Side-Chain Interactions between Sulfur-Containing Amino Acids and Phenylalanine in α-Helices

Ana Rosa Viguera and Luis Serrano*

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

Received February 1, 1995; Revised Manuscript Received March 22, 1995[⊗]

ABSTRACT: The side-chain−side-chain interaction between Phe residues and sulfur-containing residues (Cis and Met) in the two possible orientations at positions i, $i + 4$ of α-helices is described. We have analyzed the contribution to helical stability of the above interactions by studying eight polyalanine-based peptides differing at the residues at positions 9 and 13. These two positions were independently mutated from Ala (AA), to Cys (AC and CA), Met (AM and MA), and Phe (AF and FA) and to the pairs Phe-Met (FM), Met-Phe (MF), Phe-Cys (FC), and Cys-Phe (CF). The intrinsic helical propensities of Cys, Met, and Phe were found to be those previously described in the algorithm AGADIR. NMR analysis of the FM, MF, FC, and CF peptides showed the formation in aqueous solution of contacts between the aromatic ring and the side chains of Cys or Met, at the two i, $i + 4$ orientations. CD studies demonstrated the important contribution of two of these interactions (FM and FC) to α-helix stability (up to 2 kcal mol^{-1} in the Phe-Cys pair). Statistical analysis of the protein database provides a rationale for the stereospecificity and free energies of the interactions. The very favorable interaction between an aromatic ring and a sulfur-containing amino acid explains why in the protein database around 50% of the sulfur atoms are contacting aromatic rings (Reid *et al.*, 1985).

Currently, attempts to describe the energetics of systems formed by short polypeptide chains have been addressed in α-helices. The model most often used for that purpose, the helix−coil transition, is based on the statistical mechanics theory and needs the assumption of certain simplifications (Zimm & Brag, 1959; Lifson & Roig, 1961). Recently, all the available experimental information for α-helix formation and stability has been gathered and included in an algorithm, AGADIR, based on the helix−coil transition theory (Muñoz & Serrano, 1994a, 1995a). This algorithm successfully predicts, within the limits of experimental error, the helical content at a residue level of monomeric peptides in aqueous solution at different pHs and temperatures (Muñoz & Serrano, 1995b). There are 400 possible side-chain−side-chain interactions at positions i, $i + 3$ and i, $i + 4$ of an α-helix. These interactions were initially estimated from the NMR analysis of peptides in 30% TFE, following a naive approach in which the residues were grouped according to their chemical properties (Muñoz & Serrano, 1994a; Muñoz *et al.*, 1995). Later they were refined by the study of the helical content, determined by circular dichroism (CD),[1] of a large peptide database. This approach allows one to obtain precise values for the common side-chain−side-chain interactions, but not for interactions between amino acids which are not frequent in the database. In AGADIR the interaction between an aromatic residue at position i and a Met and/or a Cys at position $i + 4$ is considered favorable (−0.4 and −0.2 kcal mol^{-1}, respectively) (Muñoz & Serrano, 1995a). The opposite orientation is considered to be less favorable (−0.2 kcal mol^{-1} for the Met-Phe pair), or even repulsive (0.1 kcal mol^{-1} for the Cys-Phe pair) (Table 1). It was pointed out several years ago that in proteins there are groups of alternating aromatic and sulfur-containing side chains that could stabilize proteins through strong noncovalent interactions between sulfur groups and the π electron system of aromatic rings (Morgan *et al.*, 1978; Reid *et al.*, 1985). Experimental (Rodner *et al.*, 1980) and theoretical (Nemethy & Scheraga, 1981) analysis of the interaction between sulfur and aromatic compounds showed it to be very favorable (up to −3.3 kcal mol^{-1} in the most favorable orientation of a dimethyl sulfide−benzene complex; Nemethy & Scheraga, 1981). The large discrepancy between this number and those assigned by AGADIR for these interactions could be due to either the influence of other factors (i.e., entropic cost of fixing the amino acid side chains or steric hindrance in α-helices) or an underestimation due to the scarcity of peptides with these interactions, or both. In this work the interaction between Phe and Cys or Met residues and its contribution to helix stability have been experimentally analyzed for the first time, by CD and nuclear magnetic resonance (NMR). The experimental results have been compared to the statistical analysis of the protein database.

MATERIALS AND METHODS

Peptide Synthesis. The peptides were synthesised at the EMBL peptide synthesis service by solid-phase synthesis methods. Peptide homogeneity, composition, and molecular weight were checked by analytical HPLC, amino acid analysis, and matrix-assisted laser desorption time-of-flight mass spectrometry.

The concentrations of the peptide samples were determined by ultraviolet absorbance (Gill & von Hippel, 1989).

^{1}H-NMR Spectroscopy. All the peptides (3 mM) were analyzed in water at pH 3.0 and 278 K. DQFCOSY,

[⊗] Abstract published in *Advance ACS Abstracts,* June 15, 1995.

[1] Abbreviations: TSP, sodium 3-(trimethylsylyl)[2,2,3,3-$^2H_4$]propionate; HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; 2D, two-dimensional; ppm, parts per million; COSY, 2D scalar correlated spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; CD, circular dichroism.

0006-2960/95/0434-8771$09.00/0

TOCSY, and NOESY (mixing time = 200 ms) spectra were performed in a BRUKER AMX 500MHz spectrometer using standard procedures (Wütrich, 1986). Sodium 3-(trimethylsilyl)[(2,2,3,3-$^2H_4$]propionate (TSP) was used as an internal reference. The proton resonances were assigned by the sequential assignment procedure. The $C_\alpha H$ proton conformational shifts were obtained by subtracting the random-coil chemical shift values (Wütrich, 1986), from those measured in the peptide.

Circular Dichroism Spectroscopy. CD spectra were recorded on a Jasco-710 instrument calibrated with (1*S*)-(+)-10-camphorsulfonic acid. Single scans in the range between 180 and 250 nm were obtained at a temperature of 278 K by taking points every 0.2 nm with a 1-s integration time and a 1-nm slit width. Cells with path lengths of 0.01 and 0.5 cm were used for the analysis of peptide concentrations of 500 and 10 μM, respectively. Each spectrum is the average of 20−30 scans smoothed by using the reverse Fourier transform noise reduction software from JASCO, after baseline subtraction.

Aggregation Tests of the Peptides. The possible aggregation of the three peptides was analyzed by CD. At pH 7.0 the peptides aggregated. At pH 3.0 no significant changes in the far-UV CD spectra were found in the 10 μM to 0.5 mM range, indicating that the peptides are basically monomeric in water.

Calculation of the Helical Content. The helical content of the peptides was estimated by two different methods. The first one is based on the empirical equation developed by Chen *et al.* (1974):

$$\% \text{ helical content} = 100[\theta_{222}^{obs}/(-39\,500(1 - 2.57/n))] \quad (1)$$

where n is the number of peptide bonds and is θ_{222}^{obs} the ellipticity of the peptides at 222 nm. For a peptide of 15 residues, $n = 14$. In the second method the helical content was estimated from the ellipticity at 193 nm using eq 1

$$\% \text{ helix} = 18.782 + 0.00072555\theta_{193}^{obs} \quad (2)$$

where θ_{193}^{obs} is the mean residue ellipticity at 193 nm. This equation was obtained by correlating the ellipticity changes at 193 nm with the helical contents obtained from the method of Chen *et al.* (1974) for the peptides without Phe and for peptide AA in the presence of different TFE concentrations (Figure 4A).

Statistical Analysis of the Protein Database. The probability of finding a certain fingerprint sequence ($P_{fingerprint}$) in a helical conformation if there is a random distribution is calculated by multiplying the individual probabilities of each residue type (P_{ind}) of the fingerprint.

An individual probability is the probability of finding a specific residue type in a certain position when this position is in a helical conformation. To calculate the probabilities, we have used a protein database with 279 structures [Hobohm *et al.*, 1992; the full list is described in Muñoz and Serrano (1994b)], included in the program WHATIF (Vriend, 1990). The protein database has been divided into three-residue segments, and we have counted the number of those in which the central position is helical ($N_{conf} = 20\,744$). 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 N_{conf}:

$$P_{ind} = N_{res}/N_{conf} \quad (3)$$

For a Phe residue (F in Table 3), we find 830 cases ($P_{ind} = 0.04$). For Cys, (C in Table 3), we find 245 cases ($P_{ind} = 0.012$). Finally, for Met (M in Table 3), we find 548 cases ($P_{ind} = 0.026$). With these individual probabilities we can calculate the probability of a particular fingerprint. This fingerprint in our case is made up of five residues in a helical conformation in which the first and last positions are occupied by one of the residues we are analyzing (Phe, Cys, or Met) and the three in the center can be occupied by any residue (X in Table 3). The individual probability of the three central residues is 1.

The number of cases expected ($N_{expected}$) for each sequence fingerprint in a five-residue helical segment (see Table 3) is calculated from the following equation:

$$N_{expected} = P_{fingerprint} N_{total} \quad (4)$$

where N_{total} is the total number of five-residue segments adopting the helical conformation (13 164 cases).

RESULTS

Peptide Design. To experimentally assess the favorable contribution of the Phe-Cys, Cys-Phe, Met-Phe, and Met-Phe pairs, we have designed a template peptide with the sequence Tyr-Gly-Gly-Ser-Ala-Ala-Glu-Ala-X-Ala-Lys-Ala-Y-Ala-Arg-NH_2, where X at position 9 and Y at position 13 represent the two positions mutated to Cys, Met, or Phe (Ala in the reference peptides). This peptide has a capping box motif at the N-terminus to favor helix nucleation (Lyu *et al.*, 1992; Muñoz & Serrano, 1994a, 1995a), and the C-terminus is blocked to eliminate the dipole destabilization effect (Muñoz & Serrano, 1995a). There is a Tyr residue at the N-terminus to measure protein concentration (Gill & von Hippel, 1989), which is separated by two Gly residues from the rest of the peptide to eliminate the Tyr contribution to the far-UV CD spectrum (Chakrabartty *et al.*, 1993). The mother peptide (containing Ala at positions X and Y) was designed by AGADIR to have a helical content around 35% at pH 7.0. Helical contents in the range 15 to 70% are very sensitive to small changes in energy (Muñoz *et al.*, 1995a), and therefore it is easier to detect the interaction between Phe and Cys or Met. Also, the shape of the CD spectrum for helical contents in the 15−40% range is quite indicative for the helix population, while at values higher than 40% it is much less sensitive.

NMR Analysis. The peptides containing Cys at position 9 and Phe at position 13 (CF) or Met at position 9 and Phe at position 13 (MF), and those with opposite orientations (FC and FM), have been studied by nuclear magnetic resonance (NMR), in water. Figure 1A−D shows the NOE summary for these peptides. Although there is some signal overlapping, the existence of several $i, i + 3$ NOEs, as well as the upfield chemical shifts (Figure 1 E,F; the chemical shift values of the peptides are available as supporting information), shows the existence of a significantly populated α-helical conformation starting at Ser4, which extends up to Arg15. There are several NOE cross peaks between the side chains of Ser4 and Glu7, which are typical of the formation of a capping box motif (Lyu *et al.*, 1992), but most interestingly we find strong NOE cross peaks between the side chain of Phe at position N5 and those of Cys or

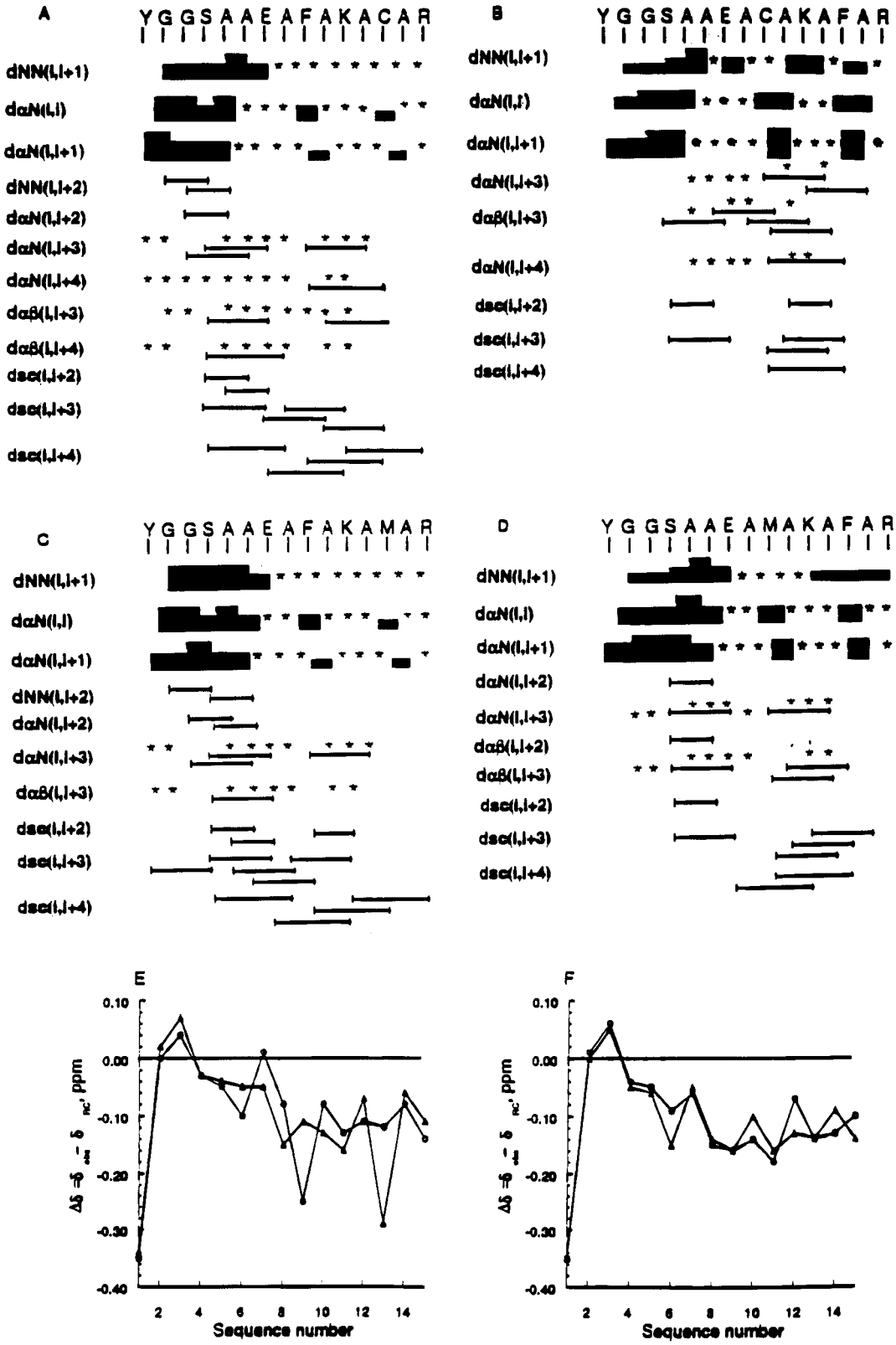

FIGURE 1: Summary of the NOEs observed for 3 mM peptide samples in water, pH 3.0, 278 K: (A) FC, (B) CF, (C) FM and (D) MF. NOEs involving side chains are grouped as dsc. An asterisk indicates NOEs that could not be observed due to resonance overlap; for the nonsequential NOEs it is placed at the position of the i^{th} residue. (E) Plot of the conformational shifts of the $C_\alpha H$ protons defined as $\delta_{measured} - \delta_{coil}$ in water, of the FC peptide (circles) and the CF peptide (triangles). (F) Plot of the conformational shifts of the $C_\alpha H$ protons defined as $\delta_{measured} - \delta_{coil}$ in water, of the FM peptide (circles) and the MF peptide (triangles).

Met at position N9 and vice versa (Figure 2A−D). This indicates that the side chain of Phe interacts in aqueous solution with the side chains of Cys and Met, at the two possible i, $i + 4$ orientations. In the CF and FC peptides the chemical shift of the Cα proton of the Cys residue is significantly shifted upfield (Figure 1E). This can be the result of the proximity of the aromatic ring to the Cα group (Giessner-Prettre & Pullman, 1981), due to a favorable interaction with the Cys side chain. On the other hand, in the FM and MF peptides this upfield shift of the Cα proton

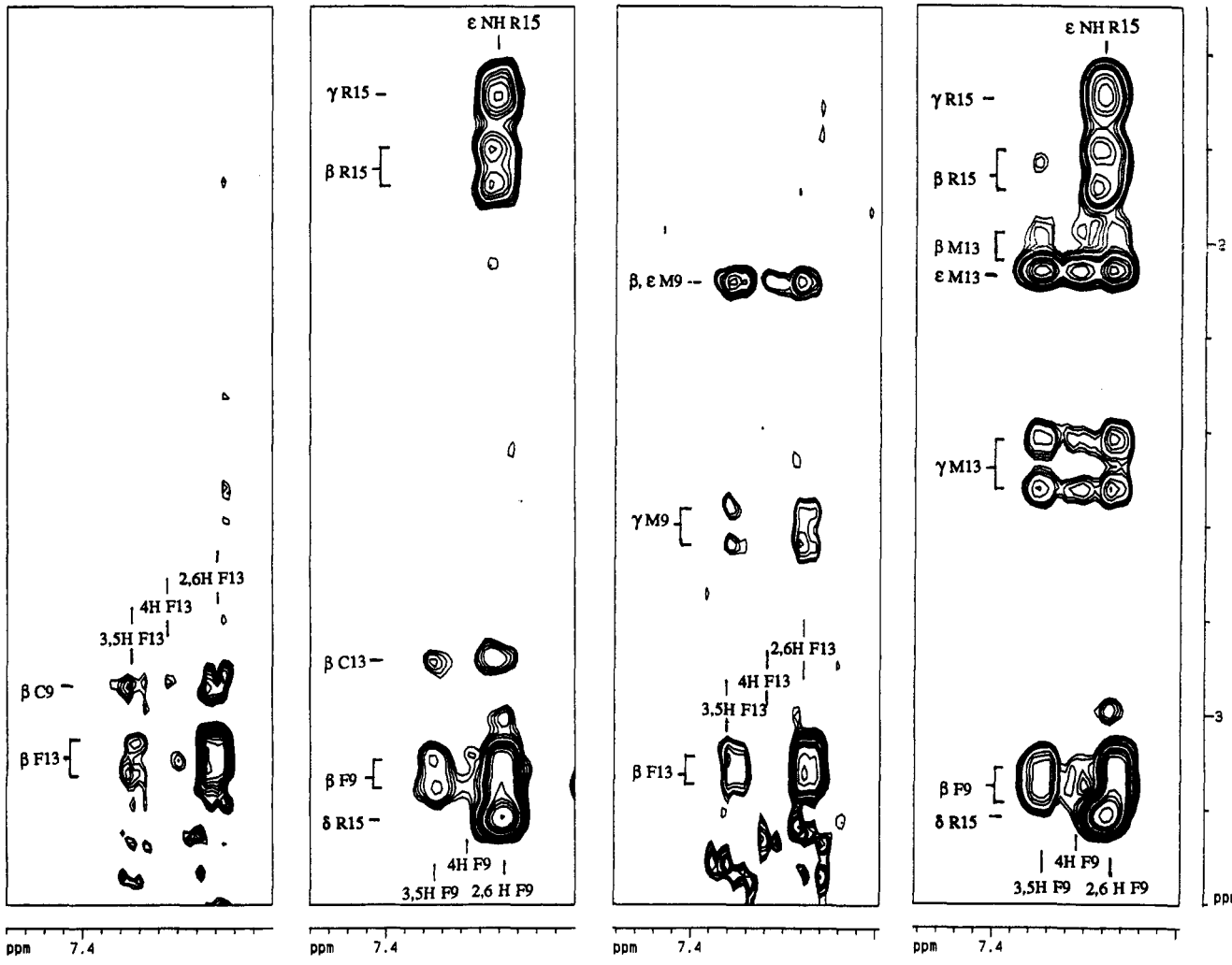

FIGURE 2: Selected regions of the NOE spectra showing the NOEs between the side chains of Cys and Met with Phe. From left to right: CF peptide, FC peptide, MF peptide, and FM peptide.

of the Met residue is not seen, indicating that the aromatic ring is far away from the Cα, as expected if the aromatic ring is close to the sulfur atom. This is corroborated by the low intensity of the NOEs between the Cβ of Met and the aromatic ring in the FM peptide (Figure 2D; in the MF peptide this cannot be seen due to signal overlapping with the Cε protons).

CD Analysis. The energetics of the interactions have been analyzed through the CD analysis of the helical content in a series of mutant peptides (Figure 3). As expected from their helical propensities (O'Neil & DeGrado, 1990; Lyu *et al.*, 1990; Horovitz *et al.*, 1992; Blaber *et al.*, 1993; Chakrabartty *et al.*, 1994; Muñoz & Serrano, 1995a), introduction of a Met or a Cys residue at position 9 decreases the helical content of the peptides with respect to Ala (Figure 3A). The same happens, although to a lesser extent, when they are introduced at position 13 (Figure 3B). This difference is expected due to the fact that normally in peptides the helix ends are frayed, and consequently introduction of an unfavorable residue in the middle of the helix decreases the helical population to a larger extent than when the residue is placed at the N- or C-terminus (Chakrabartty *et al.*, 1991). Placing a Phe residue at position 9 or 13 also decreases the helical content and modifies the CD spectrum with respect to the other peptides (Figure 3C,D) (Woody, 1978; Chakrabartty *et al.*, 1993).

Estimation of the Helical Content. The error in the estimation of the helical content of peptides in the absence of aromatic residues (if they are not separated by at least two Gly residues from the helical region; Chakrabartty *et al.*, 1993), or alternative secondary structure conformations, is mainly dependent on the imprecision in the determination of peptide concentration. The presence of a Tyr residue allows obtaining an accuracy of $\pm 1-2\%$ in the estimation of peptide concentration (Shoemaker *et al.*, 1990). This results in an average error of around $\pm 0.5\%$ in the determination of the helical population in our peptides (Table 1). In the case of the peptides containing a Phe residue in the helical region, the aromatic ring has a significant positive contribution around 222 nm, preventing use of the method of Chen *et al.* (1974) to calculate the helical content (Woody, 1978; Chakrabartty *et al.*, 1993). One way to solve this problem is to find a wavelength in which the aromatic contribution to the far-UV CD spectrum is nil. In the model dipeptide, Lys-Phe, the far-UV CD spectrum has zero ellipticity at 193 nm (Brahms & Brahms, 1980). However, it is not clear whether this will be the case in a helical enviroment. To determine the aromatic contribution of Phe in a helical conformation, it is necessary to have two peptides with the same helical content, differing only in the presence or absence of this residue. In our case, this can be done by adding trifluoroethanol (TFE) to the peptides in saturating

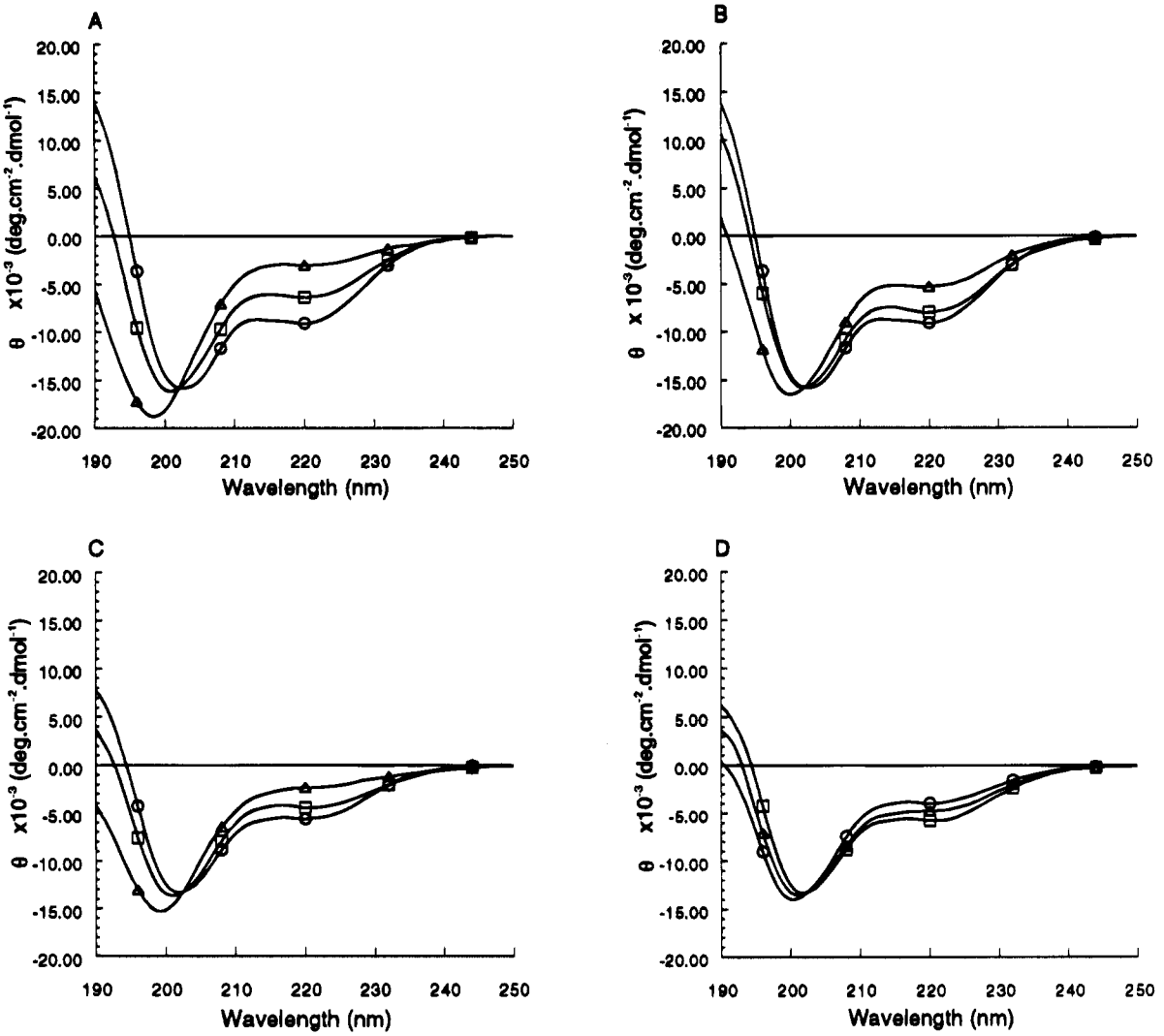

FIGURE 3: CD analysis of the peptides. (A) CD spectra of peptides AA (circles), CA (triangles), and MA (squares). (B) Peptides AA (circles), AC (triangles), and AM (squares). (C) Peptides AF (circles), CF (triangles), and MF (squares). (D) Peptides FA (circles), FC (filled triangles), and FM (open squares). Conditions: [peptide] = 0.50 mM; 278 K; pH = 3.0 in water.

Table 1: CD Analysis of the Peptides and Determination of the Free Energy of Interaction between the Different Side Chains Being Considered

peptide	% helix[a] (193 nm)	%helix[b] (222 nm)	AGADIR[c]	AGADIR[d] (new)	ΔG_{int}[e] (kcal mol^{-1})
AA	31 ± 3	30 ± 3	30	30	
FA	19 ± 3	13 ± 3	16	16	
AF	22 ± 3	19 ± 3	22	22	
AM	27 ± 3	26 ± 3	25	25	
FM	25 ± 3	20 ± 3	18	22	−0.65 ± 0.15 (−0.4)
MA	21 ± 3	21 ± 3	22	22	
MF	21 ± 3	15 ± 3	19	19	−0.20 ± 0.2 (−0.2)
AC	19 ± 3	18 ± 3	20	20	
FC	21 ± 3	16 ± 3	12	18	−2.0 ± 0.4 (−0.3)
CA	10 ± 3	10 ± 3	13	13	
CF	14 ± 3	8 ± 3	9	11	−0.60 ± 1.7 (0.1)

[a] Helical content calculated from the ellipticity at 193 nm (see Materials and Methods). The error is calculated from the ±2% error in the concentration determination (Shoemaker *et al.*, 1990). [b] Helical content calculated by the method of Chen *et al.* (1974). The error is calculated from the ±2% error in the concentration determination (Shoemaker *et al.*, 1990). [c] Helical content predicted by AGADIR, without the new values for the interactions studied here. [d] Helical content calculated by AGADIR after introducing the values for the interactions shown in column 6. [e] Free energy of interaction introduced into AGADIR to reproduce the experimental results in column 2, within a ±3% error. The error number is obtained by calculating the free energy of interaction that will produce a predicted value which is 3% higher or lower than the value shown in column 5. In the case of the CF peptide, it is not possible to decrease the helical content of the peptide below 9%. This is due to alternative helical conformations with the Cys as the N-cap. Therefore, we have calculated the error by just considering the free energy of interaction required to raise the predicted helical content by 3%. In parentheses we show the values considered before (Muñoz *et al.*, 1995a).

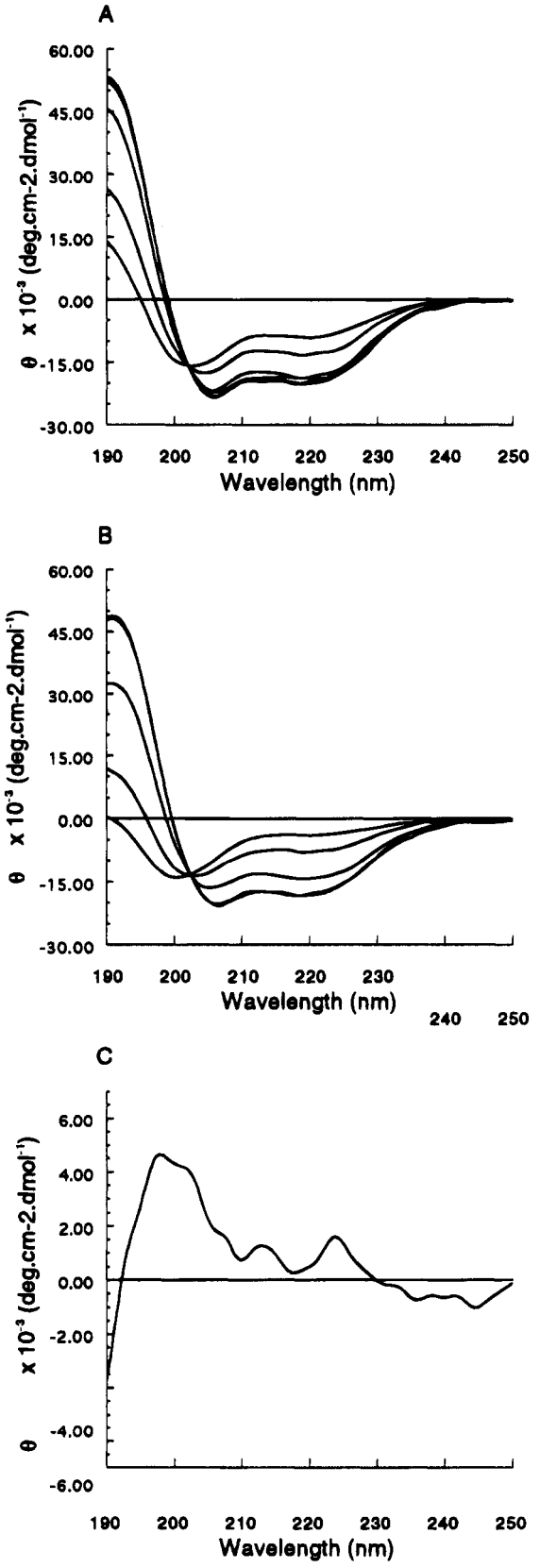

FIGURE 4: TFE titration of the AA and FA peptides. (A) CD spectra of peptide AA with different TFE concentrations (0, 5, 14, 25, 33, and 40%). (B) CD spectra of peptide FA with different TFE concentrations (0, 5, 14, 25, 33, and 40%). (C) CD difference spectrum of peptide FA minus peptide AA, in 40% TFE.

concentrations. Titration of the AA and FA peptides with TFE shows that they reach similar maximum helical contents at around 30%, v/v, TFE (Figure 4A,B). Subtraction of the CD spectrum of the AA peptide from that of the FA peptide, at 40%, v/v, TFE should correspond to the far-UV spectrum of the Phe aromatic contribution. This spectrum has zero ellipticity at 193 nm (Figure 4C), as does the model Lys-Phe peptide (Brahms & Brahms, 1980). The same happens when the spectra in 30% TFE are used (data not shown). It is then possible to estimate the helical content of the peptides without Phe, using the method of Chen *et al.* (1974) (Table 1), and to correlate them with the changes in the ellipticity at 193 nm (see Materials and Methods). From this correlation it is possible to derive an empirical equation that estimates the helical content of the peptides from the ellipticity at 193 nm (see Materials and Methods) and use it to calculate the helical population of the peptides with a Phe residue (Table 1). The difference spectrum of two peptides with the same helical content determined as indicated above (MA and MF peptides) is very similar to that obtained from the difference in the TFE spectra of peptides AA and FA. Moreover, there is a good correlation of the helical contents determined by this method and the position of the ellipticity minimum in the different peptides (data not shown).

Table 2: Determination of the Intrinsic Helical Propensities of Cys, Met, and Phe

residue	AGADIR[a]	O'Neil[b]	Lyu[c]	Horovitz[d]	Blaber[e]	Chakrabartty[f]
M	0.21	0.31	0.33	0.31	0.10	0.51
F	0.47	0.36		0.69	0.37	0.93
C	0.60	0.54		0.81	0.54	0.81

[a] Intrinsic helical propensities for Met, Phe, and Cys, with respect to Ala, were obtained from the analysis by AGADIR of more than 450 peptides (Muñoz *et al.*, 1995a). These values reproduce the CD experimental data in this work. [b] Values from O'Neil and DeGrado (1990). [c] Values from Lyu *et al.* (1990). [d] Values from Horovitz *et al.* (1992). [e] Values from Blaber *et al.* (1993). [f] Values from Chakrabartty *et al.* (1994).

Intrinsic Helical Propensities. We have used the CA, MA, FA, AC, AM, and AF peptides to determine the intrinsic helical propensities of Cys, Met, and Phe. This can be done in these peptides since, with respect to position 9, there are only Ala residues at positions $i - 4$, $i - 3$, $i + 3$, and $i + 4$, and the same happens with respect to position 13, but in this case there are no residues at positions $i + 3$ and $i + 4$. AGADIR correctly predicts the helical content of the peptides having these residues at positions 9 and 13 (within a ±3% error) (Table 1), thus showing that the intrinsic propensities of these amino acids were correctly estimated in the algorithm. The intrinsic values for these residues are closer to those obtained by Lyu *et al.* (1990), O'Neil & DeGrado (1990), Horovitz *et al.* (1992), and Blaber *et al.* (1993) than to those from Baldwin and co-workers (Chakrabartty *et al.*, 1994), as previously reported (Muñoz & Serrano, 1995a; Muñoz *et al.*, 1994b) (Table 2).

Free Energies of Interaction. To calculate the free energies of interaction between the side chains of Phe and those of Cys or Met, we could perform a double mutant cycle (Serrano *et al.*, 1990). However, in the case of helical peptides we do not have a two-state transition, and consequently this approximation is not valid (Lyu *et al.*, 1990), since it renders significantly smaller values. We have estimated the free energies of interaction by using the helix/coil transition theory incorporated in the algorithm AGADIR (Muñoz & Serrano, 1994, 1995a,b). To do so, we have fitted

Table 3: Statistical Preferences for the Interactions between Aromatic and Sulfur-Containing Residues

sequence[a]	database[b]	α_5[c]	α_{5expec}[d]	$freq_{\alpha 5}$[e]	contact[f]	freq[g]
FXXXC	42	7	6	1.17	5	0.72 ± 0.23
FXXXM	59	32	13	2.46	16	0.50 ± 0.03
CXXXF	37	13	6	2.17	3	0.23 ± 0.03
MXXXF	46	17	13	1.31	2	0.12 ± 0.01

[a] Sequence fingerprint: X is any amino acid, F is Phe, C is Cys, and M is Met. [b] Number of five-residue segments which have each sequence fingerprint in the protein structure database. [c] Number of five-residue segments in helical conformation which have each sequence fingerprint. [d] Number of expected five-residue segments in column 3, if the residue distribution were random. [e] Normalized frequency of occurrence for the conformation indicated in column 3. [f] Number of cases in which the distance between the van der Waals radius of the side chains being considered is less than 0.25 Å. [g] Ratio between the number of five-residue segments in helical conformation (column 3) that have each sequence fingerprint and those in which the side chains are in contact with each other (column 6). The error is calculated by adding or subtracting two cases to or from column 3 and then calculating the difference of the resulting values with the real value. The search was done in a protein database of 279 proteins with less than 50% homology filtered for the quality of data (Hobohm *et al.*, 1992) with the program WHATIF (Vriend, 1990). The proteins in which the distance between the van der Waals radii of the side chains being considered is less than 0.25 Å are the following: The pair Phe-Cys was found in 5rub (164−168), 1nar (242−246), 1bgc (6−10), 1prc (153−157), and 1mup (134−138). The pair Cys-Phe was found in 135l (30−34), 2ccy (120−124), and 1min (94−98). The pair Phe-Met was found in 1csh (222−226, 260−264), 1lid (16−20), 3cla (50−54), 1osa (141−145), 1byb (433−437), 1pbe (342−346), 4fxn (99−103), 2lhb (134−138), 2scp (75−79), 1gox (325−329), 1bpb (185−189), 1prc (70−74, 230−234, 269−273), and 1col (52−56). Finally, the pair Met-Phe was found in two proteins: 3sdh (36−40) and 5rub (352−356).

the value of the energy contribution of the i, $i + 4$ side-chain−side-chain interaction to obtain the experimental helical content within a 3% error (Table 1). Two of the interactions analyzed here (FC and FM) turned out to be significant and favorable (Table 1), especially the one between the aromatic ring of Phe and a Cys at position $i + 4$, (-2 ± 0.4 kcal mol^{-1}). The free energies of interaction for the opposite orientations (CF and MF), although favorable, are smaller and within the fitting error with AGADIR.

Statistical Analysis. We have performed a statistical analysis of the protein database, looking at the frequency of the simultaneous appearance of different residues at positions i and $i + 4$ (see Materials and Methods). If any pair of residues at these positions in α-helices is more abundant than expected from a random distribution, it suggests the existence of a favorable interaction between them (Sippl, 1990). In Table 3 we show the statistical preferences for the different orientations of the Phe-Cys and Phe-Met pairs in α-helices in proteins. The pair in which the Phe is at position i and the Met at position $i + 4$ seems to be the most favorable, while that with a Cys at position $i + 4$ is the least favorable. The last result is in contradiction with the experimental data shown above. On the other hand, if we compare the number of cases in which the side chain of the Phe residue contacts the side chain of Cys or Met residues (less than 0.25 Å between their van der Waals radii) with the total number of cases in the protein database, we find a good correlation with the experimental results.

Geometry of the Interaction. In Table 4 we show the average dihedral angles, obtained from the protein database, for the four pairs of i, $i + 4$ helical residues analyzed here (Phe-Cys, Cys-Phe, Phe-Met, and Met-Phe), when the van der Waals radii of their side chains are less than 0.25 Å apart. The rotamers of the aromatic ring in the experimentally more favorable orientation (Phe-Cys and Phe-Met) correspond to those which are more frequent in α-helices ($-160 < \chi_1 > 160$ and $60 < \chi_2 > 100$; ~50% in α-helices). The same applies to Cys ($-40 < \chi_1 > -80$, ~60%) and to the χ_1 angle of Met ($-40 < \chi_1 > -80$; ~50%). On the other hand, in the opposite orientations (Cys-Phe and Met-Phe), there are not any preferred rotamers for the side chains (Table 4), and these are very rare in α-helices, <5%. The strained conformation of the Phe ring in the Cys-Phe and Met-Phe pairs is translated to the ψ angle, which is smaller on average than in the favorable orientation (Table 4). The statistical analysis of the average distances between Phe rings and the side chains of Cys and Met, when they are contacting each other, indicates that in the most favorable orientation the sulfur atom is located above the edge of the aromatic ring, being closer to the meta and para carbons (Table 5). This is similar to what is found for the interaction between sulfur and aromatic compounds in solution (Rodner *et al.*, 1990) and between aromatic rings and Met or Cys side chains (Reid *et al.*, 1985).

DISCUSSION

The interactions between aromatic side chains and sulfur-containing amino acids, initially identified from the analysis of protein structures (Morgan *et al.*, 1978; Reid *et al.*, 1985) and the interaction of sulfur with aromatic compounds (Rodner *et al.*, 1980; Nemethy & Scheraga, 1981), seem to form a real local specific interaction that stabilizes α-helices through a combination between a hydrophobic interaction between the side chains of the aromatic ring and those of Cis and Met and, probably more importantly, a strong electrostatic interaction between the electronegative sulfur atom and the positively charged aromatic ring hydrogens

Table 4: Average Dihedral Angles of the Different Residue Pairs Studied Herein When in a Helical Conformation with Their Side Chains Contacting Each Other

pair	residue	ϕ	ψ	χ_1	χ_2	χ_3
FC	F	−62 ± 4.4	−45 ± 2.5	−176 ± 5.3	76 ± 5.6	
	C	−63 ± 5.4	−44 ± 8.6	−79 ± 12		
CF	F	−87 ± 35	−26 ± 20	−101 ± 53	3.3 ± 93	
	C	−63 ± 2.0	−44 ± 7.8	−138 ± 65		
FM	F	−62 ± 8.2	−45 ± 8.0	173 ± 70	55 ± 11	
	M[a]	−62 ± 7.3	−37 ± 4.4	−68 ± 5.0	−75 ± 33	−85 ± 34
MF[b]	F	−66	−35	−53	−37	
	M	−54	−50			−37

[a] Thirteen of the 17 FM pairs have negative values for the χ_1, χ_2, and χ_3 angles. The other four have mixed values for the three χ angles and are not shown here. [b] There are only two MF pairs in the database with the above requirements. When the corresponding angles differ by more than 20°, we do not show the data.

Table 5: Average Distances of the Different Residue Pairs Studied Herein When in a Helical Conformation with Their Side Chains Contacting Each Other[a]

pair	group	CG	CD_a	CD_b	CE_a	CE_b	CZ
FM	$M_{C\beta}$			5.4 ± 0.4		5.0 ± 0.3	5.3 ± 0.5
	M_{CG}	4.8 ± 0.5	5.0 ± 0.6	4.3 ± 0.7	4.6 ± 0.7	3.9 ± 0.7	4.1 ± 0.7
	M_{SD}	5.5 ± 0.4		4.8 ± 0.4	5.4 ± 0.6	4.3 ± 0.5	4.2 ± 0.6
	M_{CE}			5.3 ± 0.8	4.4 ± 0.8	5.0 ± 0.6	4.3 ± 0.7
FC	$C_{C\beta}$			5.5 ± 0.4		5.3 ± 0.2	5.5 ± 0.4
	C_{SH}	4.7 ± 0.1	4.7 ± 0.4	4.2 ± 0.2	4.4 ± 0.5	3.7 ± 0.1	3.8 ± 0.3

[a] Average distances between the side-chain groups of Cys and Met and the side-chain groups of Phe. Only those cases in which the average distance is smaller than 5.5 Å are shown. The ± indicates the standard deviation. In the two pairs shown here, FC and FM, one side of the ring is always closer to the side-chain groups of Cys and Met. To reflect this, we have switched the CD1 and CD2 as well as the CE1 and CE2 groups, so that the smallest distances are always assigned to the CD2 and CE2 groups. This is why we have changed the nomenclature to CD_a, CD_b, CE_a, and CE_b. We have searched with the program WHATIF[26] for those helical residue pairs at positions i, $i + 4$ that involve a Phe residue and either Cys or Met, in which the distance between the van der Waals' radii of their side chains is less than 0.25 Å.

(Rodner *et al.*, 1980; Nemethy & Scheraga, 1981). The favorable electrostatic interaction between the sulfur group and the aromatic ring should result in the latter being closer to the Cα of Cys than to that of Met. This is seen in the larger upfield shifts of the Cα protons of the Cys residue, compared to those of the Met in the FM and MF peptides. That the interaction between the aromatic side chain of Phe and the side chain of Cys or Met is taking place in a helical conformation is shown by the chemical shifts with respect to random coil values and the presence of several i, $i + 3$ and i, $i + 4$ NOEs in the interaction region.

The free energies of interaction of some of the pairs analyzed here have been obtained through the use of an algorithm based on the helix/coil transition theory. This algorithm, when applied to peptides containing other interactions analyzed by site-directed mutagenesis in proteins (i.e., salt bridges, capping interactions), renders values similar to those found experimentally (Muñoz & Serrano, 1995a). One problem in this approach is that the quantification of the helical content needs to be relatively precise. There are three possible sources of error in the quantification of the helical content in peptides: (i) The aromatic contribution in the far-UV CD spectrum. We have used the ellipticity at 193 nm to quantify the helical content of the peptides containing Phe residues, since the contribution of the Phe aromatic ring at this wavelength is nil in random coil peptides (Brahms & Brahms, 1980) or in a helical conformation (see above). This eliminates the aromatic contribution and allows precise measurements to be obtained. (ii) The experimental error in the determination of peptide concentration is small, ±2% (Shoemaker *et al.*, 1990) and results in an error around ±0.5% in the helical content. There is another error due to the baseline and the measurement of the ellipticity by the apparatus, which results in an error around 1−2% in the estimation of the helical content. Therefore, we can safely assume that the helical measurements are going to be precise within a ±3% error. (iii) The presence of alternative secondary structure conformations with different far-UV CD spectra (i.e., β-turns). This is very difficult to determine, but it is very unlikely that in polyalanine-based peptides there could be alternative conformations, significantly populated.

In peptides with a significant helical content (>15%), a small error in the quantification of the helical content does not result in significant errors in the determination of the interaction energies (Table 1). The opposite is true for peptides with little helical content (<15%). This is so because of the cooperative nature of the helix/coil transition theory, which requires relatively high energies to increase the helical content of a peptide when the original value is low (Muñoz & Serrano, 1995a). It is clear that the error in the estimation of the MF and CF interactions is very high and does not allow obtaining precise measurements of their interaction energies, although they seem to be favorable. Following the policy previously established in AGADIR (Muñoz *et al.*, 1995), we have considered those values which resulted in the prediction of a helical content, within a 3% difference with the experimental values. This prevents the overestimation of some interactions due to experimental errors and the possible presence of unknown factors in certain peptides.

As we showed in Results, the value resulting from the comparison between the number of cases in the protein database for the Phe-Cys pair and the expected ones does not correspond to the interaction energy calculated here. However, when looking only at those pairs in which the two side chains are in contact with each other, we found a good correspondence with the experimental data. The discrepancy could be due to a statistical error caused by the small number of expected Phe-Cys pairs or because of packing reasons (i.e., burial of the Cys residue in this conformation could be unfavorable). The analysis of the rotamers for Phe, Cys, and Met in the pairs analyzed here provides a thermodynamic explanation for the experimental free energies. In the Phe-Cys pair, the rotamers of the Phe and Cys side chains are very favorable in α-helices, and consequently the energy cost of making the interactions is low. The same will apply to the Phe-Met pair, although this case is more complicated due to the higher flexibility of the Met side chain, which should result in a larger entropic cost, the larger hydrophobic surface buried by the Phe-Met interaction, and the lower negative charge density of the sulfur atom. In the opposite orientation the Phe residue needs to adopt a high-energy rotamer to contact the Cys or Met residues, and the helix needs to be slightly distorted at the Phe position (smaller ψ angle), with the corresponding energy cost.

The strong contribution of these interactions in the favorable orientations to α-helix stability, plus the fact that they are present in aqueous solution, in the absence of tertiary interactions, suggests a role for them in the early stages of protein folding by promoting the formation of α-helices. These results explain why in proteins about half of the sulfur atoms are in contact with an aromatic ring (Reid *et al.*, 1985) and indicates that the interaction between an aromatic residue and a Cys or Met residue could be quite important for protein folding and stability in general.

SUPPORTING INFORMATION AVAILABLE

Four tables showing chemical shifts of the proton resonances in the FC, FM, CF, and MF peptides (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Blaber, M., Zang, X., & Matthews, B. (1993) *Science 260*, 1637−1640.

Brahms, S., & Brahms, J. (1980) *J. Mol. Biol. 138*, 149−178.

Chakrabartty, A., Kortemme, T., Padmanabhan, S., & Baldwin, R. L. (1993) *Biochemistry 32*, 5560−5565.

Chakrabartty, A., Schellman, J. A., & Baldwin, R. L. (1991) *Nature 351*, 586−588.

Chakrabartty, A., Kortemme, T., & Baldwin, R. L. (1994) *Protein Sci. 3*, 843−847.

Chen, Y. H., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350−3359.

Giessner-Prettre, C., & Pullman, B. (1981). *Biochem. Biophys. Res. Commun. 101*, 921−926.

Gill, S. C., & Hippel, P. H. v. (1989) *Anal. Biochem. 182*, 319−326.

Hobohm, U., Scharf, M., Schneider, R., & Sander, C. (1992) *Protein Sci. 1*, 409−417.

Horovitz, A., Matthews, J., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560−568.

Lifson, R., & Roig, A. (1961) *J. Chem. Phys. 34*, 1963−1974.

Lyu, P. C., Zhou, H. X., Jelveh, N., Wemmer, D. E., & Kallenbach, N. R. (1992) *J. Am. Chem. Soc. 114*, 6560−6562.

Lyu, P. C., Liff, L. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669−673.

Morgan, R. S., Tatsch, C. E., Gusahrd, R. H., McAdon, J. M., & Warme, P. K. (1978) *Int. J. Pept. Protein Res. 11*, 209−217.

Muñoz, V. & Serrano, L. (1994a) *Nature Struct. Biol. 1*, 399−409.

Muñoz, V., & Serrano, L. (1994b) *Proteins: Struct., Funct., Genet. 20*, 301−311.

Muñoz, V., & Serrano, L. (1995a) *J. Mol. Biol. 245*, 275−296.

Muñoz, V., & Serrano, L. (1995b) *J. Mol. Biol. 245*, 297−308.

Muñoz, V., Jimenez, M. A., Rico, M., & Serrano, L. (1995) *J. Mol. Biol.* (in press).

Nemethy, G., & Scheraga, H. A. (1981) *Biochem. Biophys. Res. Commun. 94*, 807−813.

O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646−650.

Reid, K. S. C., Lindley, P. F., & Thornton, J. M. (1985) *FEBS Lett. 190*, 209−213.

Rodner, B. L., Jackman, L. M., & Morgan, R. S. (1980) *Biochem. Biophys. Res. Commun. 94*, 807−813.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., & Fersht, A. R. (1990) *Biochemistry 29*, 9343−9352.

Shemaker, K. R., Fairman, R., Schultz, D. A., Robertosn, A. D., York, E. J., Stewart, J. M., & Baldwin, R. L. (1990) *Biopolymers 29*, 1−11.

Sippl, M. (1990) *J. Mol. Biol. 213*, 859−883.

Vriend, G. (1990) *J. Mol. Graph. 8*, 52−56.

Woody, R. W. (1978) *Biopolymers 17*, 1471−1476.

Wütrich, K. (1986) NMR of macromolecules, John Wiley & Sons, New York.

Zimm, B. H., & Brag, J. K. (1959) *J. Chem. Phys. 31*, 526−535.

BI9502295