A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins

C. Nick Pace and J. Martin Scholtz
Department of Medical Biochemistry and Genetics, Department of Biochemistry and Biophysics, and Center for Macromolecular Design, Texas A&M University, College Station, Texas 77843-1114 USA

ABSTRACT The average globular protein contains 30% α-helix, the most common type of secondary structure. Some amino acids occur more frequently in α-helices than others; this tendency is known as helix propensity. Here we derive a helix propensity scale for solvent-exposed residues in the middle positions of α-helices. The scale is based on measurements of helix propensity in 11 systems, including both proteins and peptides. Alanine has the highest helix propensity, and, excluding proline, glycine has the lowest, ~1 kcal/mol less favorable than alanine. Based on our analysis, the helix propensities of the amino acids are as follows (kcal/mol): Ala = 0, Leu = 0.21, Arg = 0.21, Met = 0.24, Lys = 0.26, Gln = 0.39, Glu = 0.40, Ile = 0.41, Trp = 0.49, Ser = 0.50, Tyr = 0.53, Phe = 0.54, Val = 0.61, His = 0.61, Asn = 0.65, Thr = 0.66, Cys = 0.68, Asp = 0.69, and Gly = 1.

INTRODUCTION

An interesting article by Linus Pauling describing his discovery of the α-helix in the early 1950s was recently published (Pauling, 1996). Since its discovery, the α-helix has justifiably dominated our thinking about protein structure and protein folding. The α-helix probably plays an important role in the early stages of protein folding and is the most prevalent type of secondary structure found in proteins (Stickle et al., 1992).

When the first crystal structures became available, it appeared that certain amino acids were found more frequently in α-helices (e.g., Ala, Leu, and Glu), and others were found less frequently (e.g., Pro, Gly, and Asp) (Davies, 1964; Guzzo, 1965; Prothero, 1966; Cook, 1967; Ptitsyn, 1969). Based on this sort of information, Chou and Fasman (1974) derived a helix propensity scale, and proposed a method of predicting where α-helices (and β-sheets) would occur in folded proteins, using just the amino acid sequence. This has been an active area of research for over 20 years, and the regions of α-helix can now be predicted with a success rate of over 60% (Defay and Cohen, 1995; Frishman and Argos, 1997).

Brown and Klee (1971) showed that a 13-residue peptide from the C-terminus of RNase A forms a significant amount of α-helix in solution. The same residues are α-helical in the intact protein. This surprising observation stimulated studies of the α-helix to random coil transitions of peptides as models for the energetics of interactions that stabilize proteins. Peptides are particularly well suited to the study of helix propensity, which measures how the side chain of a particular amino acid affects the helix-to-coil equilibrium. Helix propensities have now been studied in amino acid polymers (Wójcik et al., 1990), peptides of defined length and sequence (Park et al., 1993; Chakrabartty and Baldwin, 1995; Muñoz and Serrano, 1995; Kallenbach et al., 1996; Rohl et al., 1996; Yang et al., 1997), coiled-coils of α-helices (O'Neil and DeGrado, 1990), and intact proteins (Horovitz et al., 1992; Blaber et al., 1994). Recently we measured the helix propensities of the amino acids in an α-helix in an intact protein, ribonuclease T1, and in a 17-residue peptide with sequences identical to those of the α-helices in the protein (Myers et al., 1997a,b). In general, the propensities were in excellent agreement, and there were reasonable explanations in the few cases where differences were observed (Myers et al., 1997b).

In this paper we derive a helix propensity scale based on the available experimental data on proteins and peptides noted above. This scale is in excellent agreement with a scale based only on data from peptides that was developed by Muñoz and Serrano (1995) using a completely different approach. It is also in excellent agreement with a structure-based helix propensity scale developed by Luque et al. (1996). It is in reasonably good agreement with helix propensity scales based on the frequency of occurrence of amino acids in α-helices in proteins (Chou and Fasman, 1978; Williams et al., 1987; Richardson and Richardson, 1988).

AN EXPERIMENTALLY BASED HELIX PROPENSITY SCALE

Our helix propensity scale is based on the 11 sets of experimental data given in Table 1. The results are presented as $\Delta(\Delta G)$ values relative to Ala, which has been set to zero because it is usually the amino acid with the most favorable helix propensity. The first five columns are results based on studies of four proteins and the coiled-coil used by O'Neil and DeGrado (1990) (see also Lovejoy et al., 1993), and the

Received for publication 18 June 1997 and in final form 16 September 1997.

Address reprint requests to Dr. C. Nick Pace, Department of Medical Biochemistry and Genetics, Texas A&M University, College Station, TX 77843-1114. Tel.: 409-845-1788; Fax: 409-847-9481; E-mail: pace@bioch.tamu.edu.

0006-3495/98/07/422/06 $2.00

TABLE 1 Experimental scales for helix propensity

AA	T1–21	Ba–32	T4–44	T4–131	C–Coil	T1-Pep	AK/AQ	KEAKE	K_2AE_2	E_4K_4	AGADIR
Gly	0.90	0.91	0.96	0.94	0.77	0.97	1.97	1.95	1.05	0.74	1.10
Ala	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Val	0.66	0.88	0.33	0.26	0.63	0.64	1.04	1.25	0.73	0.47	0.46
Ile	0.44	0.81	0.12	0.10	0.54	0.34	0.71	0.78	0.38	0.42	0.35
Leu	0.13	0.35	0.04	0.17	0.15	0.22	0.37	0.31	0.36	0.19	0.19
Met	0.15	0.31	0.10	0.13	0.27	0.15	0.52	0.45	0.53	0.24	0.21
Phe	0.57	0.69	0.37	—	0.36	0.60	1.00	0.45	1.11	—	0.47
Tyr	0.39	0.82	0.24	—	0.60	0.38	0.69	0.78	1.95	—	0.47
Trp	0.30	0.98	0.38	—	0.32	0.08	0.96	0.62	1.11	—	0.47
Ser	0.45	0.41	0.43	0.30	0.42	0.47	0.79	1.01	0.76	0.51	0.52
Cys	0.74	1.00	0.54	—	0.54	0.52	0.91	0.85	1.17	—	0.60
Thr	0.57	0.79	0.42	0.38	0.66	0.65	1.22	1.25	0.89	0.56	0.57
Lys^+	0.51	0.19	0.23	—	0.12	0.38	0.29	0.20	0.59	—	0.15
Arg^+	0.41	0.14	0.19	—	0.09	0.47	0.22	−0.01	0.41	—	0.06
Gln	0.33	0.48	0.16	—	0.44	0.30	0.55	0.62	0.80	0.33	0.32
Glu^0	−0.05	—	0.43	0.06	—	0.17	0.48	0.32	—	—	—
Glu^-	0.69	0.55	—	—	0.50	0.31	0.62	0.39	0.21	—	0.34
Asn	−0.34	0.66	0.57	—	0.70	0.62	0.96	1.12	0.85	0.60	0.60
Asp^0	−0.33	—	0.54	0.17	—	0.66	0.79	1.10	—	—	—
Asp^-	0.71	0.71	—	—	0.62	0.68	0.81	1.10	0.78	—	0.59
His^0	0.17	—	—	—	0.71	0.67	0.84	1.17	0.54	—	0.62
His^+	0.56	0.78	0.39	—	—	1.20	1.11	1.17	—	—	—
Pro	—	4.08	3.46	—	3.77	1.10	4.07	—	2.56	—	2.72

The values given are $\Delta(\Delta G)$ values in kcal/mol relative to alanine. They are based on experimental data from these articles: T1–21 and T1–Pep, Myers et al. (1997b); Ba–32, Horovitz et al. (1992); T4–44 and T4–131, Blaber et al. (1994); C-Coil, O'Neil and DeGrado (1990); AK/AQ, Rohl et al. (1996); KEAKE, Park et al. (1993), as analyzed by Chakrabartty and Baldwin (1995); K_2AE_2, Yang et al. (1997); E_4K_4, Kallenbach et al. (1996); AGADIR, Muñoz and Serrano (1995).

The sequences of the four residues on either side of guest site in the helices are T1–21 and T1–Pep, −STAQ**X**AAYK−; Ba–32, −KSAQ**X**LG− (the X + 3 and X + 4 residues are not in the α-helix); T4–44, −QAAK**X**ELDK−; T4–131,−DEAA**X**NLA− (the X + 4 residue in not in the α-helix); C-Coil, −AALE**X**KLQA−; AK/AQ, −AAK(or Q)A**X**AAK(or Q)A−; KEAKE, −AKEA**X**AKEA−; K_2AE_2, −AKKA**X**AEEA−; E_4K_4, −KKKX**X**XEEE−; AGADIR, based on 423 peptide sequences.

last six columns are results based on studies of α-helix to random coil transitions of peptides. The last column gives a helix propensity scale from Muñoz and Serrano (1995) that is based on the helical content measured in 423 peptides. Thus it includes some of the results from the other studies of peptides, but was derived using an analysis completely different from those used in the individual studies.

The correlations among the various scales are generally good, but the magnitude of the $\Delta(\Delta G)$ values differs significantly in some cases. This can be seen by comparing the $\Delta(\Delta G)$ values for Gly. On most of the scales, Ala has the highest helix propensity, and Gly has the lowest (excluding Pro). Note that for AK/AQ and KEAKE, the $\Delta(\Delta G)$ value for Gly is almost twice as large as it is for the other systems, and for E_{4K4} and the coiled-coil systems, the $\Delta(\Delta G)$ for Gly is ~25% less than for the other systems. We will discuss this below. For the other seven scales in the table, the $\Delta(\Delta G)$ values for Gly are approximately the same: 0.98 ± 0.07. (If all 11 $\Delta(\Delta G)$ values for Gly are averaged, 1.12 ± 0.43 is obtained.) To derive our helix propensity scale, we adjusted all of the scales to fit a range in which the Gly-Ala difference is exactly 1.0. These scales are shown in Table 2.

In our studies of helix propensity with the α-helix in RNase T1, the results showed that some specific interactions occur that were not foreseen (Myers et al., 1997b). For example, with Asn or uncharged Asp present at the guest position, the protein was considerably more stable than the peptide, giving these amino acids a greater helix propensity than Ala (Table 1). We think this is due to the side chains of these residues donating a hydrogen bond to the Oγ of Ser^{17} in the protein, but this interaction does not occur to the same extent in the peptide because of helix fraying (Myers et al., 1997b). Because of the possibility of specific interactions such as this and to minimize the contribution of outliers, for each amino acid we eliminate the two extreme values before averaging. Thus the number of values averaged ranged from 4 to 9, depending on the amino acid. The average values for each amino acid are given in Table 2, and these are our experimentally based helix propensities.

FEATURES OF THE HELIX PROPENSITY SCALE

The helix propensities from Table 2 are grouped in Table 3 to illustrate the structural similarities of the amino acids with similar helix propensities. After Ala, uncharged Glu (Glu^0) is the best helix former, and it is substantially better (0.24 kcal/mol) than charged Glu (Glu^-). Similarly, Asp^0 is 0.26 kcal/mol better than Asp^-. This may indicate that the charged carboxyls can form stronger hydrogen bonds with the amide hydrogens in the peptide backbone in the random coil, and that this lowers their helix propensities. Leu and Met have long been recognized as strong helix formers on the basis of their frequent occurrence in helices (Chou and

TABLE 2 Relative helix propensities

AA	T1–21	Ba–32	T4–44	T4–131	C–Coil	T1-Pep	AK/AQ	KEAKE	K_2AE_2	E_4K_4	AGADIR	AVERAGE*	AVE DEV
Gly	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00
Ala	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Val	0.73	0.97	0.34	0.28	0.82	0.66	0.53	0.64	0.70	0.64	0.46	0.61	0.11
Ile	0.49	0.89	0.13	0.11	0.70	0.35	0.36	0.40	0.36	0.57	0.32	0.41	0.12
Leu	0.14	0.38	0.04	0.18	0.19	0.23	0.19	0.16	0.34	0.26	0.17	0.21	0.05
Met	0.17	0.34	0.10	0.14	0.35	0.15	0.26	0.23	0.50	0.32	0.19	0.24	0.07
Phe	0.63	0.76	0.39	—	0.47	0.62	0.51	0.23	1.06	—	0.43	0.54	0.11
Tyr	0.43	0.90	0.25	—	0.78	0.39	0.35	0.40	1.85	—	0.43	0.53	0.18
Trp	0.33	1.08	0.40	—	0.42	0.08	0.49	0.32	1.06	—	0.43	0.49	0.16
Ser	0.50	0.45	0.45	0.32	0.55	0.48	0.40	0.52	0.72	0.69	0.47	0.50	0.06
Cys	0.82	1.10	0.56	—	0.70	0.54	0.46	0.43	1.11	—	0.55	0.68	0.17
Thr	0.63	0.87	0.44	0.40	0.86	0.67	0.62	0.64	0.84	0.76	0.52	0.66	0.10
Lys^+	0.57	0.21	0.24	—	0.16	0.39	0.15	0.10	0.56	—	0.14	0.26	0.12
Arg^+	0.46	0.15	0.20	—	0.12	0.48	0.11	0.00	0.39	—	0.05	0.21	0.12
Gln	0.37	0.53	0.17	—	0.57	0.31	0.28	0.32	0.76	0.45	0.29	0.39	0.09
Glu^0	−0.06	—	0.45	0.06	—	0.18	0.24	0.17	—	—	—	0.16	0.05
Glu^-	0.77	0.60	—	—	0.65	0.32	0.31	0.20	0.20	—	0.31	0.40	0.15
Asn	−0.38	0.73	0.59	—	0.91	0.64	0.49	0.57	0.81	0.81	0.55	0.65	0.10
Asp^0	−0.37	—	0.56	0.18	—	0.68	0.40	0.56	—	—	—	0.43	0.14
Asp^-	0.79	0.78	—	—	0.81	0.70	0.41	0.56	0.74	—	0.54	0.69	0.09
His^0	0.19	—	—	—	0.92	0.69	0.43	0.60	0.51	—	0.56	0.56	0.07
His^+	0.62	0.86	0.41	—	—	1.24	0.56	0.60	—	—	—	0.66	0.10
Pro	—	4.48	3.60	—	4.90	1.13	2.07	—	2.44	—	2.47	3.01	0.82

The helix propensities from Table 1 have been adjusted so that $\Delta(\Delta G) = 0$ for Ala and $\Delta(\Delta G) = 1$ for Gly, by dividing the $\Delta(\Delta G)$ values in each column in Table 1 by the $\Delta(\Delta G)$ value for Gly.

*For each row, the lowest and highest values were eliminated, and then the average and the average deviation were calculated using the remaining values.

Fasman, 1978), and this is borne out by the results in Table 3. Arg^+ and Lys^+ resemble Leu and Met in structure near the backbone, and are equally good helix formers. Gln and Glu^- are good helix formers, substantially better than Asn and Asp^-. This probably reflects the difference in side-chain length: it costs less conformational entropy to form specific interactions in the random coil with the shorter side chains, so they have lower helix propensities than the longer side chains with the same polar groups. The aliphatic side chains Leu, Ile, Val, and Ala are interesting: Ala is 0.21 kcal/mol better than Leu, Leu is 0.20 kcal/mol better than Ile, which is 0.20 kcal/mol better than Val. It is clear that Ala is favored over Leu, and Leu is favored over Ile and Val, mainly because of conformational entropy (Hermans et al., 1992; Creamer and Rose, 1994). It is less clear why Ile is favored over Val. As previously suggested, the Cδ of Leu and Ile seems to have a favorable interaction in the helix that buries some nonpolar surface and thereby increases their helix propensity (Horovitz et al., 1992; Blaber et al., 1994; Chakrabartty and Baldwin, 1995). The three aromatic residues have similar helix propensities and are about equidistant between Gly and Ala. This suggests that that the hydrogen bonding capabilities of Trp and Tyr do not affect their helix propensities. The amino acids that are the worst helix formers all have polar groups separated from the backbone by only a single $-CH_2-$ group. The effect of an added $-CH_2-$ group is remarkably similar in comparable residues: Asn–Gln = 0.26 kcal/mol, Asp^0–Glu^0 = 0.29, and Asp^-–Glu^- = 0.27. Again, this probably reflects the cost of freezing an additional $-CH_2-$ bond to form a specific interaction in the random-coil state. (Based on several different studies, Doig and Sternberg (1995) estimate the mean conformational free energy change to be 0.5 kcal/mol per bond frozen.) Pro is clearly the worst helix former, substantially below Gly. Because it lacks an amide hydrogen for main chain hydrogen bonding and because of its unique geometry, Pro is clearly a special case and will not be included in the remainder of the discussion.

COMPARISON OF THE EXPERIMENTAL HELIX PROPENSITY SCALE WITH OTHER SCALES

In Table 4, the amino acids are listed in order, using the helix propensities from Table 3. Asp^0 and Glu^0 are omitted because they are only relevant at low pH. Because the pKs of His residues are generally 7 ± 1, the charge status of a given His is not known unless the pK values have been measured. Consequently, we have used the average of the values for His^0 and His^+ for His, 0.61 kcal/mol.

The Baldwin laboratory has carefully chosen the peptides they use to measure helix propensities to minimize specific interactions. Their helix propensities are given in column AK/AQ in Table 4. It can be seen that the correlation between their scale and the scale reported here is excellent. The correlation coefficient is 0.99 for the nonpolar amino acids and 0.94 for all of the amino acids. However, the $\Delta(\Delta G)$ for Gly is twice as large for the AK/AQ peptides than it is for our scale, and all of the amino acids have greater $\Delta(\Delta G)$ values in their scale.

The average bulkiness of the side chains at X ± 1 (the residues closest to the guest site in the random coil), and at

TABLE 3 A helix propensity scale based on experimental studies of proteins and peptides

Amino acid	Helix propensity (kcal/mol)
Ala	0.00
Glu^0	0.16
Leu	0.21
Met	0.24
Arg^+	0.21
Lys^+	0.26
Gln	0.39
Glu^-	0.40
Ile	0.41
Asp^0	0.43
Ser	0.50
Trp	0.49
Tyr	0.53
Phe	0.54
Val	0.61
Thr	0.66
His^0	0.56
His^+	0.66
Cys	0.68
Asn	0.65
Asp^-	0.69
Gly	1.00
Pro	3.16

The values given are the average values from Table 2. Note that the amino acids are grouped to highlight structural similarities, and they are not ranked in strict order of decreasing helix propensity, as they are in Table 4.

TABLE 4 Comparison of helix propensity scales

AA	Exptl	AK/AQ	AGADIR	Williams	C. & F.	Luque
Ala	0	0	0	0	0	0
Arg	0.21	0.22	0.06	0.20	0.44	0.15
Leu	0.21	0.37	0.19	0.07	0.21	0.15
Met	0.24	0.52	0.21	0.11	−0.03	0.18
Lys	0.26	0.29	0.15	0.18	0.26	0.11
Gln	0.39	0.55	0.32	0.14	0.31	0.30
Glu	0.40	0.62	0.34	−0.18	−0.09	0.37
Ile	0.41	0.71	0.35	0.32	0.34	0.48
Trp	0.49	0.96	0.47	0.39	0.34	0.35
Ser	0.50	0.79	0.52	0.84	0.65	0.48
Tyr	0.53	0.69	0.47	0.67	0.73	0.46
Phe	0.54	1.00	0.47	0.25	0.29	0.35
Val	0.61	1.04	0.46	0.43	0.36	0.36
His	0.61	0.97	0.62	0.36	0.42	0.62
Asn	0.65	0.96	0.6	0.65	0.75	0.52
Thr	0.66	1.22	0.57	0.65	0.59	0.59
Cys	0.68	0.91	0.6	0.75	0.72	0.57
Asp	0.69	0.81	0.59	0.42	0.41	0.47
Gly	1.00	1.97	1.10	0.98	0.85	0.79
For the seven nonpolar amino acids (Gly, Ala, Val, Ile, Leu, Met, Phe)[#]						
Corr coef		0.99	0.98	0.96	0.93	0.94
Slope		1.91	1.05	0.97	0.84	0.74
Intercept		−0.02	−0.06	−0.11	−0.07	0.01
For all 19 amino acids[#]						
Corr coef		0.94	0.97	0.78	0.74	0.93
Slope		1.75	1.06	1.05	0.86	0.81
Intercept		−0.07	−0.08	−0.12	−0.01	0.00

The helix propensity scales are from the following: Exptl, Table 3; AK/AQ, Rohl et al. (1996); AGADIR, Muñoz and Serrano (1995); Williams, Williams et al. (1987); C. & F., Chou and Fasman (1978); Luque, Luque et al. (1996).

[#]The helix propensity values from the five scales are compared to our new experimental scale, and the correlation coefficient, slope, and *y* intercept are shown for the nonpolar residues only and the entire set of 19 amino acids.

X ± 3 and X ± 4 (the residues closest to the guest site in the α-helix) is substantially less in the AK/AQ and KEAEK peptides (Table 1), and we previously suggested that this was the cause of the substantial difference between the Ala-stabilized peptides and the other systems (Myers et al., 1997b). It has been shown in α-helices in T4 lysozyme that replacing amino acids with larger side chains by alanine leads to an increase in the accessibility of main-chain amide and carbonyl groups to water (Blaber et al., 1995). However, the Kallenbach laboratory recently measured a complete set of helix propensities in a peptide with exactly the same residue bulkiness as the KEAEK peptides and found a $\Delta(\Delta G)$ value for Gly of 1.05 (peptide K_2AE_2 in Table 1) (Yang et al., 1997). At present we are aware of no reasonable explanation for the large difference between the $\Delta(\Delta G)$ values from the AK/AQ and KEAEK peptides and the other systems.

The measured helicities of 423 peptides were used by Muñoz and Serrano (1995) to derive parameters characterizing the helix-to-coil transition of peptides in solution. They separate helix propensities from capping effects and other specific interactions, as the experimental studies try to do by selecting favorable systems. They use these parameters in a algorithm named AGADIR to predict the α-helix content of peptides with impressive accuracy. A perusal of Table 4 shows that the helix propensities from their analysis are in excellent agreement with those derived here. The agreement between AGADIR, which is based entirely on experimental results from peptides, and the scale here, which is based on experimental results from both peptides and proteins, shows that studies of the energetics of helix formation in proteins and peptides give equivalent results, in favorable cases.

Further support for this idea is obtained when the five protein systems and the six peptide systems are averaged separately to determine the $\Delta(\Delta G)$ values for the nonpolar amino acids, in which specific interactions will be minimal. For Leu, Met, Ile, Phe, and Val, the proteins give $\Delta(\Delta G)$ values of 0.19 ± 0.08, 0.22 ± 0.10, 0.46 ± 0.28, 0.56 ± 0.13, and 0.63 ± 0.25, and the peptides give 0.22 ± 0.05, 0.28 ± 0.09, 0.39 ± 0.06, 0.57 ± 0.22, and 0.60 ± 0.07, respectively. Finally, the excellent agreement between the helix propensities measured in helices in an intact protein, ribonuclease T1, and in α-helical peptides with identical sequences offers further strong support that combining results from studies of proteins and peptides to derive a helix propensity scale is reasonable. It also shows that analyses of peptide energetics in terms of Lifson-Roig (1961) theory and protein energetics in terms of a two-state analysis of thermal or urea unfolding studies lead to comparable results. Thus we think the experimentally based helix propensity

scale in Table 3 provides a good measure of helix propensity for solvent-exposed amino acid residues at the middle positions (noncapping) of an α-helix.

Next we compare our helix propensity scale to two scales based on the frequency of occurrence of amino acids in α-helices in globular proteins. The most popular scale over the years has been that of Chou and Fasman (1978). It is based on a sample of 29 proteins. The correlation with our scale is good (0.93) when just the nonpolar amino acids are included, and fair (0.74) when all of the amino acids are included (Table 4). There are two important differences between the two scales. For the experimental studies used in Table 1, the guest sites were chosen so that they would be solvent exposed and near the center of the α-helices. In contrast, the sample used by Chou and Fasman (1978) includes residues at the ends of helices and residues that are partially or completely buried in the interior of the proteins. It is known that residues at the ends of α-helices have propensity scales that are different from those in the middle of α-helices (Doig and Baldwin, 1995; Aurora et al., 1997; Doig et al., 1997), and it is unlikely that the factors that determine helix propensity would apply to buried residues. Experimental studies of peptides in solution show that the bulk solvent environment can override the secondary structure propensities (Waterhous and Johnson, 1994). Given this, the agreement between our scale and the Chou-Fasman scale should probably be regarded as good.

Williams et al. (1987) based their helix propensity scale on the frequency of occurrence of the amino acids in α-helices found in the structures of 75 proteins. The correlation with our scale is good (0.96) for the nonpolar amino acids, and fair (0.78) when all of the amino acids are considered. Williams et al. (1987) did not consider solvent exposure, but did subdivide the residues into those at the ends of helices and those in the middle of helices. It is surprising and puzzling that the correlation coefficient drops to 0.92 for the nonpolar amino acids when only amino acids in the middle of helices are considered. This is identical to the correlation coefficient obtained when our scale is compared to that of the Richardsons (1988), who also analyzed the middle residues in the helices separately from those at the ends. (Their scale is based on the frequency of occurrence of amino acids in 215 α-helices in a sample of 45 proteins.) The major outlier in these comparisons is Glu^-. It is observed more frequently in helices in protein than would be expected based on our scale. The same is true, but to a smaller extent, for Asp^- and Phe. In contrast, Ser is observed less often in helices than would be expected based on our scale.

In the last column of Table 4, we compare our helix propensities with those from a scale reported by Luque et al. (1996). The correlations are excellent. In their approach, they make use of the differences in accessible surface area between the α-helical and randomly coiled states, and empirical thermodynamic parameters based on experimental studies to derive a helix propensity scale. With this approach, it is most difficult to model the accessibility of the randomly coiled state (Creamer et al., 1997). Consequently, they used four of the data sets in Table 1 to derive accessibilities for the coiled state that minimized the differences between predicted and experimental $\Delta(\Delta G)$ values. Given this, it is reassuring but not so surprising that there is such good agreement between the two scales.

FACTORS THAT CONTRIBUTE TO HELIX PROPENSITY

An α-helix is stabilized mainly by a favorable enthalpic contribution of $\sim$1 kcal/mol per residue from the formation of the backbone hydrogen bonds, and from van der Waals interactions that are more favorable in the helix than the coil (Scholtz et al., 1991; Yang and Honig, 1995). The randomly coiled state is greatly favored by conformational entropy (Aurora et al., 1997). The entropic cost of fixing the backbone dihedral angles in forming an α-helix is in the range of 1.5–2 kcal/mol at 25°C (Nemethy et al., 1966; Yang and Honig, 1995; Wang and Purisma, 1996). If the α-helix is composed of Ala, the favorable interactions win and the helix is more stable than the coil, but if the α-helix is composed of Gly, the unfavorable interactions win and the coil is more stable than the helix. The 1 kcal/mol difference in helix propensity between Gly and Ala is due mainly to the large reduction in phi-psi space available to residues when the H of Gly is replaced by a CH_3 in Ala; this is almost entirely an effect of the CH_3 group on the conformational entropy of the backbone in the coil (Nemethy et al., 1966; Hermans et al., 1992; Luque et al., 1996).

In contrast to the effect on the backbone, the CH_3 group of Ala experiences no loss in conformational entropy when a randomly coiled polypeptide folds to an α-helix (Hermans et al., 1992). All of the other side chains, however, experience an unfavorable change in side-chain conformational entropy when a helix is formed, and the loss is greater for residues such as Val and Ile, where the β carbon is branched, than for residues such as Leu and Met, where it is not (Hermans et al., 1992; Aurora et al., 1997). Estimating just the contribution of conformational entropy, Creamer and Rose (1994) find $\Delta(\Delta G)$ values of 0.13 for Met, 0.15 for Leu, 0.22 for Phe, 0.28 for Ile, and 0.39 for Val. Comparing these with the values in Table 3 suggests that conformational entropy can account for many but not all of the helix propensities of the nonpolar amino acids. A similar conclusion was reached in other analyses (Hermans et al., 1992; Avbelj and Moult, 1995; Wang and Purisma, 1996). It is not clear what other factors are important in determining helix propensities, but a number have been suggested: burying a hydrophobic surface in the helix (Blaber et al., 1994; Yang and Honig, 1995), unfavorable steric contacts in the helix (Hermans et al., 1992), screening of electrostatic effects (Avbelj and Moult, 1995), and favorable side chain-to-helix van der Waals interactions (Wang and Purisma, 1996).

In summary, we have derived a helix propensity scale based on studies of the stability of proteins and the α-helix-to-coil transition of peptides. It is reassuring that the two

sets of measurements lead to similar results. Helix propensities make a small but important contribution to protein stability and are crucially important to the stability of α-helical segments in unfolded proteins.

We dedicate this paper to the memory of Dr. Gregorio Weber. For many of us, his research awakened an appreciation of the power of fluorescence spectroscopy and the dynamics of protein structure.

We thank Bea Huyghues-Despointes, Doug Laurents, and Jeff Myers for helpful discussions. We thank Irene Luque for sending us information used in Table 4.

This research was supported by grants from the National Institutes of Health (GM37039 to CNP and GM52483 to JMS), and the Robert A. Welch Foundation (A-1060 to CNP and A-1281 to JMS). JMS is an American Cancer Society Junior Faculty Research Awardee (grant JFRA-577). CNP is supported by the Tom and Jean McMullin Professorship.

REFERENCES

Aurora, R., T. P. Creamer, R. Srinvasan, and G. D. Rose. 1997. Local interactions in protein folding: lessons from the α-helix. *J. Biol. Chem.* 272:1413–1416.

Avbelj, F., and J. Moult. 1995. Role of electrostatic screening in determining protein main chain conformational preferences. *Biochemistry.* 34: 755–764.

Blaber, M., W. A. Baase, N. Gassner, and B. W. Matthews. 1995. Alanine scanning mutagenesis of the α-helix 115–123 of phage T4 lysozyme: effects on structure, stability, and the binding of solvent. *J. Mol. Biol.* 246:317–330.

Blaber, M., X. J. Zhang, J. D. Lindstrom, S. D. Pepoit, W. A. Baase, and B. W. Matthews. 1994. Determination of α-helix propensity within the context of a folded protein. *J. Mol. Biol.* 235:600–624.

Brown, J. E., and W. A. Klee. 1971. Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry.* 10:470–476.

Chakrabartty, A., and R. L. Baldwin. 1995. Stability of α-helices. *Adv. Protein Chem.* 46:141–176.

Chou, P. Y., and G. Fasman. 1974. Conformational parameters of amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry.* 13:211–222.

Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* 47:251–276.

Cook, D. A. 1967. The relation between amino acid sequence and protein conformation. *J. Mol. Biol.* 29:167–171.

Creamer, T. P., and G. D. Rose. 1994. α-Helix-forming propensities in peptides and proteins. *Proteins Struct. Funct. Genet.* 19:85–97.

Creamer, T. P., R. Srinivasan, and G. D. Rose. 1997. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry.* 36:2832–2835.

Davies, D. R. 1964. A correlation between amino acid composition and protein structure. *J. Mol. Biol.* 9:605–609.

Defay, T., and F. E. Cohen. 1995. Evaluation of current techniques for ab initio protein structure predictions. *Proteins Struct. Funct. Genet.* 23:431–445.

Doig, A. J., and R. L. Baldwin. 1995. N- and C-capping preferences for all 20 amino acids in α-helical peptides. *Protein Sci.* 4:1325–1336.

Doig, A. J., M. W. Macarthur, B. J. Stapley, and J. M. Thornton. 1997. Structures of N-temini of helices in proteins. *Protein Sci.* 6:147–155.

Doig, A. J., and M. J. E. Sternberg. 1995. Side-chain conformational entropy in protein folding. *Protein Sci.* 4:2247–2251.

Frishman, D., and P. Argos. 1997. The future of protein secondary structure prediction accuracy. *Folding Design.* 2:159–162.

Guzzo, A. V. 1965. The influence of amino acid sequence on protein structure. *Biophys. J.* 5:809–822.

Hermans, J., A. G. Anderson, and R. H. Yun. 1992. Differential helix propensity of small apolar side chains studied by molecular dynamics simulations. *Biochemistry.* 31:5646–5653.

Horovitz, A., J. M. Matthews, and A. R. Fersht. 1992. Alpha-helix stability in proteins II. Factors that influence stability at an internal position. *J. Mol. Biol.* 227:560–568.

Kallenbach, N. R., P. Lyu, and H. Zhou. 1996. CD spectroscopy and the helix-coil transition in peptides and polypeptides. *In* Circular Dichroism and the Conformational Analysis of Biomolecules. G. D. Fasman, editor. Plenum Press, New York. 201–259.

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

Lovejoy, B., S. Choe, D. Cascio, D. K. McRorie, W. F. DeGrado, and D. Eisenberg. 1993. Crystal structure of a synthetic triple-stranded α-helical bundle. *Science.* 259:1288–1293.

Luque, I., O. L. Mayorga, and E. Freire. 1996. Structure-based thermodynamic scale of α-helix propensities in amino acids. *Biochemistry.* 35: 13681–13688.

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

Myers, J. K., C. N. Pace, and J. M. Scholtz. 1997a. A direct comparison of helix propensity in proteins and peptides. *Proc. Natl. Acad. Sci. USA.* 94:2833–2837.

Myers, J. K., C. N. Pace, and J. M. Scholtz. 1997b. Helix propensities are identical in proteins and peptides. *Biochemistry.* 36:10923–10929.

Nemethy, G., S. J. Leach, and H. A. Scheraga. 1966. The influence of amino acid side-chains on the free energy of helix-coil transitions. *J. Phys. Chem.* 70:998–1004.

O'Neil, K. T., and W. F. DeGrado. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science.* 250:646–651.

Park, S. H., W. Shalongo, and E. Stellwagen. 1993. Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. *Biochemistry.* 32:7048–7053.

Pauling, L. 1996. The discovery of the alpha helix. *Chem. Intelligencer.* January:32–38.

Prothero, J. W. 1966. Correlation between the distribution of amino acids and alpha helices. *Biophys. J.* 6:367–370.

Ptitsyn, O. B. 1969. Statistical analysis of the distribution of amino acid residues among helical and non-helical regions in globular proteins. *J. Mol. Biol.* 42:501–510.

Richardson, J. S., and D. C. Richardson. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science.* 240:1648–1652.

Rohl, C. A., A. Chakrabartty, and R. L. Baldwin. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* 5:2623–2637.

Scholtz, J. M., S. Marqusee, R. L. Baldwin, E. J. York, J. M. Stewart, M. Santoro, and D. W. Bolen. 1991. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. USA.* 88:2854–2858.

Stickle, D. F., L. G. Presta, K. A. Dill, and G. D. Rose. 1992. Hydrogen bonding in globular proteins. *J. Mol. Biol.* 226:1143–1159.

Wang, J., and E. O. Purisma. 1996. Analysis of thermodynamic determinants in helix propensities of nonpolar amino acids through a novel free energy calculation. *J. Am. Chem. Soc.* 118:995–1001.

Waterhous, D. V., and W. C. Johnson, Jr. 1994. Importance of environment in determining secondary structure in proteins. *Biochemistry.* 33: 2121–2128.

Williams, R. C., A. Chang, D. Juretic, and S. Loughran. 1987. Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta.* 916:200–204.

Wójcik, J., K. H. Altman, and H. A. Scheraga. 1990. Helix-coil stability constants for the naturally occurring amino acids in water XXIV. Half-cysteine parameters from random poly(hydroxybutylglutamine-co-*S*-methylthio-L-cysteine). *Biopolymers.* 30:121–134.

Yang, A.-S., and B. Honig. 1995. Free energy determinants of secondary structure formation. I. α-Helices. *J. Mol. Biol.* 252:351–365.

Yang, J., E. J. Spek, Y. Gong, H. Zhou, and N. R. Kallenbach. 1997. The role of context on alpha-helix stabilization: host-guest analysis in a mixed background peptide model. *Protein Sci.* 6:1–9.