Article No. mb982145 *J. Mol. Biol.* (1998) **284**, 173–191

JMB

Elucidating the Folding Problem of α-Helices: Local Motifs, Long-range Electrostatics, Ionic-strength Dependence and Prediction of NMR Parameters

Emmanuel Lacroix, Ana Rosa Viguera and Luis Serrano*

EMBL, Meyerhofstrasse 1
Heidelberg D-69117, Germany

The information about the conformational behavior of monomeric helical peptides in solution, as well as the α-helix stability in proteins, has been previously utilized to derive a database with the energy contributions for various interactions taking place in an α-helix: intrinsic helical propensities, side-chain–side-chain interactions, main-chain–main-chain hydrogen bonds, and capping effects. This database was implemented in an algorithm based on the helix/coil transition theory (AGADIR). Here, we have modified this algorithm to include previously described local motifs: hydrophobic staple, Schellman motif and Pro-capping motif, new variants of these, and newly described side-chain–side-chain interactions. Based on recent experimental data we have introduced a position dependence of the helical propensities for some of the 20 amino acid residues. A new electrostatic model that takes into consideration all electrostatic interactions up to 12 residues in distance in the helix and random-coil conformations, as well as the effect of ionic strength, has been implemented. We have synthesized and analyzed several peptides, and used data from peptides already analysed by other groups, to test the validity of our electrostatic model. The modified algorithm predicts, with an overall standard deviation value of 6.6 (maximum helix is 100%), the helical, content of 778 peptides of which 223 correspond to wild-type and modified protein fragments. To improve the prediction potential of the algorithm and to have a direct comparison with nuclear magnetic resonance data, the algorithm now predicts the conformational shift of the $C^{\alpha}H$ protons, $^{13}C^{\alpha}$ and $^{3}J^{\alpha}N$ values. We have found that for those peptides correctly predicted from the point of view of circular dichroism, the prediction of the NMR parameters is very good.

Keywords: α-helix stability; protein folding; secondary structure prediction; biotechnology

**Corresponding author*

Introduction

The energetics of systems formed by short polypeptide chains has been described for α-helices. The helix/coil transition theory integrates that description within the framework of statistical mechanics, and lays the foundation for algorithms that quantify the helical tendency of amino acid sequences. Its simplest version, postulated by Zimm & Bragg (1959), used equilibrium constants characteristic of each amino acid to stand for the nucleation and elongation of helical segments. Later versions of the helix/coil transition theory and algorithms include detailed interaction terms such as capping interactions, side-chain–side-chain interactions, $i,i+3$ and $i,i+4$ electrostatic effects, and interaction of charged groups with the helix dipole (reviewed by Muñoz & Serrano, 1995a; Chakrabartty & Baldwin, 1995). These terms were introduced either to follow experimental data (Muñoz & Serrano, 1994, 1995b; Chakrabartty & Baldwin, 1995; Lifson & Roig, 1961; Lomize & Mosberg, 1997; Andersen & Tong, 1998) or to correspond to the statistical analysis of the protein database (Misra & Wong, 1997).

Our helix/coil transition program, AGADIR, predicted the global helical behavior of peptides in solution. A calculation performed with 423 peptides gave results comparable to the data obtained

E-mail address of the corresponding author: Serrano@embl-heidelberg.de

0022–2836/98/460173–19 $30.00/0

by circular dichroism. AGADIR also roughly described the helicity at a residue level, and reproduced the tendencies obtained from nuclear magnetic resonance (NMR) data (Muñoz & Serrano, 1995a,b). This algorithm was used to design mutations in protein α-helices and successful improvements of the helical tendencies of both the isolated peptides and the proteins were achieved (Villegas *et al.*, 1996; Muñoz *et al.*, 1996; Viguera *et al.*, 1996). The engineered proteins showed higher resistance towards chemical and thermal denaturations.

Two new approaches to model the formation of helical segments, the multiple sequence approximation (AGADIRms) and the one-sequence approximation (AGADIR1s) were extensively compared with the original (Muñoz & Serrano, 1997). While AGADIR1s and AGADIRms are virtually identical (for polyalanine peptide sequences of less than 50 residues), the new formalisms are more precise than the residue-partition function of AGADIR. They also perform equally to the Lifson-Roig formalism, despite the differences in the treatment of helix/coil cooperativity, when an equivalent set of parameters is used (Muñoz & Serrano, 1997).

However, details of the energetics are not yet properly implemented in helix/coil transition algorithms. First, electrostatic interactions are partially treated, i.e. only $i,i+1$, $i,i+2$, $i,i+3$ and $i,i+4$ interactions are considered, while charges separated by more than 20 Å in a helical peptide contribute to helix stability (Sitkof *et al.*, 1994). Second, the ionic strength is not taken into consideration, although experiments established its strong effect on helix stability, even in neutral peptides (Scholtz *et al.*, 1991). Third, the helical propensity of the 20 amino acid residues depends on their position in the α-helix (Petukhov *et al.*, 1998), and helix/coil transition algorithms use, in contrast, a single set of parameters for all the positions of the helix (except the capping residues). Finally, recently identified sequence motifs stabilize the helical conformation, because contacts between helical and non-helical residues are formed (the "hydrophobic staple"; Muñoz *et al.*,1995; Muñoz & Serrano, 1995a; Creamer & Rose, 1995; the "Schellman motif"; Viguera & Serrano, 1995a; and the "Pro-capping motif"; Prieto & Serrano, 1997). Since a residue either before the N-cap, or after the C-cap, is involved, classical helix/coil transition algorithms do not deal with these motifs.

We introduced ionic-strength dependence, long-range electrostatics and local motifs into our helix/coil transition algorithm, AGADIR1s. This results in a slight modification of the values for some of the parameters previously described (Muñoz & Serrano, 1995b), and improves the predictive power of the algorithm. The new version, AGADIR1s-2, predicts the helical behavior of peptides, in aqueous solution, for any pH, temperature and ionic-strength conditions. AGADIR1s-2 also predicts the conformational shifts of the individual C^{α} protons, and other NMR parameters, offering predictions at higher resolution. The quality of the prediction can be put to use for global (CD) and residue-level (NMR) estimations.

Results

Background (AGADIR1s)

The new version of AGADIR uses the classical one-sequence approximation of AGADIR1s. A direct comparison with the Lifson-Roig theory and parameters can be found by Muñoz & Serrano (1997). The free energy of any helical segment, $\Delta G_{\text{helical-segment}}$, is the difference in free energy between the random-coil and helical states. In AGADIR1s, we considered $\Delta G_{\text{helical-segment}}$ as the following summation:

$$\Delta G_{\text{helical-segment}} = \Delta G_{\text{Int}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{SD}} + \Delta G_{\text{nonH}} + \Delta G_{\text{dipole}} \quad (1)$$

where ΔG_{Int} is the summation of the intrinsic tendencies of the n residues in the $n+2$ helical segment (N- and C-cap are considered separately), to adopt helical dihedral angles (Muñoz & Serrano, 1994, 1995d) and reflects the loss of conformational entropy between the helix and random-coil states. ΔG_{Hbond} is the sum of the main-chain–main-chain enthalpic contributions, which include the formation of main-chain–main-chain $i,i+4$ hydrogen bonds. ΔG_{SD} sums the net contributions, with respect to the random-coil state, of all side-chain–side-chain interactions located at positions $i,i+3$ and $i,i+4$ in the helical region. ΔG_{nonH} represents the sum of the net contributions to helix stability of all the residues that are not in a helical conformation (only the N and C-capping residues). ΔG_{dipole} represents the interaction of charged groups with the helix macrodipole.

Development of AGADIR1s-2

In this section, we describe the modifications effected on the terms of equation (1). Throughout the text we refer to the different positions in the α-helix using the nomenclature published by Richardson & Richardson (1988):

```
N″ N′ Ncap N1-N2-N3-N4-N5.....C5-C4-C3-C2-C1-Ccap C′ C″
STC STC STC -He-He-He-He-He-He-He-He-He-He-He-STC STC STC
```

where STC (S, strand; T, turn; and C, coil) indicates a non-helical conformation and He is a helical residue.

ΔG_{Int}

In AGADIR1s, we considered this term to be purely entropic and to reflect the energy required to fix an amino acid residue in helical angles. Other

components, however, must be included here, i.e. the changes in solvation and van der Waals' contacts of the side-chain with the helix. These components vary according to helix position, as shown theoretically and experimentally for some amino acid residues at the first helical turn of an α-helix (Petukhov *et al.*, 1998; M. Petukhov & L.S., unpublished results). AGADIR1s-2 allows different intrinsic helical propensities at positions N1, N2, N3 and N4, for different amino acids, according to the experimental evidence obtained in model poly-alanine-based peptides (Petukhov *et al.*, 1998; M. Petukhov & L.S., unpublished results). The different helical propensities of titratable amino acid residues, when they are either neutral or charged (Chakrabartty *et al.*, 1994), are also included in AGADIR1s-2.

ΔG_{nonH}

AGADIR1s-2 distinguishes between the neutral and charged forms of Cys, His and Asp at capping positions. Neutral Cys is a poor hydrogen-bond donor or acceptor and therefore not a good N-capping residue. When charged, however, it can interact with the helix macrodipole and simultaneously make a charged hydrogen bond with the amide group of residue N4. The opposite is true for His, which cannot make a hydrogen-bond with the helix N terminus and has a strong repulsion from the helix macrodipole when it is charged (Sancho *et al.*, 1992). From the experimental data of Kortemme & Creighton (1995) for Cys, and of Armstrong & Baldwin (1993) for His, we evaluated that the N-capping contribution of these two residues is 1 kcal/mol more favorable when Cys is charged or His is neutral. The C-cap properties of Asp were modified to include similar pH dependence. When this residue is not charged, it can make a side-chain hydrogen bond to the carbonyl group of residue C3, as does Asn. Under these conditions, Asp has the C-capping value of Asn.

In AGADIR1s-2, the contribution of local sequence motifs involving the interaction of a residue outside the α-helix with a helical residue is added to ΔG_{nonH}. First, a free energy term (see Tables 2 and 3 of the supplementary material) adds up to the energy of helical segments that contain a hydrophobic staple (Muñoz *et al.*, 1995; Muñoz & Serrano, 1995e) or a Schellman motif (Schellman, 1980; Aurora *et al.*, 1995; Viguera & Serrano, 1995a). Second, the capping property of the N-cap is modified if a "capping box" (Harper & Rose, 1993) is present; that of the C-cap if there is a Pro-capping motif (Prieto & Serrano, 1997). In addition, the following combination described by Petukhov *et al.* (1996), contributes −1 kcal/mol to the stability of a helical segment: free N terminus, capping box motif and an Asp or a Glu at position N4. The stabilization is due to a strong interaction between residue N4, the N-capping residue and the charged N-terminal group.

Strong position dependence for pairs of residues were found using the program WHATIF (Vriend, 1990). They could reflect the formation of local motifs that were not previously described. We found three positive cases in our peptide database:

(1) A variant of the capping box motif in which the side-chain of a Thr at position N3 can make a hydrogen bond to the side-chain of Asp, Ser or Asn at position N-cap (Figure 1). The abundance of these pairs is twice what is expected from a random distribution (determined as indicated by Prieto & Serrano, 1997; data not shown). The experimental analysis of a protein fragment corresponding to the α-helix of the B1 domain of protein G, with and without this motif, enabled us to assign the value of a hydrogen bond energetic contribution to this interaction (Blanco *et al.*, 1997). Similar side-chain–side-chain interactions could be established with the side-chain of a Ser at position N3. The expected numbers of cases from a random distribution correspond, however, to the observed numbers (data not shown), and we do not have any experimental evidence for a stabilizing effect.

(2) A local interaction, akin to the hydrophobic staple motif but involving a charged residue, takes place between a Lys or Arg at position N4 and the main-chain carbonyl group of residue N′, when there is a Ser or a Thr at the N-cap (Figure 2). Due to the capping by Ser or Thr, the carbonyl group of the preceding residue points towards residue N4, and is at the right distance to form a hydrogen bond with the side-chain of the basic residues. We assigned a value of −0.3 kcal/mol to the interaction based on the experimental analysis of a peptide series described below.

(3) Histidine residues at position C1 or the C-cap are frequently guarded by an aromatic residue at C5 (11 cases out of 33) or C4 (12 cases out of 56), respectively, while in the helix center only eight histidine residues out of 165 cases are paired with an aromatic residue at position $i-4$. The optimal packing of Phe or Tyr side-chains with a histidine residue at position $i+4$ requires that the residue at position i adopts a χ_1 rotamer of approximately 180°, and the $i+4$ residue of approximately −60°. This corresponds to the C5/C1 and C4/C-cap cases (in the helix center the favorite rotamer for His is aproximately 180°). Using the peptides analysed by Armstrong *et al.* (1993), in which an $i,i+4$ Phe-His pair is placed at different positions of a polyalanine helix, we determined that the aromatic residue-His pair is three times stronger when the histidine residue is the C-cap or C1 residue.

ΔG_{SD}

This term included side-chain–side-chain interactions between non-charged residues plus the $i,i+1$, $i,i+3$ and $i,i+4$ electrostatic interactions. In AGADIR1s-2, we only consider side-chain–side-chain interactions inside the α-helix (ΔG_{SD}), excluding the electrostatic component between charged groups ($\Delta G_{\mathrm{electrost}}$). In some cases, when a hydrogen bond between two side-chains is made that

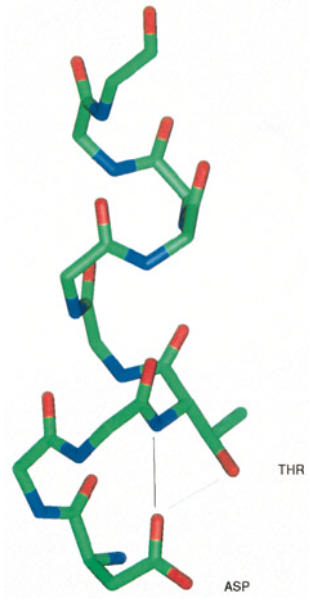

Figure 1. The N terminus of an α-helix from protein 1igd (residues 27 to 31), with an Asp-Thr pair at positions N-cap and N3 (nomenclature of Richardson & Richardson, 1988). The hydrogen-bonding network between the groups mentioned in the text is shown by thin lines.

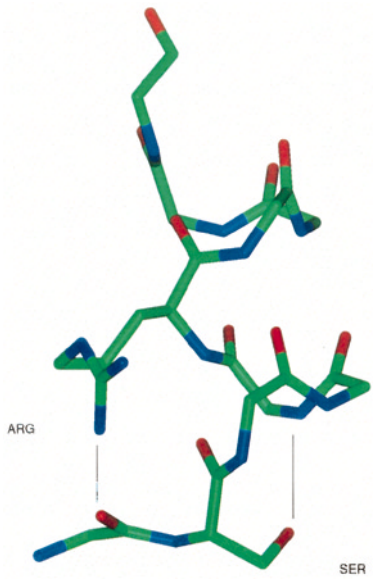

Figure 2. The N terminus of an α-helix from protein 2hbg (residues 2 to 7), with a Ser-Arg pair sequence at positions N-cap and N4. The side-chains of the rest of the residues have been removed. The hydrogen-bonding network between the groups mentioned in the text is shown by thin lines.

involves a titratable amino acid residue (i.e., Gln-Asp, Gln-Glu; Huyghues-Despointes *et al.*, 1995; Huyghues-Despointes & Baldwin, 1997; Smith & Scholtz, 1998), the strength of the interaction is pH dependent but independent of salt concentration. In those cases a constant term is added to the side-chain–side-chain energy when the corresponding residue is charged and this term is taken into consideration when calculating its pK_a.

The values of some side-chain–side-chain interactions were modified, according to new peptide studies by different groups (Padmanabhan & Baldwin, 1994a,b; Viguera & Serrano, 1995b; Stapley *et al.*, 1995; Huyghues-Despointes & Baldwin, 1997; Huyghues-Despointes *et al.*, 1995; Stapley & Doig, 1997; Smith & Scholtz, 1998). Some interactions that were assumed to be attractive based on chemical similarity between amino acid residues, are now repulsive, i.e. the $i,i+4$ interactions between aromatic side-chains (including His) and β-branched aliphatic residues (Val and Ile). The repulsion comes from the steric incompatibility between these side-chains in the middle of an α-helix (Creamer & Rose, 1995). We also added the interaction between an aromatic group and a Cys side-chain (Viguera & Serrano, 1995b).

ΔG_{dipole}

This term included all electrostatic interactions between the helix macrodipole or the free N and C termini and groups located in the α-helix. AGADIR1s-2 also considers interactions of the helix macrodipole with charged groups located outside the helical segment. We assigned half a charge to the helix macrodipole (positive at the N terminus and negative at the C terminus) and we considered that it is not titratable by pH.

There is an effect of ionic strength on helix stability that is due to the helix macrodipole but that is independent of the presence of charged amino acids residues. Baldwin and co-workers suggested that increasing the ionic strength should stabilize the α-helix by shifting the equilibrium between the helical conformation (which should have a large dipole moment) and the random-coil (which has a very small dipole moment, due to the random orientation of the dipoles carried by the peptide bonds). Experimental characterization of a neutral soluble peptide, under different ionic-strength conditions, supported this assumption (Scholtz *et al.*, 1991). For low ionic strengths (below 0.15 M), the effect on ΔG seems to be linear and similar for different salts. Above this value the dependence follows an exponential until approximately 1 M salt. AGADIR1s-2 takes this effect into account (see Materials and Methods).

$\Delta G_{electrost}$

This new term includes all electrostatic interactions between two charged residues inside and outside the helical segment (see the electrostatics

model below). Cys and Tyr are now correctly treated as titratable amino acid residues.

The free energy of a helical segment, $\Delta G_{\text{helical-segment}}$, in AGADIR1s-2 is described in equation (2):

$$\Delta G_{\text{helical-segment}} = \Delta G_{\text{Int}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{SD}} + \Delta G_{\text{dipole}} + \Delta G_{\text{nonH}} + \Delta G_{\text{electrost}} \quad (2)$$

Electrostatic model

Determination of the distance between charges

Electrostatic interactions are distance dependent. The precise estimation of the free energy of interaction between two charged side-chains in an α-helix depends therefore on atomic details, i.e., all possible rotamers should be considered, with proper weights, to calculate the distances between the charged groups. Practically this is not possible with algorithms that use no structural details. To obviate this problem the protein database has been used to extract pseudo-energies that correlate well with experimental free energies, e.g., for helical propensities (Muñoz & Serrano, 1995d) and ϕ distributions in the random-coil state (Serrano, 1995; Smith *et al.*, 1996). Based on these results, we have assumed that the average distance between two charged amino acid residues at different positions of an α-helix in the protein database reflects the average distance found in a helical peptide. The same applies to the random-coil. In the case of the interaction of a charged group with the helix dipole, we have measured the average distances of charged groups at different positions in the helix and the first four amide groups or the last four carbonyl groups.

We previously neglected the electrostatic interactions of residues outside the α-helix conformation with the helix macrodipole and/or residues within the helix. However, the maximum distance between an Asp side-chain at position N′ and the first four amide groups of the helix is short enough to create electrostatic interactions. To take into account these interactions we have empirically introduced a linear dependence of the distance with the number of residues separating the non-helical charged residue and the cap positions. For residues N′ and C′, the distance is 6 Å. That separation distance increases by 3 Å for every extra position after the N′ or C′ positions. In the case of the interaction of charged non-helical residues with charged helical residues, we only considered the caps and residues N′ and C′.

Regarding non-blocked peptides, we have considered a linear dependence of the distance between the free termini charged groups, and the helix first turns that contain the partial charges due to the helix macrodipole, with the number of residues separating the terminal free end and the capping residue (2.1 Å when the capping residue has the charged group, then 4.1 Å, 6.1 Å,... for every extra residue). Using these distances, we obtain similar repulsive free energies as the ones described in AGADIR1s (for the free N terminus at positions N-cap, N′, N″ and N‴, they are: 0.81, 0.36, 0.21 and 0.14 kcal/mol, respectively, at 50 mM ionic strength, pH 7.0 and 273 K; the same values apply to the free C terminus).

Determination of the electrostatic free energy contribution

The model used to calculate the electrostatic contribution to helix stability is explained in detail in the Materials and Methods section. We describe here how electrostatic terms are computed. The electrostatic interactions obviously change with the degree of ionization and consequently with the pH of the solution, while the pK_a of ionizable groups in a peptide change from their standard values depending on the electrostatic environment. We considered in AGADIR1s-2 all electrostatic interactions (this involves charged side-chain groups, free N-terminal and C-terminal main-chain groups, and the succinyl blocking molecule if the peptide is succinilated) to compute the electrostatic environment of the amino acid residues in the random-coil and helical segments, taking into account the ionic strength and assuming full charges ($q_i = q_p = 1$ in equation 6). Then, the pK_a of these residues in the random-coil and helical states are calculated and the populations of the charged forms calculated. ΔG_{SD}, the side-chain–side-chain non-charged interaction term, is decreased to remove the contribution from the pairs of charged residues, while the electrostatic terms are scaled down to reflect the same populations of charged pairs ($q_i q_p$).

N and C termini blocking groups

In the previous versions, we considered that the residue immediately following the acetyl group at the N terminus (ac + 1), or that preceding the amide group at the C terminus (am − 1), could only be considered as capping residues. However, NMR studies of protected peptides (Shalongo *et al.*, 1994) have shown that these residues can adopt helical angles. In AGADIR1s-2, residues ac + 1 and am − 1 can be helical, the acetyl and amide groups performing as capping residues. With respect to the succinyl group, we considered that it has a similar N-capping stabilizing contribution as the acetyl group, but that it can make electrostatic interactions with the helix dipole and other charged groups.

Testing the new model

Electrostatic interactions

To test the validity of our electrostatic model we have performed a pH titration, monitored by CD at 222 nm, on a peptide series (the KR series), designed to analyze electrostatic interactions

within the helix, as well as with residues outside the helical conformation:

```
KR-1c   NH3-Y-G-G-S-A-G-A-G-A-G-A-K-R-G-A-A-am
KR-1a     Ac-Y-G-G-S-A-A-A-A-A-A-A-K-R-A-A-A-am
KR-1f   NH3-Y-G-G-S-A-A-A-A-A-A-A-K-R-A-A-A-am
KR-2f   NH3-Y-G-G-S-A-A-A-A-A-A-K-A-R-A-A-A-am
KR-3f   NH3-Y-G-G-S-A-A-A-A-A-K-A-A-R-A-A-A-am
KR-4f   NH3-Y-G-G-S-A-A-A-A-K-A-A-A-R-A-A-A-am
KR-5f   NH3-Y-G-G-S-A-A-A-K-A-A-A-A-R-A-A-A-am
KR-6f   NH3-Y-G-G-S-A-A-K-A-A-A-A-A-R-A-A-A-am
KR-7f   NH3-Y-G-G-S-A-K-A-A-A-A-A-A-R-A-A-A-am
KR-8f   NH3-Y-G-G-S-K-A-A-A-A-A-A-A-R-A-A-A-am
```

The C terminus is amidated (am) to eliminate charge effects of the free end groups (Muñoz & Serrano, 1995b). The N terminus is free (NH_3) to study the effect on helix stability, except in the control peptide KR-1a (Ac). There is a serine residue to nucleate the helix as the N-capping residue, far away from the free N terminus and a tyrosine residue to determine peptide concentration, which is separated by two glycine residues from the rest of the peptide to minimize the contribution of the aromatic ring to the ellipticity at 222 nm (Chakrabartty *et al.*, 1993). We have chosen lysine and arginine residues to favor peptide solubility and also because they have different pK_as that allow us to titrate them separately. Peptide KR-1c, with four glycine residues in the middle, is used as a control, while peptides KR-1f and KR-1a are used as references. NMR analysis of peptides KR-1c, KR-1a and KR-1f shows that the control peptide is essentially unstructured, while in peptides KR-1a and KR-1c the helix conformation mainly extends from residues Ser4 or Ala5 to the end (data not shown). In any of the three peptides we could find non-sequential nuclear Overhauser effects (NOEs) between the side-chain of Tyr1 and the rest of the molecule.

The control peptide KR-1c is essentially unstructured whatever the pH considered, due to the presence of four glycine residues intercalated in the poly-Ala. Calculation of the expected ellipticity at 222 nm using AGADIR1s-2 indicates that the changes in ellipticity at 222 nm upon pH titration (Figure 3A) cannot be explained by changes in helix content. Therefore they can be mainly attributed to changes in the absorption properties of Tyr upon titration of the OH group that affect the far-UV CD spectrum. We find two titratable groups with apparent pK_a values of 7.1 ± 0.1 and 9.4 ± 0.1 (Table 1), which are close to the values obtained in

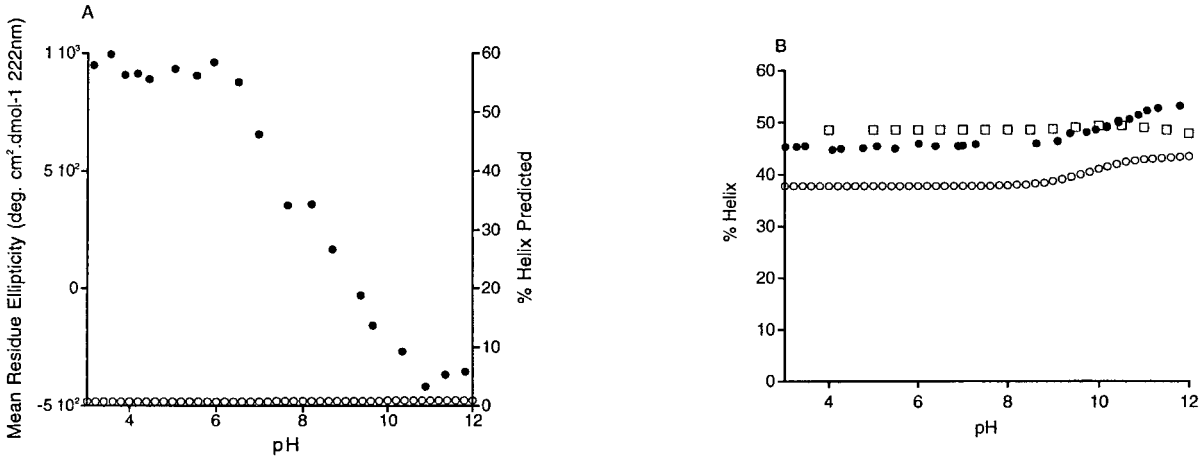

Figure 3. pH titration of the ellipticity at 222 nm for peptides. A, Change in mean residue ellipticity at 222 nm with pH for peptide KR-1c. The change in ellipticity in the control peptide KR-1c is assumed to be due to an effect of the aromatic ring on the ellipticity at 222 nm. Filled circles, left *y* axis, experimental values; open circles, right *y* axis, AGADIR1s-2 predicted values. B, Dependence of the helical content determined from the ellipticity at 222 nm for peptide KR-1a. Filled circles, experimental values; open circles, AGADIR1s-2 predicted values; open squares, AGADIR1s values. The error in the experimental values due to concentration determination is around 3%. All peptides were analysed at 278 K in aqueous solution and the pH was adjusted with HCl and NaOH.

Table 1. Determination of the pK_a values for the titratable groups in the KR series of peptides

Position	pK_{NH}	%ΔHelix	pK_{Tyr}	%ΔHelix	pK_{Lys}	%ΔHelix
KR-1c	7.1 ± 0.11	*	9.4 ± 0.1	2.4 ± 0.3	*	*
KR-1a	*	*	9.2 ± 0.3	2.8 ± 0.7	10.6 ± 0.1	5.4 ± 0.7
KR-1f	7.4 ± 0.05	7.9 ± 0.2	9.3 ± 0.3	2.6 ± 2.0	10.3 ± 0.1	4.9 ± 0.3

The data shown in this table come from the fitting of the curves in Figures 3 and 4, as indicated in Materials and Methods. pK_{NH}, pK_a of the free amino-terminal group; pK_{Tyr}, pK_a of the Tyr side-chain OH group; pK_{Lys}, pK_a of the Lys side-chain NH group; %ΔHelix, increment in helical content after titrating the corresponding charged group. The experimental conditions are those indicated in the legend to Figure 4. *, Not determined, either because it is not present or because it is not observed.

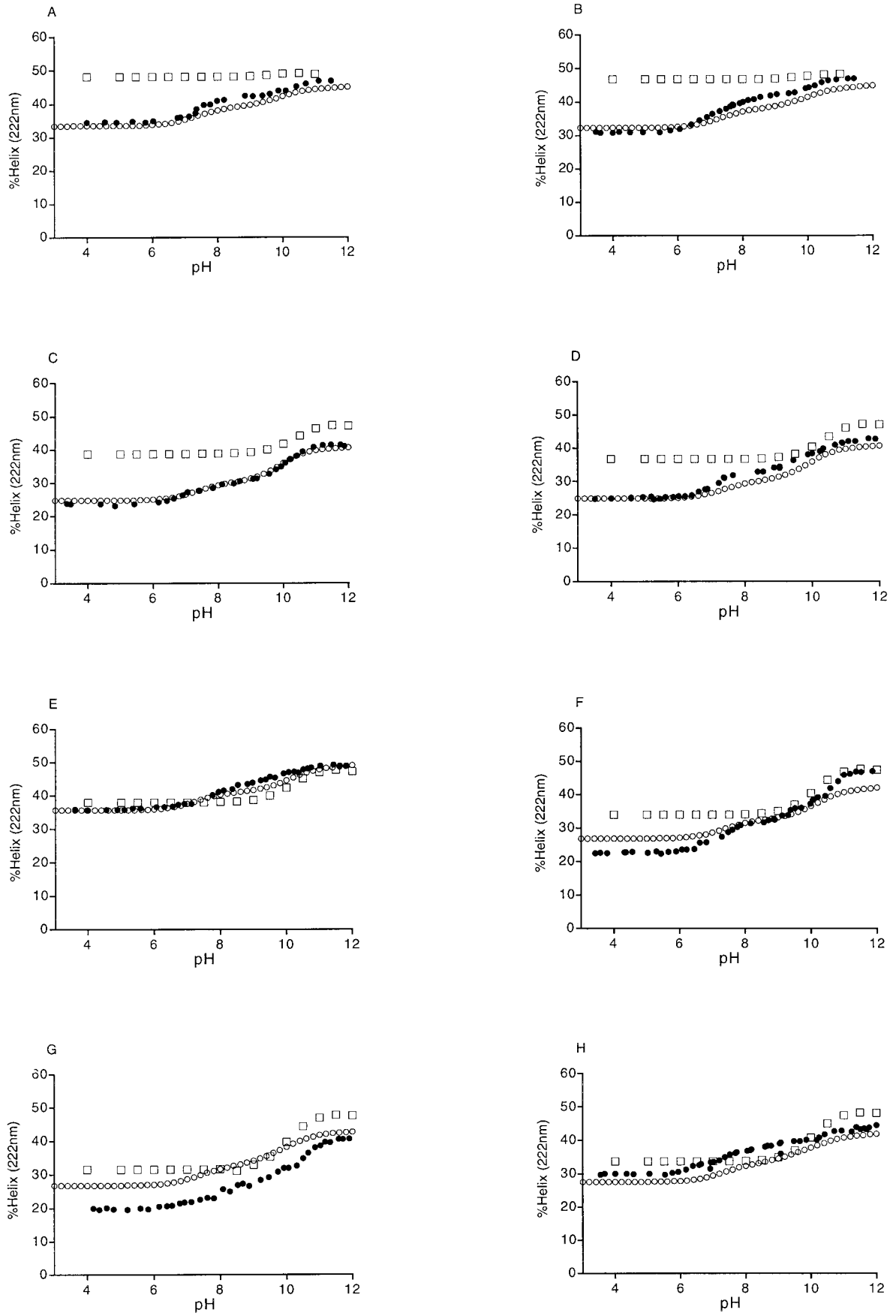

Figure 4. pH titration of the ellipticity at 222 nm for peptides. A, KR-1f; B, KR-2f; C, KR-3f; D, KR-4f; E, KR-5f; F, KR-6f; G, KR-7f; H, KR-8f. Filled circles, experimental values; open circles, AGADIR1s-2 predicted values; open squares, AGADIR1s values. The error in the experimental values due to concentration determination is around 3%. All peptides were analysed at 278 K in aqueous solution and the pH was adjusted with HCl and NaOH.

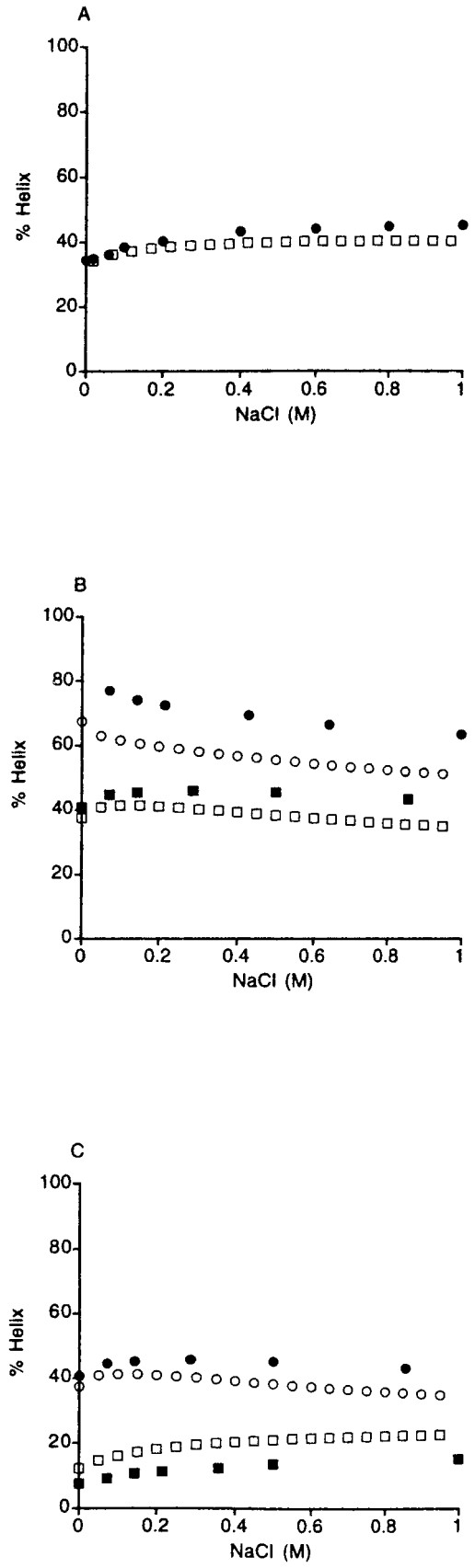

Figure 5. Comparison between the predicted (open symbols) and experimental (filled symbols) average helical values for different peptides under increasing ionic strength. A, Peptide KR-1f (this work). B, circles, peptide DR4 (ADAAARDAAARDAAARY;. squares, peptide DR3 (ADAARADAARADAARY; Huyghues-Despointes *et al.*, 1993). C, circles, peptide RD4 (ARAAADRAAADRAAADY;. squares, peptide RD3 (ARAADARAADARAADY; Huyghues-Despointes *et al.*, 1993).

model peptides for the free amino and hydroxyl groups of Tyr. The changes in ellipticity for the two titrations are of a similar magnitude (approximately -670 deg. cm^2 dmol^{-1}).

The pH titrations of peptides KR-1a (Figure 3B) and KR-1f (Figure 4A) show a broad transition around pH 10.0 that can be fitted considering two titratable groups with apparent pK_a values of ~10.5 and ~9.2 (Table 1), that can be attributed to the Lys and Tyr side-chains, respectively. In the case of peptide KR-1f, another transition with an apparent pK_a of ~7.4 can be attributed to the free N-terminal charge. The neutralization of these three groups results in increases in helical content that range from ~2.8% in the case of Tyr1 (this value is obtained after taking into consideration the changes in the absorption properties of Tyr upon titration of the OH group) to ~ 8% in the case of the free N-terminal charge (Table 1). These results confirm the necessity to consider electrostatic effects due to charged groups not only inside the helical conformation, but also outside it.

In Figure 4 (A to H) we show the pH titration of the ellipticity at 222 nm for all the non-acetylated KR peptides. At pH 8.5, in which the first titration due to the free N-terminal groups is essentially finished, we find differences in the helical content of these peptides, illustrating that the position of the Lys side-chain determines the helical content. These differences remain but decrease at pH 12 when the Lys side-chain should be neutral, therefore indicating that they are partly electrostatic in nature. Other factors like side-chain–side-chain interactions, position dependent helix propensities or capping motifs could, however, play a role. This could be the case for peptide KR-5f that has the highest helix content both at pH 8.5 and 12. In this peptide the Lys side-chain at position N4 can make a hydrogen bond to the main-chain carbonyl group of residue N′ (Gly 3), as it is often found in the protein database (Figure 2).

The previous version of AGADIR cannot reproduce the pH dependence of the different peptides, while AGADIR1s-2 reproduces both the shape of the titration curves and the actual value of helical content (Figure 4), at least within the experimental error due to peptide concentration determination (±3%). The main exception is peptide KR-7f for which the algorithm reproduces the shape of the pH titration, indicating that the electrostatic approximation used in AGADIR1s-2 is correct, but that the helical content is overpredicted through all the transition (Figure 4G). One possibility to explain this result is that the intrinsic helical propensity of Lys is position dependent in the first helical turn as we have found for Gly, Ala, Val, Ile

Table 2. Peptides analysed by NMR in this work

	pH	Temp. (K)	Ionic strength	Sequence	%Helix
SAAE	6	280	0.00	Ac-GGSAAEGG-Am	0
SAAD	6	280	0.00	Ac-GGSAADGG-Am	0
SADE	6	280	0.00	Ac-GGSADEGG-Am	0
DAAE	6	280	0.00	Ac-GGDAAEGG-Am	0
Hyman	5.5	278	0.00	YGTAQRAALGNISNVVRTAGAGSKKG	8
AR	2.7	277	0.00	Ac-GGFPAVEAAAARAKAAAARGY-Am	2
peptGTP	5.5	278	0.00	GTTGHVDSGKSALTARIKAHAAEKYG	8
βα1	5	278	0.005	KGFTVTGAGQAGRELYAKLHAGK	10
βα2	5	278	0.005	KGFTVTGAGQAARELYAKLHAGK	26
βα3	5	278	0.005	KGFTVTGASPAARELYAKLHAGK	26

The corresponding chemical shifts are available upon request. Ac, acetylated peptide; Am, amidated peptide; %Helix. percentage of α-helix determined from the ellipticity at 222 nm; Temp., temperature in Kelvin.

and Leu (Petukhov *et al*., 1998). AGADIR's electrostatic model includes interactions between two helical, or one helical and one non-helical, charged groups, the helix macrodipole and charged helical and non-helical residues. Elimination of any of these different electrostatic interactions results in the incorrect reproduction of the experimental data for some or all of the peptides (data not shown).

Ionic-strength dependence

Figure 5 shows the comparison between the experimental and predicted values for different peptides analyzed under different NaCl concentrations (0 to 1 M). Figure 5A shows the salt titration of peptide KR-1f which contains two positively charged groups at $i,i+1$ positions and a free N-terminal group. Figure 5B and C show more complicated cases with several ionic pairs placed at different positions and orientations (Huyghues-Despointes *et al*., 1993). Although there are discrepancies between the predicted and experimental values, the trends for the complex changes produced by salt in the different peptides are correctly reproduced by our electrostatic model.

Prediction of average helical content

A total of 778 peptides (including those described here, Table 2) have been included in our database to refine and test the predictive power of the new AGADIR version. From these, 223 peptides correspond to wild-type and modified protein fragments and 555 peptides correspond to designed sequences, generally polyalanine-based peptides.

Figure 6 illustrates the correlation between the helical population for the 778 different peptides estimated from CD (the percentage helical population was calculated using the method of Chen *et al*., 1974), and the predicted average helical content. The overall correlation coefficient is extremely good ($r = 0.96$). The slope is 1.0 and the line intersects the origin of the x and y axes. Similar results are obtained when doing a blind test consisting only of the peptides corresponding to protein fragments and mutants of these ($r = 0.92$, slope of 0.94 ± 0.03 and intersection in the origin -0.2 ± 0.4; data not shown).

The prediction for the original dataset used in previous versions of AGADIR does not reveal a significant improvement, as expected due to the good performance of AGADIR on this set. Running AGADIR1s and AGADIR1s-2 on the blind dataset, using only those peptides analysed under low ionic strength and that have no local motifs, shows a slight improvement in the prediction for AGADIR1s-2 ($r = 0.92$ *versus* 0.90). The introduction of all the new peptides, studied under different salt conditions and carrying new local motifs (aproximately 300 new peptides), however, significantly lowers the quality of the prediction for AGADIR1s ($r = 0.87$), thus indicating the necessity to introduce the electrostatic model and the new interactions.

From a practical point of view it is interesting to know the precision of the calculated helical values for a given sequence. Statistical analysis of the differences between the calculated and experimental helical values for the 778 peptides of the database shows that the data could be reasonably fitted to a normal distribution with a standard deviation

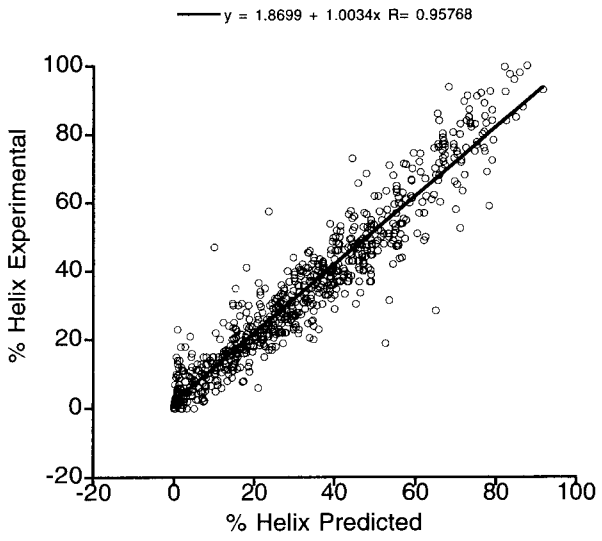

Figure 6. General correlation analysis between the average helical content determined by far-UV CD and the values predicted by AGADIR1s-2.

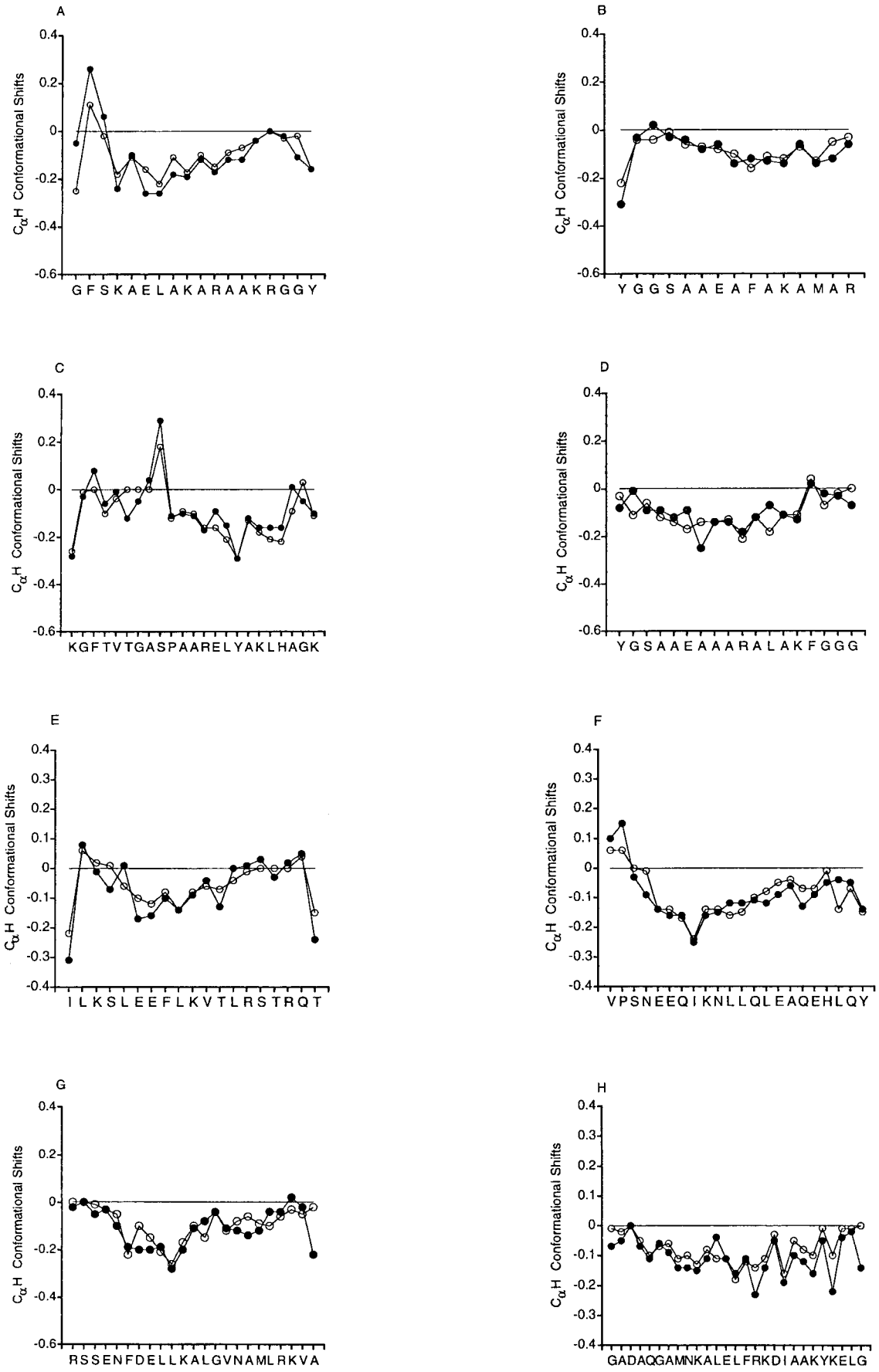

Figure 7A–H (*legend opposite*)

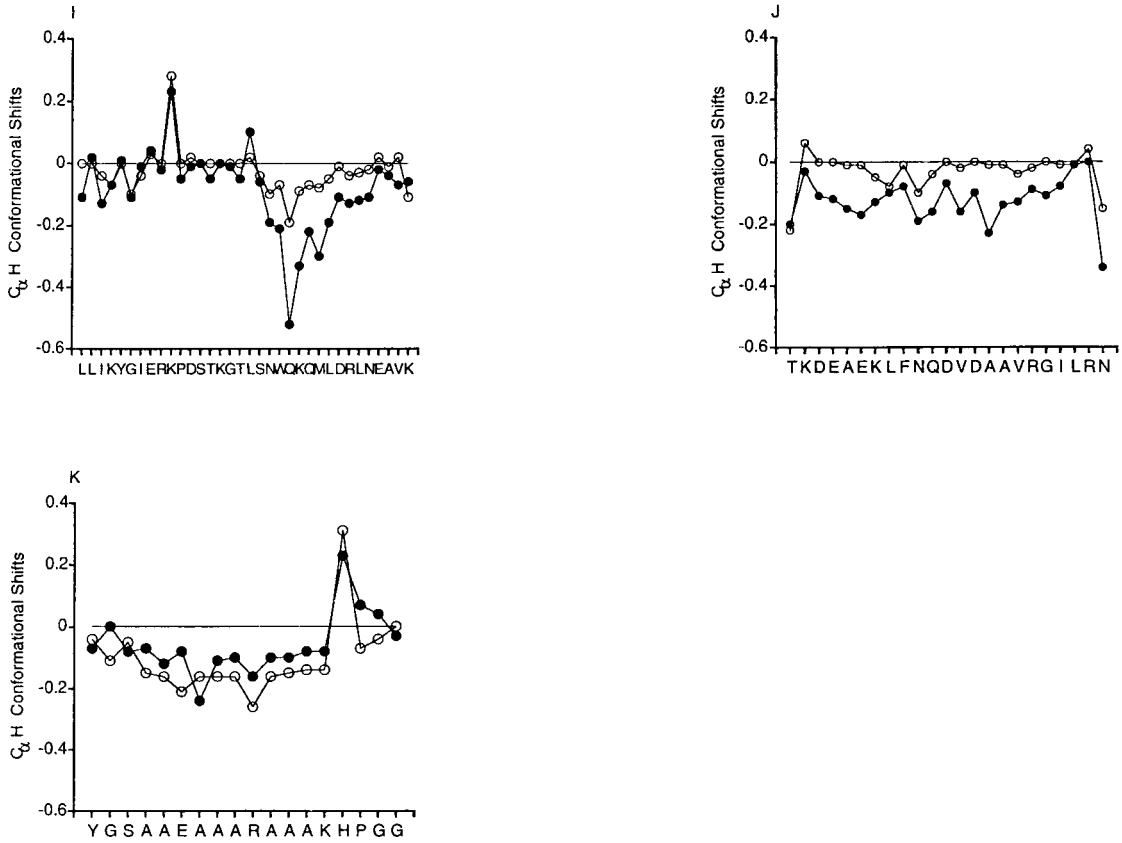

Figure 7. Comparison between the predicted $C^{\alpha}H$ conformational shifts (open circles) and the experimental ones (filled circles) for several peptides with a significant helical content as determined by far-UV CD. In A to H the difference between the average helical population predicted by AGADIR1s-2 and the CD experimental values differ by less than 4%. In I to K, the differences were larger than 10%. A, FL peptide (Muñoz *et al.*, 1995). B, FM peptide (Viguera & Serrano, 1995a). C, peptide βα3 analysed here (Table 2). D, peptide SCLF (Prieto & Serrano, 1997). E, a fragment of IL6 (Morton *et al.*, 1994). F, Ada2H-2 (Villegas *et al.*, 1996). G, a protein fragment containing a Schellman motif (Sukummar & Gierasch, 1997). H, helix H of myoglobin (Shin *et al.*, 1993). I, synthetic calmodulin-binding peptide (Munier *et al.*, 1993). J, C-helix from T4 lysozyme (McLeish *et al.*, 1994). K, peptide CAHP (Prieto & Serrano, 1997).

value of 6.6. This means that in 95% of the cases, we can calculate the helical content of a peptide with an error of ±14 (i.e., the prediction is $x \pm 14\%$ helix). The main deviation from a normal distribution occurs for peptides that have more than 80% helix content. In those cases AGADIR1s-2 tends to underpredict the helix content, as did previous versions of AGADIR (Muñoz & Serrano, 1994, 1995a).

Prediction of nuclear magnetic resonance parameters

AGADIR1s-2 predicts the $C^{\alpha 1}H$ and $^{13}C^{\alpha}$ conformational shifts, plus the $^{3}J\alpha N$ coupling constants. The comparison of the experimental and predicted data constitutes a blind test for AGADIR1s-2, since no NMR data were ever used to refine the helix/coil parameters. For our NMR peptide database (112 peptides, either polyalanine-based peptides, designed non-polyalanine-based peptides, natural fragments or mutants of those; see also Table 2), AGADIR1s-2 predicts reasonably well the experimental $C^{\alpha 1}H$ conformational shift values ($r = 0.75$ and the slope is 0.89 ± 0.02; data not shown). To illustrate the successful residue-level prediction we show in Figure 7A to H a comparison between the predicted $C^{\alpha 1}H$ conformational shift values and the experimental ones for several peptides for which the far-UV CD helical content was predicted with a ±4% error. Where the error is above 10% for the global helical content, the predicted $C^{\alpha 1}H$ conformational shifts have similar errors (Figure 7I to K).

Discussion

The work presented here and in previous publications (Muñoz & Serrano, 1994, 1995b,c, 1997) puts the large amount of experimental information available in an appropriate framework, to explain the helical behavior of any linear monomeric peptide devoid of tertiary interactions. Starting from the standard one-sequence approximation algorithm (Muñoz & Serrano, 1997), we improved our helix/coil transition program, AGADIR.

Helical segments

The new version, AGADIR1s-2, samples more helical windows than previous versions for blocked peptides, allowing the residue following the N-terminal acetyl group or preceding the C-terminal amide group to adopt a helical conformation. It is consequently more similar to the implementation of the Lifson-Roig theory by Baldwin and co-workers (Chakrabartty & Baldwin, 1995), that mainly differs in the quantification of the parameters. Because there is a larger number of helical segments, we slightly increased the enthalpic main-chain–main-chain contribution (−0.895 kcal/mol) and modified the entropic cost of fixing the 20 amino acid residues in helical angles.

Local motifs

AGADIR1s-2 identifies the sequence patterns that correspond to local motifs known to stabilize helical conformations (hydrophobic staple, Schellman motif and Pro-capping motif) through the interaction of a side-chain external to the helix with a helical residue. In addition, the algorithm includes other motifs described here. The program adds a free-energy term for the formation of the motif to the corresponding helix segments. Residues at positions of the helix that are not restricted in the sequence patterns can modulate the contribution of these motifs to helix stability. In the hydrophobic staple the residue at position N′ could interact not only with residue N4, but also with residue N7 (Muñoz *et al.*, 1995a). Therefore, the values used in AGADIR1s-2 are only estimates that should be modified once more data is available. The energy contribution of the local motifs has been added as a term that depends on temperature or pH in a similar manner as other interactions. In the case of the capping box motif the temperature dependence is similar to that of normal N-capping involving a hydrogen bond and there is a pH dependence when there is a Glu or Asp at position N4. Similarly in the hydrophobic staple motif we have introduced a temperature dependence similar to that for hydrophobic side-chain–side-chain interactions. In those cases in which experimental evidence relating to temperature or pH dependence of these motifs is available, AGADIR1s-2 reproduces satisfactorily the data.

In a recent review, Aurora & Rose (1998) enumerate several local motifs found at the N and C termini of helices in proteins, and include motifs involving residues at positions further apart from the helix than N′ or C′. The impact of these long-range capping motifs on helix stability is, however, not yet known and cannot be computed in AGADIR until experimental data are obtained.

Electrostatic interactions

One of the major modifications and improvements of the new version of AGADIR is the treatment of long-range electrostatic interactions in the model, obtained from the statistical analysis of the protein database. The sequences and local structural environments of specific pairs of charged groups in the protein database are sufficiently diverse to consider that the average distances observed between the charged groups are only consequences of the corresponding amino acid residue types and their sequence separation. In other words, these average distances can be used to represent the distances in peptides. The contribution to helix stability of a pair of charged groups is calculated as the change in the electrostatic potential corresponding to the variation in the charge-to-charge distance between the ground state (any conformation) and the helix. The values obtained with this approximation agree with those obtained for $i,i+3$ and $i,i+4$ interactions in model peptides. Our approximation cannot be correct for all cases, since selective packing and interaction with other side-chains can displace the separation distances from the measured average values. However, AGADIR1s-2 reproduced reasonably well the behavior of several peptides with complex electrostatic interactions.

Another modification with respect to AGADIR1s is the complete consideration of electrostatic interactions within the helix and with charged groups outside the helical conformation to determine the pK_as of titratable amino acids residues.

The use of distances between charges in the helix and random-coil states also allows us to consider the screening effect of ions and to take into account the solvent ionic strength, which improves the predictive capability of the algorithm. To implement the ionic-strength dependence, we have used the theoretical approximation described by Sitkoff *et al.* (1994). Electrostatic interactions scale down, exponentially, with increasing charge-to-charge distances and higher ionic strength. Salts also tend to screen the charges due to the helix macro-dipole, and thus stabilize helices. Different salts at high ionic strength have different effects and the only calibration was done for NaCl (on a neutral peptide) by Baldwin and co-workers (Scholtz *et al.*, 1991). This means that AGADIR's prediction should be accurate for all salts at low ionic strength (less than 0.1 M) and for NaCl up to 1 M. Other contributions, such as the effect of salt on hydrophobic interactions, are considered of lesser importance and are not included. The changes in helical content due to salts are roughly reproduced by AGADIR1s-2.

Prediction of helical content and comparison with NMR parameters

Our new model accurately predicts the average helical behavior of 778 peptides, including 223 peptides corresponding to natural protein fragments and mutants of these. In about two-thirds of the cases the discrepancies between the predicted helical content and the experimental one is less than 7%. The prediction of the global helical contents of

Table 3. Conformational shifts for the C^{α} proton and $^{3}J^{\alpha}N$ random-coil values for the 20 amino acid residues

	N′	N-cap	Middle	C1 (Gly)	C-cap	C′	$^{3}J^{\alpha}N$
A	0.15	0	−0.20	−0.2	0.0	0.15	5.67
V	0.15	0	−0.56	−0.2	0.0	0.15	7.44
L	0.15	0	−0.38	−0.2	0.0	0.15	6.58
I	0.15	0	−0.52	−0.2	0.0	0.15	7.24
P	0.15	0	−0.30	−0.2	0.0	0.15	6.00
F	0.15	0	−0.47	−0.2	0.0	0.15	7.06
Y	0.15	0	−0.42	−0.2	0.0	0.15	7.18
W	0.15	0	−0.38	−0.2	0.0	0.15	6.70
M	0.15	0	−0.33	−0.2	0.0	0.15	6.52
S	0.15	0.05	−0.24	−0.2	0.0	0.15	6.61
T	0.15	0.05	−0.45	−0.2	0.0	0.15	7.46
C	0.15	0	−0.47	−0.2	0.0	0.15	7.10
N	0.15	0.1	−0.33	−0.2	0.0	0.15	7.04
D	0.15	0.1	−0.22	−0.2	0.0	0.15	6.60
Q	0.15	0	−0.36	−0.2	0.0	0.15	6.39
E	0.15	0	−0.31	−0.2	0.0	0.15	6.07
K	0.15	0	−0.32	−0.2	0.0	0.15	6.41
R	0.15	0	−0.32	−0.2	0.0	0.15	6.46
H	0.15	0	−0.51	−0.2	0.0	0.15	7.05
G	0.15	0	−0.26	−0.2	0.1	0.15	5.48

The differences in chemical shift for the C^{α} proton between the expected value in the corresponding helical X position and the random-coil values of Merutka *et al.* (1995) are shown in the five middle columns. The nomenclature used for the different helix positions is that of Richardson & Richardson (1988).

N′, residue before the N-cap; this residue adopts an extended conformation when a hydrophobic staple motif (Muñoz *et al.*, 1995) is formed. Based on experimental data on different peptides bearing this motif we have assigned a downfield conformational shift value of 0.15. This value is corrected by the probability of a hydrophobic staple motif being made. The probability is determined from the free energy of interaction (shown in Table 2 of the supplementary material) between residue N′ and N + 3 in a hydrophobic staple motif. N-cap, N-cap residue; only good N-cap residues (Ser, Thr, Asn and Asp) will tend to have a fixed conformation and that is why they have values different from 0. Middle, helical residues located between position N-cap and C-1. C-1(Gly), when Gly is the C-cap residue, the amino acid residue at position C − 1 normally has a ϕ value of around −90°, resulting in a smaller conformational shift than in the middle of the helix. Since we do not have statistical data for the $C^{\alpha}H$ chemical shift value in proteins at this position we have determined this value considering a ϕ value −90° and using the corresponding equation mentioned by Wishart *et al.* (1995). C-cap, C-cap residue. We have assigned a value of zero since many different dihedral angles are found for the different amino acid residues at this position in the protein database; the only exception is Gly which normally adopts positive dihedral angles which result in a small downfield chemical shift. C′, residue after the C-cap. This residue adopts an extended conformation when a Schellman motif (Aurora *et al.*, 1995; Viguera & Serrano, 1995a) is formed. Based on experimental data on different peptides bearing this motif we have assigned a downfield conformational shift value of 0.15. This value is corrected by the probability of a Schellman motif being made. The probability is determined from the free energy of interaction (shown in Table 3 of the supplementary material) between residue C′ and C-3 in a Schellman motif. $^{3}J^{\alpha}N$ Coupling constant values of the 20 amino acid residues in the random-coil state (Serrano, 1995).

protein fragments or their experimental determination by circular dichroism are, however, not always adequate because they do not provide any information regarding the distribution of the helix population along the sequence. This information is essential to identify the positions where mutations should be done to increase helix stability in proteins (Villegas *et al.*, 1996; Viguera *et al.*, 1996; Muñoz *et al.*, 1996) and for *ab initio* folding calculations. Nuclear magnetic resonance is the main technique that produces parameters related to the structural populations at a residue level. The direct prediction of these parameters by the helix/coil transition algorithm offers a stronger test for the accuracy of its prediction, and provides valuable information.

To reproduce the $C^{\alpha 1}H$ proton conformational shifts we used, as reference for random-coil values the peptides analysed by Merutka *et al.* (1995) and Wishart *et al.* (1995). For the 100% helix population we adopted the values described by Wishart *et al.* (1991) for proteins. This resulted in an overprediction of the conformational shifts. In fact, the analysis of several peptides in 30% (v/v) TFE with helical populations close to 100% (data not shown), shows that on average the change in the $C^{\alpha 1}H$ chemical shift for the residues is around 0.1 ppm less than what we should expect from the values obtained in proteins (Table 3). This could be due to a higher mobility of peptides compared to proteins and to the fact that many of the residues analysed by Wishart *et al.* (1991) in proteins were buried.

The effect of aromatic residues (ring-current effects) in the random-coil and helical states, on the $C^{\alpha 1}H$ conformational shifts has been introduced in an additive manner. We used constant terms (Table 4) despite the possible variations due to side-chain packing. In the case of the random-coil we have only considered the aromatic effect on residues $i-2$, $i-1$, $i+1$ and $i+2$. In the helical conformation we only considered the effects on residues at positions $i-4$, $i-3$, $i+1$, $i+3$ and $i+4$. We have also included the downfield shift that Pro induces on the preceding residue in the random-coil state. Finally, we consider that the N and C-cap residues adopt non-random-coil conformations, as well as residues N′ and C′, if they are involved in a hydrophobic staple or Schellman motif, respectively (Table 3). We have found in general a very good agreement between observed

Table 4. Effect of neighboring amino acid residues on the chemical shift of residue X in the helical and random-coil states

Helix	$i-4$	$i-3$	$i+1$	$i+3$	$i+4$
Trp			−0.3		
Aromatic				0.15	−0.10
Random coil	$i-2$	$i-1$	$i+1$	$i+2$	
Phe, Tyr	−0.04	−0.07	−0.1	−0.04	
Trp	−0.04	−0.09	−0.1	−0.16	
Pro			0.28		
Free-N	0.06	−0.22			
Free-C			0.04	−0.16	

All the values are in ppm. Aromatic residues are Phe, Tyr, Trp and His. Helix, the influence that the aromatic side-chain could have on the conformational shifts of the $C^{\alpha}H$ protons of the neighboring helical residues, was determined using the program of Osapay & Case (1991) on a modeled 13 residue standard polyalanine-based helix with the aromatic residue at position 6. We considered the rotamers found in the database and applied a weight to each rotamer proportional to their relative abundance. The obtained values were checked against our peptide database and we found that the calculated relative magnitudes and signs of the aromatic effect on neighboring residues fitted our data but that the absolute numbers were smaller. Tyr, Phe and Trp will produce a shift of −0.15 ppm on a His at position $i+4$ and the His will shift the aromatic residue at position i by −0.15 ppm. Random coil, aromatic residues in the random-coil conformation induce an upfield shift of residues $i-2$, $i-1$, $i+1$ and $i+2$ (Merutka *et al*, 1995; Viguera & Serrano, 1995a; M. Rico, personal communication). The presence of a proline residue in the random-coil conformation induces a downfield shift of the preceding residue of around 0.28 ppm (Wishart *et al*., 1995). Finally, in non-protected peptides with charged free ends there is an upfield shift of the first and last residues and a downfield shift of the second and previous to last residues (average values of −0.22 ppm, −0.16 ppm, 0.06 ppm and 0.04 ppm; values obtained from our peptide database and M. Rico, personal communication). These values are pH dependent and become null when the corresponding terminal group is non-charged.

and predicted values, indicating that our algorithm is able to describe not only the average helical properties of a peptide sequence, but also the behavior of individual residues.

In some cases we have found that the algorithm overestimates, or underestimates, the helical content both by CD and NMR criteria (Figure 7I to K). There are several possible explanations for this:

(1) For many of the interactions found in helices there are no experimental data. This is especially important for the local motifs at the N and C termini of helices that contribute significantly to helix stability.

(2) There is no cooperativity for side-chain–side-chain interactions in the algorithm. Essentially, the energetics are described as the sum of two-body interactions, although the additivity is questionable when several favorable or unfavorable $i-4$, $i-3$, $i+3$ and $i+4$ interactions coincide. This could result, in some cases, in severe overprediction when trying to engineer a peptide with the maximum number of favorable interactions. An example of this can be found in Figure 7K. In this peptide a Glu is part of the capping box motif (Ser-Ala-Ala-Glu) and also establishes a good electrostatic interaction with an Arg at $i+4$. Formation of the local motif will drive the Glu away from the Arg and therefore the $i,i+4$ electrostatic interaction should be weaker than the value assigned by AGADIR1s-2.

(3) Most of the experimental parameters included in AGADIR are not derived from single energy contributions, but from a combination of them. For the simplest polyalanine-based peptide (that contains polar residues for solubility) at least four parameters need to be considered (hydrogen bond, intrinsic helical tendency, capping properties and the charges or blocking groups at the end of the peptide). There are thus several closely related solutions for parameters such as hydrogen-bond contribution or helix propensities, all of them giving approximately similar results.

(4) We assume that there is no energy coupling, other than electrostatic, between residues in the random-coil reference state (although there are examples to the contrary, e.g. β-turn conformations are found in short peptides in aqueous solution).

(5) Some of the peptides described in the literature could be partly aggregated, or have tertiary interactions (i.e. those peptides with higher helical content).

Future improvements

The possibility of comparing the CD and NMR experimental values with the predicted ones constitutes a powerful tool to improve AGADIR1s-2, which still has many limitations as described above. AGADIR could be improved when new peptides with new interactions or motifs are described, but only an algorithm that uses a detailed structural representation and an appropriate force-field, and takes into account all possible amino acid residue arrangements and experimental conditions, would offer a definitive tool. Despite all the limitations, AGADIR1s-2 predicts, with a very good accuracy, the helix content, on average and at a residue level, of the majority of the monomeric peptides described in the literature. Thanks to its good performance, AGADIR could be used in protein and peptide design, and as a reference in the development of the next generation of algorithms.

AGADIR1s-2 can be run on the WWW (http://www.embl-heidelberg.de/Services/index.html#5) and is available upon request.

Materials and Methods

Peptide synthesis

The solid-phase synthesis of the peptides was performed on an Abimed AMS422 multiple peptide synthesizer using Fmoc chemistry and PyBOP activation at a 0.025 mmol scale. After synthesis was completed, protecting groups were removed and the peptide chains were cleaved from the resin with a mixture of TFA (10 ml), phenol (0.75 g), EDT (0.25 ml), thioanisole

(0.5 ml) and water (0.5 ml) for three hours. The peptides were purified on a Vydac C-18 reverse phase column (20 mm × 250 mm, 0.01 mm particle size) at a flow rate of 10 ml/minute. Peptide homogeneity (>98%) was determined by HPLC using an acetonitrile gradient of 0.7% per minute. The peptide composition was confirmed by amino acid analysis and the molecular mass was checked by matrix-assisted laser desorption time-of-flight mass spectrometry.

Peptide concentration

The concentration of the different peptides was determined by amino acid analysis, or UV absorbance using the method of Gill & Hippel (1989). For those peptides which do not contain tyrosine or tryptophan residues, the concentration was determined by amino-acid analysis. The error in both cases was around 10%.

Circular dichroism analysis

Circular dichroism spectra were recorded on a Jasco-710 instrument at a temperature of 5°C. The peptides (roughly 1.5 mg) were dissolved in 1 ml of deionized water with 2.5 mM sodium phosphate buffer (pH 7.0), unless otherwise indicated. The pH, when different, was adjusted with HCl or NaOH. To check for concentration dependence of the CD spectra, different dilutions of the peptides (10 to 750 μM), using cuvettes with different pathlengths (0.1 mm to 0.5 cm), were scanned. CD spectra in the range 190 to 250 nm were obtained using the continuous-scan option (100 nm/minute scan speed), with a one second response time, and taking points every 0.1 nm. For every sample we took 30 scans and the experiment was repeated three times on different days. The ellipticity was calibrated using D-10-camphorsulfonic acid.

Determination of the helical percentage from the CD spectra

In order to estimate the helical population of the different peptides by CD we used the mean residue ellipticity at 222 nm, taking into account the peptide length (Chen *et al.*, 1974):

$$\%\text{helical content} = 100[Q_{222}^{\text{obs}}/(-39,500(1-2.57/n))] \quad (3)$$

where n is the number of peptide bonds and Q_{222}^{obs} is the ellipticity of the peptides at 222 nm.

pH titration of the peptides by CD

The titration CD curve can be fitted using the following equation:

$$X = X_{\text{pH3}} + \sum_{1}^{i}(\Delta X_{\text{p}K_a^{\text{app}_i}}/(1+10^{(\text{p}K_a^{\text{app}_i}-\text{pH})})) + \sum_{1}^{j}(\Delta X_{\text{p}K_a^{\text{app}_j}}/(1+10^{(\text{pH}-\text{p}K_a^{\text{app}_j})})) + \cdots \quad (4)$$

where X is the property measured at a given pH. X_{pH3} is the quantity of this property at pH3. $\Delta X_{\text{p}K_a^{\text{app}_i}}$ is the increment in this property when an acidic group is titrated. $\Delta X_{\text{p}K_a^{\text{app}_j}}$ is the increment when a basic group is titrated. $\text{p}K_a^{\text{app}_i}$ is the apparent pK_a of the acidic groups and $\text{p}K_a^{\text{app}_j}$ of the basic ones.

Nuclear magnetic resonance

NMR samples were prepared by dissolving the lyophilised peptides in 0.5 ml of water with 10% (v/v) ^2H_2O, using milli-Q water from a Millipore water system and ^2H_2O from Cambridge Isotope Chemicals at a concentration of around 2.5 mM and 0.1 mM. Minute amounts of HCl and NaOH were added in order to adjust the pH of the samples this was measured with an Ingold combination electrode (Wilmad) inside the NMR tube and isotope effects were not corrected. Sodium 3-trimethylsylyl (2,2,3,3-$^2H_4$) propionate (TSP) was used as an internal reference at 0.00 ppm. NMR experiments were performed on a Bruker AMX-500 spectrometer at 278 K. The data were processed with the program X-WINNMR from Bruker.

Calculation of average distances between charged groups using the protein structures database or computer modeling

The database of 3-D structures was obtained following the principles described by Hobohm *et al.* (1992). This database has been filtered for quality of the data and consists of 279 proteins with less than 50% sequence homology for a total of 59,117 amino acid residues (the proteins are listed by Muñoz & Serrano, 1995d), and it is currently included in the program WHATIF (Vriend, 1990). The search was conducted using the WHATIF Scan3D option and the search query was either H_n, H_n/(S-T-C)$_1$, (S-T-C)$_1$/H_n or (*)$_n$, when looking for interactions between a helix residue and a residue at the N-cap (also between the helix dipole and a charged residue), a helical residue and a residue at the C-cap (also between the helix dipole and a charged residue), and between residues in the random-coil conformation, respectively (where H is helix, S is strand, C is coil, T is turn, * is any type of secondary structure and n is the number of residues). To prevent helix end effects, in all the cases in our protein database search we looked for helical segments of different length having three helical residues before the amino acid residue i being investigated and three residues after amino acid residue $i+x$ (where x goes from 1 to 12). For the random-coil, we just measured the average distance in a segment of the same length, but without a defined conformation (search for S-T-C). We considered as potential charged amino acid residues: Asp, Glu, Tyr, Cys, His, Lys and Arg. For some of these residues the location of the charge is not restricted to one group but could be distributed among several: Asp, Glu, Arg and His. In these cases we used the group closer to the branching point (Asp C^{γ}; Glu C^{δ}; Arg C^{ζ}; His C^{ε_1}) to measure the corresponding average distances.

For some particular pairs (interaction inside the helix between Cys or Tyr with other residues and His with His, Cys and Tyr residues and interactions between a charged amino acid residue located at the N-cap, C-cap, N′ or C′ positions, and another one inside the α-helix), there where not enough cases (less than five). In these cases we modeled α-helices using the Insight/Discover program (Byosim Technologies, Inc., San Diego, CA) and adopted the strategy of Honig and co-workers (Sitkoff *et al.*, 1994). For the α-helical conformation the φ and ψ angles were fixed to −60° and −45°, respectively. The

N-cap position was fixed to ϕ and ψ angles of −167° and 147°, respectively. At the C terminus of the α-helix we assumed one of the most frequent conformations found in helices, in which residue C1 adopts distorted helical angles (ϕ −90°, ψ 0°), and residue C-cap if it is a Gly was modeled with positive dihedral angles (ϕ 72°, ψ 27°). In the case of non-Gly residues at the C-cap position we adopted extended non-helical angles (ϕ −120°, ψ 120°). For residues N′ and C′, we adopted an extended conformation (ϕ −120°, ψ 120°), except when residue C-cap is a Gly (phi angle of −90°). To obtain a reasonable approximation to the rotamer conformation we just model pairs of Arg side-chains at the different positions of an α-helix: N′, N-cap, N1, N2, N3.....C3, C2, C1, C-cap and C′, using an extended rotamer. The same was done between pairs of aspartate residues. The distance between the methyl groups immediately preceding the charged groups was measured for every Arg and/or Asp pair and then averaged for equivalent Arg and Asp pairs. These values were then assigned to the different cases mentioned.

In the case of the interaction between helical charged residues and the helix macrodipole, we first used the approach indicated above, although in this case the distances in the random-coil are not used since there is no helix dipole in this conformation. Using WHATIF we have determined the average distances of the groups mentiond above to the first four amide groups or the last four carbonyl groups of the protein α-helices. The average of the four values has been considered to represent the effective distance between the charged groups and the helix dipole. In this case, we could find enough data for all titratable amino acid residues.

Determination of the electrostatic free energy contribution

The electrostatic contribution of two charged side-chains on the surface of an α-helix ($\Delta G_{electrost}$), is the difference between the electrostatic interaction in the helical (G_{Hel}) and random-coil (G_{RC}) states (equation (5)):

$$\Delta G_{electrost} = (G_{Hel} - G_{RC}) \quad (5)$$

The electrostatic interaction is determined using Coulomb's equation and the charge screening by salt is introduced by the Debye-Huckel factor (equation (6)):

$$G = \frac{e^2 q_i q_p}{3\pi\varepsilon_0\varepsilon_r r_{ip}} \times \exp(-K \times r_{ip}) \quad (6)$$

where e is the charge of the electron, q_i and q_p are the charges of the two residues involved, r_{ip} is the distance between the two charges, ε_0 is the vacuum permitivity, ε_r is the dielectric constant of the medium and K is the Debye-Huckel parameter (equation (7)):

$$K = (8\pi e^2 N_A I/1000kT)^{1/2} \quad (7)$$

where I is the ionic strength of the solution, N_A is Avogadro's number, k is the Boltzmann's constant and T is the temperature. This approximation has been demonstrated to be valid for the random-coil state, as well as for solvent-exposed isolated helices (Sitkof *et al.*, 1994).

To determine q_i and q_p, the pK_a of the residues in the random-coil, p$K_{a(RC)}$, and in the corresponding helical segment, p$K_{a(Hel)}$, are calculated using equations (8) and (9):

$$pK_{a(RC)} = pK_{a(ref)} + G_{RC}/2.3RT \quad (8)$$

$$pK_{a(Hel)} = pK_{a(ref)} + G_{Hel}/2.3RT \quad (9)$$

Once the pK_a values are known, the degrees of ionization are obtained either from equation (10) (acidic amino acid residues) or (11) (basic amino acid residues):

$$q = 1/(1 + (10^{pH - pK_a})) \quad (10)$$

$$q = 1/(1 + (10^{pK_a - pH})) \quad (11)$$

Contribution to helix stability of ionic strength through the interaction with the helix macrodipole

We have empirically fitted the dependence of $\Delta\Delta G$ with salt, obtained from Scholtz *et al.* (1991), to the following equation:

$$\Delta\Delta G_{hel} = -\alpha(1 - \exp(-\beta I)) \quad (12)$$

where $\Delta\Delta G_{hel}$ is the difference in folding free energy of a particular α-helix segment in a solution with ionic strength I and in pure water. The values of α and β are, respectively, 0.15 and 6. This energy has been added to every helical segment in AGADIR1s-2.

Prediction of NMR parameters

To predict any NMR parameter we need to have the appropriate reference for their random-coil and helix values.

Random coil values

In the case of the C^α proton ($\delta C^\alpha H_{RC}$) and $^{13}C^\alpha$ shifts, NMR analysis of Gly-Gly-X-Gly-Gly pentapeptides has provided a reference for the random-coil values for the 20 amino acid residues (Merutka *et al.*, 1995; Wishart *et al.*, 1995). For the random-coil $^3J^\alpha N$ coupling values we have used the numbers obtained from the statistical analysis of the protein database, which have proven to reproduce the experimental values of many amino acid residues in an unstructured conformation (Serrano, 1995; Smith *et al.*, 1996; Ramírez-Alvarado *et al.*, 1996, 1997). The random-coil $C^\alpha H$ values are affected by the presence of aromatic residues at positions: $i-2$, $i-1$, $i+1$ and $i+2$, or when a proline residue is found at position $i+1$. The changes in the chemical shift ($\Delta\delta C^\alpha H_{RC-Aro}$), produced by neighboring aromatic or proline residues in the random-coil are shown in Table 4.

Helical values

We have used as reference for 100% helix the C^α proton values ($\delta C^\alpha H_{Hel}$) compiled by Wishart *et al.* (1991) plus 0.1 ppm (see discussion). In the case of the $^{13}C^\alpha$ shifts we have used the values of Beger & Bolton (1997). With respect to the $^3J^\alpha N$ coupling constants we have calculated the helical value for the 20 amino acid residues using the Karplus (1959) equation on a standard α-helix (ϕ −60°, ψ −45°). The helical C^α proton values of the 20 amino acid residues are modified by the presence of neighboring aromatic residues at positions $i-4$, $i-3$, $i+3$ and $i+4$. The changes in the chemical shift

($\Delta\delta C^{\alpha}H_{Hel\text{-}Aro}$), produced by neighboring helical aromatic residues are shown in Table 4.

Knowing the parameters for the helical and random-coil states, the partition function for all helical segments, and assuming a fast exchange between the helical and non-helical conformations, we can determine the NMR parameters with respect to random-coil values for every residue X in the peptide as shown in equation (13) for the C^{α} proton conformational shifts:

$$\Delta C^{\alpha}H - X = \sum_{i}^{n}([\delta C^{\alpha}H_{Hel} - \delta C^{\alpha}H_{RC}]W^{i}_{hel} + [\Delta\delta C^{\alpha}H_{Hel-Aro}]W^{i}_{hel}) + \sum_{j}^{m} W^{j}_{RC}[\Delta\delta C^{\alpha}H_{RC-Aro}] \quad (13)$$

where n is the number of helical segments containing residue X (includes positions N′, N-cap, C-cap and C′) and m is the number of aromatic-X, or Pro-X, pairs. W^{i}_{hel} is the weight of helical segment i (Muñoz & Serrano, 1997). W^{j}_{RC} are the weights of the random-coil conformations of the different aromatic-X or Pro-X pairs. This is calculated by multiplying the random-coil probability of residue X by that of the corresponding aromatic and/or Pro neighbor. The helix-random-coil or random-coil-helix combinations between residue X and the neighboring aromatic, and/or proline residues, are excluded. The reason for this is that we do not have any experimental information for these cases.

Parameter refinement

Parameter refinement was done as described (Muñoz & Serrano, 1994, 1995b). Essentially, the original values described in AGADIR1s were used as starting points. The main difference between this refinement and the previous one is the separation of the capping effect into proper capping and local motifs (see above) and the introduction of long-range electrostatic interactions. The introduction of these extra parameters results in a slight modification of some of the previous values (Muñoz & Serrano, 1994, 1995b). The new values for the different refined parameters are shown in the supplementary material (Tables 1 to 7).

Acknowledgments

A.R.V. is a recipient of a fellowship from the Spanish science ministry (MEC). This work has been partly funded by an EU biotechnology grant (BIO4-CT97-2086). We are very grateful to Dr G. Vriend for modifying the program WHATIF in order to obtain the statistical data required for determining the distances between charged groups in the protein database.

References

Andersen, N. H. & Tong, H. (1998). Empirical parametrization of a model for predicting peptide helix/coil equilibrium populations. *Protein Sci.* **6**, 1920–1936.

Armstrong, K. M. & Baldwin, R. L. (1993). Charged histidine affects α-helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Natl Acad. Sci. USA,* **90**, 11337–11340.

Armstrong, K. M., Fairman, R. & Baldwin, R. L. (1993). The $(i,i+4)$ Phe-His interaction studied in an alanine-based α-helix. *J. Mol. Biol.* **230**, 284–291.

Aurora, R. & Rose, G. D. (1998). Helix capping. *Protein Sci.* **7**, 21–38.

Aurora, R., Srinivasan, R. & Rose, G. D. (1995). Rules for α-helix termination by glycine. *Science,* **264**, 1126–1129.

Beger, R. D. & Bolton, P. H. (1997). Protein ϕ and ψ dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J. Biomol. NMR,* **10**, 129–142.

Blanco, F. J., Ortiz, A. R. & Serrano, L. (1997). Role of a non-native interaction in the folding of protein G B1 domain as inferred from the conformational analysis of the α-helix fragment. *Folding & Design,* **2**, 123–133.

Chakrabartty, A. & Baldwin, R. L. (1995). Stability of α-helices. *Advan. Protein Chem.* **46**, 141–176.

Chakrabartty, A., Kortemme, T., Padmanabhan, S. & Baldwin, R. L. (1993). Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry,* **32**, 5560–5565.

Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino-acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**, 843–852.

Chen, Y. H., Yang, J. T. & Chau, K. H. (1974). Determination of the helix and β-form of proteins in aqueous solution by circular dichroism. *Biochemistry,* **13**, 3350–3359.

Creamer, T. P. & Rose, G. D. (1995). Interactions between hydrophobic side-chains within α-helices. *Protein Sci.* **4**, 1305–1314.

Gill, S. C. & Hippel, P. H. (1989). Calculation of protein extintion coefficients from amino acid sequence data. *Anal. Biochem,* **182**, 319–326.

Harper, E. T. & Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry,* **32**, 7605–7609.

Hobohm, U., Scharf, M., Schneider, R. & Sander, C. (1992). Selection of representative protein data sets. *Protein Sci.* **1**, 409–417.

Huyghues-Despointes, B. M. P. & Baldwin, R. L. (1997). Ion-pair and charged hydrogen-bond interactions between histidine and aspartate in a peptide helix. *Biochemistry,* **36**, 1965–1970.

Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1993). Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci.* **2**, 80–85.

Huyghues-Despointes, B. M. P., Scholtz, J. M. & Baldwin, R. L. (1995). Measuring the strength of side-chain hydrogen bonds in peptide helices: The Gln-Asp $(i,i+4)$ interaction. *Biochemistry,* **34**, 13267–13271.

Karplus, M. (1959). Contact electron-spin coupling of nuclear magnetic moments. *J. Phys. Chem.* **30**, 11–15.

Kortemme, T. & Creighton, T. E. (1995). Ionisation of cysteine residues at the termini of model α-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family. *J. Mol. Biol.* **253**, 799–812.

Lifson, R. & Roig, A. (1961). On the theory of helix-coil transitions in biopolymers. *J. Chem. Phys.* **34**, 1963–1974.

Lomize, A. L. & Mosberg, H. I. (1997). Thermodynamic model of secondary structure for α-helical peptides and proteins. *Biopolymers,* **42**, 239–269.

McLeish, M. J., Nielsen, K. J., Najbar, L. V., Wade, J. D., Lin, F., Doughty, M. B. & Craik, D. J. (1994). Conformation of a peptide corresponding to T4 lysozyme residues 59-81 by NMR and CD spectroscopy. *Biochemistry,* **33**, 11174–11183.

Merutka, G., Dyson, H. J. & Wright, P. E. (1995). 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR,* **5**, 14–24.

Misra, G. P. & Wong, C. F. (1997). Predicting helical segments in proteins by a helix-coil transition theory with parameters derived from a structural database of proteins. *Proteins: Struct. Funct. Genet.* **28**, 344–359.

Morton, C. J., Simpson, R. J. & Norton, R. S. (1994). Solution structure of synthetic peptides corresponding to the C-terminal helix of interleukin-6. *Eur. J. Biochem.* **219**, 97–107.

Munier, H., Blanco, F. J., Precheur, B., Dieisis, E., Nieto, J. L., Craescu, C. T. & Barzu, O. (1993). Characterisation of a synthetic calmodulin-binding peptide derived from *Bacillus authracis* adenylate cyclase. *J. Biol. Chem.* **268**, 1695–1701.

Muñoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* **1**, 399–409.

Muñoz, V. & Serrano, L. (1995a). Helix design, prediction and stability. *Curr. Opin. Biotech.* **6**, 382–386.

Muñoz, V. & Serrano, L. (1995b). Elucidating the folding problem of α-helical peptides using empirical parameters, II: helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**, 275–296.

Muñoz, V. & Serrano, L. (1995c). Elucidating the folding problem of α-helical peptides using empirical parameters III: temperature and pH dependence. *J. Mol. Biol.* **245**, 297–308.

Muñoz, V. & Serrano, L. (1995d). Intrinsic secondary structure propensities of the amino acids, using statistical ϕ-ψ matrices: comparison with the experimental scales. *Proteins: Struct. Funct. Genet.* **20**, 301–311.

Muñoz, V. & Serrano, L. (1995d). Analysis of $i,i+5$ and $i,i+8$ hydrophobic interactions in a helical model peptide bearing the hydrophobic staple motif. *Biochemistry,* **34**, 15301–15306.

Muñoz, V. & Serrano, L. (1997). Development of the multiple sequence approximation within the AGADIR model of α-helix formation. Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers,* **41**, 495–509.

Muñoz, V., Blanco, F. & Serrano, L. (1995). The ''hydrophobic-staple'' motif. A role for loop-residues in α-helix stability and protein folding. *Nature Struct. Biol.* **2**, 380–385.

Muñoz, V., Cronet, P., López-Hernández, E. & Serrano, L. (1996). Analysis of the effect of local interactions in protein stability. *Folding & Design,* **1**, 167–178.

Osapay, J. & Case, D. A. (1991). A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* **113**, 9436–9444.

Padmanabhan, S. & Baldwin, R. L. (1994a). Helix-stabilizing interaction between tyrosine and leucine or valine when the spacing is $i,i+4$. *J. Mol. Biol.* **241**, 706–713.

Padmanabhan, S. & Baldwin, R. L. (1994b). Tests for helix-stabilizing interactions between various nonpolar side-chains in alanine-based peptides. *Protein Sci.* **3**, 1992–1997.

Petukhov, M., Yumoto, N., Murase, S., Ohmura, R. & Yoshikawa, S. (1996). Factors that affect the stabilization of α-helices in short peptides by a capping box motif. *Biochemistry,* **35**, 387–397.

Petukhov, M., Muñoz, V., Yumoto, Y., Yoshikawa, Y. & Serrano, L. (1998). Position dependence of nonpolar amino acid intrinsic helical propensities. *J. Mol. Biol.* **278**, 279–289.

Prieto, J. & Serrano, L. (1997). C-capping and helix stability: the Pro C-capping motif. *J. Mol. Biol.* **274**, 276–288.

Ramírez-Alvarado, M., Blanco, F. J. & Serrano, L. (1996). *De novo* design and structural analysis of a model β-hairpin peptide system. *Nature Struct. Biol.* **3**, 604–612.

Ramírez-Alvarado, M., Blanco, F. J., Niemann, H. & Serrano, L. (1997). Role of β-turn residues in β-hairpin formation and stability. *J. Mol. Biol.* **273**, 898–912.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science,* **240**, 1648–1652.

Sancho, S., Serrano, L. & Fersht, A. R. (1992). Histidine residues at the N- and C-termini of α-helices: perturbed pK_as and protein stability. *Biochemistry,* **31**, 2253–2258.

Schellman, C. (1980). In *Protein Folding* (Jaenicke, R., ed.), Elsevier/North-Holland, New York.

Scholtz, J. M., York, E. J., Stewart, J. M. & Baldwin, R. L. (1991). A neutral water-soluble, α-helical peptide: the effect of ionic strength on the helix-coil equilibrium. *J. Am. Chem. Soc.* **113**, 5102–5104.

Serrano, L. (1995). Comparison between the ϕ distribution of the amino acids in the protein database and NMR data indicates that amino acids have various ϕ propensities in the random-coil conformation. *J. Mol. Biol.* **254**, 322–333.

Shalongo, W., Dugad, L. & Stellwagen, E. (1994). Distribution of helicity within the model peptide acetyl $(AAQAA)_3$ amide. *J. Am. Chem. Soc.* **116**, 8288–8293.

Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L. L., Dyson, H. J. & Wright, P. E. (1993). Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry,* **32**, 6356–6364.

Sitkof, D., Lockhart, D. J., Sharp, K. A. & Honig, A. B. (1994). Calculation of electrostatic effects at the amino terminus of an α-helix. *Biophys. J.* **67**, 2251–2260.

Smith, J. S. & Scholtz, J. M. (1998). Energetics of polar side-chain interactions in helical peptides: salt effects on ion and hydrogen bonds. *Biochemistry,* **35**, 7292–7297.

Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996). Analysis of main-chain torsion angles in proteins. Prediction of NMR coupling constants for native and random-coil conformations. *J. Mol. Biol.* **255**, 494–506.

Stapley, B. J. & Doig, A. J. (1997). Hydrogen bonding interactions between glutamine and asparagine in α-helical proteins. *J. Mol. Biol.* **272**, 465–473.
Stapley, B. J., Rohl, C. A. & Doig, A. J. (1995). Addition of side-chain interactions to modified Lifson-Roig helix-coil theory: application to energetics of phenylalanine-methionine interactions. *Protein Sci.* **4**, 2383–2391.
Sukummar, M. & Gierasch, L. M. (1997). Local interactions in a Schellman motif dictate interhelical arrangement in a protein fragment. *Folding & Design*, **2**, 211–222.
Viguera, A. R. & Serrano, L. (1995a). Experimental analysis of the Schellman motif. *J. Mol. Biol.* **251**, 150–160.
Viguera, A. R. & Serrano, L. (1995b). Side-chain interactions between sulfur-containing amino acids and Phe in α-helices. *Biochemistry*, **34**, 8771–8779.
Viguera, A. R., Villegas, V., Aviles, F. X. & Serrano, L. (1996). Favourable native-like helical local interactions can accelerate protein folding. *Folding & Design*, **2**, 23–33.
Villegas, V., Viguera, A. R., Aviles, F. X. & Serrano, L. (1996). Stabilisation of proteins by rational design of α-helix stability using helix/coil transition theory. *Folding & Design*, **1**, 29–34.
Vriend, G. (1990). WHATIF: a molecular modeling and drug design program. *Mol. Graph.* **8**, 52–56.
Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333.
Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995). ^{1}H, ^{13}C and ^{15}N random-coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.
Zimm, B. H. & Bragg, J. K. (1959). Theory of the phase transition between helix and random-coil. *J. Chem. Phys.* **31**, 526–535.

Edited by A. R. Fersht

(Received 3 April 1998; received in revised form 15 July 1998; accepted 6 August 1998)

http://www.hbuk.co.uk/jmb

Supplementary material for this paper comprising seven Tables and further References is available from JMB Online