition of one of the structures of the F-L peptide, with the two unique α-helices that have a similar sequence fingerprint (Phe-Thr/Ser-X-X-Glu-Leu, structures 1alc and 1rec of the protein database), produces root-mean-square deviations of 0.1 nm and 0.08 nm for the backbone (Fig. 3*b*). This agreement suggests that the presence of the two motifs determine the local

conformation at the N terminus of α-helices not only in our template peptide but also in proteins.

The reversal of the direction of the polypeptide chain plus the fact that the motif is stable by itself in aqueous solution, suggests an important role in the early stages of protein folding for loop residues preceding an α-helix. The interaction between the side chains of the loop residue N' and the α-helical residue N4 may promote the encounter of the α-helix with the preceding secondary structure element, accelerating protein folding and defining a folding pathway.

Methods

Protein database analysis. The protein database used is based on the one described by Hobohm *et al.*[24], including 279 proteins with less than 50 % homology[25]. This database is actually implemented in the program WHATIF[26]. The conformational searches were done with the SCAN3D option of the same program using the Kabsch & Sander definition of secondary structure[27].

Calculation of the probability of finding different fingerprint sequences in a helical conformation. The probability of finding a certain fingerprint sequence in a helical conformation if there is a random distribution, is calculated by multiplying the individual probabilities of each residue type of the fingerprint ($P_{fingerprint}$).

A individual probability is the probability of finding a specific residue type in a certain position when this position adopts a helical conformation. To calculate them, we have divided the protein database in three-residue segments and counted the number of those in which the central position is helical (N_{conf} = 19,422). The individual probability of a specific residue type (P_{ind}) is the number of these segments which contain this residue-type in the central position (N_{res}), divided by the total number (N_{conf}).

$$P_{ind} = N_{res}/N_{conf} \quad (1)$$

For a hydrophobic residue (Leu,Ile,Val,Met and Phe; H in Table 1), we find 5,830 cases (P_{ind} = 0.30). For Asp, Asn,Ser and Thr, (C in Table 1), we find 3,489 cases (P_{ind} = 0.18). Finally, for Glu and Gln (E, in Table 1), we find 2,533 cases (P_{ind} = 0.13). Whenever there is no specific residue-type at one position of the nine-residue segment in a particular fingerprint (X in Table1), its individual probability is 1.

The length of the fingerprints we have analyzed is nine-residues. The specific residue-types are located in the first six positions and the extra three residues are included to prevent any sequence dependence due to proximity of the helix C-cap. The number of cases expected ($N_{expected}$), for each sequence fingerprint in a nine-residue helical segment (see Table 1), is calculated by the following equation:

$$N_{Expected} = P_{fingerprint} * N_{total} \quad (2)$$

where N_{total} is the total number of nine-residue segments adopting the helical conformation (6,738 cases).

Table 3 Mean residue ellipticity at 222 nm obtained from the far-UV CD spectra of the corresponding peptides and percentage helical population

Peptide	Ellipticity 222 nm	% Helix 222 nm[1]	R1[2]
G–A	–6224	18.4	18
A–A	–6236	18.4	18
G–L	–7011	20.7	19
A–L	–8248	24.4	24
V– L	–7964	23.5	24
I–L	–8391	24.8	26
L–L	–13310	39.3	42
M–L	–13891	41.0	40
F–L	–14393	42.5	52

Each peptide is denoted with a label in which the first position corresponds to the N' residue (residue 2 of the peptide) and the last position corresponds to the N4 residue (residue 7 of the peptide).
[1]The helical content was determined by the method of Chen *et al.*[29].
[2]Helical content determined by the method of Bruch *et al.*[30]

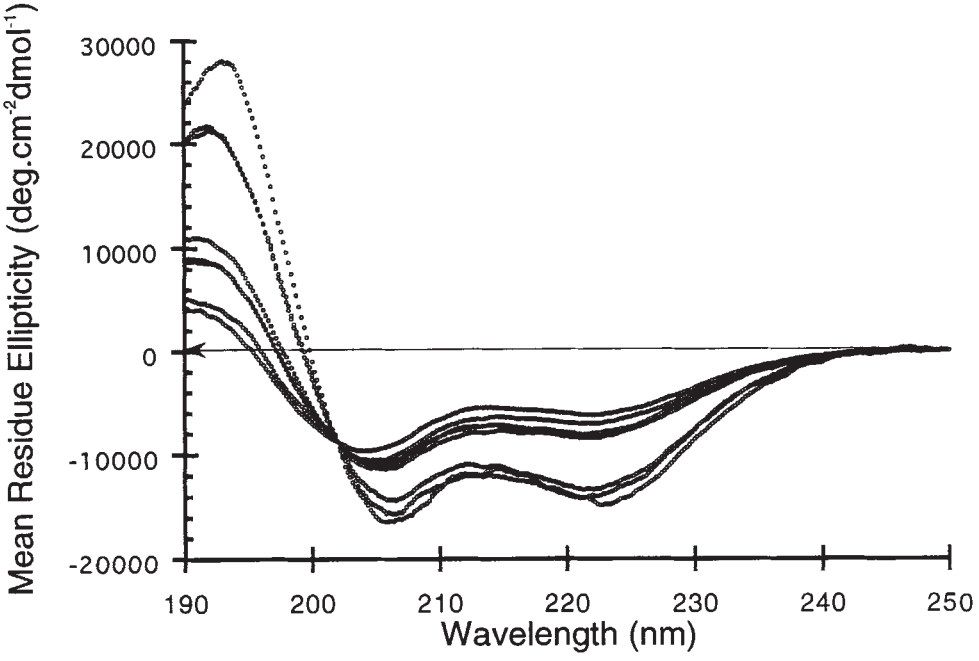

Fig. 4 Far-UV CD spectra of the peptides G-A, G-L, A-A, A-L, V-L, I-L, L-L, M-L and F-L (in order of decreasing ellipticity at 222nm).

Calculation of the probability of finding different fingerprint sequences, when the two first residues are not in a helical conformation and the following seven are helical. The calculations of the probabilities for the different fingerprints have been carried out in a similar way as described above. The only difference resides in the individual probabilities of the first two amino acids, since they are not in a helical conformation. To calculate them, we have divided the protein database in three-residue segments and counted the number of those in which the central position is non-helical (N_{conf} = 40,228). The individual probability of a specific residue type in a non-helical conformation (P_{ind}) is the number of these segments which contain this residue-type in the central position (N_{res}), divided by the total number (N_{conf})(eq. 1).

For a hydrophobic residue (Leu,Ile,Val,Met and Phe; H in Table 1), we find 10,225 cases (P_{ind} = 0.254). For Asp, Asn,Ser and Thr, (C in Table 1), we find 10,249 cases (P_{ind} = 0.255).

The number of cases expected ($N_{expected}$), for each fingerprint in a nine-residue segment in which the first two residues are

in non-helical angles (see Table 1) is calculated using equation 2. In this case N_{total} is 1,183.

Calculation of the normalized pairwise residue frequencies for positions N' and N4 in the hydrophobic staple motif. The normalized frequency of occurrence, *f*, is calculated using the following equation:

f = ((number of specific pairs)/(number of all pairs)) / ((number of specific pairs in database)/(number of all pairs in database)). Where number of specific pairs is the number of times a particular hydrophobic pair (Leu-Leu), is found in our database at positions N' and N4 of α-helices. Number of all pairs is the number of all the N'-N4 pairs in helices. The number of specific pairs in the database is the number of times a particular hydrophobic pair of residues at positions i and i+5 at any conformation is found in the protein database. Number of all pairs in database, is the number of i and i+5 pairs at any conformation in the protein database. The error bars were calculated by adding or substracting two cases to the number of times a particular hydrophobic pair (that is Leu-Leu), is found in our database at positions N' and N4 of a-helices. The resulting values were used to calculate two *f* values (*f–2* and *f+2*). The difference between the two *f* values provides a rough estimation of the reliability of the data. In the case of Met, the error calculated was nearly as great as the calculated *f* value (data not included).

NMR analysis. NMR samples were prepared in a $H_2O/^2H_2O$ 9:1 (by vol.) mixture, at pH 6.1, 278 K. DQFCOSY, TOCSY and NOESY spectra were performed in a BRUKER AMX 500MHz spectrometer using standard procedures[13]. Sodium 3-trimethylsyl (2,2,3,3-$^2H_4$)propionate was used as an internal reference. The proton resonances were assigned by the sequential assignment procedure[13] NOESY spectra (mixing time=300 ms) of 10 mM and 1 mM peptide samples were identical within the experimental error; no significant differences were found in the chemical shifts, line widths and relative NOE intensities. The circles refer to the values for the two CαH of Gly. The CαH proton conformational shifts were obtained by substracting the random-coil chemical shift values[13] from those measured in the peptide.

Calculation of the structure. NOE crosspeak intensities were classified by visual inspection of the NOESY spectrum into four categories: strong, medium, weak and very weak. These categories were translated into upper limit distance constraints of 0.3 nm, 0.35 nm, 0.4 nm and 0.5 nm, respectively, with the proper pseudoatom correction for the stereotopic atoms. 150 structures were generated using a set of 26 sequential and 46 non-sequential distance restraints plus 3 φ angle restraints (derived from $J^3_{HN\alpha}$< 5Hz) with the distance geometry program DIANA[28]. The fourteen best structures (no distance restraint violations greater than 0.01 nm and no van der Waals contact exceeding by more than 0.002 nm the sum of the atomic radii), were chosen as a representative ensemble of the folded peptide. The r.m.s. deviation for the Phe 2 to Arg 11 residues, is 0.05 nm for the backbone atoms. The following side chain-side chain NOEs were included as restrictions: βE6–βF2; βE6-2,6H F2; βE6–γL7; βE6–βγK9; δL7-2,3,5,6H F2; δL7 –βF2; δL7–βA10; γL7–β F2; βA10-2,3,5,6 H F2.

CD analysis. The CD spectra were acquired in a JASCO-710 dichrograph, using the continuous mode with 1 nm bandwith, 1 s response and a scan speed of 50 nm min^{-1}. The samples were prepared in 50 mM pH 7.0 phosphate buffer at 278 K. To check the concentration dependence of the ellipticity at 222 nm, two spectra at 20 μM (5 mm path length cuvette) and 1mM (0.1 mm path length cuvette) peptide concentration were done.

Received 19 January; accepted 27 February 1995.

Acknowledgements
We are very grateful to T. Creighton for critical reading of the manuscript. F. J. Blanco was supported by an HCMP postdoctoral fellowship from the Commission of the European Communities.

1. Anfinsen, C. B. The kinetics of formation of of native ribonuclease during oxydation of the reduced polypeptide chain. *Science* **181**, 223–230 (1973).
2. Chreighton, T. E. Protein folding. *Biochem. J.* **270**,1–16 (1990).
3. Karplus, M. & Weaver, D. L. Protein folding dynamics. *Biopolymers* **18**, 1421–1437 (1979).
4. Dill, K. A., Fiebig, K.M. & Chan, H.S. Cooperativity in protein-folding kinetics. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1942–1496, (1993).
5. Harper, J. & Rose, G.D. Helix stop signals in proteins and peptides: the capping box. *Biochemistry* **32**, 7605–7609 (1993).
6. Richardson, J.S. & Richardson, D.C. Amino acid preferences for specific locations at the ends of α-helices. *Science* **240**, 1648–1652 (1988).
7. Dasgupta, S. & Bell, J.A. Design of helix ends. *Int. J. Peptide Res.* **41**, 499–511 (1993).
8. Seale, J.W., Srinivasan R. & Rose, G.D. Sequence determinants of the capping-box, a stabilizing motif at the N-termini of α-helices. *Prot. Sci.* **3**, 1741–1745, (1994).
9. Sippl, M. Calculation of conformational 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).
10. Jiménez, 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 Letts* **221**, 320–324 (1987).
11. Wishart, D.S., Sykes, B.D. & Richards, F.M. Relationship between nuclear magnetic resonance chemical shifts and protein secondary structure. *J. molec. Biol.* **222**, 333–344 (1991).
12. 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).
13. Wüthrich, K. *NMR of proteins and nucleic acids*. (John Wiley & sons, New York, 1986).
14. Lyu, P.C., Liff, M.I., Marky, L.A. & Kallenbach, N.R. Side chain contributions to the stability of α-helical structures in peptides. *Science* **250**, 669–673 (1990).
15. 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).
16. Serrano, L., Neira, J.L., Sancho, J. & Fersht, A.R. Effect of alanine versus glycine in α-helices on protein stability. *Nature* **356**, 453–455 (1992).
17. Horovitz, A. Matthews, J. & Fersht, A.R. α-helix stability in proteins. II. Factors that influence stability at an internal position. *J. molec. Biol.* **227**, 560–568 (1992).
18. Blaber, M., Zang, X. & Matthews, B. Structural basis of amino acid α-helix propensity. *Science* **260**, 1637–1640 (1993).
19. Chakrabartty, T. Kortemme & R. L. Baldwin. Large differences in the helix propensities of alanine and glycine. *Prot. Sci.* **3**, 843–847 (1994).
20. Jiménez, M.A, Muñoz, V., Rico, M. & Serrano, L. Helix stop and start signals in peptides and proteins: The capping box does not necessarily prevent helix elongation. *J. molec. Biol.* **242**, 487–496 (1994).
21. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. *Nature struct. Biol.* **1**,399–409 (1994).
22. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters, II: helix macrodipole effects and rational modification of the helical content of natural peptides. *J. molec. Biol.* **245**, 275–296. (1995).
23. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters, III: Temperature and pH dependence. *J. molec. Biol.* **245**, 297–308. (1995).
24. Hobohm, U., Scharf, M., Schneider, R. & Sander, C. Selection of representative protein datasets. *Prot. Sci.* **1**, 409 (1992).
25. Muñoz, V. & Serrano, L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Prot. Struct. Funct. Genet.* **20**, 301–311 (1994).
26. Vriend, G. WHATIF: A molecular modeling and drug design program. *J. molec. Graph.* **8**, 52–56 (1990).
27. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern of recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
28. P. Güntert, W. Braun, K. Wüthrich. Improved efficiency of protein structure calculations from NMR using the program DIANA with redundant dihedral angle constraints. *J. molec. Biol.* **217**, 517–530 (1991).
29. Chen, Y.H., Yang, J.T. & Chow, K.H Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).
30. Bruch, M.D., Dhingra, M. M. & Gierasch, L.M. Side chain backbone hydrogen bonding contributes to helix stability in peptides derived from an α-helical region of carboxypeptidase A. *Prot. Struct. Func. Genet.* **10**, 130 (1991).