Determination of Free Energies of N-Capping in α-Helices by Modification of the Lifson–Roig Helix–Coil Theory To Include N- and C-Capping†

Andrew J. Doig,* Avijit Chakrabartty,* Tod M. Klingler, and Robert L. Baldwin

Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305

Received November 8, 1993; Revised Manuscript Received January 11, 1994®

ABSTRACT: We have previously shown that varying the N-terminal amino acid in α-helical peptides can cause large variations in helix content (Chakrabartty et al., 1993a). The Lifson–Roig theory for the helix–coil transition predicts, however, that substitutions at the N-terminus in an unacetylated peptide should have no effect on α-helix stability. We have therefore modified the theory to include these N-capping effects by assigning a statistical weight (the "*n*-value") to the amino acid immediately preceding a stretch of helical residues. The *n*-value measures the N-capping propensity of an amino acid, and like the helix propensity (*w*-value), it is independent of neighboring residues or positions in sequence. The new theory was used, with the experimental data for these substitutions, to calculate *n*-values and, hence, free energies for N-capping for the amino acids Gln, Ala, Val, Met, Pro, Ile, Leu, Thr, Gly, Ser, and Asn as well as for the acetyl group, which is commonly used to cap peptides. The free energies vary by ≈2 kcal mol^{-1} from Gln (worst) to Asn (best), and the acetyl group is nearly as effective as Asn. N-Capping free energies were also found for Leu, Thr, Gly, Ser, and Asn when the N-terminus is charged at pH 5. The unfavorable effect of protonation of the N-terminus in an α-helix was found to be ≈0.5 kcal mol^{-1}. Our results agree well with a survey of N-capping preferences from protein crystal structures and are compared to results from site-directed mutagenesis of N-caps in proteins.

Theory for the transition between α-helix and random-coil states in polypeptides was developed in the 1950's and 1960's and has proved successful in dealing with experimental studies of the helix–coil transition until the present time. The two main approaches are those of Lifson and Roig (1961) and Zimm and Bragg (1959), whose differences originate in the way statistical weights are assigned to different residue conformations. Helix–coil theories have been reviewed by Poland and Scheraga (1970) and Qian and Schellman (1992). It has been known for some time, however, that helix–coil theories ignore certain aspects of isolated helix formation, such as charged group–helix macrodipole interactions (Shoemaker et al., 1985, 1987) and side chain–side chain interactions (Marqusee et al., 1987). Several modified helix–coil theories which consider these types of interactions have been developed (Vásquez et al., 1987, 1988; Gans et al., 1991; Scholtz et al., 1993). Another important aspect of isolated helix formation is the phenomenon of N-capping (Chakrabartty et al., 1993a; Forood et al., 1993; Lyu et al., 1993). Here, we describe the Lifson–Roig model in detail and the Zimm–Bragg model briefly and argue that both are unsatisfactory in attempting to account for the experimental observations of N-capping described in the previous paper (Chakrabartty et al., 1993a). We show how Lifson–Roig theory can be modified to include capping interactions. We then use the modified theory to obtain capping free energies from the experimental data, and in so doing, we are able to compare, quantitatively, the energetic contributions of helix capping in peptides with those in proteins.

Helix content in peptides can be quantified by the circular dichroism (CD)[1] signal at 222 nm. If the sequence of the peptide is altered, resulting in a change in CD signal, then it is necessary to use a helix–coil theory to convert this to a free-energy change. The CD signal on its own can only say whether the change is stabilizing or destabilizing to the α-helix.

RESULTS AND DISCUSSION

Lifson–Roig Theory

In Lifson–Roig theory, the number of residues is defined as the number of α-carbons which are flanked by peptide (CONH) units on both sides. Thus, in the peptide NH_3^+-ABCDEF-COO^-, only B, C, D, and E count as residues. If the N-terminus is acetylated and if the C-terminus is replaced by an amide group, A and F therefore also count as residues. When the N-terminus is an amino group, the first amino acid A is not considered in the theory and can make no contribution to the partition function. Lifson–Roig theory therefore predicts that modification of the first amino acid will have no effect on helix content. The experiments in the previous paper (Chakrabartty et al., 1993a) show that this is false. Large differences in helix content are found when the first amino acid is altered in an unacetylated peptide that forms a helix of marginal stability. The Lifson–Roig theory is incapable of accounting for this phenomenon, and a modified theory is thus necessary.

In Lifson–Roig theory, each residue in a polypeptide sequence is considered to exist in one of two states, helix or coil. The helix state (h-state) is defined by having its ϕ and ψ dihedral angles restricted to an α-helical geometry (centered around $\phi = -57°$ and $\psi = -47°$). The coil state (c-state) is defined as the remainder. A single polypeptide conformation can thus be described by a sequence made up of h- and c-states,

† This work was supported by NIH Grant GM 31475 and a CAMIS grant from the National Library of Medicine, LM05305. A.J.D. is a recipient of a Science and Engineering Research Council NATO Postdoctoral Fellowship. A.C. is a recipient of a Medical Research Council of Canada Postdoctoral Fellowship.

* Authors to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* February 15, 1994.

[1] Abbreviations: CD, circular dichroism; $\Delta\Delta G$, change in the free energy of protein folding.

0006-2960/94/0433-3396$04.50/0

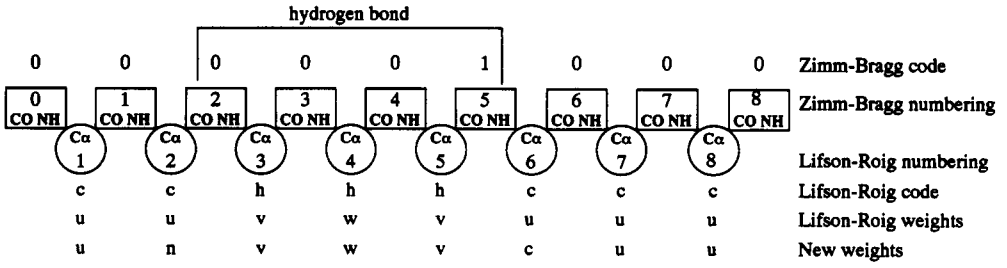

FIGURE 1: Codes, numberings, and statistical weights in a partially helical polypeptide for the Zimm–Bragg, Lifson–Roig, and modified Lifson–Roig models. (See, also, Qian & Schellman, 1992).

for example, ccccchhhhhccchhcc. All sequences are listed from the N- to C-terminus. Lifson–Roig theory gives each residue in a sequence a statistical weighting based upon its own state and the states of the two residues to either side (i.e., a triplet). These weightings are defined as: u, coil state; v, helix state adjacent to one or two coil states (or one or no helix states); and w, helix state flanked on both sides by two helix states.

Each run of h-states in a conformation is thus bounded at either end by a single residue with a v-weighting. The conformation

c c c c c h h h h h c c c h h c c

thus has weightings of

$u\ u\ u\ u\ u\ v\ w\ w\ w\ v\ u\ u\ u\ u\ v\ v\ u\ u$

Helical states are partitioned in this way into v- and w-weightings because each w-weighting is associated with a hydrogen bond in the helix while a v-weighting is not (Figure 1). Thus, a run of N consecutive h-states will have a weighting of v^2w^{N-2} ($N > 2$) and will contain $N - 2$ helical hydrogen bonds (Figure 1). In other words, the cooperative nature of helix formation results from the fact that it is necessary to orient three successive residues so that a single hydrogen bond can form. This unfavorable event is accounted for by a v-value much less than unity. The c-state is chosen as the reference state, and thus, u-weightings are set to 1. The Lifson–Roig definitions can be rewritten as a matrix (**m**) as follows:

$$\mathbf{m} = \begin{array}{c} \\ \bar{\mathrm{h}}\bar{\mathrm{h}} \\ \bar{\mathrm{h}}\bar{\mathrm{c}} \\ \bar{\mathrm{c}}\bar{\mathrm{h}} \\ \bar{\mathrm{c}}\bar{\mathrm{c}} \end{array} \begin{array}{c} \begin{array}{cccc} \bar{\mathrm{h}}\bar{\mathrm{h}} & \bar{\mathrm{h}}\bar{\mathrm{c}} & \bar{\mathrm{c}}\bar{\mathrm{h}} & \bar{\mathrm{c}}\bar{\mathrm{c}} \end{array} \\ \begin{pmatrix} w & v & 0 & 0 \\ 0 & 0 & 1 & 1 \\ v & v & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

In each triplet, the state of the leftmost residue is shown at the start of each row. The state of the rightmost residue in each triplet is shown at the end of the top of each column. The state of the center residue of each triplet is shown as a $\bar{\mathrm{c}}$ or $\bar{\mathrm{h}}$, both in the rows and columns. When the states of the center residue of the triplet in the rows and columns differ, the weighting of that triplet must be zero. The Lifson–Roig matrix can be used to generate a partition function (Z) for a polypeptide of N residues with eq 1:

$$Z = (0\ \ 0\ \ 1\ \ 1)\mathbf{m}^N \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad (1)$$

The end vectors ensure that the first and last residues cannot have a w-weighting. If an end residue is in an h-state, it can only have a v-weighting; an end residue in a c-state will have a weighting of 1.

Zimm–Bragg Theory

Before showing how the Lifson–Roig theory may be modified to include capping effects, we first wish to discuss the alternate helix–coil theory, namely Zimm–Bragg (1959). The differences in the two theories originate in the manner in which units in the polypeptide chain are classified. The units in Zimm–Bragg theory are peptide groups, rather than residues, and they are classified on the basis of whether their NH groups participate in hydrogen bonds within the helix, rather than whether the residue occupies helical ϕ,ψ space. The Zimm–Bragg coding is shown in Figure 1. A unit is given a code of 1 (e.g., peptide unit 5 in Figure 1) if its NH group forms a hydrogen bond and a code of 0 otherwise. The first hydrogen-bonded unit proceeding from the N-terminus has a statistical weight of σs, successive hydrogen-bonded units have weights of s, and non-hydrogen-bonded units have weights of 1. The s-value is thus a propagation parameter, analogous to w in Lifson–Roig, and σ is an initiation parameter, analogous to v^2 (but not identical; see Qian & Schellman, 1992).

The σ-value physically means the probability of three successive residues being restricted into an α-helical geometry with no intrahelix hydrogen-bond formation (Figure 1). An important conceptual problem with the Zimm–Bragg theory is that σ is not a property of a single peptide group or residue. The value of the σ-parameter depends on several residues. For the helix shown in Figure 1, these include residues 3–5. The assignment of a σ-value to a particular amino acid is not possible in a heteropolymer as the unit for the sequence is the peptide group, not the residue. Similarly, the elongation parameter s is associated with a peptide group, not a specific residue. If a single residue substitution is made in the middle of a peptide sequence, it is not clear whether this leads to a change in the s-value of the preceding residue, the following residue, or both. The Lifson–Roig model treats such a substitution straightforwardly as simply a change in the parameters of that residue. When dealing with heteropolymers, it is therefore easier to use, and modify, Lifson–Roig rather than Zimm–Bragg.

Incorporation of Capping Effects into Lifson–Roig Theory

The definitions specifying which residues contribute to the partition function must be altered from the Lifson–Roig definitions as we have shown that the first residue in a peptide with a free N-terminus must be included. In the modified theory, every amino acid residue is counted and makes a contribution to the partition function. The peptide NH_3^+-

ABCDEF-COO$^-$ therefore has all six amino acid residues included in the partition function, whereas Lifson–Roig theory would include only the central four.

Acetylation of the N-terminus and amidation of the C-terminus both produce peptides which could have an extra hydrogen bond in the helix and so both increase the number of units in the parition function by one. The peptide Ac-ABCDEF-CONH$_2$ is therefore considered to have eight units, six for the amino acids A–F and one each for the acetyl and amide groups. The term residue is used in the modified Lifson–Roig theory to refer to any amino acid residue, whether or not it has flanking peptide groups on both sides. Weights are assigned to each of these groups. In Lifson–Roig theory, this peptide would have six units, namely the amino acid residues A–F.

N- and C-capping effects can be included in the Lifson–Roig theory by modifying the triplet definitions given above. The N-cap residue is the c-state immediately preceding an h-state (i.e., on the N-terminal side) and is given a weighting of n. The C-cap residue is the c-state immediately after an h-state and is given a weighting of c. A c-state between two h-states is assumed, for the present, to have the geometric mean of the n- and c-weightings, namely $\sqrt{nc}$. (This may not be accurate, but experimental measurement of the weighting of the hch-state is probably not possible, as it will be highly improbable. The introduction of a single c-state into the middle of a stretch of h-states has a probability of occurrence of $v^2\sqrt{nc}/w^3$, relative to maintaining all h-states. As $n \approx c \approx w \approx 1$ and $v \approx 0.05$, $v^2\sqrt{nc}/w^3 \approx 0.0025$, so this is improbable. Physically, this corresponds to a single residue in the middle of a helix not forming a helical hydrogen bond, thus introducing a kink in the helix. This has been observed with prolines (Barlow & Thornton, 1988), so Pro may have a relatively high weighting when it is the center residue in an hch-triplet. It may also occur to a significant extent with Gly as it has a low w-value.) Other weightings are unchanged. The weightings of center residues of the eight possible triplet conformations are now:

triplet state	statistical weight of center residue
hhh	w
hhc	v
hch	$\sqrt{nc}$
hcc	c
chh	v
chc	v
cch	n
ccc	1

Note that if $n = 1$ and $c = 1$, the modified theory reduces to the Lifson–Roig theory, where all c-states have the same weightings. The consequences of giving different weightings to center residues in chh-, hhc-, and chc-conformations and to breaks between two helical segments have been examined by Bixon and Lifson (1967). These modified definitions give the new matrix **l**:

$$\mathbf{l} = \begin{array}{c} \\ \bar{\text{h}}\text{h} \\ \text{h}\bar{\text{c}} \\ \bar{\text{c}}\text{h} \\ \bar{\text{c}}\text{c} \end{array} \begin{array}{c} \begin{array}{cccc} \bar{\text{h}}\text{h} & \bar{\text{h}}\text{c} & \bar{\text{c}}\text{h} & \bar{\text{c}}\text{c} \end{array} \\ \begin{pmatrix} w & v & 0 & 0 \\ 0 & 0 & \sqrt{nc} & c \\ v & v & 0 & 0 \\ 0 & 0 & n & 1 \end{pmatrix} \end{array} \tag{2}$$

This matrix is not singular and cannot be reduced to a 3 × 3 matrix (unlike the Lifson–Roig matrix). The partition function Z for an N-residue sequence is

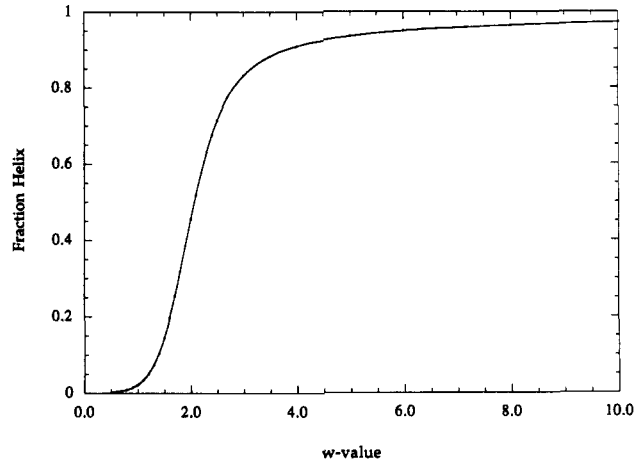

FIGURE 2: Variation of fraction helix with w-value in a 12-residue homopolymer. $n = 1$, $c = 1$, $v = 0.048$.

$$Z = (0 \quad 0 \quad 0 \quad 1)\mathbf{l}^N \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3}$$

The initial vector ensures that the N-terminal residue can only take the weightings n (if it precedes an h-state) or 1 (if it precedes a c-state). Similarly, the end vector allows weightings of only c or 1 for the final residue. The end units cannot be in an h-state as this requires both ϕ and ψ dihedral angles to be restricted into an α-helical geometry. An N-terminal amino acid with a free N-terminus does not have ϕ restricted (its Cα-NH$_3^+$ bond); similarly, a C-terminal amino acid does not have ψ restricted (its Cα-COO$^-$ bond). Terminal acetyl and amide groups do not have ϕ or ψ dihedral angles, as the required atoms do not exist, so they cannot be in an h-state either. The modified theory has increased the number of units in the partition function by two compared to Lifson–Roig theory; however, these additional units, one at each end, can only be in a c-state. The maximum number of h-states (and hence the maximum number of helical hydrogen bonds) is therefore unchanged, as it should be.

From the partition function, many average properties of the system can be calculated. Perhaps the most important of these is the average number of helical hydrogen bonds/molecule which is given by:

$$\langle n_h \rangle = \frac{\partial \ln Z}{\partial \ln w} \tag{4}$$

$\langle n_h \rangle / N_h$, where N_h is the maximum possible number of hydrogen bonds in a helix, is often taken to be proportional to the fractional helicity (or average helix content) of a peptide, which is also proportional to the CD signal at 222 nm. Residues with a v-weighting can also be classed as helical. Calculations from the helix–coil theory essentially all, therefore, involve comparing the calculated fraction helix with the CD signal. Other properties which can be calculated from Z include the mean number of residues with v-, c-, n-, and u-weightings (though these are not directly experimentally measurable by CD) and the average length of helices.

Figures 2, 3, and 4 show calculations from the modified Lifson–Roig theory of how helix content varies with changing one of w, v, n, or c in a 12-residue peptide. The parameters remaining constant have the values $w = 1.4$, $v = 0.048$, $n =$

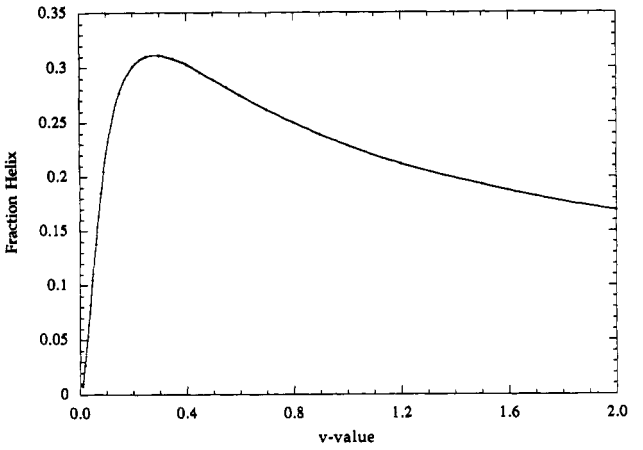

FIGURE 3: Variation of fraction helix with v-value in a 12-residue homopolymer. $n = 1$, $c = 1$, $w = 1.4$.

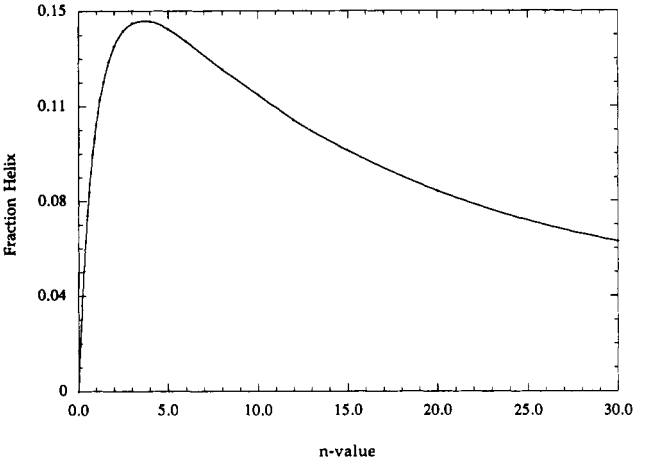

FIGURE 4: Variation of fraction helix with n-value in 12-residue homopolymer. $c = 1$, $w = 1.4$, $v = 0.048$.

1, and $c = 1$. The plots for varying n and c are identical due to symmetry. The plots where w and v are varied both have $n = 1$ and $c = 1$; they therefore give identical results to the Lifson–Roig theory. In Figure 2, increasing w always results in an increase of helix content from 0% to 100%. When v, n, or c is varied, however, helix content increases from zero to a maximum and then decreases. This is because these three parameters relate to the probability of occurrence of the end of a helical sequence. When they are very low, the most likely conformation is all coil, and increasing the favorability of end formation means that a single helical segment becomes more likely. When v, n, or c is large, the most common conformations have more than one helical sequence. An increase of the probability of end formation results in many short helical segments with fewer residues having a w-weighting and thus lower helix content.

The Lifson–Roig theory is highly suitable for further modification, should this become necessary in the future. For example, additional experimental evidence may demonstrate that the center residues of the triplets chh, chc, and hhc, which at present all the have the same weighting of v, may in fact differ in energy and require separate parameters. Changes such as this can be readily incorporated into the theory, though we have not found it necessary.

Heteropolymers

The partition function for a heteropolymer can be calculated provided that w, v, n, and c are known for each residue in the sequence. Each residue provides a matrix (**l**) with its own parameters. These are multiplied in order with eq 5 to give Z.

$$Z = (0\ \ 0\ \ 0\ \ 1)\prod_{j=1}^{j=N}\mathbf{l}_j\begin{pmatrix}0\\0\\0\\1\end{pmatrix} \tag{5}$$

For a particular amino acid (unit i), the probability that it has a w-weighting (and hence a hydrogen bond) in a sequence of N units is given by:

$$\frac{\partial \ln Z}{\partial \ln w_i} = \frac{1}{Z}(0\ \ 0\ \ 0\ \ 1)$$

$$\left(\prod_{j=1}^{j=i-1}\begin{pmatrix} w_i & v_j & 0 & 0\\ 0 & 0 & \sqrt{n_j c_j} & c_j\\ v_j & v_j & 0 & 0\\ 0 & 0 & n_j & 1\end{pmatrix}\right)\begin{pmatrix} w_j & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

$$\left(\prod_{j=i+1}^{j=N}\begin{pmatrix} w_j & v_j & 0 & 0\\ 0 & 0 & \sqrt{n_j c_j} & c_j\\ v_j & v_j & 0 & 0\\ 0 & 0 & n_j & 1\end{pmatrix}\right)\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

By summing these probabilities, the helical content of a heteropolymer can be found. The probability that a particular unit (i) has a weighting of v is given by:

$$\frac{\partial \ln Z}{\partial \ln v_i} = \frac{1}{Z}(0\ \ 0\ \ 0\ \ 1)$$

$$\left(\prod_{j=1}^{j=i-1}\begin{pmatrix} w_i & v_j & 0 & 0\\ 0 & 0 & \sqrt{n_j c_j} & c_j\\ v_j & v_j & 0 & 0\\ 0 & 0 & n_j & 1\end{pmatrix}\right)\begin{pmatrix} 0 & v_i & 0 & 0\\ 0 & 0 & 0 & 0\\ v_i & v_i & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

$$\left(\prod_{j=i+1}^{j=N}\begin{pmatrix} w_j & v_j & 0 & 0\\ 0 & 0 & \sqrt{n_j c_j} & c_j\\ v_j & v_j & 0 & 0\\ 0 & 0 & n_j & 1\end{pmatrix}\right)\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

The probability that a particular unit has a weighting of n or c can be found similarly. These equations have been incorporated into a computer program. The program inputs an amino acid sequence and a list of w-, v-, n-, and c-values for each amino acid and outputs the probability of each amino acid being in a particular conformation. The mean fraction helix is found from the sum of the individual residue probabilities of having a w- or v-weighting. This program, written in C or FORTRAN, is available from us by anonymous FTP from cmgm.stanford.edu.

Application of the Modified Lifson–Roig Theory to Estimate N-Capping Free Energies

We showed previously that variation of the N-terminal residue in unacetylated 12- and 17-residue peptides can have large effects on helix content (Chakrabartty et al., 1993a). These effects cannot be accounted for by conventional Lifson–Roig theory. The theory outlined above is adequate, however.

Table 1: Sequences and Helix Contents of Peptide in Data Set

sequence	fraction helix: neutral N-terminus	fraction helix: charged N-terminus
NH_2-AAKAAAAKAAAAKAAGY-$CONH_2$	0.378[a]	0.232[c]
NH_2-GAKAAAAKAAAAKAAGY-$CONH_2$	0.525[a]	0.422[c]
Ac-YGAAKAAAAKAAAAKAA-COOH	0.490[a]	
Ac-YGAAKAAAAKAAAAKAG-COOH	0.481[a]	
Ac-YGGKAAAAKAAAAKAAAAK-$CONH_2$	0.677[b]	
Ac-YGGKAAGAKAAAAKAAAAK-$CONH_2$	0.373[b]	
Ac-YGGKAAAAKAAGAKAAAAK-$CONH_2$	0.256[b]	
Ac-YGGKAAAAKAAAAKAAGAK-$CONH_2$	0.465[b]	
NH_2-IAKAAAKAAAAKAAGY-$CONH_2$	0.419[a]	0.308[c]
NH_2-LAKAAAAKAAAAKAAGY-$CONH_2$	0.449[a]	0.317[c]
NH_2-SAKAAAAKAAAAKAAGY-$CONH_2$	0.525[a]	0.384[c]
NH_2-TAKAAAAKAAAAKAAGY-$CONH_2$	0.457[a]	0.358[c]
NH_2-MAKAAAAKAAAAKAAGY-$CONH_2$	0.402[a]	0.299[c]
NH_2-VAKAAAAKAAAAKAAGY-$CONH_2$	0.375[a]	0.265[c]
NH_2-NAKAAAAKAAAAKAAGY-$CONH_2$	0.595[a]	0.501[c]
NH_2-QAKAAAAKAAAAKAAGY-$CONH_2$	0.293[a]	0.205[c]
NH_2-PAKAAAAKAAAAKAAGY-$CONH_2$	0.404[a]	0.234[c]
Ac-AKAAAAKAAAAKAAGY-$CONH_2$	0.578	

[a] 1 M NaCl, pH 9.55, 273 K (Chakrabartty et al., 1993b). [b] 1 M NaCl, pH 7, 273 K (Chakrabartty et al., 1993a). [c] 1 M NaCl, pH 5, 273 K (Chakrabartty et al., 1993b).

We have no experimental evidence for differences in C-capping preferences, though statistical preferences in protein structures do show significant variation (Richardson & Richardson, 1988), but have included C-capping in the theory for completeness. We expect that variations in helix content due to C-capping will be observed in peptides in due course and that the theory described here will be appropriate for their analysis.

n-Values for the 11 uncharged, nonaromatic amino acids can be determined using the measured helix contents of the peptides having the sequence H_3N^+-XAKAAAAKAAAA-KAAGY-$CONH_2$ (Chakrabartty et al., 1993a), where X is the amino acid being varied. These peptides have 18 units in the modified Lifson–Roig theory, as they have 17 amino acids and an amide group at the C-terminus. We have additionally synthesized and measured the helix content of the peptide Ac-AKAAAAKAAAAKAAGY-$CONH_2$, as before (Chakrabartty et al., 1993a). This peptide also has 18 units in the modified theory, so that the *n*-value of the acetyl group can be determined. It is also necessary to know the *w*-values of Ala, Lys, and Gly and the *n*-values of Lys and Gly if these sequences are used. It has recently been demonstrated that if aromatic residues are included in a peptide for concentration determination, they should be placed at the end of the sequence after a helix-breaking Gly spacer to ensure that the aromatic ring does not affect the CD signal (Chakrabartty et al., 1993b). We have therefore used data only from peptides with this Gly-Tyr terminal sequence, as they are more reliable. The complete set of peptide sequences and helix contents used are listed in Table 1.

With the available information, it is not possible to obtain absolute values of *n* and *c* for each residue. This is because every run of consecutive helical states is bounded by both an *n*- and *c*-weighting. Thus, for each conformation, the number of residues with an *n*-weighting is equal to the number with a *c*-weighting. It is not, therefore, possible with these data to measure *n* independently of *c*. We therefore define the neutral alanine residue (i.e., Ala at the uncharged N-terminus or at a nonterminal position) to have *n*- and *c*-weightings of 1. Alanine at a charged N-terminus will have a different *n*-value (as will other residues). The *n*- and *c*-weightings of other residues can be calculated relative to neutral Ala. Free-energy differences for N-capping of a particular amino acid, relative to that of alanine, can be calculated exactly from the relative *n*-weightings. For example, $-RT\ln(n(\text{Val})$ gives the free-energy change of a Ala → Val substitution at an N-cap, and if the substitution Gln → Val is made at an N-cap, the free-energy change is $-RT\ln(n(\text{Val})/n(\text{Gln}))$. In the future, it may be possible to obtain absolute values of *n* and *c* by comparing the degrees of fraying at the N- and C-termini. The required experimental data are not available at present, however.

Table 2: *n*-Values at pH 9.55

residue	*n*-value	$-RT\ln(n)$ (kcal mol^{-1})
Gln	<0[a]	
Ala	1	0
Val	0.77	0.14
Met	1.22	–0.11
Pro	1.25	–0.12
Ile	1.54	–0.23
Leu	2.25	–0.44
Thr	2.44	–0.48
Gly	6.22	–0.99
Ser	4.97	–0.87
Asn	10.25	–1.26
acetyl	8.94	–1.19

[a] The program fits to a negative *n*-value for Gln, which is not possible. A *n*-weighting for Gln at pH 9.55 could not therefore be obtained.

The *n*-values of the uncharged, nonaromatic residues and the *w*-values of Ala and Gly were found using the data in Table 1 and a computer program which found the least-squares best fit to the experimental helix contents by floating these parameters. We do not have sufficient data to independently calculate the *w*- and *n*-values of Ala and Lys; we therefore found a weighted mean value of *w*(Ala/Lys) ($\langle w \rangle$) and fixed *n*(Ala/Lys) to 1. The fitting program calls on the modified helix–coil theory program to calculate the fraction helix of a peptide from the current parameter set of *w*- and *n*-values. The *v*-value of every residue was fixed to 0.048 (Rohl et al., 1992) and not floated. The *c*-values of every residue were fixed at 1. The *n*-values obtained by this procedure for the neutral amino acids are given in Table 2. $\langle w \rangle$ was found to be 1.42, and *w*(Gly) was 0.037, in good agreement with previous results (Chakrabartty et al., 1991, 1993b). If the *n*-values do not vary with temperature, N-capping free energies can be found using $-RT\ln(n)$ at any temperature. This will be precise if *n* is entirely entropic and there is no change in the change in heat capacity upon unfolding (which is unlikely). A rationalization for the rank order of N-cap preferences we found was given in the previous paper (Chakrabartty et al., 1993a), with the exception of the acetyl group. The acetyl group has a high *n*-value as it does not have a NH_2 group which presumably interacts unfavorably with nearby NH groups near the N-terminus and with the positively charged helix dipole (Hol et al., 1978).

When the N-terminus is charged (NH_3^+), the helix content drops (Chakrabartty et al., 1993a), presumably because of an unfavorable interaction between the NH_3^+ group and the helix dipole, which is positive at the N-terminal end. *n*-Values of charged residues are therefore lower. We found the *n*-values of residues when they are at a charged N-terminus using the peptides listed in Table 1 and the helix contents at pH 5. The *w*-values of Ala and Gly were fixed to those found previously, and the *n*-values were found relative to the *n*-value of neutral Ala (defined to be 1). The following *n*-values for the amino acids with charged N-termini were found, Leu (0.03), Thr (0.53), Gly (1.62), Ser (0.92), and Asn (3.85). Negative *n*-values were found for Gln, Ala, Val, Met, Pro, and Ile as well as for neutral Gln. Negative *n*-values may result from

experimental error in the measurement of helix contents; however, a more likely explanation is that the true *n*-value of Ala is greater than 1. If the *n*-value of Ala was higher, a greater range of *n*-values smaller than that of neutral Ala would be possible. We are, however, unable to determine the absolute *n*-value of Ala at present. In the future, if this information is available, possibly by studying asymmetry in the fraction helix distribution along a peptide, it may be possible to determine the complete range of absolute *n*-values. There may also be inaccuracies in *w*- and *v*-values which lead to negative *n*-values.

There is growing evidence that the distribution of mean helicity in peptide α-helices is asymmetric, with residues toward the N-terminus having a greater mean helicity than those toward the C-terminus (Lyu et al., 1990; Miick et al., 1993; C. A. Rohl & R. L. Baldwin, unpublished). Asymmetry can readily be explained by residues toward the N-terminus of the peptides having greater *n*-values than the *c*-values of residues toward the C-terminus. This can easily be accomplished, for example, by acetylation.

The *n*-values found where the N-terminus is uncharged (at pH 9.55) are expected to be more relevant to protein stability than those found when the N-terminus has a positive charge. This is because an N-cap residue in a protein will have its NH group as part of a neutral amide group (unless the N-cap residue is the first residue in the protein sequence). For this reason, our uncharged N-capping free energies (pH 9.55) are compared to those in other systems.

Free-Energy Change for Protonation of N-terminus of α-Helix

Positive *n*-values were found for charged Leu, Thr, Ser, Asn, and Gly. They correlate well with those found at pH 9.55 ($R = 0.85$; plot not shown). These can be converted to free energies of N-capping ($-RT\ln(n)$, where $T = 273$ K) and can thus be compared to *n*-values for the amino acids at a neutral N-terminus. These give the free-energy change for going from a charged (NH_3^+) N-terminus to an uncharged (NH_2) N-terminus at the end of an α-helix. These following results are found, Leu (–2.34 kcal mol^{-1}), Thr (–0.83 kcal mol^{-1}), Gly (–0.73 kcal mol^{-1}), Ser (–0.92 kcal mol^{-1}), and Asn (–0.53 kcal mol^{-1}). The anomalously high value for Leu is likely to be a severe overestimate. This is because the *n*-value for Leu (0.03) is very low, so that the free energy calculated from *n* is very sensitive to any inaccuracy. If this value is disregarded, these results show that the unfavorable effect on helix stability of protonation of the N-terminus is 0.5–0.9 kcal mol^{-1}.

We have found the free-energy change for protonation of the N-terminus for only 5 residues out of the 11 for which we have data. This is because Gln, Ala, Val, Met, Pro, and Ile have negative *n*-values at pH 5 that cannot be compared with their *n*-values at pH 9.55. In the fitting procedure which generated these results, *v* for all residues was fixed to 0.048. If *v* is fixed to a smaller value, the range of positive *n*-values (zero to infinity) gives a wider range of helix contents. The helix content thus becomes more sensitive to the *n*-value of the first residue if *v* is smaller. We have therefore refitted all the data with *v* fixed to 0.006. This gives positive *n*-values for all the N-terminal amino acids we have examined at both pH 5 and 9.55. Their ratios can thus be compared, and the free-energy change for protonation of the N-terminus can be found for all residues. These results are given in Table 3. Changing *v* in this way cannot be justified by experiment; indeed, experimental evidence supports the higher *v*-value we used earlier. The purpose of this procedure is solely for the purpose of determining the ratios of positive *n*-values for all residues. The *n*-values quoted in Table 3 should not be used for any other purpose. The value of *v* used here (0.006) was the smallest that gave positive *n*-values for all residues, to one significant figure. We have also calculated *n*-values with *v* fixed in the range 0.002–0.048 and found that the ratios of *n* at pH 9.55–5, and hence the free energies for protonation, were fairly insensitive to the actual *v*-value used, though the individual *n*-values changed considerably.

Table 3: Free-Energy Change of α-Helix Stability Resulting from Protonation of the N-Terminus

N-terminal amino acid	*n*-value with *v*(Ala) = 0.0006		*n*-value at pH 9.55/ *n*-value at pH 5	free-energy change (kcal mol^{-1})
	pH 9.55	pH 5		
Gln	1.55	0.08	19.6[a]	–1.61[a]
Ala	3.72	0.47	7.99[a]	–1.13[a]
Val	3.62	1.62	3.55	–0.69
Met	4.45	1.69	2.64	–0.53
Pro	4.59	0.50	9.16[a]	–1.20[a]
Ile	5.07	1.86	2.72	–0.54
Leu	6.41	2.05	3.13	–0.62
Thr	6.95	3.10	2.24	–0.44
Gly	12.3	5.69	2.16	–0.42
Ser	11.8	3.84	3.08	–0.61
Asn	23.1	9.77	2.24	–0.47

[a] Likely to be an overestimate. See text for discussion.

Table 4: N-Capping Free Energies Relative to That of Alanine in Different Systems (kcal mol^{-1})

residue	$-RT\ln(n_i)$ at pH 9.55	T4 lysozyme position 59		barnase positions 6 and 26[b]
		pH 2[a]	pH 6.5[a]	
Gln				–0.42
Ala	0	0	0	
Val	0.14	0	0	0.15
Met	–0.11			
Pro	–0.12			0.87
Ile	–0.23			
Leu	–0.44			
Thr	–0.48	–2.8	–1.5	–2.05
Gly	–0.99	–0.6	0.1	–0.69
Ser	–0.87	–2.1	–1.3	–1.64
Asn	–1.26	–2.2	–0.4	–0.86
Asp		–1.9	–0.3	–2.02
Glu				–0.25
His				–0.16

[a] Bell et al. (1992). [b] Table 7 in Serrano et al. (1992). Mean of results from two N-cap sites (residues 6 and 26).

The free-energy changes for protonation of the N-terminus for Gln, Ala, and Pro (Table 3) are likely to be overestimates as *n* at pH 5 is small. The other results suggest that a typical value for this free-energy change is ≈0.5 kcal mol^{-1}. It is not clear whether there are any significant differences between residues.

Comparison with Results from Crystal Structures

Residue preferences for specific locations within α-helices have been studied by Richardson and Richardson (1988) by examining 215 helices from 45 distinct protein structures. A capping residue is defined by them as the last (or first) residue whose α-carbon lies within a cylinder formed by the helix backbone and along the helix spiral path. Alternate definitions are possible, defined by φ and ψ angles or hydrogen bonding. Their N-cap preferences, which are scaled relative to the

Table 5: Correlation Coefficients between N-Capping Free Energies in Different Systems

	$-RT\ln(n_1)$ at pH 9.55	T4 lysozyme position 59		barnase positions 6 and 26[b]	data of Richardson and Richardson
		pH2[a]	pH 6.5[a]		
$-RT\ln(n_1)$ at pH 9.55	1				
T4 lysozyme position 59 pH 2	0.61	1			
T4 lysozyme position 59 pH 6.5	0.24	0.82	1		
barnase positions 6 and 26	0.58	0.89	0.74	1	
data of Richardson and Richardson	0.77	0.75	0.42	0.61	1

[a] Bell et al. (1992). [b] Table 7 in Serrano et al. (1992). Mean of results from two N-cap sites (residues 6 and 26).

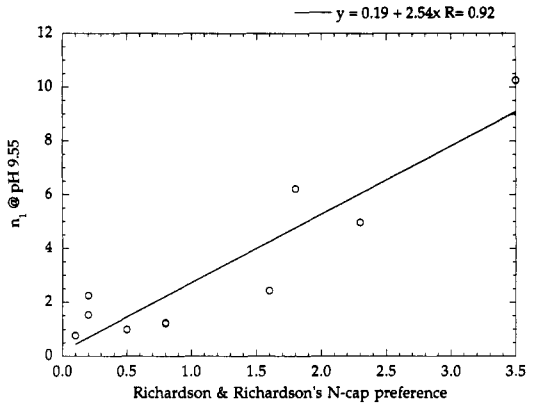

FIGURE 5: Correlation of N-cap value at pH 9.55 with N-cap preferences of Richardson and Richardson (1988).

probability of occurrence of that residue at any site, can be compared to the n-values found here. Figure 5 shows the n-values at pH 9.55 plotted against the preferences of Richardson and Richardson. The correlation is high (R = 0.92), suggesting that similar effects occur within proteins as within peptides. For the rare N-caps, the total number of occurrences found by Richardson and Richardson is small, making their results for these residues less reliable. It should be noted that the quantities being plotted here are not identical. Our n-values give the likelihood of a residue being in an N-cap state relative to an unfolded, coil state. The N-cap preferences of Richardson and Richardson give the likelihood of a residue being in an N-cap state rather than in a different secondary structural element in a folded protein, such as a β-sheet or an internal position in a helix. A similar difficulty is present when the data of Richardson and Richardson are compared to results obtained by mutating proteins (discussed below).

Comparison with Results from Other Peptide and Protein Systems

Capping in α-helical peptides has been examined by Forood et al. (1993) in a 12-residue peptide and by Lyu et al. (1993) in a 20-residue peptide. At the N-terminus, Forood et al. found the order of the ability of residues to increase helix content to be Asp > Asn > Ser > Glu > Gln > Ala, in good agreement with our results. Lyu et al. studied a helical peptide, where the third residue was assumed to be the N-cap, and found the N-cap preference order to be Ser > Asn > Gly > Ala.

Results from substituting N-cap residues in proteins can be readily compared with our results as the change in free energy of protein unfolding is measured to analyze the mutations. Table 4 lists our results and those from mutagenesis of N-cap residues in two proteins. Note that the free energies were determined under different conditions (of temperature, salt concentration, and denaturant concentration) and by different methods so that a precise agreement is not expected. The results are given relative to Ala in all cases for straightforward comparison.

Seven residues have been compared in an N-cap site (position 59 which is Thr in the wild-type) in T4 lysozyme by Bell et al. (1992). $\Delta\Delta G$ was determined from the change in melting temperature of the mutant protein relative to wild-type. Eleven different N-cap residues have been examined at positions 6 and 26 (both Thr in the wild-type) in barnase by Serrano et al. (1992). $\Delta\Delta G$ for each mutation was determined from the change in urea concentration required to unfold the protein. Table 5 gives the correlation coefficients for the straight line of best fit when all these N-capping free energies are plotted against each other.

CONCLUSION

We have shown that the Lifson–Roig model for the helix–coil transition in peptides is incapable of accounting for variations in helix content with changes in the N-terminal amino acid. The theory, however, can be simply modified to include capping effects by including terminal amino acids and acetyl and amide blocking groups as units that participate in the helix–coil transition and by assigning statistical weights to the coil units adjacent to the terminal helical residues. Application of the theory to the results given recently (Chakrabartty et al., 1993a) gives N-capping free energies differing over a range of 2 kcal mol^{-1} from Gln to Asn. The results are compared with those obtained by surveying the protein data base and with those obtained by mutations at N-cap sites in T4 lysozyme and barnase. The importance of N- and C-terminal helical boundaries for predicting helical regions in proteins has been discussed by Presta and Rose (1988) and by Lattman and Rose (1993). The modified theory together with the experimental capping values should prove useful in improving predictions of these regions in proteins and helicity in peptides. Improved estimates of w- or s-values in helical peptide systems can be made after including capping effects.

ACKNOWLEDGMENT

We would like to thank John Schellman, Hong Qian, Marty Scholtz, Tanja Kortemme, and Carol Rohl for many helpful discussions. Steve Mayo, Carol Rohl, and Tanja Kortemme are thanked for computer programming. A.J.D. would like to thank the S.E.R.C. (U.K.) for a NATO Postdoctoral Fellowship. We thank the Mass Spectrometry Facility, University of California, San Francisco (supported by an NIH grant), for verification of peptide identity.

REFERENCES

Barlow, D. J., & Thornton, J. M. (1988) *J. Mol. Biol. 201*, 601–619.

Bell, J. A., Becktel, W. J., Sauer, U., Baase, W. A., & Matthews, B. W. (1992) *Biochemistry 31*, 3590–3596.

Bixon, M., & Lifson, S. (1967) *Biopolymers 5*, 309–314.

Chakrabartty, A., Schellman, J. A., & Baldwin, R. L. (1991) *Nature 351*, 586–588.
Chakrabartty, A., Doig, A. J., & Baldwin, R. L. (1993a) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11332–11336.
Chakrabartty, A., Kortemme, T., Padmanabhan, S., & Baldwin, R. L. (1993b) *Biochemistry 32*, 5560–5565.
Forood, B., Feliciano, E. J., & Nambiar, K. P. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 838–842.
Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., & Kallenbach, N. R. (1991) *Biopolymers 31*, 1605–1614.
Hol, W. G. J., van Duijnen, P. T., & Berendsen, H. J. C. (1978) *Nature 273*, 443–446.
Lattman, E. A., & Rose, G. D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 439–441.
Lifson, S., & Roig, A. J. (1961) *J. Chem. Phys. 34*, 1963–1974.
Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669–673.
Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J., & Kallenbach, N. R. (1993) *Biochemistry 32*, 421–425.
Marqusee, S., Robbins, V. H., & Baldwin, R. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5286–5290.
Miick, S. M., Casteel, K. M., & Milhauser, G. L. (1993) *Biochemistry 32*, 8014–8021.
Poland, D., & Scheraga, H. A. (1970) *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York, London.
Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632–1641.
Qian, H., & Schellman, J. A. (1992) *J. Phys. Chem. 96*, 3987–3994.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M., & Baldwin, R. L. (1992) *Biochemistry 31*, 1263–1269.
Scholtz, J. M., Qian, H., Robbins, V. H., & Baldwin, R. L. (1993) *Biochemistry 32*, 9668–9676.
Serrano, L., Sancho, J., Hirshberg, M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 544–559.
Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., & Baldwin, R. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2349–2353.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature 326*, 563–567.
Vásquez, M., & Scheraga, H. A. (1988) *Biopolymers 27*, 41–58.
Vásquez, M., Pincus, M. R., & Scheraga, H. A. (1987) *Biopolymers 26*, 351–371.
Zimm, B. H., & Bragg, J. K. (1959) *J. Chem. Phys. 31*, 526–535.