Biol Res 29: 213-225 (1996)

Interactions that regulate the helical fold in proteins

HILDA CID, FELIPE GAZITUA and MARTA BUNSTER

Departamento de Biología Molecular, Facultad de Ciencias Biológicas,
Universidad de Concepción, Concepción, Chile

Several factors that may contribute to the stabilization of the helical structure in proteins, detected in studies made on short synthetic peptides, have been reported. Some of them are: presence of alanine or leucine, ionic-pair bonding, stabilization of the helical dipole moment by appropriate charges at the helix N- and C-caps, and aromatic interactions of amino acids located at positions i, i + 4. An analysis of 54 helical structures from 12 proteins showed that all these stabilizing factors were also present in proteins, but the influence of any of them had a different weight, according to the distribution of the hydrophobic and hydrophilic amino acid residues in the helical sequence. The role of non-sequence depending interactions in helical stability, such as presence of disulfide bridges, or bonding of helical residues to substrate and/or cofactors, was also analysed.

***Key words**: amphipathic helices; helical structure stabilization; helix-forming amino acids; protein helices; secondary structure prediction.*

INTRODUCTION

Among the secondary structure elements, the helical structure has always deserved a special interest, since the first protein structure was solved in 1960. Today, it is believed that the helical structure, acting as folding nuclei, could play a crucial role in the folding process of a protein. If it can be proved that the helical stability mainly depends on interactions among amino acid residues close enough in the sequence, the folding of parts of the polypeptide chain in a helical structure would not need to wait for the synthesis of the complete protein. Thus, the elucidation of the factors responsible for helical folding is an important step in the knowledge of the folding-unfolding process in proteins.

Several approaches have been used to solve the above problem. Some of them are:

- Secondary structure prediction methods, which try to locate the position of the helical zones in the amino acid sequence (Fasman, 1989). These methods show different degrees of success when applied to proteins with a known tertiary structure, usually not better than 70%.
- Studies of model synthetic peptides, to test the dependence of the helical folding on the presence and/or the location of certain amino acid residues in the sequence (Chou *et al*, 1972; DeGrado and Lear, 1985; Marqusee and Baldwin,1987; Marqusee *et al*, 1989; Merutka and Stellwagen, 1990; Forood *et al*, 1993). The extension of the results obtained from these studies to proteins might be questioned, with the argument that the limited number of different amino acid residues involved in these peptides makes them very different from a protein sequence.
- Studies of the conformation of synthetic peptide analogues to known helical sequences in proteins. They overcome the doubts on the validity of the results pointed above, but are necessarily restricted to only a few and short sequences (Shoemaker *et al*, 1985, 1990; Epand *et al*, 1987; Strehlow and Baldwin, 1989).

Correspondence to: Dra Hilda Cid, Laboratorio de Biofísica Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Casilla 4077, Correo 3, Concepción, Chile. Phone: (56-41) 234-986, ext 2587. Fax: (56-41) 245-975.

- Studies of the role of certain amino acid residues on helix stability. They are done by site-specific mutagenesis of sequences with helical structures treated as an integral part of a protein (Serrano *et al*, 1991, 1992; Horovitz *et al*, 1992).

We have chosen to study a sample of 54 helical structures on a group of 12 different proteins, representing the 3 structural classes that contain helices. The tertiary structure of all these proteins has been solved to a resolution of 2 Å or better, and their coordinates are available in the Brookhaven Protein Data Bank (Bernstein *et al*,1977), thus the precise location of the helical structures, based on their hydrogen bonds (Kabsch and Sander, 1983), is known. We have applied 4 different secondary structure prediction methods to this sample, and we have tried to spot in the sequences that failed to be predicted as helices, the presence of the stabilizing factors detected by studies on synthetic peptides, usually not considered in the prediction methods. A positive result would explain the success or failure of the prediction methods, and would point to the variety of interactions responsible for the stability of helical structures in proteins.

METHODS

Database

The proteins and helices included are listed on Table I. The same proteins used by Presta and Rose (1988) to test the presence of helix initiation or termination signals were selected. The Brookhaven Protein Data Bank coordinates and programs (Bernstein *et al*, 1977) were used for all calculations.

Definition of the amphipathic character of helical structures

An amphipathic helix has been defined as an α-helix with opposing polar and non-polar faces, oriented parallel to the helix axis (Segrest *et al*, 1990). It is known that more than 50% of helical structures in globular proteins are located at the protein surface. Thus, part of the helical surface (mainly hydrophilic) is in contact with the solvent, whereas the rest, mainly hydrophobic, faces the protein core. The amphipathic helices show a clear periodicity of hydrophobic and hydrophilic residues in their sequence, which can be easily recognized by several prediction methods, such as the three that follow:

Helical Wheels

The helical wheel diagram is a representation of protein sequences, as the projection of the side chains of the amino acid residues, on a plane perpendicular to the helical axis. The detection of amphipathic α-helices is achieved simply by plotting the 'suspected helical' amino acid sequence, at angles of 100°, at constant distances of an origin: hydrophilic and hydrophobic residues will

TABLE I

Proteins and helical regions used in this study*

Code**	Protein name	Helical regions***	Structural class
5CPA	carboxipeptidase A	15-28, 74-89, 94-100, 113-121, 174-186, 216-230, 254-260, 286-305	α/β
3CPV	parvalbumin	8-15, 26-32, 40-50, 80-87, 99-107	αα
4CYT	cytochrome c	3-12, 50-53, 61-69, 71-74, 88-101	αα
4DFR	dihydrofolate reductase B	25-35, 44-50, 78-83, 97-103	α/β
4FXN	flavodoxin	11-25, 66-72, 94,104, 125-135	α/β
1LZ1	lysozyme, human	5-14, 25-35, 90-99, 110-114	α+β
1MBO	myoglobin, sperm whale oxidized	4-17, {21-35, [37-42]}, 52-56, 59-76, 83-95, 101-118, 125-148	αα
5PTI	pancreatic trypsin inhibitor	48-55	α+β
1PPT	avian pancreatic peptide	14-31	**α
5RSA	ribonuclease A, bovine	4-12, 25-32, {51-55, [56-57]}	α+β
1SN3	scorpion neurotoxin	{23-29, [30]}	α+β
1TIM	triose phosphate isomerase	18-30, 47-54, 80-86, 96-101, 106-118, 131-135, 139-153, 178-203, 216-221, 239-244	α/β

* Resolution of X-ray structures of all proteins is 2 Å or better, and crystallographic R factors are less than 20%.
** Brookhaven Protein Data Bank four-character name (Bernstein *et al*, 1977).
*** [] is a segment of 3_{10} helix; {} denote a segment considered as a single helix.

be clustered on two clearly distinguishable faces, clearly distinguished (Segrest *et al*, 1990). Examples of an amphipathic and a non-amphipathic helices, detected by this method, are shown in Figure 1.

Method of Cid et al.

This method gives the relative position of the polypeptide chain with respect to the protein surface, making use of the linear correlation between this distance and a hydrophobicity coefficient defined by Ponnuswamy *et al* (1980). The "hydrophobicity profile" is simply a plot of this coefficient $<H_f>$, versus the amino acid number in the sequence. Basic profiles have been defined for four secondary structure elements: exposed helical structure, ß-turns, buried and exposed ß-strands (Cid *et al*, 1982). The identification of these basic patterns in the hydrophobicity profile of the protein yields the predicted secondary structure. The profile defined as an exposed helical structure would correspond to an amphipathic helix (Figure 2). Ponnuswamy's hydrophobicity scale has been recently recalculated, on an extended database, for each of the four protein structural classes (Cid *et al*, 1992).

Method of the hydrophobic moment

The helical amphipathicity of a protein sequence can have a mathematical expression in terms of the hydrophobic moment (Eisenberg *et al*, 1984; Parker and Song, 1990), defined as follows:

$$<\mu_H> = (1/N)\ \{[\sum_{n=1}^{n=N} H_n \operatorname{sen}(n\delta)]^2 + [\sum_{n=1}^{n=N} H_n \cos(n\delta)]^2\ \}^{1/2}$$

$<\mu_H>$ is the average hydrophobic moment of a protein sequence of N amino acid residues; H_n is the hydrophobicity coefficient of the *n*th aminoacid residue, in the Eisenberg's hydrophobicity scale, and δ is the angle between two consecutive side chains projected on a plane perpendicular to the helix axis. In the Parker and Song's method the angle δ has a constant value of 100° and N is the observation window of 11 amino acid residues. According to this method, a value higher than 0.36 for the average hydrophobic moment would suggest the presence of an amphipathic helical sequence.

Chou and Fasman's secondary structure prediction method

This method is based on empirical probabilities: it defines conformational parameters P_α, P_β and P_t for the 20 natural amino acids, which represent the normalized frequency of occurrence of each amino acid residue in a particular type of secondary structure, as obtained from a data base of 29 fully

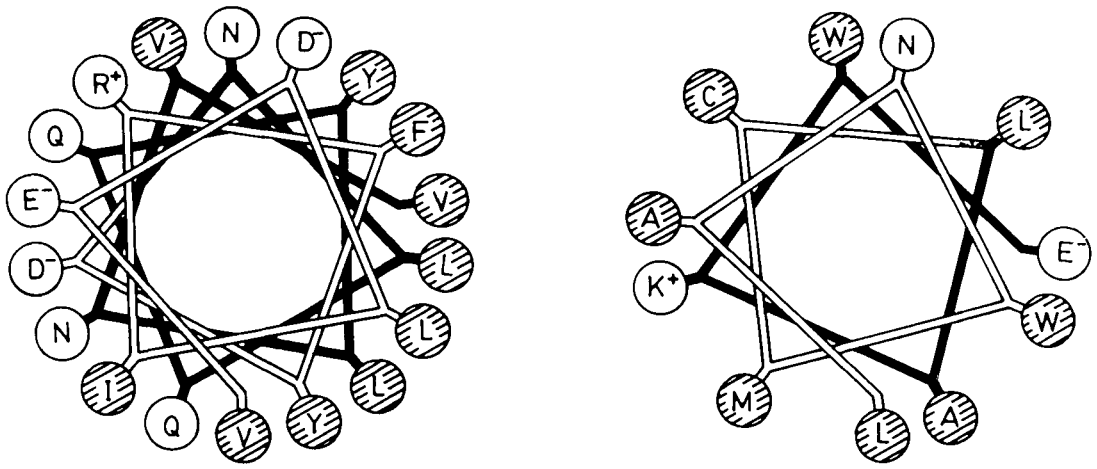

Fig 1. Determination of the amphipathic character of helical sequences by the method of helical wheels. A) Avian pancreatic hormone (1PPT), the amphipathic helix 14- 31. B) Human lysozyme (1LZ1), the non-amphipathic helix 25-35. The hydrophobic residues are hatched.

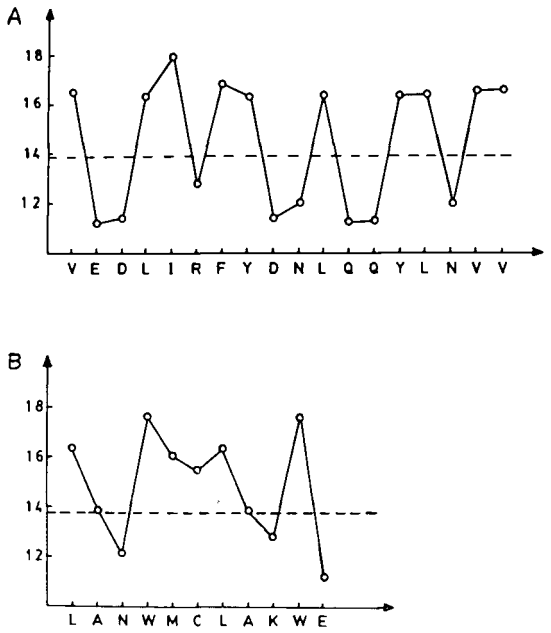

Fig 2. Hydrophobicity profiles of: A) the amphipathic helix 14-31 from 1PPT and B) the non-amphipathic helical sequence 25-35 from 1LZ1 (B). Ponnuswamy's averaged hydrophobicity scale (Cid *et al*, 1992) was used.

determined protein structures (Chou and Fasman, 1978). An average <P$_j$> greater than 1.0, for a group of amino acids taken in sequence (6 for a helix, 5 for a ß-strand and 4 for a ß-turn) is an indication that the structure type j is likely to occur in that region of the sequence. The performance of this method does not depend on the amphiphilic character of the helical structures.

RESULTS AND DISCUSSION

Amphipathic helices

We defined as amphipathic helices those structures showing two different hydropathic faces, parallel to the helix axis and clearly distinguishable in the helical wheel plot; less than 20% of the amino acid residues in a face and in the whole sequence, could be "misplaced" in an amphipathic helix. If this number was between 20% and 30%, the helices were designed as semi-amphipathic, and non-amphipathic if the "misplaced" residues were 30% or more. Our amphipathic and semi-amphipathic helices coincide with the definition of amphipathicity of Segrest *et al* (1990). Figure 1 presents examples in the form of helical wheel diagrams of the two extreme groups, and Figure 2 their corresponding hydrophobicity profiles. Table II presents their distribution in our database, using the three hydrophobicity scales defined above (Ponnuswamy *et al*, 1980; Cid *et al*, 1992; Eisenberg *et al*, 1984).

When the prediction methods of Cid *et al* (1982), Parker and Song (1990), and Chou and Fasman (1978), as originally reported, were applied to these three groups of helices, the predictions showed a different success. On Table III are listed the k_δ values (percentage of helical residues correctly predicted) for each of the three groups of helical structures, obtained with the different secondary structure prediction methods. As expected from the definition of the methods, the results present a good correlation between amphipathic character and prediction success for the method of Cid *et al* and low correlation for the Chou and Fasman's method. The method of Parker and Song showed good predictions for the amphipathic and semi-amphipathic helices, but did not discriminate when the helical structures were non-amphipathic, probably because the results of the method depend not only on the amphipathic character of the amino acids in the sequence,

TABLE II

Hydropathic character(*) of the 54 helical structures considered in the database

Hydrophobicity Scale	Amphipathic		Semi-Amphipathic		Non-Amphipathic	
	Number	%	Number	%	Number	%
Eisenberg's	14	26	11	20	29	54
Ponnuswamy's	15	28	13	24	**26**	48
Id. by Structure Classd	20	37	11	20	23	43

(*) Amphipathic helix defined as a structure with clearly distinguishable polar and nonpolar faces (Segrest *et al*, 1990), and in addition, no more than 20% of the aminoacids in any of the faces can have a wrong hydropathic character. A non-amphipathic helix could still show two faces with different hydrophobic character, but it has 30% or more of its amino acids misplaced.

TABLE III

Correlation between helix amphipathic character and prediction success*

Amphipathic character**	Prediction method Chou & Fasman			Cid *et al***			Parker & Song**		
	kα < 50%	50%<kα<70%	kα > 70%	kα < 50%	50%<kα<70%	kα >70%	kα< 50%	50%< kα < 70%	kα >70%
amphipathic	33%	20%	47%	33%	0%	67%	27%	6%	67%
semi-amphipathic	54%	23%	23%	54%	5%	31%	23%	23%	54%
non-amphipathic	46%	12%	42%	81%	11%	8%	31%	23%	46%

* Prediction success ka indicated as percentage of aminoacid residues correctly predicted as helical.
**Ponnuswamy's average hydrophobicity scale was used.

but on the vector sum of the hydrophobicity values involved.

The non-amphipathic helices of our database, either presented a "wrong" periodicity in the location of their hydrophilic and hydrophobic amino acid residues, or, they showed zones where one amphiphilic character was predominant.

Helical structures with hydrophobic zones

Figure 3 shows the "hydrophobic" helical structures of our database, which represent 26% (14/54) of the sample; sequences of 4 or more hydrophobic residues, according to Ponnuswamy's average hydrophobicity scale (Cid *et al*, 1992), characterize these helices. On a 100% of their hydrophobic regions alanine (Ala) is present, and, on a 64% of them, also leucine (Leu) is found. Several studies have recognized both amino acid residues as strong helix formers. Correlations between the amount of helical structure in proteins and their amino acid composition, expressed as a percentage, have been reported (Davies, 1964; Kriegbaum and Knutton, 1973). These results, corroborated by statistical studies on the participation of amino acids in protein helical structures, always point to alanine and leucine, together with glutamic acid (Glu) and lysine (Lys) as the strongest helix formers (Chou *et al*, 1972; Chou and Fasman, 1978). The fundamental role of alanine in the helix stabilization in short alanine-based peptides, has been clearly established (Marqusee *et al*, 1989). Also, the substitution of Ala by glycine (Gly) in the C-peptide helix (peptide with the amino acid sequence of the C-helix of ribonuclease A) has shown the influence that the presence of alanine has in the helix stability, independent of the position of this amino acid in the helical sequence, with the exception of both helical ends (Strehlow and Baldwin,1989).

What are the special characteristics of Ala and Leu that make them so strong helix formers? Several explanations have been suggested. Among them, the influence of straight chains in non-polar amino acids, seems to be one of the most coherent (Padmanabhan and Baldwin, 1985). Recently, a statistical study of the distribution of the dihedral angles ϕ, ψ for each one of the 20 amino acid residues in a database of 67 protein structures (Niefind and Schomburg, 1991) has shown that the distributions of the ϕ and ψ angles of Ala, Leu, Glu and Lys, share the greater correlation, and that their experimentally determined Ramachandran plots present an absolute maximum located at a position close to the ideal $\phi = -58°$, $\psi = -47°$ angles, accepted for a right handed α helical structure. The similarity between the Ramachandran plots and the conformational energy maps suggests that these ϕ and ψ values correspond to a minimum energy configuration for the "helix former" amino acids, and they tend to acquire this configuration independently of their location in the amino acid sequence.

"Hydrophilic" helices

We found that 15 out of the 54 helical structures considered (28%), presented hydrophilic zones of 4 or more hydrophilic residues in sequence, and /or the number of hydrophilic residues clearly exceed the hydrophobic ones. These helices, listed in Fi-

Protein Code	Structure Class	Helix Location	Sequence
4 CYT	$\alpha\ \alpha$	88 - 101	$K^+GE^-R^+Q\ D^-$ [L V A Y L] K^+S [A]
1 MBO	$\alpha\ \alpha$	4 - 17	E^-GE^- [W] Q [L V L H V W A] K^+ [V]
		101 - 118	[I] K^+ [Y L] E^- [F I] SE^- [A I I H V L H] SR^+
3 CPU	$\alpha\ \alpha$	8 - 15	D^- [A] D^- [I A A A L]
		26 - 32	[H] K^+ [A F F A] K^+
		40 - 50	[A] D^-D^- [V] K^+K^+ [A F A I I]
1 LZ1	$\alpha + \beta$	25 - 35	[L A] N [W M C L A] K^+ [W] E^-
		90 - 99	[A] D^- [A V A C A] K^+R^+ [V]
5 CPA	α / β	15 - 28	[L] D^-E^- [I Y] D^- [F M] D^- [L L V A] Q
		74 - 89	[I] T Q [A] T Q [V M F A] K^+K^+ [F] TE^-N
		216 - 230	K^+TE^- [L] N Q [V A] K^+S [A V A A L]
4 DFR	α / β	25 - 35	P [A] D^- [L A W F] K^+R^+NT
1 TIM	α / β	106 - 118	D^-E^- [L I] G Q K^+ [V A H A L A]
		178 203	P Q Q [A] Q E^- [V H] E^-K^+ [L] R^+G [W L] K^+T [H V] S D^- [A V A V] Q

Fig 3. The helical sequences with hydrophobic regions. Hydrophobic amino acid residues are boxed. Note the presence of alanine in all and leucine in most hydrophobic zones.

gure 4, show one or more of the following features:

- The presence of charged amino acid residues, specially Glu and Lys in their sequences. Since Glu and Lys are strong helix formers, they seem to fulfill in hydrophilic helical sequences, the same role that Ala and Leu have in the hydrophobic helices.
- The location of residues with opposite charges in positions i, i + 4 or i, i +3 , thus suggesting the possibility of "salt bridges" or ionic-pair interactions. Studies using synthetic peptides (Marqusee and Baldwin, 1987) have indicated that this could be a helical stabilization factor, with a higher probability for the pair i, i + 4. In our database we found 21 possible ion pairs of the type i, i + 3 and 22 of the type i, i + 4 which, when checked in the Brookhaven Data Base, with a cutoff limit of 4.0 Å, gave us only 6 distances below this limit, as shown on Table IV. From a study on 38 protein structures, known at a resolution of 2.5 Å or better (Barlow and Thornton, 1983), it was deduced that a distance less or equal to 4.0 Å between oppositely charged groups indicated the existence of an ion-pair interaction. However, the presence of these ionic-pairs in 44% of the helices of the sample suggests that they should have a positive influence in helix stabilization. This favourable ionic pair interaction could explain the helical character of sequences which show, simultaneously, a "wrong helical periodicity" (WP) and a "low helical potential" (LHP). This is the case with the C-terminal region of helix 74-89 or the N-terminal region of helix 174-186 of carboxypeptidase A (5CPA), the helical configuration of helix 50-53 in cytochrome c (4CYT), and several others in the sample.
- The presence of negative charges at the N-terminal and positive charges at the C-terminal of the helix. These charges may stabilize the helix dipole moment, as found in synthetic peptides (Shoemaker *et al*, 1985; Forood *et al*, 1993). Table V

Protein Code	Structure Class	Helix Location	Sequence
4 CYT	$\alpha\alpha$	3 - 12	V A [K⁺G K⁺K⁺T] F V [Q]
4 CYT	$\alpha\alpha$	50 - 53	[D⁻] A [N K⁺]
		61 - 74	[N D⁻T] L M [E⁻] Y L [E⁻N P K⁺K⁺] Y
		88 - 101	[K⁺G E⁻R⁺Q D⁻] L V A Y L [K⁺S] A
1 MBO	$\alpha\alpha$	37 - 42	[P E⁻T] L [E⁻K⁺]
3 CPU	$\alpha\alpha$	80 - 87	[G E⁻T K⁺] F L [K⁺]
4 DFR	α/β	25 - 35	[P] A [D⁻] L A W F [K⁺R⁺N T]
4 FXN	α/β	11 - 25	[N T E⁻K⁺] M A [E⁻] L I A [K⁺G] I I [E⁻]
4 FXN	α/β	94 - 104	[K⁺] W M [R⁺D⁻] F [E⁻E⁻R⁺] M [N]
		125 - 135	[E⁻Q D⁻] C I [E⁻] F [G K⁺K⁺] I
1 TIM	α/β	18 - 30	[R⁺K⁺S] L [G E⁻] L I H [T] L [D⁻G]
		96 - 101	[S E⁻R⁺R⁺] H V
		131 - 135	L [D⁻E⁻R⁺E⁻]
		139 - 153	[T E⁻K⁺] V V F [Q E⁻T K⁺] A I A [D⁻N]
1 SN3	α/β	23 - 30	[E⁻G] C [D⁻T E⁻] C [K⁺]

Fig 4. The «hydrophilic" helical sequences. Hydrophilic amino acid residues are boxed. Note the presence of glutamic acid and lysine in the hydrophilic regions, and the preference of negatively charged residues for the N-cap and of positively charged ones for the C-cap.

TABLE IV

Salt Bridges with distance between ion-pair < 4 Å

Protein	Sequence	Ion pair	N-O distance	Position
5CPA	174-186	K177-D181	3.04 Å	i, i + 4
1PPT	14-31	E15-R19	2.80 Å	i, i + 4
4DFR	97-103	R98-E101	2.56 Å	i, i + 3
1MBO	52-56	E52-K56	2.78 Å	i, i + 4
	101-118	k102-E105	2.86 Å	i, i + 3
3CPV	40-50	D42-K45	2.77 Å	i, i + 3

TABLE V

Distribution of charged aminoacid residues at the N- and C-caps of helical structures

N to N + 2	C to C-2	Number	%
−	+	14	26
−	0	11	20
−	−	7	13
0	−	3	6
0	0	3	6
0	+	4	7
+	−	4	7
+	0	2	4
+	+	6	11

−: negatively charged; +: positively charged;
0: no charge, at neutral pH.

shows the charge distribution at the N and C-caps of the 54 helical structures. As expected (Richardson and Richardson, 1988) the negatively charged amino acid residues prefer the N-cap (59%), and the positive ones have a weaker preference for the C- cap (44%). Only on 14 (26%) of the helices was the combination of a negative N-cap with a positive C-cap found; of these, 8 were "hydrophilic helices and 2 were "hydrophobic" helices. It is also interesting to note the presence of ionic-pairs with the same charge at positions i, i + 4 at the N- and C-caps of the helical sample under study. If we accept a "helix stabilizing influence" of ionic-pairs with opposite charge (SIP), we must also accept a "destabilizing influence" of ion-pairs with the same charge (DIP). Twelve DIP including a N-cap amino acid residue, and seven including a C-cap residue were detected in the helical sample, thus suggesting that this, as well as the presence of proline or glycine, or a sequence of at least three amino acids with

a "low helix potential", can be used as a helix-termination signal (Table VI).

The presence of possible aromatic-aromatic interactions in the helical sequences was also examined, since it has been reported that they may contribute to the stabilization of the first helical structure of barnase (Serrano *et al*, 1991). However, a study on 34 proteins (Burley and Petsko, 1985) has shown that 80% of these interactions would contribute to stabilize tertiary structure and the remaining 20%, the quaternary structure of proteins. The helical sequences of our database where such interactions might occur are listed in Table VII.

The effect of sequence-independent factors, which would influence the environment for certain amino acid residues, such as involvement in S=S bridges, or proximity to cofactors or to enzyme active sites, has been also surveyed. Tables VIII and IX present a list of these cases, and in order to decide their specific role in helix stability, the results of two prediction methods, one depending on the amino acid composition and the other on the hydrophatic character, are also indicated. Only on two cases, both methods which ignore these interactions, simultaneously predicted correctly less than 50% of the amino acids as belonging to a helical structure.

Finally, we are aware that the protein sample used in this study is a limited one, even though it fulfills the conditions to be representative of all known globular protein structures with helical regions: it contains the same number of proteins of each structural class, and the proteins included had different sizes, functions and origins (Table I). But probably the best test is to verify that the conclusions obtained from this database are found outside it. We chose, as a test protein, the *Staphylococcus aureus* ß-lactamase (Herzberg and Moult, 1987) which presents eleven helical structures. The hydrophobicity profiles of these helices are shown in Fig. 5; the location of the helical sequence determined by X-ray diffraction is also marked. Only five out of the eleven sequences (α3,

TABLE VI

Helix termination signals in protein structures

Protein*	Helix	N-cap	C-cap	Protein*	Helix	N-cap	C-cap
5CPA	15-28	DIP(D16-D20); LHP	P30; LHP	1MBO	4-17	G5; WP	WP
	74-89	LHP	G91; WP; LHP		21-35	G23; LHP	P37; LHP
	94-100	P94; WP; LHP	LHP		37-42	P37; SIP ($E38^{-}$-$K42^{+}$)	SIP (E38)
	113-121	P113; G115; WP	WP; DIP (H120-R124)		52-56	SIP ($E52^{-}K56^{+}$)	SIP ($E52^{-}$-$K56^{+}$)
	174-186	WP; LHP	DIP (H186-K190)		59-76	WP	WP
	216-230	P214; WP; LHP	?		83-95	G80; LHP	DIP (H93-H97); WP
	254-260	G253; G252; LHP	G262; LHP		101-118	P100; DIP (K98-K102)	P120; G121; LHP
	286-305	P288; LHP	WP; N307 C-term.		125-148	G124; LHP	G150; LHP
3CPV	8-15	LHP	?	4CYT	3-12	G1 (N-term.)	LHP
	26-32	WP; LHP	G34; LHP		50-53	LHP; WP	LHP; WP
	40-50	WP; LHP	WP; LHP		61-69	LHP	P71
	80-87	G80; WP; LHP	DIP (K83-K87)		71-74	P71	P76; G77; LHP
	99-107	G98; WP; LHP	A108 is C-terminal		88-101	G89; DIP (K87-R91)	LHP
4DFR	25-35	P25; WP; LHP	WP; LHP	4FXN	11-25	G10; G8	G27; DIP (E25-D29)
	44-50	G43-LHP	G51		66-72	DIP (E63-E67); P68	LHP
	78-83	?	LHP; G86		94-104	G93; G91	G105; G107
	97-107	G95; G96; G97	P105		125-135	E123-D127; D122-E125	C-term. I138
1TIM	17-29	G15; DIP (K12-R17)	G29	1LZ1	5-14	DIP (K1-R5)	G16
	46-53	P43; LHP	DIP (K53-K57)		25-35	LHP	G37-LHP
	79-85	P79; LHP	G86; LHP		90-99	DIP (D87-D91)	DIP (K97-R101)
	95-100	G93; DIP (H94-R98)	G102		110-114	LHP	LHP
	105-117	G102; WP; LHP	G119; G121	5RSA	4-12	N-terminal	LHP
	130-134	DIP (K129-R133)	G136		25-32	WP; LHP	WP; LHP
	138-152	G136; WP; LHP	WP; LHP		51-57	DIP (E49-D53)	WP; LHP
	177-202	P177; WP; LHP	WP; LHP	5PTI	48-55	WP; LHP	G56; G57; LHP
	215-220	G213; G214	G222	1PPT	14-31	DIP (D11-E15)	LHP
	238-243	P237; WP	WP	1SN3	23-30	G24	WP

* Brookhaven Protein Data Bank four-character name. **DIP**, destabilizing ionic pair; **LHP**, low helical potential (*i.e.*, 3 consecutive non helix-former residues); **WP**, "wrong periodicity" (non-amphipathic periodicity of hydrophilic and hydrophobic residues); **SIP**, stabilizing ionic pair (it is the only stabilization factor responsible for that specific helical sequence). Location of accepted helix disruptors proline and glycine in the vicinity of N- and C-caps also indicated.

TABLE VII

Possible aromatic interactions in helical structures

Protein	Helix	Aminoacids	Interaction	Distance*	Position
1MBO	101-118	Y103-F106	CA-CD1	3.75 Å	i, i + 3
			CD2-CD1	3.98 Å	
			CD2-CE1	3.87 Å	
5CPA	74-89	F82-F86	CB-CZ	3.94 Å	i, i + 4
4FXN	66-72	F66-F69			i, i + 3
	94-104	W95-F99			i, i + 4
4DFR	97-103	Y100-F103			i, i + 3
1TIM	47-54	Y47-F50			i, i + 3
1LZ1	110-114	W109-W112			i, i + 3

* A cut-off distance of 4 Å was used.

TABLE VIII

Effect of not considering the presence of disulfide bridges in helix prediction

Protein	Helix	S = S bridge	K_α Method C & F	P & S	Helix character
1LZ1	5-14	cys6-cys128	60	100	Semi-amphipathic
	25-35	cys30-cys116	73	100	Non-amphipathic
	90-99	cys77-cys95	80	100	Non-amphipathic
1SN3	23-30	cys25-cys46 cys29-cys48	63	0	Non-amphipathic
5PTI	48-55	cys5-cys55 cys30-cys51	75	100	Amphipathic
5RSA	25-32	cys26-cys84	88	0	Amphipathic

kα = % helical aminoacids correctly predicted.
C & F = Chou and Fasman's prediction method.
P & S = Parker and Song's prediction method.

α5, α6, α10, α11) did show the alternance of hydrophilic and hydrophobic amino acid residues which characterise the amphipathic helices (Fig 2). The presence of the helix stabilizing factors or helix termination signals, previously discussed, in the eleven helical sequences is analysed below:

α1 (33 40)

Stabilizing factors: A negative charge at the N-terminal; 4/7 "helix-formers" residues (boxed); a "stabilizing ion-pair" (SIP) D35-K39.
Helix termination signals: K31 is the N-terminal residue of the secreted enzyme; a destabilizing ionic interaction (DIP) K39-H43 at the C-cap.

α2 (72-82)

Stabilizing factors: Amphipathic periodicity 72-77 interrupted by the hydrophobic sequence A78-L82, which contains 3/4 "helix-former" residues.
Helix termination signals: 3 non-helix formers residues are found before the N-cap and after the C-cap.

α3 (107-113)

Stabilizing factors: Characteristic amphipathic sequence of 2 hydrophobic, 2 hydrophilic amino acid residues.
Helix termination signals : P107 at the N-terminal of the helix and G114, at the C-cap, where also a DIP K111-K115 is present.

α4 (119-127)

Stabilizing factors: The helical sequence 119-127 presents 6 out of 9 helix forming residues and a SIP K120-E124.
Helix termination signals: Three non-helix former residues before the N- and after the C-terminal.

α5 (132-142) and α6 (145-154)

Stabilizing factors: A long amphipathic sequence is interrupted by two consecutive helix disruptors G143 and G144.
Helix termination signals: 4 non-helix former amino acid residues at the N-cap; G155 at the C-end and a DIP K149-K153 may play against the SIP K153-D157 that would continue the helix.

α7 (166-177)

Stabilizing factors: This helix, which shows the characteristic hydrophobicity profile of an exposed ß-strand (Cid *et al*, 1982) might

TABLE IX

Impact of not considering the presence of substrates and cofactors in helical structure prediction

Protein	Helix	aa residue		Bonded to	Ka Method C & F	Ka Method P & S
LZ1	23-35	glu	35	substrate	73	18
5RSA	4-12	his	12	cofactor [PO4]	100	100
4DFR	25-35	asp	27	substrate	82	0
		leu	28	id.		
		trp	30	id.		
		phe	31	id.		
		lys	32	id.		
	44-50	ile	50	substrate	0	100
		arg	44	cofactor NADH		
		his	45	id.		
		thr	46	id.		
	78-83	val	78	id.	83	100
	97-103	gly	97	id.	0	100
		arg	98	id.		
		val	99	id.		
		tyr	100	id.		
		gln	102	id.		
4CYT	61-69	leu	68	cofactor HEM	56	100
4FXN	11-25	thr	12	cofactor HEM	80	100
		asn	11	id.		
1MBO	59-76	his	64	cofactor HEM	44	44
		thr	67	id.		
		val	68	id.		
		ala	71	id.		
		leu	72	id.		
		ser	92	id.		
5CPA	113-121	his	120	substrate	67	56
		ser	121	id.		
	254-260	ser	254	id.	0	0

k_α = % helical amino acids correctly predicted.
C & F= Chou and Fasman's prediction method.
P & S = Parker and Song's prediction method.

be originated by the SIP R164-E168; also, 3 out of 6 residues are "helix formers".
Helix termination signals: Three "non-helix formers" residues are found before the N-terminal, and after the C-terminal of the helix.

α8 (183-193)

Stabilizing factors: Two alanines in its hydrophobic region, and an amphipathic sequence after that, probably not interrupted in G187 due to the presence of strong helix-former residues such as A84, A85 and K88.
Helix termination signals: P183 is the N-terminal, preceded by 3 non-helix formers. The helix does not continue after residue 193 probably due to a DIP (K188-K192).

α9 (201-213) and α10 (218-224)

Stabilizing factors: The hydrophilic N-terminal region of helix α9 presents 4 out of 5 helix forming residues and the following hydrophobic region presents 4 out of 6 helix forming residues, and a possible SIP interaction (K205-D209). α10 has an amphipathic profile and a SIP D218-K222.
Helix termination signals: α9 has DIP K201-K205 at the N-cap and a hydrophilic region of 7 amino acid which does not fulfill any of the requirements pointed above, at the C-end.

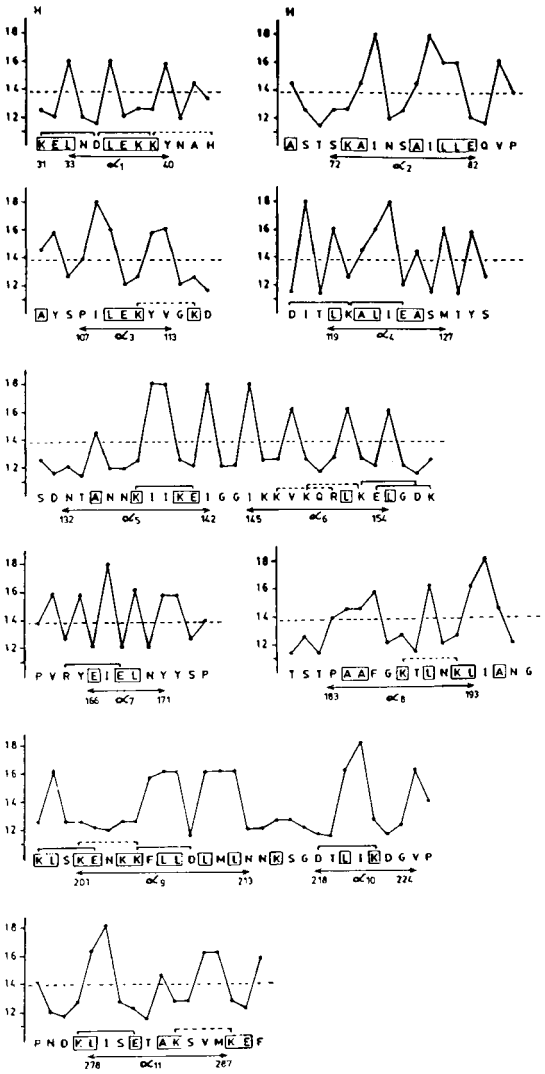

Fig 5. Hydrophobicity profiles of the eleven helical sequences of the *Staphylococcus aureus* ß-lactamase. The averaged hydrophobicity scale defined by Cid *et al* (1992) was used. The helical regions obtained by X-ray diffraction studies are indicated (Herzeberg and Moult,1987). The "helix former residues" are boxed. The stabilizing ion-pair interactions are indicated by a full line and the destabilizing ion-pair interactions by a broken line. Helices α_3 and α_7 are 3_{10} helices (Brookhaven Data Bank).

α10 has G217 and D218 at the N-terminal, and G224 and P226 at the C-cap.

α11 (278-287)

Stabilizing factors: An amphipathic helix profile, a SIP K277-E281.
Helix termination signals: Three non-helix former residues at the N-cap. A DIP K284-K288 may prevent the helix from continuing up to the C-terminal residue F290.

CONCLUSIONS

Several factors, by themselves or in combinations, seem to regulate the helical fold in protein structures. These are:

Stabilizing factors, in order of importance

1. Periodicity in the distribution of hydrophilic and hydrophobic residues is determinant in the amphipathic helices, which, according to our definition, represent 28% of the sample (52% with the semi-amphipathic group). The sample shows several examples of amphipathic helices formed mainly by amino acids with a "low helical potential": helix 286-305 (5CPA) which even includes P288 and G296, helix 97-103 (4DFR) and helix 48-55 (5PTI) are some of them.

2. When the amphipathic periodicity is not present, amino acids with a strong helix potential are capable of maintaining the helical structure:

- In hydrophobic regions, the presence of "strong helix formers" Ala and Leu is necessary.
- In hydrophilic regions, the presence of "strong helix formers" Glu and Lysine may be important.
- In regions with alternating hydrophilic and hydrophobic amino acid residues, often found in exposed ß-strands (Cid *et al*, 1982), the presence of strong helix formers will produce a helical fold, such as helix 52-56 and the N-terminal of helix 83-95 (1MBO).

3. The helical fold in hydrophilic regions can be also "forced" by the presence of stabilizing ion-pairs, specially if they involve Glu and Lys which are "strong helix formers".

4. The presence of negative and positive charges at the N- and C-caps respectively, of the helical region, seems to help in the helix stabilization, specially in hydrophilic helical structures, but does not seem to be, by itself, a determinant factor to induce the helical fold. At least 4 helices of our sample present the opposite distribution of charges at the N- and C-caps.

5. The possibility of aromatic-aromatic in teractions between residues in positions i, i + 4 and i, i + 3 was detected in two of the 54 helical structures.

Helix termination signals

The helix termination at the N- or C-caps seems to be regulated by the following signals, which may or not appear at the same time:

- Presence of the accepted "helix disruptors" proline or glycine.
- A sequence of three amino acid residues with a "low helical potential" in a region without an amphipathic periodicity.
- The presence of a "helix-destabilizing" ionic pair, formed by two amino acids with the same charge at positions i,i + 4, in the vicinity of any of the helix-caps.

Summary

The existence of a variety of factors responsible for the stabilization or disruption of the helical fold, and the fact that sometimes the combination of several of them is required, can explain the variable success of helix prediction methods, which usually consider only one or two of these factors. That is the case of the Chou and Fasman method (1978), which is based on the amino acid "helix potential", or the Cid *et al* (1982) and the Parker and Song (1990) methods, designed for amphipathic helical structures, or the Presta and Rose method (1988), which considers only the character of donor or acceptor of hydrogen bonds of the amino acid residues, located respectively at the beginning and end of helical sequences.

All the stabilization factors, as well as the helix termination signals, mentioned above are "sequence-dependent" and only involve short-range interactions between amino acid residues. The influence of sequence-independent factors, such as involvement in S=S bridges, or proximity to cofactors or to an enzyme active site, was not clearly confirmed, since in most of the cases where they were present, other sequence-dependent factors coexisted with them. Therefore, all results presented here do confirm that helices may act as folding nuclei in the protein folding process, since helix stabilization does not depend on long-range interactions requiring an advanced state of protein folding.

ACKNOWLEDGEMENTS

This work was partially supported by research grant 93.31.51-1 from the University of Concepción and by a DAAD's fellowship to one of us (MB).

REFERENCES

BERNSTEIN FC, KOETZLER TF, WILLIAMS GJ, MEYER EF Jr, BRICE MD, KENNARD O, SHIMANOUCHI T, TASOUMI M (1977) The protein data bank: a computer based archival file for macromolecular structures. J Mol Biol 112: 335-542

BARLOW DJ, THORNTON JM (1983) Ion-pairs in proteins. J Mol Biol 168: 867-885

BURLEY SK, PETSKO GA (1985) Aromatic-aromatic interactions: a mechanism of protein structure stabilization. Science 229: 23-28

CHOU PY, WELLS M, FASMAN GD (1972) Conformational studies on copolymers oxypropyl-L-glutamine and L-leucine. Biochemistry 11: 3028-3043

CHOU PY, FASMAN GD (1978) Empirical predictions of protein conformation. Annu Rev Biochem 47: 251-276

CID H, BUNSTER M, ARRIAGADA E, CAMPOS M (1982) Prediction of secondary structure of proteins by means of hydrophobicity profiles. FEBS Lett 150: 247-254

CID H, BUNSTER M, CANALES M, GAZITUA F(1992) Hydrophobicity and structural classes in proteins. Protein Engineering 5: 373-375

DAVIES DR (1964) A correlation between amino acid composition and protein structure. J Mol Biol 9: 605-609

DE GRADO WF, LEAR JD (1985) Induction of peptide conformation at apolar water interfaces. A study with model peptides of defined hydrophobic periodicity. J Am Chem Soc 107: 7684-7689

EISENBERG D, WEISS RM, TERWILLIGER TC (1984) The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci USA 81: 140-144

EPAND RM, GAWISH A, IQBALD M, GUPTA KB, CHEN CH, SEGREST JP, ANANTHARAMAIAH GM (1987) Studies of synthetic peptide analogues of the amphipathic helix. J Biol Chem 262: 9389-9396

FASMAN GD (ed) (1989) Prediction of Protein Structure and the Principles of Protein Conformation. NY: Plenum Press

FOROOD B, FELICIANO E, NAMBIAR KP (1993) Stabilization of α-helical structures in short peptides via end capping. Proc Natl Acad Sci USA 90: 838-842

HERZBERG O, MOULT J (1987) Bacterial resistance to ß-lactam antibiotics: Crystal structure of ß-lactamase from *Staphylococcus aureus* PC1 at 2.5 Å resolution. Science 236: 694-701

HOROVITZ A, MATHEWS JM, FERSHT A (1992) α-Helix stability in proteins II. Factors that influence stability at an internal position. J Mol Biol 227: 560-568

KABSCH W, SANDER C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637

KRIEGBAUM WR, KNUTTON SP (1973) Prediction of the amount of secondary structure in a globular protein from its amino acid composition. Proc Natl Acad Sci USA 70: 2809-2813

MARQUSEE S, BALDWIN RL (1987) Helix stabilization by Glu^- ... Lys^+ salt bridges in short peptides of de novo design. Proc Natl Acad Sci USA 84: 8898-8902

MARQUSEE S , ROBBINS V, BALDWIN RL (1989) Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci USA 86: 5286-5290

MERUTKA G, STELLWAGEN E (1990) Positional independence and additivity of aminoacid replacements on helix stability in monomeric peptides. Biochemistry 29: 894-898

NIEFIND K, SCHOMBURG D (1991) Amino acid similarity coefficients for protein modeling and sequence alignment derived from main-chain folding angles. J Mol Biol 219: 481-497

PADMANABHAN S, BALDWIN RL (1985) Straight-chain non-polar amino acids are good helix formers in water. J Mol Biol 219: 135-137

PARKER W , SONG P-S (1990) Location of helical regions in tetrapyrrole-containing proteins by a helical hydrophobic moment analysis. J Biol Chem 265: 17568-17575

PONNUSWAMY PK, PHRABHAKARAN M, MANAVALAN P(1980) Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. Biochim Biophys Acta 623: 301-316

PRESTA LG, ROSE GD (1988) Helix signals in proteins. Science 240: 1632-1641

RICHARDSON JS, RICHARDSON DC (1988) Amino acid preferences for specific locations at the end of α-helices. Science 240: 1648-1652

SEGREST JP, DE LOOF H, DOHLMAN J, BROUILLETTE CG, ANANTHARAMAIAH GM (1990) Amphipathic helix motif: Classes and properties. Proteins 8: 103-117

SERRANO L, BYCROFT M, FERSHT A (1991) Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. J Mol Biol 218: 465-475

SERRANO L, SANCHO J, HISHBERG M, FERSHT A (1992) α-Helix stability in proteins I. Empirical correlations concerning substitution of side-chain at the N and C-cap and the replacement of alanine by glycine or serine at solvent-exposed surfaces. J Mol Biol 227: 544-559

SHOEMAKER KR, KIM PS, BREMS DN, MARQUSEE S, YORK EJ, CHAIKEN IM, STEWART JM, BALDWIN RL (1985) Nature of the charged-group effect on the stability of the C-peptide helix. Proc Natl Acad Sci USA 82: 2349-2353

SHOEMAKER KR, FAIRMAN R, SCHULZ DA, ROBERTSON AD, YORK EJ, STEWART JM, BALDWIN RL (1990) Side-chain interactions in the C-peptide helix: Phe8 ... $His12^+$. Biopolymers 29: 1-11

STREHLOW KG, BALDWIN RL (1989) Effect of the substitution Ala-Gly at each of the five residue positions in the C-peptide helix. Biochemistry 28: 2130-2133