ELSEVIER

Biophysical Chemistry 55 (1995) 127–135

Biophysical Chemistry

# α-Helix formation by peptides of defined sequence

Robert L. Baldwin *

*Department of Biochemistry, Stanford Medical Center, Stanford, CA 94305-5307, USA*

**Abstract**

The factors controlling α-helix formation in water by peptides of defined sequence are beginning to be understood. The field is close to the point where the extent of helix formation can be predicted for peptides of any sequence. Our own approach to the problem, and the main results obtained by following this approach, are summarized below. The chief reason for studying α-helix formation by peptides is to understand precisely and in detail one part of the protein folding problem. Questions about peptide helix formation can be answered at a fundamental level, in terms of the physico-chemical mechanisms involved.

*Keywords:* Helix-stabilizing interaction; Helix propensity; N-cap

## 1. Early studies of helices formed by protein fragments

The study of helix formation by peptide fragments of proteins began in 1971 with the report [1] by Brown and Klee that C-peptide (residues 1–13 of ribonuclease A, containing the N-terminal helix, residues 3–13) shows partial helix formation (about 25%) in water at 0° C. Studies in other laboratories reported, however, that peptides from myoglobin [2] and from staphylococcal nuclease [3] show no measurable helix formation at 25° C, and the subject languished until 1982 when our laboratory [4] confirmed the report by Brown and Klee and showed further that two critical ionizing residues are involved in stabilizing the C-peptide helix. They were later found to be Glu $2^-$ [5,6] and His $12^+$ [6], at opposite ends of the helix.

Because the locations and charges of Glu $2^-$ and His $12^+$ are consistent with stabilization of the helix by charge interaction with the helix macrodipole [7] (i.e., with the partial charges on the backbone NH and CO groups), we tested for this interaction by varying the charge on the N-terminal residue [8] and found that the charge–helix dipole interaction does indeed make an important contribution to helix stability. This conclusion might have been drawn earlier from the beautiful studies by Ooi and coworkers [9,10] of placing several charged residues at either end of an alanine helix. Further work on C-peptide showed, however, that Glu $2^-$ and His $12^+$ are in fact involved in specific helix-stabilizing interactions with other charged side chains: there is a Glu $2^-$ · Arg $10^+$ salt bridge interaction [11] and also a Phe 8 · His $12^+$ pseudo H-bond interaction [12] whose properties are described consistently [13] by the model of Levitt and Perutz [14].

Both the Glu 2 · Arg 10 salt bridge and the Phe 8 · His 12 interaction are visible in the X-ray structure of ribonuclease A [15]. Consequently, because these interactions are also important in stabilizing the isolated C-peptide helix, this helix provides the first

* Corresponding author.


*SSDI* 0301-4622(94)00146-4

example of an autonomous folding unit [16], in which the information needed for correct folding is encoded in the local sequence. Moreover, NMR studies [17,18] showed that in the longer S-peptide (residues 1–20 of ribonuclease A) the helix stops near Met 13, the C-terminal residue of the helix in the folded protein: a "helix stop" signal appears to be encoded in the sequence (see below).

Although substitution experiments in C-peptide succeeded in identifying the three helix-stabilizing interactions mentioned above, they also showed that helix experiments with protein fragments are complicated because additional interactions may occur with other amino acid side chains besides the interaction under study and because, contrary to expectations based on the host–guest values for helix propensities (see below), differences in helix propensities are a major factor affecting substitution experiments. This situation is illustrated in Fig. 1, with experiments made in C-peptide on the Phe 8 · His 12 interaction [12]. Interchanging Ala 11 with His 12 breaks the interaction, which is specific for an $i$, $i+4$ spacing between Phe and His, and so the helix content of the peptide with His 11, Ala 12 is substantially lower than that of the peptide with Ala 11, His 12 (Fig. 1a) at pH 5.0, where histidine is ionized, although both peptides have the same helix content at pH 8.0, where the His residue is uncharged. On the other hand, replacing either Phe 8 or His 12 by Ala, in peptides where the Phe 8 · His 12 interaction has been broken by prior substitution of Phe 8 or His 12, produces an even larger increase in helix content (Fig. 1b).

Experiments such as these led to the finding by Marqusee et al. [19] that alanine has an unusually high helix propensity, and that short alanine-based peptides, containing a few $Lys^+$ or $Glu^-$ residues for solubility in water, form moderately stable, monomeric helices. At this point, the stage was set for determining the factors that control peptide helix stability, such as helix propensities and side-chain interactions, in alanine-based peptides that allow only minimal interference from unwanted side-chain interactions between the substituted residue and the host.

## 2. Tests of helix-coil theory in alanine-based peptides

When it became possible to study helix formation in simple alanine-based peptides, the next task was to determine whether classical helix-coil theory adequately describes helix-formation by short peptides in water. It was already known that side-chain interactions can make important contributions to helix stability, and these are not included in the classic

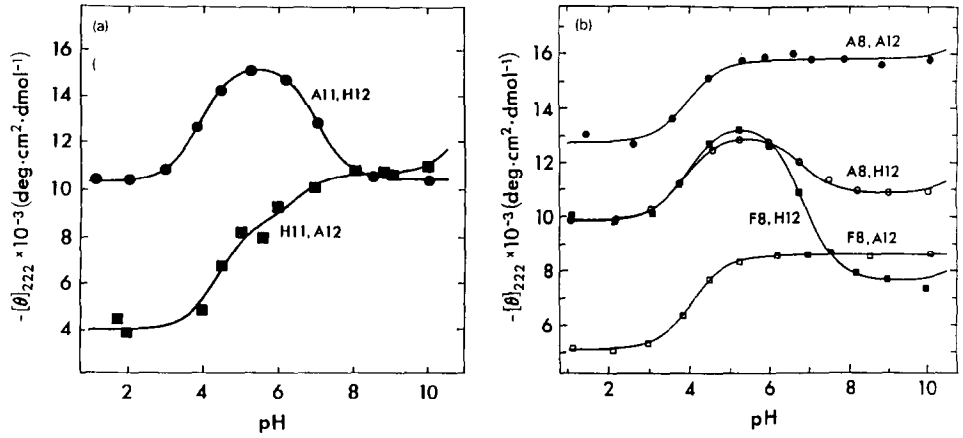


Fig. 1. Study of the Phe 8 · His $12^+$ interaction in a derivative of C-peptide, illustrating the effects of the helix propensities of Phe, His and Ala in substitution experiments. In part A, the Phe · His interaction is broken by interchanging His 12 and Ala 11. In part B, the Phe · His interaction is broken by substituting either Phe 8 or His 12 with Ala. Part B also shows the effects of making the substitutions Phe 8 → Ala and His 12 → Ala when the Phe · His interaction has been broken by prior substitution. The ordinate shows the mean residue ellipticity; 100% helix corresponds to about $-[\theta]_{222} = 34\,000$. Data from [12].

theories of Zimm and Bragg [20] and Lifson and Roig [21]. Moreover, it was evident that the "host–guest" values for the helix propensities of the amino acids [22], obtained with random sequence copolymers using hydroxybutyl- or hydroxypropyl-L-glutamine as the host, cannot explain stable helix formation by alanine peptides. Instead, there must be a "context dependence" of these values, i.e. a dependence on neighboring residues. It was important to find out how general this problem is, and whether it affects the use of current helix-coil theories to obtain meaningful parameters for helix formation.

The standard procedure, dating back to the classic 1959 study by Zimm and coworkers [23] for obtaining the parameters of helix-coil theory and also for testing the theory, is to measure thermal unfolding curves (by circular dichroism, or CD) for a set of peptides with uniform sequences and varying chain lengths. The helix nucleation constant ($\sigma$ in the Zimm–Bragg (or ZB) theory, $v^2$ in the Lifson–Roig (or LR) theory), enters just once for each helical sequence, whereas the propagation parameter ($s$ in the ZB theory, $w$ in the LR theory) enters repeatedly, once for every stabilized helical residue. Thus, the chain length dependence of the helix content, at constant temperature, contains the information needed to separate the nucleation constant, $\sigma$ or $v^2$, from $s$ or $w$. Relations allowing the ZB parameters to be computed from the LR parameters, and vice versa, have been derived by Qian and Schellman [24]. If $\sigma$ and also the enthalpy of helix formation ($\Delta H^\circ$) are independent of temperature, then only three helix-coil parameters are needed to fit the family of thermal unfolding curves: $\sigma$, $\Delta H^\circ$ and $s$ at 0° C.

Fig. 2 shows the results of fitting CD-monitored unfolding curves for a set of alanine-based peptides with the repeating sequence $(AEAAKA)_n$ and with chain lengths varying from 14 to 50 residues [25]. The lines show results of fitting to the ZB theory with $\sigma = 3.3 \times 10^{-3}$, $\Delta H^\circ = 0.95$ kcal/mol residue, and $\langle s \rangle = 1.35$ at 0° C. Parameters describing the dependence of the mean residue ellipticity on chain length and temperature, for the helix and the coil, also enter into the computed curves. The fit of theory to experiment appears to be reasonably good, and it has been tested in two ways: by measuring $\Delta H^\circ$ calorimetrically and by determining $\sigma$ by an independent method. Even though a 50-residue peptide was studied, its calorimetric value of $\Delta H^\circ$ could be determined only within broad limits ($0.9 < -\Delta H^\circ < 1.3$ kcal/mol res.) [26], because its thermal unfolding curve covers too wide a temperature range to allow direct fitting of the baselines. In the older literature there are good calorimetric values of $\Delta H^\circ$ for poly-L-glutamate and poly-L-lysine ($-1.1$ and $-0.9$ respectively kcal/mol res.) [27] and the agreement of these values with the result for alanine-based peptides indicates that the enthalpy change upon helix formation results from changes in the peptide backbone when the helix is formed. It is important that helix formation is an enthalpy-driven reaction and that the enthalpy change is nearly constant when different side chains are studied.

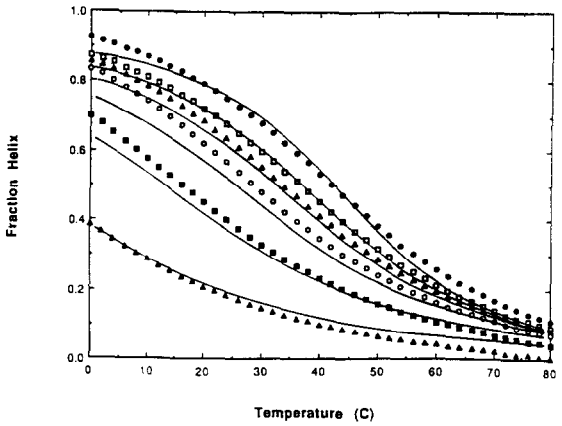


Fig. 2. Determining the parameters of the helix-coil transition by analyzing thermal unfolding curves monitored by CD for a series of peptides with chain lengths of 14 (bottom curve), 20, 26, 32, 38 and 50 (top curve) residues. The repeating sequence is AEAAKA, flanked by two marker residues: Ac–Tyr (N-terminal) and Phe–$NH_2$ (C-terminal). All unfolding curves are fitted by the same values of three adjustable parameters of helix-coil theory: the nucleation constant, the enthalpy change per residue, and the average value of the helix propagation parameter. Data from [25].

The second method used to determine $\sigma$ was hydrogen exchange: $\sigma = 0.0019$ was obtained, in reasonable agreement with the value found from thermal unfolding transitions. A similar set of alanine-based peptides was studied [28]: this set has the repeating sequence $(AAKAA)_n$ and chain lengths varying from 6 to 51 residues. The base-catalyzed exchange rate of each peptide NH proton is proportional to the fraction of the proton that is free to

exchange because it is not hydrogen-bonded in the helix. This fraction varies strongly with position in the helix because the ends of the helix are frayed. The fraction free at each position can be computed from the LR theory if the parameters of the theory are known; conversely, the helix-coil parameters can be computed from the kinetic exchange curves measured for a series of peptides with uniform sequences and varying chain lengths.

Fig. 3a shows the individual exchange curves measured by $^{15}$N-edited $^{1}$H-NMR at 7 different positions in a 21-residue peptide [29]: each curve follows a single exponential decay, and the strong dependence of exchange rate on position reflects the fraying of the helix ends. Fig. 3b shows the overall exchange curves for the series of peptides with chain lengths varying from 6 to 51 residues [28] measured by proton NMR, using the integrated area of all peptide NH resonances. The dry peptides are dissolved in $^{2}H_2O$ at zero time, in conditions (0° C, pH 1.5–4.0) where exchange is slow enough to be monitored directly by 1D NMR, even for unstructured peptides. The lines in Fig. 3b have been fitted to LR theory using just three adjustable parameters: average values for $\sigma$ and $s$ at 0° C (ZB notation for the parameters) and for the exchange rate in the coil form.

With values in hand for the average helix-coil parameters in these two series of alanine-based peptides, it is possible to determine the individual helix propensities of the 20 amino acids by substitution experiments. As it turns out, the differences in helix propensity are large enough that a single central substitution (Ala → X) in a 17-residue peptide is sufficient in many cases to determine the $s$-value of X. For those other amino acids with $s$-values close to Ala, three substitutions are made at sites spaced $i$, $i + 5$ apart, so that the substituted residues are on opposite sides of the helix and cannot interact with each other. In any case, it is desirable to make peptides in which the same substitution (Ala → X) occurs at different positions, in order to cross-check the results.

Fig. 4 shows measurements of the change in helix content when the single substitution Ala → Gly is made at several different positions in individual 17-residue peptides [30]. The helix content changes from about 75% in the host peptide (no substitution)

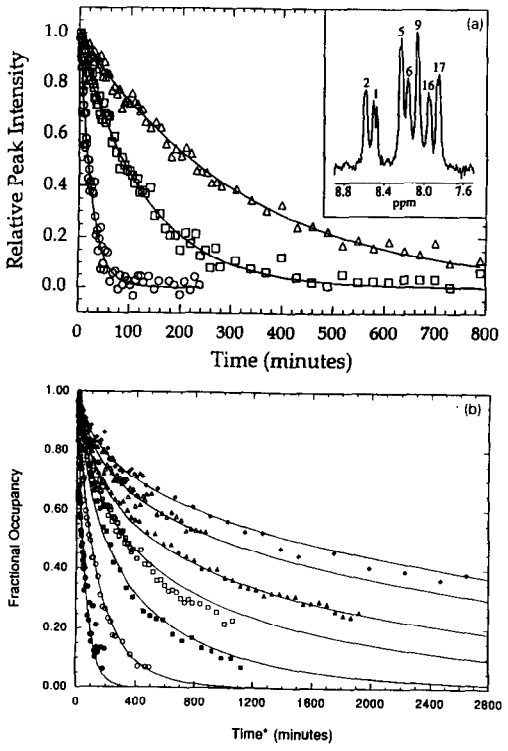


Fig. 3. Hydrogen exchange ($^{1}$H–$^{2}$H) kinetics used to determine the parameters of helix-coil transition theory. Part A tests the assumption that the rate of hydrogen exchange depends on the fraction of the peptide NH proton that is free, not hydrogen-bonded in the helix. The exchange curves of three individual NH protons are shown, at residue positions 1, 5 and 9 in a 21-residue peptide, measured by $^{15}$N-edited $^{1}$H-NMR (data from [29]). The inset figure shows the spectrum of a peptide that is $^{15}$N-labeled at 7 different residue positions. Part B shows the overall exchange curves measured for a set of peptides of chain lengths 6 (fastest exchange), 16, 21, 26, 31, 41 and 51 (slowest exchange) residues. The repeating sequence is AAKAA, flanked by an acetyl blocking group at the N-terminus and Tyr–$NH_2$ at the C-terminus. All kinetic curves are fitted by the same values of three parameters: the helix nucleation constant, the average value of the helix propagation parameter, and the average exchange rate in the coil form. Data from [28].

to about 20% when the Ala → Gly substitution is made at the central position, residue 9. The solid line shows the position dependence expected from the LR theory. Current work, which takes account of two factors not included in the original 1991 study [30] (see below), indicates that the ratio of $s$-values for Ala to Gly is about 30:1.

## 3. Inclusion of an N-cap parameter in helix-coil theory

The N-cap and C-cap residues are hydrogen-bonded in the helix, but in the Lifson–Roig theory they are defined as coil residues because their backbone $\phi$, $\varphi$, angles are not restricted to helical values. As coil residues, they are not expected to have any effect on helix stability, but experiments show otherwise. The possible significance of the N-cap and C-cap residues for helix stability in proteins was suggested in 1988 by Richardson and Richardson [31], who observed that particular amino acids occur with unusual frequencies at the N-cap and C-cap positions (notably the neutral polar amino acids Asn and Ser are found at N-cap), and by Presta and Rose [32], who proposed that side chain-main chain H-bonds can be important determinants of stability at the ends of helices, where there are 4 unsatisfied NH groups at the N-terminus and 4 unsatisfied CO groups at the C-terminus. Directed mutagenesis experiments in barnase [33] and in T4 lysozyme [34] confirmed the importance of the N-cap amino acid. The experiments of Serrano et al. [35] were particularly striking because the substitution Gly → Ala proved to be destabilizing at the N-cap position although it is strongly stabilizing at interior positions.

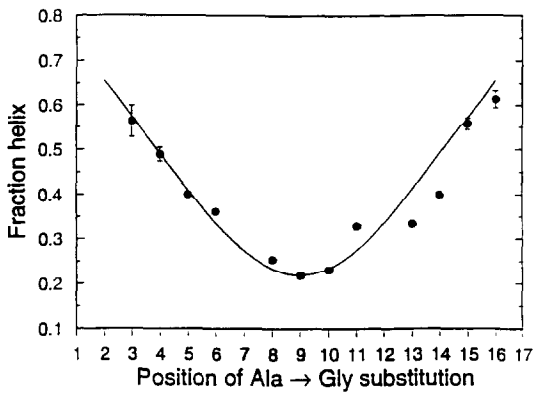


Fig. 4. Position dependence of the effect of substituting a helix-breaking amino acid. The helix contents of a series of alanine-based peptides, each containing a single Ala → Gly substitution at a different position, are shown. The helix content of the unsubstituted host peptide is 73%. The solid line is calculated from the Lifson–Roig theory, based on known values for the helix nucleation constant and average helix propagation parameter in the host peptide, and on the value determined from this experiment (see text) for the ratio of helix propensities of alanine to glycine. Data from [30].

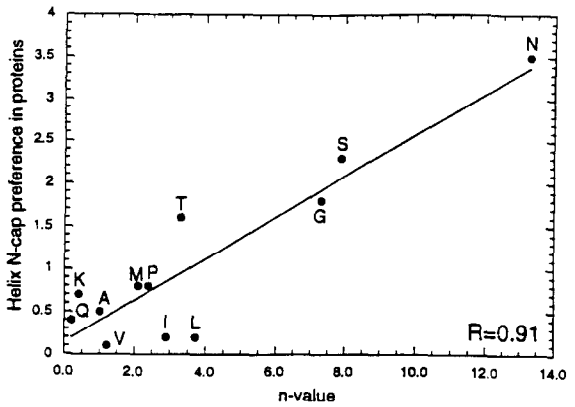


Fig. 5. Correlation of the N-cap propensity ($n$-value) determined in alanine-based peptide helices [40] with the N-cap preference (relative amino acid frequency) in protein helices [31].

Experiments from the laboratories of Gierasch [36], Kallenbach [37] and Nambiar [38] showed that N-cap substitutions can also affect the stability of peptide helices, especially when neutral polar residues such as Asn and Ser are used, that can H-bond to an unsatisfied main-chain NH group. We found next that all 20 amino acids have differing N-cap propensities in alanine-based peptides [39] (also Doig and Baldwin, unpublished results) and these N-cap propensities are closely correlated with the amino acid frequencies at the N-cap position in protein helices (shown in Fig. 5). This correlation indicates that there is a close connection between the factors affecting the stability of peptide and protein helices.

The peptide experiments show that a new parameter, the N-cap propensity, is required in the helix-coil theory to describe the stability of peptide helices in water. Fortunately, the Lifson–Roig theory can be extended straight forwardly to include the N-cap and C-cap parameters and this has been done [40]. The C-cap parameter is included for symmetry, although the choice of the C-cap amino acid has only a small effect on peptide helix stability. It is striking that the rank order of the N-cap propensities is quite different from that of the helix propensities: they are unmistakably different parameters. In the Lifson–Roig theory [21], the residues that nucleate the helix are

specified to be the two end residues, just interior to the N-cap and C-cap residues. Because the N-cap residue is next to a helix-nucleating residue, the involvement of the N-cap residue in helix stability can be interpreted as meaning that helix nucleation is more complex than originally envisaged in the Lifson–Roig theory. The N-terminal residue of a peptide is not the only position at which the helix can start: because the ends of a peptide helix are frayed, the helix can also begin at interior positions in a peptide, and so several different amino acids in a peptide can serve as the N-cap residue, with frequencies that depend on the extent of helix fraying.

## 4. Helix propensities of the 20 amino acids

The helix propensities of the 20 amino acids measured in alanine-based peptides [41] are given in Table 1: values for uncharged amino acids were determined in Ala–$Lys^+$ peptides while the ionized amino acids were measured in neutral Ala–Gln peptides. Values for the LR parameter $w$ are given as well as for the ZB parameter $s$. The two parameters are numerically similar and $s$ can be computed readily from $w$ and $v$ [24]. The Lifson–Roig theory is better suited to the analysis of helix formation by peptides than the Zimm–Bragg theory, because $w$ is associated with the amino acid residue whereas $s$ is associated with the peptide group, which is shared by two amino acid residues. Because the Lifson–Roig theory is commonly used to compute the results of experiments on peptide helices, it is likely that the Lifson–Roig parameters will be given in the future, although the custom until now has been to give the Zimm–Bragg parameters.

The $w$-values in Table 1 were computed from the modified LR theory [40], which takes explicit account of the N-cap amino acid. This was found to be important in determining accurate values of $w$ [41], as judged by comparing the predicted and observed values of the helix contents of all the peptides studied. A tyrosine residue is included at the N- or C-terminus of each peptide in order to determine an accurate value for the peptide concentration by tyrosine absorbance; this is needed to get an accurate value of the mean residue ellipticity. A glycine residue is inserted between the tyrosine and the rest of the peptide [42], in order to avoid an induced tyrosine CD band near 222 nm resulting from interaction of the tyrosine side chain with the helix.

Table 1
Helix propensity values measured at 273 K. Data from Ref. [41]

| Residue | $s$-value [a] | $w$-value [a] |
|---|---|---|
| Ala | 1.54 | 1.61 |
| $Arg^+$ | 1.04 | 1.09 |
| Leu | 0.92 | 0.96 |
| $Lys^+$ | 0.78 | 0.82 |
| Glu | 0.63 | 0.66 |
| Met | 0.60 | 0.63 |
| Gln | 0.53 | 0.56 |
| $Glu^-$ | 0.43 | 0.45 |
| Ile | 0.42 | 0.44 |
| Tyr | 0.37--0.50 [b] | 0.39--0.53 [b] |
| His | 0.36 | 0.38 |
| Ser | 0.36 | 0.38 |
| Cys | 0.33 | 0.35 |
| Asn | 0.29 | 0.31 |
| Asp | 0.29 | 0.31 |
| $Asp^-$ | 0.29 | 0.31 |
| Trp | 0.29 [b]–0.36 | 0.30 [b]–0.38 |
| Phe | 0.28 | 0.29 |
| Val | 0.22 | 0.23 |
| Thr | 0.13 | 0.14 |
| $His^+$ | 0.06 | 0.06 |
| Gly | 0.05 | 0.05 |
| Pro | ~ 0.001 | ~ 0.001 |

[a] Values obtained by applying Lifson–Roig theory modified to include either N-capping or charged group–helix macrodipole interactions. Conditions: 273 K, 1.0 M NaCl, pH 7 for uncharged residues, and 273 K, 10 mM NaCl for charged residues.
[b] Values corrected for error in fraction helix measurement caused by aromatic contribution to $[\theta]_{222}$.

A striking property of the helix propensities measured in alanine-based peptides is the wide range of values: the value for proline is so low as to be effectively off-scale, while the values for Gly and $His^+$ are about 30-fold lower than that of Ala, and values for the other amino acids are distributed fairly evenly between Ala and Gly. Only $Arg^+$, in addition to Ala, has a $w$-value above 1 and thus is helix-stabilizing. An immediate conclusion from these $w$-values is that short peptides ($< 20$ residues) are not expected to show helix formation in water unless they either have a high proportion of Ala or else are stabilized by specific side-chain interactions. Another conclusion is that helix propensities are extremely important in determining whether or not a

given sequence will form a peptide helix, because the helix content is effectively proportional to the product of the $w$-values of all amino acids at interior positions in the helix.

A main reason for determining helix propensities in alanine-based peptides is because these values are expected to be intrinsic helix-forming tendencies, unaffected by interactions with the host side chains. Experiments that have been done so far support this view: for example, when three substitutions are made, and the spacing is $i$, $i+5$ so that the substituted residues cannot interact with themselves, then the result for three substitutions is predicted accurately from that of making just one substitution [41]. The next question that arises is whether the same values for the helix propensities of the 20 amino acids are obtained by making substitutions in different reference peptide systems. We discussed this question recently [41]. Extensive results are available for two other peptide systems (the $E_4K_4$ and AEK systems) besides our own work with AK and AQ peptides. The $E_4K_4$ peptide [42] contains 8 Glu and 8 Lys residues, in blocks of 4 each, with a central triplet of 3 Ala residues, where the substitutions are made in blocks of 3 ($Ala_3 \rightarrow X_3$). The host peptide is stabilized by $Glu^- \cdot Lys^+$ ion pair interactions, and a variant of this peptide ($E_2K_2$) that contains the same number of Glu and Lys residues but with different spacings, does not form the helix [42]. The rank order of $s$-values is the same in the AK and $E_4K_4$ systems and there is a good linear correlation between the two sets of $s$-values [41] which are, however, numerically different because the correlation line does not pass through 0,0. In the $E_4K_4$ system, it is necessary to assign a value to the $\Delta G^\circ$ for the ion-pair interactions [41]. An incorrect value for $\Delta G^\circ$ might be the origin of the numerical differences between the two sets of $s$-values.

The AEK peptide system [43,44] is an alanine-based peptide but it contains 3 pairs of Glu, Lys residues spaced $i$, $i+4$, to make helix-stabilizing ion pair interactions. The $s$-values obtained with the AEK system include values for the ionizing amino acids, whereas the values obtained with the $E_4K_4$ system do not. The correlation line for the $s$-values obtained with the AEK system versus the AK, AQ system passes through 0,0 but shows appreciable scatter, especially for the ionized amino acids [41]. It seems likely that the $s$-values of the ionized amino acids obtained in the AEK system are affected by charge interactions with the $Glu^-$ and $Lys^+$ residues of the host peptide.

The overall conclusion from comparing $s$-values in these three peptide systems (work from three different laboratories) is that the rank order of the $s$-values is the same, with some scatter, in these three systems and so helix stability in all three systems is responding to differences in helix propensity in the same manner, viewed semi-quantitatively. Physicochemical factors can be identified as being possibly responsible for the quantitative differences that have been found, and it seems likely that further work will iron out these differences.

The situation is quite different when the helix propensities measured in the AK, AQ system [41] are compared with the host–guest $s$-values obtained using random sequence copolymers and HBQ or HPQ (hydroxybutyl- or hydroxypropyl-L-glutamine) as the host residue [22]. There are two main differences: (1) the rank order of helix propensities is very different in the two systems, and (2) the differences in $s$-value between other amino acids and alanine are relatively small (less than 2-fold, excluding proline) in the host–guest system, but these differences are large (as much as 30-fold) in the AK, AQ system. Recent experiments [45] with substitution of HBQ or HPQ residues in alanine-based peptides provide a plausible explanation for the differences found between the helix propensities determined in alanine-based peptides and in host–guest copolymers. These experiments indicate that the HBQ helix is stabilized in water by hydrophobic interactions involving the hydroxybutyl moieties of neighboring HBQ residues, as suggested in 1966 by Berger and coworkers [46]. As a result, substitution of a residue such as Ala in an HBQ helix affects helix stability not only by the difference between the helix propensities of Ala and HBQ but also by breaking side-chain interactions involving neighboring HBQ residues. The result is a strong context dependence of the apparent helix propensities determined using HBQ-guest copolymers.

Helix propensities alone are able to predict the location of the N-terminal helix of RNase A when the propensities determined in alanine-based peptides are used (Fig. 6). Thus, there is no need to invoke a

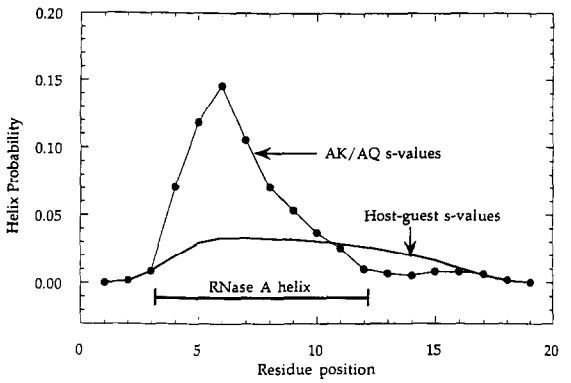


Fig. 6. Prediction of the location of the N-terminal helix of ribonuclease A, based on helix propensities determined with alanine-based peptides [41] (see Table 1) and using the modified Lifson–Roig theory [40] that includes N-capping. Unpublished data of A. Chakrabartty and R.L. Baldwin.

"helix stop" signal, other than the propensities themselves. Including the Glu 12 · Arg 10 and Phe 8 · His 12 interactions [11,12] in the prediction would increase the helix probability for those residues inside the helix.

## 5. Concluding remarks

Helix-stabilizing side-chain interactions, which were found to be important in explaining the stability of the C-peptide helix [5,6,8,11,12], can be analyzed easily and accurately in alanine-based peptides. The $(i, i+4)$ Phe · $His^+$ interaction was studied first, as a prototype [13]. The neutral alanine–glutamine host peptide [47]) which contains no ionizing groups, is ideally suited to analyzing interactions involving charged residues, either the interaction between the helix dipole and a single charged guest residue ($Asp^-$ [48], $Glu^-$ [49] and $His^+$ [50] have been studied) or the ion-pair interaction between a single pair of oppositely charged residues. A study of Glu · $Lys^+$ ion-pair interaction [49], for various spacings and both N- to C-terminal orientations of the two charged residues, showed that the ion-pair interaction is easily measurable in both orientations and for $(i, i+3)$ and $(i, i+4)$ spacings. Unexpectedly, the charged H-bond interaction ($Glu^0$ · $Lys^+$) that forms when Glu is protonated at pH 2 is nearly as strong as the ion-pair interaction at pH 7. Helix-stabilizing interactions between specific pairs of nonpolar residues are of particular interest because most protein helices have a hydrophobic face that is buried inside the protein, and so pairs of nonpolar side chains are often found in contact in these helices. Recently, a first example of this kind was reported [51]: tyrosine spaced $i$, $i+4$ from either valine or leucine makes a helix-stabilizing interaction with the other amino acid. A later study showed that limited rotational flexibility is important in producing a measurable interaction [52]: flexible, straight-chain amino acids show little or no measurable interaction with each other.

Substitution experiments in protein helices, both in barnase [33] and in T4 lysozyme [34], show qualitative agreement with the helix propensities determined in alanine-based peptides: the rank order of the helix propensities appears to be the same, when allowance is made for scatter in the protein data, resulting probably from interactions with neighboring side chains [34]. The quantitative difference in $\Delta G^\circ$ resulting from a given substitution is about 2-fold larger, however, in alanine-based peptides [41]. The origin of this effect is not known definitely, but it seems likely that burial of the helix backbone in the protein, away from water, is involved.

A basic goal of peptide helix studies is to find out if the locations of protein helices can be predicted from knowledge of helix propensities and side-chain interactions. The example of the N-terminal helix of RNase A (Fig. 6) suggests that this may be possible, but this is an isolated example, and an unusually favorable one. A goal that is closer to being achieved is the prediction of the helix contents in water of peptide sequences that correspond to helical regions in proteins. Recently such a prediction scheme was tested against a survey of all the peptide helix results in the literature, including alanine-based peptides, and was reported to show rather good agreement ($\pm 10\%$) between predicted and observed helix contents [53]. It seems likely that this goal will be reached in the near future.

## Acknowledgements

Our own work on peptide helices has been the effort of several students and postdoctoral fellows,

beginning with Andrzej Bierzynski and Peter Kim, and involving most recently Avi Chakrabartty, Andrew Doig, Beatrice Huyghes-Despointes, S. Padmanabhan, Carol Rohl and Marty Scholtz. Difficult peptides have been made by our collaborators at Denver, John Stewart and Eurice York. Valuable advice and help with helix-coil theory has been generously provided by Hong Qian and John Schellman. This research has been supported by NIH grant GM 31475.

## References

[1] J.E. Brown and W.A. Klee, Biochemistry, 10 (1971) 470.
[2] R.M. Epand and H.A. Scheraga, Biochemistry, 7 (1968) 2864.
[3] H. Taniuchi and C.B. Anfinsen, J. Biol. Chem., 244 (1969) 3864.
[4] A. Bierzynski, P.S. Kim and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 79 (1982) 2470.
[5] M. Rico, E. Gallego, J. Santoro, F.J. Bermejo, J.L. Nieto and J. Herranz, Biochem. Biophys. Res. Commun., 123 (1984) 757.
[6] K.R. Shoemaker, P.S. Kim, D.N. Brems, S. Marqusee, E.J. York, I.M. Chaiken, J.M. Stewart and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 82 (1985) 2349.
[7] W.G.J. Hol, Prog. Biophys. Mol. Biol., 45 (1985) 149.
[8] K.R. Shoemaker, P.S. Kim, E. York, J.M. Stewart and R.L. Baldwin, Nature, 326 (1987) 563.
[9] S. Ihara, T. Ooi and S. Takahashi, Biopolymers, 21 (1982) 131.
[10] S. Takahashi, E.-H. Kim, T. Hibino and T. Ooi, Biopolymers, 28 (1989) 995.
[11] R. Fairman, K.R. Shoemaker, E.J. York, J.M. Stewart and R.L. Baldwin, Biophys. Chem., 37 (1990) 107.
[12] K.R. Shoemaker, R. Fairman, D.A. Schultz, A.D. Robertson, E.J. York, J.M. Stewart and R.L. Baldwin, Biopolymers, 29 (1990) 1.
[13] K.M. Armstrong, R. Fairman and R.L. Baldwin, J. Mol. Biol., 230 (1993) 284.
[14] M. Levitt and M.F. Perutz, J. Mol. Biol., 201 (1988) 751.
[15] A. Wlodawer, L.A. Svensson, L. Sjölin and G.L. Gilliland, Biochemistry, 27 (1988) 2705.
[16] K.R. Shoemaker, R. Fairman, P.S. Kim, E.J. York, J.M. Stewart and R.L. Baldwin, Cold Spring Harbor Symp. Quant. Biol., 52 (1987) 391.
[17] M. Rico, J.L. Nieto, J. Santoro, F.J. Bermejo, J. Herranz and E. Gallego, FEBS Lett., 162 (1983) 314.
[18] P.S. Kim and R.L. Baldwin, Nature, 307 (1984) 329.
[19] S. Marqusee, V.H. Robbins and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 86 (1989) 5286.
[20] B.H. Zimm and J.K. Bragg, J. Chem. Phys., 31 (1959) 526.
[21] S. Lifson and A. Roig, J. Chem. Phys., 34 (1961) 1963.
[22] J. Wójcik, K.-H. Altmann and H.A. Scheraga, Biopolymers, 30 (1990) 121.
[23] B.H. Zimm, P. Doty and K. Iso, Proc. Natl. Acad. Sci. USA, 45 (1959) 1601.
[24] H. Qian and J.A. Schellman, J. Phys. Chem., 96 (1992) 3987.
[25] J.M. Scholtz, H. Qian, E.J. York, J.M. Stewart and R.L. Baldwin, Biopolymers, 31 (1991) 1463.
[26] J.M. Scholtz, S. Marqusee, R.L. Baldwin, E.J. York, J.M. Stewart, M. Santoro and D.W. Bolen, Proc. Natl. Acad. Sci. USA, 88 (1991) 2854.
[27] J. Hermans, J. Phys. Chem., 70 (1966) 510.
[28] C.A. Rohl, J.M. Scholtz, E.J. York, J.M. Stewart and R.L. Baldwin, Biochemistry, 31 (1992) 1263.
[29] C.A. Rohl and R.L. Baldwin, Biochemistry, 33 (1994) 7760.
[30] A. Chakrabartty, J.A. Schellman and R.L. Baldwin, Nature, 351 (1991) 586.
[31] J.S. Richardson and D.C. Richardson, Science, 240 (1988) 1648.
[32] L.G. Presta and G.D. Rose, Science, 240 (1988) 1632.
[33] A. Horovitz, J.M. Matthews and A.R. Fersht, J. Mol. Biol., 227 (1992) 560.
[34] M. Blaber, X.J. Zhang and B.W. Matthews, Science, 260 (1993) 1637.
[35] L. Serrano, J.L. Neira, J. Sancho and A.R. Fersht, Nature, 356 (1992) 453.
[36] M.D. Bruch, M.M. Dhingra and L.M. Gierasch, Proteins Struct. Funct. Genet., 10 (1991) 130.
[37] P.J. Gans, P.C. Lyu, M.C. Manning, R.W. Woody and N.R. Kallenbach, Biopolymers, 31 (1991) 1605.
[38] B. Forood, E.J. Feliciano and K.P. Nambiar, Proc. Natl. Acad. Sci. USA, 90 (1993) 838.
[39] A. Chakrabartty, A.J. Doig and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 90 (1993) 11332.
[40] A.J. Doig, A. Chakrabartty, T.M. Klingler and R.L. Baldwin, Biochemistry, 33 (1994) 3396.
[41] A. Chakrabartty, T. Kortemme and R.L. Baldwin, Protein Sci., 3 (1994) 843.
[42] A. Chakrabartty, T. Kortemme, S. Padmanabhan and R.L. Baldwin, Biochemistry, 32 (1993) 5560.
[43] S.H. Park, W. Shalongo and E. Stellwagen, Biochemistry, 32 (1993) 7048.
[44] S.H. Park, W. Shalongo and E. Stellwagen, Biochemistry, 32 (1993) 12901.
[45] S. Padmanabhan, E.J. York, L. Gera, J.M. Stewart and R.L. Baldwin, Biochemistry, 33 (1994) 8604.
[46] N. Lotan, A. Yaron and A. Berger, Biopolymers 4 (1966) 365.
[47] J.M. Scholtz, E.J. York, J.M. Stewart and R.L. Baldwin, J. Am. Chem. Soc., 113 (1991) 5102.
[48] B.M.P. Huyghues-Despointes, J.M. Scholtz and R.L. Baldwin, Protein Sci., 2 (1993) 1604.
[49] J.M. Scholtz, H. Qian, V.H. Robbins and R.L. Baldwin, Biochemistry, 32 (1993) 9668.
[50] K.M. Armstrong and R.L. Baldwin, Proc. Natl. Acad. Sci. USA, 90 (1993) 11337.
[51] S. Padmanabhan and R.L. Baldwin, J. Mol. Biol., 241 (1994) 706.
[52] S. Padmanabhan and R.L. Baldwin, Protein Sci., 3 (1994) 1992.
[53] V. Muñoz and L. Serrano, Nature Struct. Biol., 1 (1994) 399.