Intrahelical side chain interactions in α-helices: poor correlation between energetics and frequency

J. Fernández-Recio, J. Sancho*

Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

Received 5 April 1998; revised version received 5 May 1998

Abstract Polypeptide sequences in proteins may increase their tendency to adopt helical conformations in several ways. One is the recruiting of amino acid residues with high helical propensity. Another is the appropriate distribution of residues along the helix to establish stabilising side chain interactions. The first strategy is known to be followed by natural proteins because amino acids with high helical propensity are more frequent in α-helices. If proteins also followed the second strategy, stabilising amino acid pairs should be more frequent than others. To test this possibility we compared empirical energies of side chain interactions in α-helices with statistical energies calculated from a data base of proteins with low homology. We find some correlation between the stability afforded by the pairs and their relative abundance in α-helices but the realisation of energetic preferences into statistical preferences is very low. This indicates that natural α-helices do not regularly use intrahelical side chain interactions to increase their stability.

Key words: α-Helix stability; α-Helix design; Side chain interaction; Protein stability; Protein folding; Protein design

1. Introduction

Soon after the determination of the first protein structures, the analysis of the amino acid composition of protein α-helices revealed a high correlation between the frequency of an amino acid in α-helices and its helix-stabilising effect in model peptides [1,2]. This correlation inspired a method to predict α-helical regions in proteins from analysis of the amino acid sequence [3]. An additional factor influencing helix stability is the ability of certain amino acid residues to form helix-stabilising interactions when favourably spaced along the helix [4]. With the exception of long charged amino acids, the spacings at which any two helical side chains can more easily interact are the *i*,*i*+3 and the *i*,*i*+4 ones. There is considerable current interest in identifying which amino acid pairs are able to form helix-stabilising interactions and, for those pairs, which are the more favourable geometries. In some recent studies on helical side chain/side chain interactions the measured energies of interaction or the observed preferred geometries have been compared to the frequency and geometry of such interactions in natural proteins. The reported results vary: in some cases a correlation was found between the stability afforded by the pair and the frequency of the pair in proteins [5–7] while in other instances there was no such correlation [8–11]. Knowing whether intrahelical side chain/side chain interactions are, as a whole, being utilised by natural proteins or rather ignored is very important because it is connected to the general problems of protein stability and the mechanism of protein folding. To clarify this point we have computed, using a data base of 285 proteins with low sequence homology [12], the expected and actual occurrence of all *i*,*i*+3 and *i*,*i*+4 helical amino acid pairs. These statistical data have been used to calculate statistical energies (according to Boltzmann's law) that have been correlated with the empirically determined energy data available in the literature. We found a low correlation between the helix-stabilising properties of amino acid pairs and their frequencies in α-helices, which is in clear contrast with the good correlation known to exist between the helix-stabilising properties of individual amino acids and their frequencies in α-helices [4].

2. Materials and methods

2.1. Statistical survey of the data base

We have analysed the α-helices contained in a set of 285 proteins [12] with low sequence homology (less than 25%) that are implemented in the program WHATIF [13]. Each amino acid of these proteins carries a secondary structure assignment previously calculated with the program DSSP [14]. Sequence and conformational searches were made with the SCAN3D option [15] implemented in the program. The proteins contain 1771 α-helices (19 422 helical residues). An analysis of the occurrences of pairs in helices has been recently published [16] using a smaller protein data base consisting of 167 non-homologous proteins.

The occurrence of α-helical pairs in the data base was calculated as follows. The number of helical *ab* (*i*,*i*+4) pairs (N_{ab}) was the number of *a*-*x*-*x*-*x*-*b* helical sequences found in the data base. The number of helical amino acids of type *a* (N_a) and the number of helical amino acids of type *b* (N_b) were taken as the number of *a*-*x*-*x*-*x*-*x* and *x*-*x*-*x*-*x*-*b* helical sequences respectively. Similarly the number of helical amino acids in the helical space (N) was taken as the number of *x*-*x*-*x*-*x*-*x* helical sequences. This way of counting the number of helical amino acids of a given type for the purpose of calculating the statistics of helical pairs is advised to avoid end effects. The statistics of (*i*,*i*+3) pairs were done as for *i*,*i*+4 pairs but, in this case, the helical sequences counted were *a*-*x*-*x*-*b*, *a*-*x*-*x*-*x*, *x*-*x*-*x*-*b* and *x*-*x*-*x*-*x*. To determine if the global occurrence of the different pairs differs significantly from a random distribution the statistical data were evaluated by a χ^2 test. Similarly, the possible randomness of the occurrence of every particular pair was examined by a χ^2 test using contingency tables (see [16] for a similar calculation). The same analysis was done to examine whether amino acid residues are randomly distributed to form polar and apolar pairs.

2.2. Calculation of statistical energies

Suppose that the relative positions of the amino acids within α-helices have been set, through evolution, to reflect the contribution of the interaction energy of each pair to the energy of the native conformation of the protein. Let our system be constituted by all positions (each one being an *i*,*i*+3 or an *i*,*i*+4 pair) in the helical space. All positions are considered identical.

According to Boltzmann's distribution law, the probability that any position contains an *ab* pair is given by:

*Corresponding author. Fax: (34) (976) 762123.
E-mail:jsancho@posta.unizar.es

PII S0014-5793(98)00569-9

$$P_{ab} = (N_a N_b e^{-E_{ab}/kT})/z \quad (1)$$

where the product $N_a N_b$ (N_a and N_b being the number of amino acids of type a and b respectively in the helical space) is the degeneracy of the state, E_{ab} is the interaction energy of the two side chains, z is the partition function of the position considered, k is the Boltzmann constant, and T the absolute temperature.

The ratio between the probabilities of finding an *ab* or a *cd* pair in any given position is:

$$P_{ab}/P_{cd} = (N_a N_b / N_c N_d) e^{-(E_{ab}-E_{cd})/kT} \quad (2)$$

If the helical space is large enough, these probabilities can be calculated by dividing the actual number of pairs of a given type by the total number of helical pairs. Eq. 2 becomes:

$$N_{ab}/N_{cd} = (N_a N_b / N_c N_d) e^{-(E_{ab}-E_{cd})/kT} \quad (3)$$

where N_{ab} and N_{cd} are the number of amino acid pairs of type *ab* or *cd* respectively than are found in the helical space. Eq. 3 can be rearranged to:

$$\Delta E_{ab-cd} = -kT \ln(N_{ab} N_c N_d / N_{cd} N_a N_b) \quad (4)$$

where the first term of the equation represents the difference in free energy when a helical position is occupied by an *ab* pair or by a *cd* pair. This difference energy will be termed $\Delta\Delta G^{stat}_{ab-cd}$ and can be easily computed from the protein data base (see above).

2.3. Empirical energies

The empirical interaction energy between two amino acids in an α-helix ($\Delta\Delta G^{emp}_{ab}$) can be experimentally measured using peptide systems. In these peptides, a two-state folding equilibrium cannot be assumed and helix-coil transition theories that include side chain interactions are used to calculate the interaction energy of the pair from the helical content of peptides that include one such *i*,*i*+3 or *i*,*i*+4 pair. Algorithms implementing these theories are available (AGADIR [17] and SCINT [6]) and they constitute a convenient tool to calculate side chain/side chain interaction energies. All empirical energy data used in this work have been published [5–7,9–11,17–23]. Whenever errors for the reported empirical energies are available we have only considered energy values that were determined with low error (±0.2 kcal/mol).

The relationship between statistical and empirical energies has been studied by linear regression analysis. Linear regressions, analyses of correlation and statistical tests have been performed using Kaleidagraph from Abelbeck Software and StatView 2.0 from Abacus Concepts, Inc. To evaluate if correlation coefficients differ significantly from zero, a level of confidence (*P* value) calculated by *F*-test is given for each regression.

3. Results

The number of *i*,*i*+4 and *i*,*i*+3 amino acid pairs of any given type that are present in our helical space vary from 10 to around 200, except for some pairs involving rare amino acids (the complete set of data is available via Internet in http://wwwbioq.unizar.es/helixsc.html). We have performed a preliminary test to determine if the distribution of helical pairs differs significantly from randomness. Our results indicate that this is the case with a level of confidence greater than 99.99% (not shown). It is thus pertinent to test whether the amino acids in α-helices tend to distribute forming specific pairs according to the helix-stabilising character of each pair.

The available empirical interaction energies of helical pairs are compared in Table 1 with their corresponding statistical energies, calculated using Eq. 4. The reported interaction energies are relative to alanine/alanine pairs (assuming a value of 0.00 kcal/mol for these pairs at both *i*,*i*+3 and *i*,*i*+4). If the interaction energy of alanine/alanine pairs differed significantly from 0.00 kcal/mol the intercepts of our plots of statistical energies versus empirical energies (see below) would be wrong but the correlation coefficients, slopes and significance levels would be the same as those reported here.

We aim to test whether the amino acids in α-helices tend to form specific pairs according to the helix-stabilising character of each pair. That being the case, a plot of the statistical energies of *i*,*i*+3 and *i*,*i*+4 helical pairs versus the empirical energies of the pairs should yield a straight line with intercept at zero and slope unity. Such a plot is shown in Fig. 1: the low correlation coefficient and the value of the slope (far from unity) indicate that the frequency of these pairs is hardly governed by their helix-stabilising character, unlike what has been described for individual amino acids in α-helices (see [4] for a review).

Although the above result seems clear, we are aware of several factors that may complicate the analysis and we have accordingly explored the potential relationship between pair energetics and frequencies using different pair subsets.

First, to see if any difference can be established between *i*,*i*+3 and *i*,*i*+4 pairs we have correlated separately the data corresponding to each kind of pair. The linear fit of *i*,*i*+3 pairs is of no significance (Table 2). The linear fit of *i*,*i*+4 pairs yields results that are similar to those of the global fit (Table 2).

Second, the possibility exists that pairs involving two apolar amino acid residues behave differently from pairs formed by two polar amino acid residues (likely to be solvent-exposed). Moreover, the amphipathic nature of many protein α-helices is likely to force an uneven distribution of polar and apolar residues and pairs. We have analysed by a χ^2 test if amino acid residues tend to group forming polar-polar and apolar-apolar pairs or they are randomly distributed. Although the grouping effect observed is weak (5763 observed polar pairs, 5167 expected; 8967 observed apolar pairs, 8235 expected) the distribution is not random (with a confidence level of 99.99%; not shown). Because of this fact, the calculation of the statistical energies of these two kinds of pairs should best be done considering two different helical spaces: one formed by amino

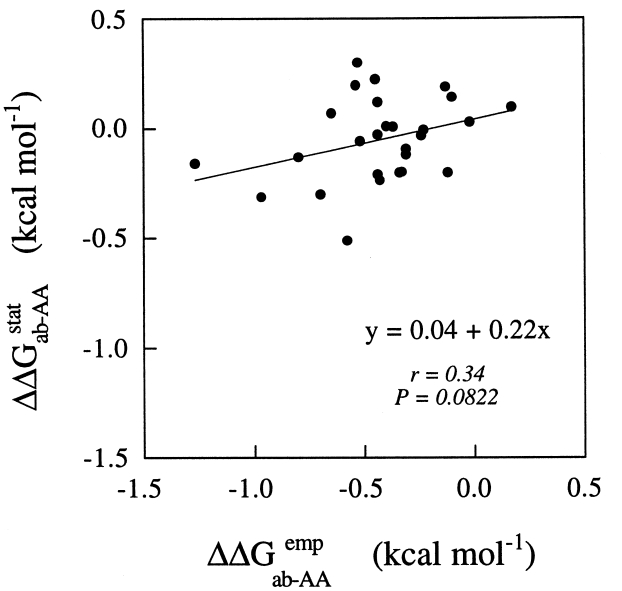

Fig. 1. Correlation between empirically measured interaction energies of amino acid pairs in α-helices (at *i*,*i*+3 or *i*,*i*+4) and the corresponding statistical energies calculated according to Eq. 4. All energies are relative to alanine-alanine pairs.

Table 1
Empirical and statistical energies for some *i*,*i*+4 and *i*,*i*+3 amino acid pairs in α-helices

Pair	Empirical[a] $\Delta\Delta G$ (kcal/mol)	Pairs observed in α-helices	Pairs[b] expected in α-helices	Statistical[c] $\Delta\Delta G$ (kcal/mol)	χ^{2d}	P^d
DK3	−0.12	89	55.6	−0.20	22.9	0.0001
DK4	−0.24	71	52.3	−0.03	7.7	0.0055
EH3	−0.23	26	22.5	0.00	0.6	0.4403
EH4	−0.10	21	20.8	0.14	0.0	0.9599
EK3	−0.38	139	87.5	−0.20	35.7	0.0001
EK4	−0.44	151	82.6	−0.21	67.2	0.0001
FH4	−1.27	16	9.5	−0.16	4.7	0.0302
FM4	−0.70	32	14.9	−0.30	21.0	0.0001
HD3	−0.53	7	10.3	0.30	1.1	0.4186
HE3	−0.45	14	18.0	0.22	1.0	0.3275
HE4	−0.54	15	16.3	0.20	0.1	0.7447
KD3	−0.40	39	34.8	0.01	0.6	0.4536
KD4	−0.58	96	31.4	−0.51	148.1	0.0001
KE3	−0.38	97	60.8	−0.20	24.7	0.0001
KE4	−0.46	107	55.7	−0.24	54.3	0.0001
KK4	0.17	63	57.8	0.10	0.6	0.4597
LY3	−0.44	59	49.4	−0.03	2.1	0.1439
LY4	−0.65	45	39.5	0.07	0.9	0.3448
MF4	−0.37	17	13.4	0.00	1.0	0.3088
QD4	−0.97	54	24.7	−0.31	38.1	0.0001
QE4	−0.31	66	43.9	−0.09	12.6	0.0004
QN4	−0.52	30	21.1	−0.06	4.0	0.0442
WH4	−0.80	6	3.8	−0.13	1.4	0.2423
YL3	−0.02	64	58.5	0.03	0.6	0.4382
YL4	−0.44	52	49.2	0.12	0.2	0.6682
YV3	−0.13	26	31.6	0.19	1.1	0.2915
YV4	−0.31	40	25.4	−0.12	9.2	0.0024

[a]Relative to an AA pair. $\Delta\Delta G$ for the following pairs are directly taken from the literature: DK3, DK4, KD3 and KD4 [22]; KK4 [18]; LY3, LY4, YV3 and YV4 [21]; QD4 [5]; WH4 [10]; $\Delta\Delta G$ for EH3 is the mean of two values: −0.30 (pH 5.5) and −0.15 kcal/mol (pH 8.5) [22]; $\Delta\Delta G$ for EH4 is the mean of two values: −0.10 (pH 5.5) and 0.0 kcal/mol (pH 8.5) [22]; $\Delta\Delta G$ for EK3 is the mean of two values: −0.38 [20] and −0.29 kcal/mol [22]; $\Delta\Delta G$ for EK4 is the mean of four values: −0.47 [20], −0.37 [18], −0.50 [19] and −0.42 kcal/mol [22]; $\Delta\Delta G$ for FH4 has been calculated using SCINT and published data from [23]; $\Delta\Delta G$ for pair FM4 is the mean of two values: −0.75 [6] and −0.65 kcal/mol [9]; $\Delta\Delta G$ for pair HD3 is the mean of two values: −0.61 (pH 5.0) and −0.45 kcal/mol (pH 9.0) [11]; $\Delta\Delta G$ for HE3 is the mean of two values: −0.50 (pH 5.5) and −0.39 kcal/mol (pH 8.5) [22]; $\Delta\Delta G$ for HE4 is the mean of two values: −0.65 (pH 5.5) and −0.43 kcal/mol (pH 8.5) [22]; $\Delta\Delta G$ for KE3 is the mean of two values: −0.38 [20] and −0.28 kcal/mol [22]; $\Delta\Delta G$ for KE4 is the mean of two values: −0.46 [20] and −0.40 kcal/mol [22]; $\Delta\Delta G$ for pair MF4 is the mean of two values: −0.54 [6] and −0.20 kcal/mol [9]; $\Delta\Delta G$ for QE4 is the mean of two values: −0.34 and −0.27 kcal/mol [20]; $\Delta\Delta G$ for QN4 is the mean of two values: −0.56 and −0.47 kcal/mol [7]; $\Delta\Delta G$ for pair YL3 is the mean of three values: 0.07, −0.15 and 0.02 kcal/mol [21]; $\Delta\Delta G$ for pair YL4 is the mean of three values: −0.36, −0.59 and −0.36 kcal/mol [21].
[b]Calculated as $N_a N_b/N$ (see Section 2).
[c]Calculated using Eq. 4. Relative to an AA pair.
[d]Each pair was analysed by a χ^2 test to evaluate if its occurrence deviates significantly from randomness (see Section 2).

acids involved in polar pairs and another space formed by amino acids involved in apolar pairs. The correlation between interaction energies of polar pairs (Table 1) and their statistical energies calculated for the polar helical space (data not shown) yields a linear fit of no significance with a slope close to zero (Table 2). This suggests that helix-stabilising polar pairs are not more frequent than others. For apolar pairs the correlation is similarly poor and the slope of only 0.2. If the energetics of polar and apolar pairs (Table 1) are compared to the statistical energies in Table 1 (which have been calculated without segregating the helical amino acids into a polar and an apolar set) similarly poor results are obtained (not shown).

Third, it is sometimes observed that data from different laboratories contain systematic deviations that could lower the correlation coefficient or change the slope when the data are pooled. To asses if this is the case here, we have selected a subset of side chain interactions that were determined either in the same laboratory or at least using the same helix-coil transition theory implemented in the same computer program [5–7,11,20,22]. For this subset the correlation coefficient and the significance are very low and the slope (0.30) is also low.

Fourth, we are concerned that the empirical data base of side chain/side chain interactions is still small. A large data base for side chain/side chain interactions that contains energy values for all amino acid pairs has recently been reported [17]. The energies have been estimated from an analysis of the helical content of many peptides of known sequence and are likely to be less accurate than the individually determined energies shown in Table 1. The large number of energy values in the Muñoz and Serrano data base, however, has prompted us to compare them with our statistical energies. In Table 2 we show the results for the correlation between the statistical energies of the pairs and the corresponding AGADIR energies [17]. The correlation coefficient is very low and the slope is close to zero. If *i*,*i*+3 and *i*,*i*+4 pairs are correlated separately or if polar and apolar pairs are correlated separately the slope never rises above 0.15.

Fifth, the calculated statistical energies of uncommon pairs may be inaccurate. We have formed a subset of abundant pairs whose calculated frequencies are non-random with a confidence level of $>99.5\%$ (as evaluated with a χ^2 test). The statistical energies of these pairs (eight polar pairs: five *i*,*i*+4 and three *i*,*i*+3; and two *i*,*i*+4 apolar pairs) have been

Table 2
Slope, r and P values for different linear regression analyses between empirical and statistical interaction energies of amino acid pairs in α-helices

Energies correlated	Number of data points	Slope	r	P
All pairs[a]	27	0.22	0.34	0.0822
i,i+3 pairs[b]	10	−0.33	0.30	0.4093
i,i+4 pairs[c]	17	0.25	0.45	0.0733
Polar pairs[d]	17	−0.08	0.10	0.7047
Apolar pairs[e]	8	0.21	0.37	0.3709
Same laboratory[f]	18	0.30	0.29	0.2375
AGADIR energies[g]	648	0.10	0.14	0.0002
Significant pairs[h]	10	0.28	0.58	0.0810

[a]Correlation between empirically measured interaction energies of amino acid pairs in α-helices (at i,i+3 or i,i+4) and the corresponding statistical energies calculated according to Eq. 4. All energies are relative to alanine-alanine pairs.
[b]Correlation between empirically measured interaction energies of i,i+3 pairs in α-helices and the corresponding statistical energies calculated according to Eq. 4.
[c]Correlation between empirically measured interaction energies of i,i+4 pairs in α-helices and the corresponding statistical energies calculated according to Eq. 4.
[d]Correlation between empirically measured interaction energies of polar pairs in α-helices and the corresponding statistical energies calculated using Eq. 4 as follows: N_a and N_b for i,i+4 (i,i+3) polar pairs are the number of *a-x-x-x-p (a-x-x-p)* and *p-x-x-x-b (p-x-x-b)* helical sequences in the helical subspace of polar pairs where p is a polar helical amino acid. Energies of polar pairs are relative to lysine-lysine pairs, assuming a value of 0.00 kcal/mol for i,i+4 and i,i+3 lysine-lysine pairs. The polar pairs considered are: DK3, DK4, EH3, EH4, EK3, EK4, HD3, HE3, HE4, KD3, KD4, KE3, KE4, KK4, QD4, QE4, QN4.
[e]Correlation between empirically measured interaction energies of apolar pairs in α-helices and the corresponding statistical energies calculated using Eq. 4 as follows: N_a and N_b are calculated as above within the helical subspace of apolar pairs. Energies of apolar pairs are relative to alanine-alanine pairs. The apolar pairs considered are: FM4, LY3, LY4, MF4, YL3, YL4, YV3, YV4.
[f]Correlation between empirically measured interaction energies of amino acid pairs in α-helices that have been determined in the same laboratory or at least using the same computer program [5–7,11,20,22] and the corresponding statistical energies calculated according to Eq. 4. All energies are relative to alanine-alanine pairs.
[g]Correlation between interaction energies of amino acid pairs in α-helices determined from ellipticity data of peptides using a helix-coil algorithm (AGADIR) [17] and the corresponding statistical energies calculated according to Eq. 4. All energies are relative to alanine-alanine pairs.
[h]Correlation between empirically measured interaction energies of amino acid pairs in α-helices and the corresponding statistical energies calculated according to Eq. 4. Data corresponding to pairs whose occurrences deviated significantly from randomness at the 99.5% level (data from Table 1 with $P < 0.005$ as evaluated by a χ^2 test). All energies are relative to alanine-alanine pairs.

correlated with their empirical energies. In this case, the correlation between statistical and empirical energies is better ($r = 0.58$; $P = 0.081$) but the slope remains at just 0.28 (Table 2).

4. Discussion

4.1. Use of side chain/side chain interactions within natural α-helices

Do proteins distribute the amino acid residues in their helices so that they tend to form i,i+3 or i,i+4 pairs according to the helix-stabilising character of the pair? If this is the case, a plot of the relative statistical energies of the pairs versus the observed relative empirical energies should yield a straight line with a slope equal to unity and an intercept of zero. We have found that such a plot (Fig. 1) has a correlation coefficient of 0.34 with $P = 0.082$. This r value does not support the hypothesis. However, since the empirical data base of side chain interactions (Table 1) is intrinsically heterogeneous, as it includes data on both polar and apolar pairs, i,i+3 and i,i+4 pairs and pairs analysed by different laboratories, we have separately analysed different subsets of the data base to see if a particular subset displays a better correlation. In most cases the correlation was similarly poor. Only the subset formed by the pairs whose frequencies show significant deviations from the random distribution displayed a better correlation between empirical and statistical energies ($r = 0.58$; $P = 0.081$). This could indicate that the worse correlations found for the other subsets were simply due to large errors in the calculated statistical energies, the errors being attributed to a small number of pairs in the data base. It is very noticeable, however, that even the plot with the better correlation shows a low slope of 0.28, similar to those found in the other plots.

The slope of these plots can be interpreted as a realisation factor (ρ) that converts empirical difference energies of helix stabilisation into statistical preference energies:

$$\Delta G^{\mathrm{stat}} = \rho \Delta G^{\mathrm{emp}} \quad (5)$$

If helical pairs appeared in α-helices according to a Boltzmann distribution the realisation factor would have the value of 1.0. The consistently low realisation factor found in our analyses, even in the only case where some correlation between statistical and empirical energies is found, indicates that the fact that a certain pair of amino acids are able to form a stabilising interaction when spaced at i,i+4 or i,i+3 in an α-helix does not make that pair much more frequent than others. This suggests that an average natural protein α-helix makes little use of i,i+3 or i,i+4 side chain interactions to increase its stability. Importantly, the little use of intrahelical side chain interactions by protein α-helices that we infer from our analyses seems to occur at both solvent-exposed and buried α-helical surfaces.

4.2. How protein α-helices attain stability

Simple reasoning identifies three sources of helix stability: the individual tendency of each amino acid to stabilise the helical conformation (whatever the physical reasons), the formation of energetically favourable side chain/side chain interactions (mainly at i,i+3 and i,i+4), and packing interactions of the helix with the rest of the protein.

The influence of amino acid composition on helix stability

is best known. The helix-stabilising properties of the amino acids have been extensively studied and there are several empirical energy scales available that quantify these effects [17,24–29]. In general, good correlations are found between statistics and energetics. We analysed the data in those energy scales in a similar way to our analyses of side chain interactions and we found that the realisation factor that converts energetic preference into statistical preference for individual amino acids is around 0.8 (see http://wwwbioq.unizar.es/helixsc.html), close to the theoretical value of 1.0. This indicates that the amino acid composition of α-helices is close to be governed by Boltzmann's law. However, given the generally low helix-stabilising character of the genetically encoded amino acids, with such an amino acid composition and in the absence of other stabilising interactions, natural α-helices would not be stable.

A means of increasing helix stability is the distribution of the helical amino acids so that they can form stabilising side chain interactions. We found that this strategy is poorly exploited by proteins as judged from the low realisation factor (lower than 0.3) that converts the stability of a pair into a statistical preference. A small contribution of intrahelical side chain interactions to helix stability has also been recently suggested [30] based on the small number of close side chain contacts in protein α-helices. Little stability is thus obtained by protein α-helices from intrahelical side chain/side chain interactions pointing to the fact that the average protein sequence that will become an α-helix when the protein folds does not contain enough intrinsic energy resources to become autonomously stabilised. If protein α-helices themselves are marginally stable then packing of the helix against the rest of the protein must provide the additional stabilising resource that allows the amino acid sequence to attain the helical conformation. In certain cases, packing interactions determine the conformation of amino acid sequences within proteins [31].

The conformational stability of natural proteins is usually low (some 5–15 kcal/mol), which may be important for functional reasons. Protein engineering experiments clearly show that the stability of proteins can easily be increased by simple point mutations [32,33]. Proteins thus seem not to be pressed to use all available strategies to increase their stability and may accordingly choose among different possibilities. The reluctance of natural proteins to use helical side chain/side chain interactions as an important source of conformational stability might simply reflect the inconvenience of this particular approach. In this respect it has been suggested that fast protein folding requires a predominance of non-local interactions [34,35] and that local interactions have a low specificity for the native state [36].

Finally, from a practical point of view, our finding that protein α-helices do not tend to form $i,i+3$ and $i,i+4$ interactions suggests that the de novo design of natural-like proteins should concentrate more on including helix-stabilising residues and in the design of packing interactions than on incorporating stabilising side chain/side chain interactions into the sequences intended to form α-helices. On the other hand, as the solvent-exposed regions of protein α-helices seem not to be tailored to maximal stability they offer an interesting scenario for stability improvement by protein engineering.

Acknowledgements: This work was supported by Grants PB94-0599 from the DGICYT (Spain) and P15/97 from the DGA (Spain). J.F. is the recipient of a PFPI fellowship from the Ministerio de Educación y Ciencia (Spain). We thank Dr A. Rey for his comments on the manuscript.

References

[1] Davies, D.R. (1964) J. Mol. Biol. 9, 605–612.
[2] Prothero, J.W. (1966) Biophys. J. 6, 367–370.
[3] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 211–222.
[4] Chakrabartty, A. and Baldwin, R.L. (1995) Adv. Protein Chem. 46, 141–176.
[5] Huyghues-Despointes, B.M.P., Klingler, T.M. and Baldwin, R.L. (1995) Biochemistry 34, 13267–13271.
[6] Stapley, B.J., Rohl, C.A. and Doig, A.J. (1995) Protein Sci. 4, 2383–2391.
[7] Stapley, B.J. and Doig, A. (1997) J. Mol. Biol. 272, 465–473.
[8] Padmanabhan, S. and Baldwin, R.L. (1994) J. Mol. Biol. 241, 706–713.
[9] Viguera, A.R. and Serrano, L. (1995) Biochemistry 34, 8771–8779.
[10] Fernández-Recio, J., Vázquez, A., Civera, C., Sevilla, P. and Sancho, J. (1997) J. Mol. Biol. 267, 184–197.
[11] Huyghues-Despointes, B.M.P. and Baldwin, R.L. (1997) Biochemistry 36, 1965–1970.
[12] Hobohm, U. and Sander, C. (1994) Protein Sci. 3, 522–524.
[13] Vriend, G. (1990) J. Mol. Graphics 8, 52–56.
[14] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[15] Vriend, G., Sander, C. and Stouten, P.F.W. (1994) Protein Eng. 7, 23–29.
[16] Klingler, T.M. and Brutlag, D.L. (1994) Protein Sci. 3, 1847–1857.
[17] Muñoz, V. and Serrano, L. (1994) Nature Struct. Biol. 1, 399–409.
[18] Stellwagen, E., Park, S.-H., Shalongo, W. and Jain, A. (1992) Biopolymers 32, 1193–1200.
[19] Lyu, P.C., Gans, P.J. and Kallenbach, N.R. (1992) J. Mol. Biol. 223, 343–350.
[20] Scholtz, J.M., Qian, H., Robbins, V.H. and Baldwin, R.L. (1993) Biochemistry 32, 9668–9676.
[21] Shallongo, W. and Stellwagen, E. (1995) Protein Sci. 4, 1161–1166.
[22] Smith, J.S. and Scholtz, J.M. (1998) Biochemistry 37, 33–40.
[23] Shoemaker, K.R., Fairman, R., Schultz, D.A., Robertson, A.D., York, E.J., Stewart, J.M. and Baldwin, R.L. (1990) Biopolymers 29, 1–11.
[24] Lyu, P.C., Liff, M.I., Marky, L.A. and Kallenbach, N.R. (1990) Science 250, 669–673.
[25] O'Neil, K.T. and DeGrado, W.F. (1990) Science 250, 646–651.
[26] Wojcik, J., Altmann, K.-H. and Scheraga, H.A. (1990) Biopolymers 30, 121–134.
[27] Horovitz, A., Matthews, J.M. and Fersht, A.R. (1992) J. Mol. Biol. 227, 560–568.
[28] Chakrabartty, A., Kortemme, T. and Baldwin, R.L. (1994) Protein Sci. 3, 843–852.
[29] Blaber, M., Zhang, X.-J., Lindstrom, J.D., Pepiot, S.D., Baase, W.A. and Matthews, B.W. (1994) J. Mol. Biol. 235, 600–624.
[30] Walther, D. and Argos, P. (1996) Protein Eng. 9, 471–478.
[31] Minor Jr., D.L. and Kim, P.S. (1996) Nature 380, 730–734.
[32] Sturtevant, J. (1994) COSB 4, 69–78.
[33] Viguera, A.R., Villegas, V., Avilés, F.X. and Serrano, L. (1996) Fold. Des. 2, 23–33.
[34] Abkevich, V.I., Gutin, A.M. and Shakhnovich, E.I. (1995) J. Mol. Biol. 252, 460–471.
[35] Govindarajan, S. and Goldstein, R.A. (1995) Proteins 22, 413–418.
[36] Muñoz, V. and Serrano, L. (1996) Fold. Des. 1, R71–R77.