

CHAPTER 1

The α -Helix as the Simplest Protein Model: Helix–Coil Theory, Stability, and Design

ANDREW JAMES DOIG

Faculty of Life Sciences, The University of Manchester, Jackson's Mill,
PO Box 88, Sackville Street, Manchester M60 1QD, UK

1.1 Introduction

Proteins are built of regular local folds of the polypeptide chain called secondary structure. α -Helices are present in nearly all globular proteins, with $\approx 30\%$ of residues found in α -helices.¹ It is such ubiquity and its structural simplicity that makes the α -helix an ideal candidate for detailed quantitative studies of the complex energetic factors involved in protein folding and stability. Here, we discuss structural features of the helix and their contributions to helix stability from studies in peptides. Some earlier reviews in this field are references 2–10.

1.2 Structure of the α -Helix

A helix combines a linear translation with an orthogonal circular rotation. In the α -helix the linear translation is a rise of 5.4 Å per turn of the helix and a circular rotation is 3.6 residues per turn. Side chains spaced $i, i + 3$, $i, i + 4$, and $i, i + 7$ are therefore close in space and interactions between them can affect helix stability. Spacings of $i, i + 2$, $i, i + 5$, and $i, i + 6$ place the side chain pairs on opposite faces of the helix avoiding any interaction. The helix is primarily stabilized by $i, i + 4$ hydrogen bonds between backbone amide groups.

The conformation of a polypeptide can be described by the backbone dihedral angles ϕ and ψ . Most ϕ, ψ combinations are sterically excluded, leaving only the broad β region and narrower α region. The residues at the N-terminus of the α -helix are called N'-N-cap-N1-N2-N3-N4 *etc.*, where the N-cap is the residue with non-helical ϕ, ψ angles immediately preceding the N-terminus of an α -helix and N1 is the first residue with helical ϕ, ψ angles.¹¹ The C-terminal residues are similarly called C4-C3-C2-C1-C-cap-C' *etc.* The N1, N2, N3, C1, C2, and C3 residues are unique because their amide groups participate in $i, i+4$ backbone-backbone hydrogen bonds using either only their CO (at the N-terminus) or NH (at the C-terminus) groups. The need for these groups to form hydrogen bonds has powerful effects on helix structure and stability.¹²

1.2.1 Capping Motifs

The amide NH groups at the helix N-terminus are satisfied predominantly by side-chain H-bond acceptors. In contrast, carbonyl CO groups at the C-terminus are satisfied primarily by backbone NH groups from the sequence following the helix.¹² The presence of such interactions would therefore stabilize helices. These interactions can be identified as specific patterns found at or near the ends of helices and are generally termed capping motifs.^{11,13-17}

A common pattern of capping at the helix N-terminus is the capping box. Here, the side chain of the N-cap forms a hydrogen bond with the backbone of N3 and, reciprocally, the side chain of N3 forms a hydrogen bond with the backbone of the N-cap.¹⁸ The definition of the capping box was expanded by Seale *et al.*¹⁹ to include an associated hydrophobic interaction between residues N' and N4 and is also known as a "hydrophobic staple".²⁰ A variant of the capping box motif is termed the "big" box with an observed hydrophobic interaction between non-polar side-chain groups in residues N4 and N'' (not N').¹⁹ The Pro-box motif involves three hydrophobic residues and a Pro residue at the N-cap.²¹

The two primary capping motifs found at helix C-termini are the Schellman and the α_L motifs.²²⁻²⁴ The Schellman motif is defined by a doubly hydrogen-bonded pattern between backbone partners, consisting of hydrogen bonds between the amide NH at C'' and the carbonyl CO at C3 and between the amide NH at C' and the carbonyl CO at C2, respectively. The associated hydrophobic interaction is between C3 and C''. In a Schellman motif, polar residues are highly favoured at the C1 position and the C' residue is typically glycine. If C'' is polar, the alternative α_L motif is observed, defined by a hydrogen bond between the amide NH at C' and the carbonyl CO at C3. As in the Schellman motif, the C' residue is typically glycine, which adopts a positive value of ϕ . However, the hydrophobic interaction in an α_L is heterogeneous, occurring between C3 and any of several residues external to the helix (C^{3'}, C^{4'}, or C^{5'}).²³

A notable difference between the N- and C-terminal motifs is that at the N-terminus, helix geometry favors side-chain-to-backbone hydrogen bonding and

selects for compatible polar residues.^{25,26} Accordingly, the N-terminus promotes selectivity in all polar positions, especially N-cap and N3 in the capping box. In contrast, at the C-terminus, side-chain-to-backbone hydrogen bonding is disfavored. Backbone hydrogen bonds are satisfied instead by post-helical backbone groups. The C-terminus need only select for C' residues that can adopt positive values of the backbone dihedral angle ϕ , most notably Gly.²³

1.2.2 Metal Binding

One way to stabilize helix conformations, especially in short peptides, is to introduce an artificial nucleation site composed of a few residues fixed in a helical conformation. For example, the calcium-binding loop from EF-hand proteins saturated with a lanthanide ion promotes a rigid short helical conformation at its C-terminus region.²⁷ This system has been used to measure enthalpic terms contributing to helical preferences of the amino acids.^{28–30} In the presence of Cd ions, a synthetic peptide containing Cys-His ligands $i, i + 4$ apart at the C-terminal region increased helicity (that is the average probability of finding dihedral angle pairs in values typical of α -helix) from 54% to 90%. The helicity of a similar peptide containing His-His ligands increased by up to 90% as a result of Cu and Zn binding.³¹ The addition of a *cis*-Ru(III) ion to a 6-mer peptide, Ac-AHAAHA-NH₂, changed the peptide conformation from random coil to 37% helix.³² An 11-residue peptide was converted from random coil to 80% helix content by the addition of Cd ions, although the ligands used were not natural amino acids but aminodiacetic acids.³³ As(III) stabilizes helices when bound to Cys side chains spaced $i, i + 4$ by -0.7 to -1.0 kcal mol⁻¹.³⁴ 19-Membered metallocyclic rings induce helix formation by covalently linking helical turns.³⁵

1.2.3 The 3_{10} -Helix

3_{10} -Helices are stabilized by $i, i + 3$ hydrogen bonds, instead of the $i, i + 4$ found in α -helices, making the cylinder of the 3_{10} -helix narrower than α and their hydrogen bonds non-linear. 3–4% of residues in crystal structures are in 3_{10} -helices.^{1,36} Most 3_{10} -helices are short, only 3 or 4 residues long, compared to a mean of 10 residues in α -helices,¹ and are commonly found as N- or C-terminal extensions to an α -helix:^{1,37,38} strong amino acid preferences have been observed for different locations within the interior³⁶ and N- and C-caps²⁵ of 3_{10} -helices in crystal structures. The 3_{10} -helix is being recognized as of increasing importance in isolated peptides and even as a possible intermediate in α -helix formation.^{39,40}

1.2.4 The π -Helix

In contrast to the widely occurring α - and 3_{10} -helices, the π -helix is extremely rare. The π -helix is unfavorable for three reasons: its dihedral angles are

energetically unfavorable relative to the α -helix,^{41,42} its three-dimensional structure has a 1 Å hole down the center that is too narrow for access by a water molecule resulting in the loss of van der Waals interactions, and a higher number of residues (four) must be correctly oriented before the first $i, i + 5$ hydrogen bond is formed, making helix initiation more entropically unfavorable than for α - or 3_{10} -helices.⁴³ π -Helices are known in both peptide and proteins, however.^{44–48}

1.3 Design of Peptide Helices

The earliest work on peptide helices was on long homopolymers of Glu or Lys which show coil-to-helix transitions on changing the pH from charged to neutral.⁴⁹ The neutral polypeptides are metastable and prone to aggregation, ultimately to β -sheet amyloid.⁵⁰ In 1971 Brown and Klee⁵¹ reported that the C-peptide of ribonuclease A, which contains the first 13 residues of the protein and which forms a helix in the protein, had high helical content at 0 °C. Work on the C-peptide showed that the replacement of interior helical residues with Ala was stabilizing, indicating that a major reason why this helix was folded in isolation was the presence of three successive alanines from positions 4–6. This led to the successful design of isolated, monomeric helical peptides in aqueous solution, first containing several salt bridges and a high alanine content, based on $(EAAAK)_n$,^{52,53} and then a simple sequence with a high alanine content solubilized by several lysines.⁵⁴ These “AK peptides” are based on the sequence $(AAKAA)_n$, where n is typically 2–5. The Lys side chains are spaced $i, i + 5$ so they are on opposite faces of the helix, giving no charge repulsion. Hundreds of AK peptides have been studied, giving most of the available results on helix stability in peptides. The alanines in the $(EAAAK)_n$ -type peptides may be removed entirely; E_4K_4 peptides, with sequences based on $(EEEEK_4)_n$ or EAK patterns, are also helical, stabilized by large numbers of salt bridges.^{5,55–57}

1.3.1 Host–Guest Studies

Extensive work from the Scheraga group has obtained helix–coil parameters using a host–guest method. Long random co-polymers were synthesized of a water soluble, non-ionic guest (poly[N^5 -(3-hydroxypropyl)-L-glutamine] (PHPG) or poly[N^5 -(4-hydroxybutyl)-L-glutamine] (PHBG)), together with a low (10–50%) content of the guest residue. Using the s and σ Zimm–Bragg helix–coil parameters (see below) for the host homopolymer, it was possible to calculate those for the guest using helix–coil theory as a function of temperature. The results obtained from the host–guest work are in disagreement with most of the results from short peptides of fixed sequence.

1.3.2 Helix Lengths

Helix formation in peptides is cooperative, with a nucleation penalty. Helix stability therefore tends to increase with length, in homopolymers at least. As the length of a homopolymer increases, the mean fraction helix will level off below 100%, as long helices tend to break in two. In heteropolymers, observed lengths are highly sequence dependent. As helices are at best marginally stable in monomeric peptides in aqueous solution, they are readily terminated by the introduction of a strong capping residue or a residue with a low intrinsic helical preference.

The length distribution of helices in proteins is very different from homo- and heteropolymers.¹ Most protein helices are short, with 5 to 14 residues most abundant. There is a general trend for a decrease in frequency as the length increases beyond 13 residues. Helix lengths longer than 25 are rare. There is also a preference to have close to an integral number of turns so that their N- and C-caps are on the same side of the helix.⁵⁸

1.3.3 The Helix Dipole

The secondary amide group in a protein backbone is polarized with the oxygen negatively charged and hydrogen positively charged. In a helix, the amides are all oriented in the same direction with the positive hydrogens pointing to the N-terminus and negative oxygens pointing to the C-terminus. This can be regarded as giving a partial positive charge at the helix N-terminus and a partial negative charge at the helix C-terminus.⁵⁹⁻⁶¹ In general, therefore, negatively charged groups are stabilizing at the N-terminus and positively charged at the C-terminus. An alternative interpretation of these results is that favored side chains are those that can make hydrogen bonds to the free amide NH groups at N1, N2, and N3 or free CO groups at C1, C2, and C3.⁶² Charged groups can form stronger hydrogen bonds than neutral groups, thus providing an alternative rationalization of the pH titration results. These hypotheses are not mutually exclusive, as a charged side chain can also function as a hydrogen bond acceptor or donor. Measurements of the amino acid preferences for the N-cap, N1, N2, and N3 positions in the helix allow a comparison to be made of the relative importance of helix dipole and hydrogen bonding interactions,⁶³⁻⁶⁶ suggesting that both charge and hydrogen-bonding interactions are important.

1.3.4 Acetylation and Amidation

A simple, yet effective, way to increase the helicity of a peptide is to acetylate its N-terminus.^{15,67} Acetylation removes the positive charge that is present at the helix terminus at low or neutral pH; this charge would interact unfavorably with the positive helix dipole and free N-terminal NH groups. The extra CO

group from the acetyl group can form an additional hydrogen bond to the NH group, putting the acetyl at the N-cap position. This has a strong stabilizing effect by approximately $1.0 \text{ kcal mol}^{-1}$ compared to alanine.^{63,68,69}

Amidation of the peptide C-terminus is structurally analogous to N-terminal acetylation: the helix is extended by one hydrogen bond and an unfavorable charge-charge repulsion with the helix dipole is removed. The energetic benefit of amidation is rather smaller, however, with the amide group being no better than Ala and in the middle if the C-cap residues are ranked in order of stabilization effect.⁶³ As most helical peptides studied to date are both acetylated and amidated, and acetylation is more stabilizing than amidation, the distribution of helicity along the peptide is generally skewed so that residues near the N-terminus are more helical than those near the C-terminus.

1.3.5 Solubility

Peptide aggregation can be assayed rigorously by sedimentation equilibrium, which determines the oligomeric state of a molecule in solution. This is difficult, however, with the short peptides often used as their molecular weights are at the lower limit for this technique. A simpler method is to check a spectroscopic signal that depends on peptide structure, most obviously circular dichroism (CD), as a function of concentration. If the signal depends linearly on peptide concentration across a large range, including that used to study the peptide structure, it is safe to assume that the peptide is monomeric. An oligomer that does not change state, such as a coiled-coil, across the concentration range cannot be excluded, however. Light scattering can detect aggregation. A monomeric peptide should have a flat baseline in a UV spectrum outside the range of any chromophores in the peptide. Stock solutions of a peptide with a single tyrosine isolated from the helix region by Gly should have $A_{300}/A_{275} < 0.02$ and $A_{250} < A_{275} < 0.2$.⁷⁰

Consideration of solubility is essential when designing helical peptides. Solubility can be achieved most easily by including polar side chains spaced $i, i + 5$ in the sequence where they cannot interact. Lys, Arg and Gln are used most often for this purpose. Gln may be preferred if unwanted interactions with charged Lys or Arg may be a problem, but some AQ peptides lack sufficient solubility and AQ peptides are less helical.

The spacing of side chains in the helix is best visualized with a helical wheel, to ensure that the designed helix does not have a non-polar face that may lead to dimerization. The following webpage provides a useful resource for this: <http://cit.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html>

1.3.6 Concentration Determination

An accurate measurement of helix content depends on an accurate spectroscopic measurement and, equally importantly, peptide concentration. This is

usually achieved by including a Tyr side chain at one end of the peptide. The extinction coefficient of Tyr at 275 nm is $1450 \text{ M}^{-1} \text{ cm}^{-1}$.⁷¹ If Trp is present, measurements at 281 nm can be used where the extinction coefficient of Trp is $5690 \text{ M}^{-1} \text{ cm}^{-1}$ and Tyr $1250 \text{ M}^{-1} \text{ cm}^{-1}$.⁷² Though the inclusion of aromatic residues is required for concentration determination, this can have the unwanted side effect of perturbing a CD spectrum, leading to an inaccurate determination of helix content. A simple solution to this problem is to separate the terminal Tyr from the rest of the sequence by one or more Gly residues.⁷³ If the aromatic residues must be included within the helical region, the CD spectrum should be corrected to remove this perturbation.⁷⁴

1.3.7 Helix Templates

A major penalty to helix formation is the loss of entropy arising from the requirement to fix three consecutive residues to form the first hydrogen bond of the helix. Following this nucleation, propagation is much more favored as only a single residue need be restricted to form each additional hydrogen bond. A way to avoid this barrier is to synthesize a template molecule that facilitates helix initiation, by fixing hydrogen bond acceptors or donors in the correct orientation for a peptide to bond in a helical geometry. The ideal template nucleates a helix with an identical geometry to a real helix. Kemp's group applied this strategy and synthesized a proline-like template that nucleated helices when a peptide chain was covalently attached to a carboxyl group.⁷⁵⁻⁷⁹ Bartlett *et al.* reported on a hexahydroindol-4-one template⁸⁰ that induces helicity in an appended hexameric peptide. Several other templates were less successful and could only induce helicity in organic solvents.⁸¹⁻⁸³ Their syntheses are often lengthy and difficult, partly due to the challenging requirement of orienting several dipoles to act as hydrogen bond acceptors or donors.

1.4 Helix-Coil Theory

Peptides that form helices in solution do not show a simple two-state equilibrium between a fully formed helix and a fully unfolded structure. Instead they form a complex mixture of all helix, all coil, or, most frequently, a distribution of helices of different lengths with increased probability at the center of the peptide (helix fraying). In order to interpret experiments on helical peptides and make theoretical predictions on helices it is therefore essential to use a helix-coil theory that deals with this distribution of helices. Recent reviews of helix-coil theory are references 84-86.

The simplest way to analyse the helix-coil equilibrium, still occasionally seen, is the two-state model where the equilibrium is assumed to be between a 100% helix conformation and 100% coil. This is incorrect and its use gives serious errors.

1.4.1 Zimm–Bragg Model

The two major types of helix–coil model are i) those which count hydrogen bonds, principally Zimm–Bragg (ZB),⁸⁷ and ii) those that consider residue conformations, principally Lifson–Roig (LR).⁸⁸ In the ZB theory the units being considered are peptide groups and they are classified on the basis of whether their NH groups participate in hydrogen bonds within the helix. The ZB coding is shown in Figure 1.1. A unit is given a code of 1 (*e.g.* peptide unit 5 in Figure 1.1) if its NH group forms a hydrogen bond and 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 a propagation parameter and σ is an initiation parameter. The difficulty of nucleating a helix is captured in the ZB model by having σ smaller than s . The statistical weight of a homopolymeric helix of N hydrogen bonds is σs^{N-1} . The cost of initiation, σ , is thus paid only once for each helix, while extending the helix simply multiplies its weight by one additional s -value for each extra hydrogen bond.

1.4.2 Lifson–Roig Model

In the LR model each residue is assigned a conformation of helix (h) or coil (c), depending on whether it has helical ϕ , ψ angles. Every conformation of a peptide of N residues can be written as a string of N c's or h's, giving 2^N conformations in total. Residues are assigned statistical weights depending on their conformations and the conformations of surrounding residues. A residue in an h conformation with an h on either side has a weight of w . This can be thought of as an equilibrium constant between the helix interior and the coil. Coil residues are used as a reference and have a weight of 1. In order to form an $i, i + 4$ hydrogen bond in a helix, three successive residues need to be fixed in a helical conformation. M consecutive helical residues will therefore have $M - 2$ hydrogen bonds. The two residues at the helix termini (*i.e.* those in the centre of chh or hlc conformations) are assigned weights of v (Figure 1.1). The ratio of w to v gives approximately the effect of hydrogen bonding ($1.7 : 0.036$ for Ala⁶⁹ or $-RT \ln (1.7/0.036) = -2.1 \text{ kcal mol}^{-1}$). A helical homopolymer segment of M residues has a weight of $v^2 w^{M-2}$ and a population in

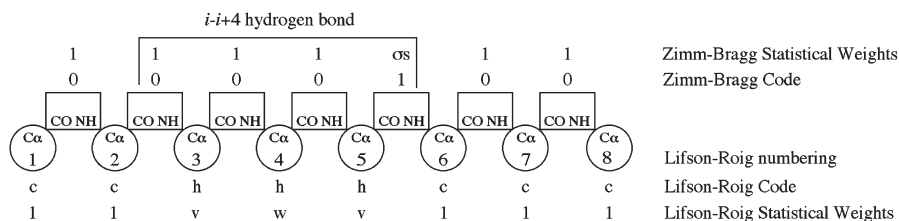


Figure 1.1 Zimm–Bragg and Lifson–Roig codes and weights for the α -helix.

the equilibrium of $v^2 w^{M-2}$ divided by the sum of the weights of every conformation (*i.e.* the partition function). In this way the population of every conformation is calculated and all properties of the helix-coil equilibrium evaluated. The LR model is easier to handle conceptually for heteropolymers since the parameters are assigned to individual residues. The substitution of one amino acid at a certain position thus changes the w - and v -values at that position. In the ZB model the initiation parameter σ is associated with several residues and s with a peptide group, rather than a residue. It is therefore easier to use the LR model when making substitutions. Indeed, most recent work has been based on this model. A further difference is that the ZB model assigns weights of zero to all conformations that contain a chc or chhc sequence. This excludes a very large number of conformations that contain a residue with helical ϕ , ψ angles but with no hydrogen bond. In LR theory, these are all considered. The ZB and LR weights are related by the following formulae:⁸⁵ $s = w/(1 + v)$; $\sigma = v^2/(1 + v)^4$.

The complete helix-coil equilibrium is handled by determining the statistical weight for every possible conformation that contains a helix plus a reference weight of 1 for the coil conformation. Each conformation considered in the helix-coil equilibrium is given a statistical weight. This indicates the stability of that conformation, with the higher the weight, the more probable the conformation. Weights are defined relative to the all-coil conformation, which is given a weight of 1. The statistical weight of a conformation can thus be regarded as an equilibrium constant relative to the coil; a weight >1 indicates the conformation is more stable than coil, <1 means less stable and $=1$ means equally stable. The population of each conformation is given by the statistical weight of that conformation divided by the sum of the statistical weights for every conformation (the partition function). Thus the greater the statistical weight, the more stable the conformation. The key to using helix-coil theory is the partition function. All the properties of a system at equilibrium are contained within the partition function, which makes it very valuable. Partition functions are extremely powerful concepts in statistical thermodynamics since they allow calculation of all properties of an equilibrium ensemble. Any property of the equilibrium can be extracted from the partition function by applying the appropriate mathematical function. In this case the properties could be the mean number of hydrogen bonds, the mean helix length, the probability that each residue is within a helix, *etc.* In particular, the mean number of residues with a weight x is given by $\frac{\partial \ln Z}{\partial \ln x}$. Circular dichroism is commonly used to give the mean helix content of a helical peptide, namely the fraction of residues that have a weight of w . LR-based models can thus be related to experimental data by equating the measured mean helix content to $\frac{\partial \ln Z}{\partial \ln w}/N$, where N is the number of residues in the peptide. Statistical weights can be regarded as equilibrium constants for the equilibrium between coil and the structure (as the reference coil weight is defined as 1). They can therefore be converted to free energies as $-RT \ln(\text{weight})$. The Lifson-Roig formalism has also been adapted to describe 3_{10} - and π -helices.^{43,89,90}

1.4.2.1 *The Unfolded State and Polyproline II Helix*

The treatment of peptide conformations is based on Flory's isolated-pair hypothesis.⁹¹ This states that while ϕ and ψ for a residue are strongly interdependent, giving preferred areas in a Ramachandran plot, ϕ , ψ pair is independent of the ϕ , ψ angles of its neighbors. Pappu *et al.* found that non-helical poly(Ala) chains mostly populated extended or fully helical conformations as many partly helical conformations are sterically disallowed. Such effects are not included in helix-coil theories. Helix-coil theories assign the same weight (1) to every coil residue; steric exclusion means that these should vary and be lower than 1 in many cases.

The polyproline II helix may well be an important conformation for unfolded proteins. Many recent papers have addressed this issue. Examples are references.⁹²⁻⁹⁴ In particular, denatured alanine rich peptides may form a polyproline II helix.⁹⁵⁻¹⁰¹ It may therefore be valid to consider residues in helical peptides to be in three possible states (helix, coil, or polyproline II), rather than two (helix or coil). No current helix-coil model takes this into account. A scale of amino acid preferences for the polyproline II helix has been published.¹⁰²

1.4.2.2 *Single Sequence Approximation*

Since helix nucleation is difficult, conformations with multiple helical segments are expected to be rare in short peptides. In the one-, or single-, helical sequence approximation, peptide conformations containing more than one helical segment are assumed not to be populated and are excluded from the partition function (*i.e.* assigned statistical weights of zero). As peptide length increases, the approximation is no longer valid since multiple helical segments can be long enough to overcome the initiation penalty. The single sequence approximation will also break down when a sequence with a high preference for a helix terminus is within the middle of the chain. Conformations with two or more helices may also often include helix-helix tertiary interactions that are ignored in all helix-coil models.

1.4.2.3 *N- and C-caps*

N-Capping has been added to LR theory by assigning a weight of n to the central residue in a cch triplet, as the N-cap is the non-helical residue preceding the start of a helical segment. Similarly, the C-cap is the first residue in a non-helical conformation (c) at the C-terminus of a helix. C-Cap weights (c -values) are assigned to central residues in hcc triplets.^{6,68,103}

1.4.2.4 *Capping Boxes*

The N-terminal capping¹⁸ includes a side-chain-backbone hydrogen bond from N3 to the N-cap ($i, i-3$). This is included in the LR model by assigning a

weight of $w*r$ to the chhh conformation, where r is the weight for the Ser backbone to Glu side-chain bond.⁶

1.4.2.5 Side-chain Interactions

As helices have 3.6 residues per turn, side chains spaced $i, i+3$ or $i, i+4$ are close in space. Side-chain interactions are thus possible when four or five consecutive residues are in a helix. They are included in the LR-based model by giving a weight of $w*q$ to hhhh quartets and $w*p$ to hhhhh quintets. The side-chain interaction is between the first and last side chains in these groups; the w weight is maintained to preserve the equivalence between the number of residues with a w weighting and the number of backbone helix hydrogen bonds.¹⁰⁴

1.4.2.6 N1, N2, and N3 Preferences

The helix N-terminus shows significantly different residue frequencies for the N-cap, N1, N2, N3, and helix interior positions.^{11,26,105,106} A complete theory for the helix should therefore include distinct preferences for the N1, N2, and N3 positions. In the original LR model, the N1 and C1 residues are both assigned the same weight, v . Shalongo and Stellwagen¹⁰⁷ separated these as v_N and v_C . Andersen and Tong¹⁰³ did the same and derived complete scales for these parameters from fitting experimental data, though some values were tentative. The helix initiation penalty is v_N*v_C and so v_N - and v_C -values are all small (≈ 0.04).

We added weights for the N1, N2, and N3 ($n1$, $n2$, and $n3$) positions as follows:¹⁰⁸ The $n1$ -value is assigned to a helical residue immediately following a coil residue. The penalty for helix initiation is now $n1.v$, instead of v^2 , as v remains the C1 weight. An N2 helical residue is assigned a weight of $n2.w$, instead of w . The weight w is maintained in order to keep the useful definition of the number of residues with a w weighting being equal to the number of residues with an $i, i+4$ main chain-main chain hydrogen bond. The $n2$ -value is an adjustment to the weight of an N2 residue that takes into account the structures that can be adopted by side chains uniquely at this position. Similarly, an N3 residue is now assigned the weight $n3.w$, instead of w .

1.4.2.7 Helix Dipole

Helix dipole effects were added to the LR model by Scholtz *et al.*,¹⁰⁹ though they used the one sequence approximation so that only one or no dipoles in total are present. In LR models helix dipole effects are subsumed within other energies. For example, N-cap, N1, N2, and N3 energies will include a contribution from the helix dipole interaction so the energy of interaction of charged groups at this position with the dipole should not be counted in addition.

1.4.3 AGADIR

AGADIR is an LR-based helix-coil model developed by Serrano, Muñoz, and co-workers. The original model¹¹⁰ included parameters for helix propensities excluding backbone hydrogen bonds (attributed to conformational entropy), backbone hydrogen bond enthalpy, side-chain interactions and a term for coil weights at the end of helical sequences (*i.e.* caps). The single sequence approximation was used. The original partition function assumed that many helical conformations did not exist, as all conformations in which the residue of interest is not part of a helix were excluded.^{104,110} These were corrected in a later version, AGADIRms, which considers all possible conformations.¹¹¹ If AGADIR and LR models are both applied to the same data, to determine a side-chain interaction energy, for example, the results are similar, showing that the models are now not significantly different.^{111,112} The treatment of the helix-coil equilibrium differs in a number of respects from the ZB and LR models and these have been discussed in detail by Muñoz and Serrano.¹¹¹ The minimal helix length in AGADIR is four residues in an h conformation, rather than three. The effect of this assumption is to exclude all helices which contain a single hydrogen bond; only helices with two or more hydrogen bonds are allowed. In practice, this probably makes little difference as chhhc conformations are usually unfavorable and hence have low populations. Early versions of AGADIR considered that residues following an acetyl at the N-terminus or preceding an amide at the C-terminus were always in a c conformation; this was changed to allow these to be helical.¹¹³

The latest version of AGADIR, AGADIR1s-2,¹¹³ includes terms for electrostatics,¹¹³ the helix dipole,^{113,114} pH dependence,¹¹⁴ temperature,¹¹⁴ ionic strength,¹¹³ N1, N2, and N3 preferences,¹¹⁵ and capping motifs such as the capping box, hydrophobic staple, Schellman motif, and Pro-capping motif.¹¹³ The free energy of a helical segment, $\Delta G_{\text{helical-segment}}$, is given by $\Delta G_{\text{helical-segment}} = \Delta G_{\text{Int}} + \Delta G_{\text{Hbond}} + \Delta G_{\text{SD}} + \Delta G_{\text{dipole}} + \Delta G_{\text{nonH}} + \Delta G_{\text{electrost}}$, which are terms for the energy required to fix a residue in helical angles (with separate terms for N1, N2, N3, and N4), backbone hydrogen bonding, side-chain interactions excluding those between charged groups, capping and helix dipole interactions, respectively. Electrostatic interactions are calculated with Coulomb's equation. Helix dipole interactions were all electrostatic interactions between the helix dipole or free N- and C-termini and groups in the helix. Interactions of the helix dipole with charged groups located outside the helical segment were also included. pH dependence calculations considered a different parameter set for charged and uncharged side chains and their pK_a values. The single sequence approximation (see above) is used again, unlike in AGADIRms. AGADIR is at present the only model that can give a prediction of helix content for any peptide sequence, thus making it very useful. It can also predict NMR chemical shifts and coupling constants.

1.4.4 Lomize–Mosberg Model

Lomize and Mosberg developed a model for calculating the stability of helices in solution.¹¹⁶ Interestingly, they extended it to consider helices in micelles or a uniform non-polar droplet to model a protein core environment. Helix stability in water is calculated as the sum of main chain interactions, which is the free energy change for transferring Ala from coil to helix, the difference in energy when replacing an Ala with another residue, hydrogen bonding and electrostatic interactions between polar side chains and hydrophobic side-chain interactions. An entropic nucleation penalty of two residues per helix is included. Different energies are included for N-cap, N1-N3, C1-C3, C-cap, hydrophobic staples, Schellman motifs, and polar side-chain interactions. Hydrophobic interactions were calculated from decreases in non-polar surface area when they are brought in contact. Helix stability in micelles or non-polar droplets is found by calculating the stability in water then adding a transfer energy to the non-polar environment.

1.5 Forces Affecting α -Helix Stability

1.5.1 Helix Interior

Different approaches have been used in order to determine the helical propensity or preference of individual amino acids. Scheraga and co-workers used a host–guest strategy (see above) to derive values for the helical preference of various amino acid residues. The host–guest system uses long random copolymers of a water soluble, non-ionic guest (poly[N⁵-(3-hydroxypropyl)-L-glutamine] (PHPG) or poly[N⁵-(4-hydroxybutyl)-L-glutamine] (PHBG)), together with a low (10–50%) content of the guest residue. The Zimm–Bragg model s and σ values of the host homopolymer are used to compute those for the guest.¹¹⁷ This work has been criticized as the host side chains can interact with each other.¹¹⁸ The introduction of a guest residue thus removes host–host interactions and replaces them with PHBG–guest or PHPG–guest side-chain interactions that may obscure the intrinsic helix propensities.

Rohl *et al.*⁶⁹ used many alanine-based peptides with the general sequences Ac-(AAKAA)_{*m*}Y-NH₂ (or with Q instead of K) to measure interior helix propensities. Substitutions in the helix interior and subsequent measures of helicity using CD spectroscopy in both water and 40% (v/v) trifluoroethanol (TFE) allowed the calculation of both the Lifson–Roig w parameter and the stabilization energy for all 20 amino acids. Kallenbach and co-workers used synthetic peptides of the form succinyl-YSEEEKAKKAXAEEAEK KKK-NH₂, where substitutions at X allowed determination of helix stabilizing energies for common amino acids.⁵⁶ Stellwagen and co-workers made substitutions in position 9 of Ac-Y(EAAAK)₃A-NH₂.⁵³

In 1998 Pace and Scholtz¹¹⁹ gathered information from many different sources and derived a scale for the propensity of each amino acid in the helix interior. This is summarized in Table 1.1. The values are in $\Delta(\Delta G)$ relative to alanine

Table 1.1 Summary of other experimental helix propensities (relative to alanine).

<i>Amino acid</i>	<i>Helix propensity</i> $\Delta(\Delta G)(\text{kcal mol}^{-1})$ <i>(taken from reference 119)</i>
Ala	0.00
Arg ⁺	0.21
Leu	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
Gly	1.00
Pro	3.16

because it is generally (though not universally) agreed that this amino acid has the highest helical propensity. (The s values from the Zimm–Bragg model, as derived by the Scheraga group,¹¹⁷ do not agree with other scales, alanine having the highest helix propensity and all other residues having lower values (a positive $\Delta(\Delta G)$ value relative to Ala)). Proline and glycine have the lowest helical propensity. The most controversial of these differences over the years has been that of alanine. Host–guest analysis showing alanine to be effectively helix-neutral has been supported by data from some other groups, notably the templated helices of Kemp and co-workers.⁷⁵ The use of template-nucleated helices has been criticized by Rohl *et al.*,¹²⁰ who argued that the low apparent helix propensity of alanine is a consequence of properties of the template–helix junction. Kemp and co-workers¹²¹ used templates to investigate the helix-forming tendency of polyalanine. Below six residues Ala had a low helix propensity, but when the limit of six was exceeded an increase was found. This suggested that there is a length-dependent term in the helicity of polyalanine. Alternatively, any destabilizing effect of the template is less significant in longer helices.

1.5.2 Caps

Some capping preferences were measured in proteins using barnase¹³ and T4 lysozyme,¹⁴ giving slightly varying results. The Kallenbach group¹²² substituted

several amino acids at the N-cap position in peptide models in the presence of a capping box. They found that Ser and Arg are the most stabilizing residues, whilst Gly and Ala are less stabilizing. The results are in agreement with the results of Forood *et al.*,¹⁶ who found that the trend in α -helix inducing ability at the N-cap is Asp > Asn > Ser > Glu > Gln > Ala. A more comprehensive work to determine the preferences for all 20 amino acids at the N-cap position used peptides with a sequence of NH₂-XAKAAAKAAAAKAAGY-CONH₂.^{15,63,68} N-capping free energies ranged from Asn (best) to Gln (worst) (Table 1.2).

We have used a similar approach using peptide models to probe the preferences at N1⁶⁴, N2,⁶⁵ and N3⁶⁶ using peptides with sequences of CH₃CO-XAAAAQAAAAQAAGY-CONH₂, CH₃CO-AXAAAAKAAAAKAAGY-CONH₂, and CH₃CO-AAXAAAAKAAAAKAGY-CONH₂, respectively. The results have given N1, N2, and N3 preferences for most amino acids for these positions (Table 1.2) and these agree well with preferences seen in protein structures, with the exception of Pro at N1. Petukhov *et al.* similarly obtained N1, N2, and N3 preferences for non-polar and uncharged polar residues by applying AGADIR to experimental helical peptide data, and found similar results.^{115,123} The complete sequences of peptides used can be seen in the table footnote. At N1, N2, and N3, Asp and Glu as well as Ala are preferred, presumably because negative side chains interact favorably with the helix dipole or NH groups, while Ala has the strongest interior helix preference.

Although it is also unique in terms of the presence of unsatisfied backbone hydrogen bonds, the C-terminal region is less explored experimentally. The C-terminus of the α -helix tends to fray more than the N-terminus, making C-terminal measurements less accurate. Zhou *et al.*¹²⁴ found that Asn is the most favored residue at the C-cap followed by Gln > Ser ~ Ala > Gly ~ Thr. Forood *et al.*¹⁶ tested a limited number of amino acids at the C-terminus (C1) finding a rank order of Arg > Lys > Ala. Doig and Baldwin⁶³ determined the C-capping preferences for all 20 amino acids in α -helical peptides. The thermodynamic propensities of some amino acids at C', C-cap, C1, C2, and C3 are also included in Table 1.2.^{125,126}

1.5.3 Phosphorylation

Phosphoserine is destabilizing compared to serine at interior helix positions.^{127,128} We investigated the effect of placing phosphoserine at the N-cap, N1, N2, N3, and interior position in alanine-based α -helical peptides, studying both the -1 and -2 phosphoserine charge states.¹²⁹ Phosphoserine stabilizes at the N-terminal positions by as much as 2.3 kcal mol⁻¹, while it destabilizes in the helix interior by 1.2 kcal mol⁻¹, relative to serine. The rank order of free energies relative to serine at each position is N2 > N3 > N1 > N-cap > interior. Moreover, -2 phosphoserine is the most preferred residue known at each of these N-terminal positions. Experimental pK_a values for the -1 to -2

T	-0.7	0.5	0.5	0.5	1.2	0.6	0.8	0.6	0.5	0.8	0.9	1.1
V	-0.1	0.6	0.5	0.4	-	0.4	0.3	0.4	0.7	0.6	0.9	1.6
W	-1.3	0.4	0.8	-	4.0	-	0.4	0.4	-	0.6	-	0.7
Y	-0.9	-	-	-	1.2	-	-	-	-	-	-2.2	-

^aNH₂-YAKAAAAKAAAAKAAAGY-CONH₂.

^bCH₃CO-YAAAAQAAAAQAAAGY₂.

^cCH₃CO-YAAAAAARAAARGGY-CONH₂.

^dCH₃CO-AYAAAAKAAAKAAAGY-CONH₂.

^eCH₃CO-AYAAAAAARAAARGGY-NH₂.

^fCH₃CO-AYAAAAKAAAKAGY-CONH₂.

^gCH₃CO-AYAAAAAARAAARGGY-NH₂.

^hNH₂-YGGSAKEAAAAAXAA-CONH₂.

ⁱSubstitution of residue 32 (C2 position) of α -helix of ubiquitin.

^jNH₂-YGGSAKEAAAAAXAY-CONH₂.

^kNH₂-YGGSAKEAAAAAX-CONH₂.

^lCH₃CO-YGAAKAAAKAAAKAX-COOH.

^mSubstitution of residue 35 (C' position) of α -helix of ubiquitin.

phosphoserine transition are in the order $N2 < N\text{-Cap} < N1 < N3 < \text{interior}$. Phosphoserine can form highly stabilizing salt bridges to Arg¹²⁸ or Lys.¹³⁰

1.5.4 Non-covalent Side-chain Interactions

Many studies have been performed on the stabilizing effects of interactions between amino acid side chains in α -helices. These studies have identified a number of types of interaction that stabilize the helix, including salt bridges,^{52,55,57,109,131–135} hydrogen bonds,^{109,135–137} hydrophobic interactions,^{104,138–140} basic–aromatic interactions,^{74,141} and polar/non-polar interactions.¹⁴² The stabilizing energies of many pairs in these categories have been measured, though some have only been analysed qualitatively. As described earlier, residue side chains spaced $i, i + 3$ and $i, i + 4$ are on the same face of the α -helix, though it is the $i, i + 4$ spacing that receives most attention in the literature, as this is stronger. A summary of stabilizing energies for side-chain interactions is given in Table 1.3. We give only those that have been measured in helical peptides with the side-chain interaction energies determined by applying helix–coil theory. Almost all are attractive, with the sole exception of the Lys–Lys repulsion.

1.5.4.1 Cooperativity

After individual side-chain interactions, the next most complex step is to study triplets, with residues A, B, and C, where B forms bonds to both residues A and C. The free energy of the triplet is often not the sum of the AB and BC bond energies. The first evaluation of the strength of an engineered complex salt-bridge in a peptide was reported by Mayne *et al.*¹⁴³ after studying a stabilizing multiple salt-bridge involving Glu3, Asp4, and Arg7 in an 11-mer α -helix. A triplet of charged Arg-Glu-Arg residues spaced $i, i + 4$, $i + 8$ or $i, i + 3$, $i + 6$ also stabilizes α -helical peptides by $-1.5 \text{ kcal mol}^{-1}$ and $-1.0 \text{ kcal mol}^{-1}$, respectively, which is more than the additive contribution of two single salt-bridges.¹⁴⁴ A similar stabilizing effect in an Arg-Phe-Met triplet in $i, i + 4$, $i + 8$ spacing was observed.¹⁴⁵ Here, the triplet energy was $0.75 \text{ kcal mol}^{-1}$ greater than the sum of the Arg-Phe and Phe-Met energies. This was attributed to both interactions favoring the same conformation of the shared central Phe. Other non-salt-bridge triplets in isolated helical peptides have also been reported, for example Glu-Phe-Arg¹⁴⁶ and Glu-Phe-Glu,¹⁴⁷ although they do not show significant effects on peptide stability. In a Glu-Lys-Glu triplet, the second potential salt-bridge provide no additional stabilization over a single interaction, as the central Lys is only able to form one bond at a time. These simple examples show that side-chain interactions can be highly non-additive when residues have the potential to form more than one bond simultaneously, and show the difficulty of predicting helix stability for typical protein sequences with multiple interactions.

Table 1.3 Summary of side-chain interaction energies from literature.

<i>Interaction</i>	$\Delta\Delta G$ ($kcal\ mol^{-1}$)	<i>Reference</i>
Ile – Lys ($i,i+4$)	-0.22	142
Val – Lys ($i,i+4$)	-0.25	142
Ile – Arg ($i,i+4$)	-0.22	142
Phe – Met ($i,i+4$)	-0.8	104
Met – Phe ($i,i+4$)	-0.5	104
Gln – Asn ($i,i+4$)	-0.5	137
Asn – Gln ($i,i+4$)	-0.1	137
Phe – Lys ($i,i+4$)	-0.14	74
Lys – Phe ($i,i+4$)	-0.10	74
Phe – Arg ($i,i+4$)	-0.18	74
Phe – Orn ($i,i+4$)	-0.4	141
Arg – Phe ($i,i+4$)	-0.1	74
Tyr – Lys ($i,i+4$)	-0.22	74
Glu – Phe ($i,i+4$)	-0.5	147
Asp – Lys ($i,i+3$)	-0.12	167
Asp – Lys ($i,i+4$)	-0.24	167
Asp – His ($i,i+3$)	> -0.63	173
Asp – His ($i,i+4$)	> -0.63	173
Asp – Arg ($i,i+3$)	-0.8	174
Glu – His ($i,i+3$)	-0.23	167
Glu – His ($i,i+4$)	-0.10	167
Glu – Lys ($i,i+3$)	-0.38	109
Glu – Lys ($i,i+4$)	-0.44	109
Phe – His ($i,i+4$)	-1.27	135
Phe – Met ($i,i+4$)	-0.7	140
His – Asp ($i,i+3$)	-0.53	135
His – Asp ($i,i+4$)	-2.38	175
His – Glu ($i,i+3$)	-0.45	167
His – Glu ($i,i+4$)	-0.54	167
Lys – Asp ($i,i+3$)	-0.4	167
Lys – Asp ($i,i+4$)	-0.58	167
Lys – Glu ($i,i+3$)	-0.38	167
Lys – Glu ($i,i+4$)	-0.46	167
Lys – Lys ($i,i+4$)	+0.17	133
Leu – Tyr ($i,i+3$)	-0.44	107
Leu – Tyr ($i,i+4$)	-0.65	107
Met – Phe ($i,i+4$)	-0.37	140
Gln – Asp ($i,i+4$)	-0.97	136
Gln – Glu ($i,i+4$)	-0.31	109
Trp – Arg ($i,i+4$)	-0.4	147
Trp – His ($i,i+4$)	-0.8	112
Tyr – Leu ($i,i+3$)	-0.02	107
Tyr – Leu ($i,i+4$)	-0.44	107
Tyr – Val ($i,i+3$)	-0.13	107
Tyr – Val ($i,i+4$)	-0.31	107
Arg ($i,i+4$) Glu ($i,i+4$) Arg	-1.5	176
Arg ($i,i+3$) Glu ($i,i+3$) Arg	-1.0	176
Arg ($i,i+3$) Glu ($i,i+4$) Arg	-0.3	176
Arg ($i,i+4$) Glu ($i,i+3$) Arg	-0.1	176
Phosphoserine – Arg ($i,i+4$)	-0.45	128

1.5.5 Covalent Side-chain Interactions

Lactam (amide) bonds formed between NH_3^+ and CO_2^- side chains can stabilize a helix, acting in a similar way to disulfide bridges in a protein by constraining the side chains to be close, reducing the entropy of non-helical states.¹⁴⁸ Lactam bridges between Lys-Asp, Lys-Glu and Glu-Orn spaced $i, i + 4$ have been introduced into analogues of human growth hormone releasing factor¹⁴⁹ and proved to be stabilizing. The same Lys-Asp $i, i + 4$ lactam was stabilizing in other helical peptide systems,^{150–153} while Lys-Glu $i, i + 4$ lactam bridges were less effective.¹⁵¹ Two overlapping Lys-Asp lactams were even more effective.¹⁵⁴ The effect of the ring size formed by the lactam was investigated by replacing Lys with ornithine or (*S*)-diaminopropionic acid. A ring size of 21 or 22 atoms was most stabilizing (a Lys-Asp $i, i + 4$ lactam is 20 atoms).¹⁴⁹ Lactams between side chains spaced $i, i + 7$ ^{155,156} or $i, i + 3$,^{156,157} spanning two or one turns of the helix, have also been reported. Disulfide bonds spaced $i, i + 7$ have been introduced into alanine-based peptides, using (D)- and (L)-2-amino-6-mercaptophexanoic acid derivatives.¹⁵⁸

Helix formation can be reversibly photoregulated. Two cysteine residues are cross-linked by an azobenzene derivative which can be photoisomerized from *trans* to *cis*, causing a large increase or decrease in the helix content of the peptide, depending on its spacing.^{159–161}

1.5.6 Capping Motifs

Although the N-terminal capping box sequence stabilizes helices by inhibiting N-terminal fraying, it does not necessarily promote elongation unless accompanied by favorable hydrophobic interactions as in a “hydrophobic staple” motif.^{162,163} The nature of the capping box stabilizing effect thus not only arises from reciprocal hydrogen bonds between compatible residues, but also from local interactions between side chains, helix macrodipole-charged residue interactions, and solvation.¹⁶⁴ Despite statistical analyses revealing that Schellman motifs are observed more frequently than expected at the helix C-terminus, this motif populates only transiently in aqueous solution, but it is formed in 30% TFE.¹⁶⁵ This might be due to the C-terminus being very frayed and so the increase of helical content contributed from this motif is small. The α_L motif seems to be more stable than the alternative Schellman motif.¹⁶³

1.5.7 Ionic Strength

Electrostatic interactions between charged side chains and the helix macrodipole can stabilize the helix.^{61,70,166} The interactions are alleviated by the screening effects of water, ions, and nearby protein atoms. The energetics of the interaction between fully charged ion pairs can be diminished by added salt and

completely screened at 2.5 M NaCl.^{134,167} Interactions of charged residues with the helix macrodipole are less affected by salt than those between charged side chains.^{167,168}

1.5.8 Temperature

Thermal unfolding experiments show that the helix unfolds with increasing temperature.^{169–171} There is no sign of cold denaturation, as seen with proteins. Enthalpy and entropy changes for the helix-coil transition are difficult to determine as the helix-coil transition is very broad, precluding accurate determination of high- and low-temperature baselines by calorimetry.¹⁶⁹ Nevertheless, isothermal titration calorimetric studies of a series of peptides that form helices when binding a nucleating La^{3+} find ΔH for helix formation to be $-1.0 \text{ kcal mol}^{-1}$,^{27,29} in good agreement with the earlier work. This system has been used to measure the enthalpic preferences of all the amino acids.³⁰

References

1. D. J. Barlow and J. M. Thornton, *J. Mol. Biol.*, 1988, **201**, 601.
2. J. M. Scholtz and R. L. Baldwin, *Ann. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 95.
3. R. L. Baldwin, *Biophys. Chem.*, 1995, **55**, 127.
4. A. Chakrabartty and R. L. Baldwin, *Adv. Protein Chem.*, 1995, **46**, 141.
5. N. R. Kallenbach, P. Lyu and H. Zhou, in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. G. D. Fasman, Plenum Press, New York, 1996, 201.
6. C. A. Rohl and R. L. Baldwin, *Meth. Enzymol.*, 1998, **295**, 1.
7. M. J. I. Andrews and A. B. Tabor, *Tetrahedron*, 1999, **55**, 11711.
8. L. Serrano, *Adv. Prot. Chem.*, 2000, **53**, 49.
9. A. J. Doig, N. Errington and T. Iqbalsyah, in *Stability and Design of α -Helices*, ed. T. Kiefhaber and J. Buchner, Weinheim, 2005.
10. J. J. Osterhout, *Prot. Pept. Lett.*, 1995, **12**, 159.
11. J. S. Richardson and D. C. Richardson, *Science*, 1988, **240**, 1648.
12. L. G. Presta and G. D. Rose, *Science*, 1988, **240**, 1632.
13. L. Serrano and A. R. Fersht, *Nature*, 1989, **342**, 296.
14. J. A. Bell, W. J. Becktel, U. Sauer, W. A. Baase and B. W. Matthews, *Biochemistry*, 1992, **31**, 3590.
15. A. Chakrabartty, A. J. Doig and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 11332.
16. B. Forood, E. J. Feliciano and K. P. Nambiar, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 838.
17. S. Dasgupta and J. A. Bell, *Int. J. Pept. Prot. Res.*, 1993, **41**, 499.
18. E. T. Harper and G. D. Rose, *Biochemistry*, 1993, **32**, 7605.
19. J. W. Seale, R. Srinivasan and G. D. Rose, *Prot. Sci.*, 1994, **3**, 1741.

20. V. Muñoz and L. Serrano, *Nat. Struct. Biol.*, 1995, **2**, 380.
21. A. R. Viguera and L. Serrano, *Prot. Sci.*, 1999, **8**, 1733.
22. C. Schellman, *Protein Folding*, 1980, 53.
23. R. Aurora, R. Srinivasan and G. D. Rose, *Science*, 1994, **264**, 1126.
24. R. Aurora and G. D. Rose, *Prot. Sci.*, 1998, **7**, 21.
25. A. J. Doig, M. W. MacArthur, B. J. Stapley and J. M. Thornton, *Prot. Sci.*, 1997, **6**, 147.
26. S. Penel, E. Hughes and A. J. Doig, *J. Mol. Biol.*, 1999, **287**, 127.
27. M. Siedlecka, G. Goch, A. Ejchart, H. Sticht and A. Bierzynski, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 903.
28. M. M. Lopez, D. H. Chin, R. L. Baldwin and G. I. Makhatadze, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 1298.
29. G. Goch, M. Maciejczyk, M. Oleszczuk, D. Stachowiak, J. Malicka and A. Bierzynski, *Biochemistry*, 2003, **42**, 6840.
30. J. M. Richardson, M. M. Lopez and G. I. Makhatadze, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 1413.
31. M. R. Ghadiri and C. Choi, *J. Am. Chem. Soc.*, 1990, **112**, 1630.
32. K. J. Kise and B. E. Bowler, *Biochemistry*, 2002, **41**, 15826.
33. F. Ruan, Y. Chen and P. B. Hopkins, *J. Am. Chem. Soc.*, 1990, **112**, 9403.
34. D. J. Cline, C. Thorpe and J. P. Schneider, *J. Am. Chem. Soc.*, 2003, **125**, 2923.
35. R. L. Beyer, H. N. Hoang, T. G. Appleton and D. P. Fairlie, *J. Am. Chem. Soc.*, 2004, **126**, 15096.
36. M. E. Karpen, P. L. De Haset and K. E. Neet, *Prot. Sci.*, 1992, **1**, 1333.
37. E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.*, 1984, **44**, 97.
38. G. Némethy, D. C. Phillips, S. J. Leach and H. A. Scheraga, *Nature*, 1967, **214**, 363.
39. G. L. Millhauser, *Biochemistry*, 1995, **34**, 3872.
40. K. A. Bolin and G. L. Millhauser, *Acc. Chem. Res.*, 1999, **32**, 1027.
41. G. N. Ramachandran and V. Sasisekharan, *Adv. Prot. Chem.*, 1968, **23**, 283.
42. B. W. Low and H. J. Grenville-Wells, *Proc. Natl. Acad. Sci. USA*, 1953, **39**, 785.
43. C. A. Rohl and A. J. Doig, *Prot. Sci.*, 1996, **5**, 1687.
44. W. A. Shirley and C. L. Brooks, *Proteins: Struct. Funct. Genet.*, 1997, **28**, 59.
45. K. H. Lee, D. R. Benson and K. Kuczera, *Biochemistry*, 2000, **39**, 13737.
46. T. M. Weaver, *Prot. Sci.*, 2000, **9**, 201.
47. D. M. Morgan, D. G. Lynn, H. Miller-Auer and S. C. Meredith, *Biochemistry*, 2001, **40**, 14020.
48. M. N. Fodje and S. Al-Karadaghi, *Prot. Eng.*, 2002, **15**, 353.
49. M. L. Tiffany and S. Krimm, *Biopolymers*, 1968, **6**, 1379.
50. E. J. Spek, Y. Gong and N. R. Kallenbach, *J. Am. Chem. Soc.*, 1995, **117**, 10773.
51. J. E. Brown and W. A. Klee, *Biochemistry*, 1971, **10**, 470.

52. S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 8898.
53. S. H. Park, W. Shalongo and E. Stellwagen, *Biochemistry*, 1993, **32**, 7048.
54. S. Marqusee, V. H. Robbins and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 5286.
55. P. C. Lyu, L. A. Marky and N. R. Kallenbach, *J. Am. Chem. Soc.*, 1989, **111**, 2733.
56. P. C. Lyu, M. I. Liff, L. A. Marky and N. R. Kallenbach, *Science*, 1990, **250**, 669.
57. P. J. Gans, P. C. Lyu, P. C. Manning, R. W. Woody and N. R. Kallenbach, *Biopolymers*, 1991, **31**, 1605.
58. S. Penel, R. G. Morrison, R. J. Mortishire-Smith and A. J. Doig, *J. Mol. Biol.*, 1999, **293**, 1211.
59. A. Wada, *Adv. Biophys.*, 1976, **9**, 1.
60. W. G. J. Hol, P. T. van Duijnen and H. J. C. Berendsen, *Nature*, 1978, **273**, 443.
61. J. Aqvist, H. Luecke, F. A. Quioco and A. Warshel, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 2026.
62. E. A. Zhukovsky, M. G. Mulkerrin and L. G. Presta, *Biochemistry*, 1994, **33**, 9856.
63. A. J. Doig and R. L. Baldwin, *Prot. Sci.*, 1995, **4**, 1325.
64. D. A. E. Cochran, S. Penel and A. J. Doig, *Prot. Sci.*, 2001, **10**, 463.
65. D. A. E. Cochran and A. J. Doig, *Prot. Sci.*, 2001, **10**, 1305.
66. T. M. Iqbalsyah and A. J. Doig, *Prot. Sci.*, 2004, **13**, 32.
67. S. M. Decatur, *Biopolymers*, 2000, 180.
68. A. J. Doig, A. Chakrabartty, T. M. Klingler and R. L. Baldwin, *Biochemistry*, 1994, **33**, 3396.
69. C. A. Rohl, A. Chakrabartty and R. L. Baldwin, *Prot. Sci.*, 1996, **5**, 2623.
70. B. M. Huyghues-Despointes, J. M. Scholtz and R. L. Baldwin, *Prot. Sci.*, 1993, **2**, 1604.
71. J. R. Brandts and K. J. Kaplan, *Biochemistry*, 1973, **10**, 470.
72. H. Edelhofer, *Biochemistry*, 1967, **6**, 1948.
73. A. Chakrabartty, T. Kortemme, S. Padmanabhan and R. L. Baldwin, *Biochemistry*, 1993, **32**, 5560.
74. C. D. Andrew, S. Bhattacharjee, N. Kokkoni, J. D. Hirst, G. R. Jones and A. J. Doig, *J. Am. Chem. Soc.*, 2002, **124**, 12706.
75. D. S. Kemp, J. G. Boyd and C. C. Muendel, *Nature*, 1991, **352**, 451.
76. D. S. Kemp, T. J. Allen and S. L. Oslick, *J. Am. Chem. Soc.*, 1995, **117**, 6641.
77. K. Groebke, P. Renold, K. Y. Tsang, T. J. Allen, K. F. McClure and D. S. Kemp, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 4025.
78. D. S. Kemp, S. L. Oslick and T. J. Allen, *J. Am. Chem. Soc.*, 1996, **118**, 4249.
79. D. S. Kemp, T. J. Allen, S. L. Oslick and J. G. Boyd, *J. Am. Chem. Soc.*, 1996, **118**, 4240.

80. R. E. Austin, R. A. Maplestone, A. M. Seffler, K. Liu, W. N. Hruzewicz, C. W. Liu, H. S. Cho, D. E. Wemmer and P. A. Bartlett, *J. Am. Chem. Soc.*, 1997, **119**, 6461.
81. T. Arrhenius and A. C. Satherthwait, *Peptides: Chemistry, Structure and Biology: Proceedings of the 11th American Peptide Symposium*, 1989, p. 870.
82. K. Muller, D. Obrecht, A. Knierzinger, C. Stankovic, C. Spiegler, W. Bannwarth, A. Trzeciak, G. Englert, A. M. Labhardt and P. Schonholzer, in *Perspectives in Medicinal Chemistry*, John Wiley & Sons, 1993.
83. D. Gani, A. Lewis, T. Rutherford, J. Wilkie, I. Stirling, T. Jenn and M. D. Ryan, *Tetrahedron*, 1998, **54**, 15793.
84. D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers*, Academic Press, New York and London, 1970.
85. H. Qian and J. A. Schellman, *J. Phys. Chem.*, 1992, **96**, 3987.
86. A. J. Doig, *Biophys. Chem.*, 2002, **101–102**, 281.
87. B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, 1959, **31**, 526.
88. S. Lifson and A. Roig, *J. Chem. Phys.*, 1961, **34**, 1963.
89. J. K. Sun and A. J. Doig, *Prot. Sci.*, 1998, **7**, 2374.
90. F. B. Sheinerman and C. L. Brooks, *J. Am. Chem. Soc.*, 1995, **117**, 10098.
91. P. J. Flory, *Statistical Mechanics of Chain Molecules*, Wiley, New York, 1969.
92. K. Chen, Z. G. Liu, C. H. Zhou, Z. S. Shi and N. R. Kallenbach, *J. Am. Chem. Soc.*, 2005, **127**, 10146.
93. A. K. Jha, A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick and K. F. Freed, *Biochemistry*, 2005, **44**, 9691.
94. A. Rath, A. R. Davidson and C. M. Deber, *Biopolymers*, 2005, **80**, 179.
95. E. W. Blanch, L. A. Morozova-Roche, D. A. E. Cochran, A. J. Doig, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 2000, **301**, 553.
96. R. V. Pappu and G. D. Rose, *Prot. Sci.*, 2002, **11**, 2437.
97. Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 9190.
98. K. Chen, Z. G. Liu and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 15352.
99. S. A. Asher, A. V. Mikhonin and S. Bykov, *J. Am. Chem. Soc.*, 2004, **126**, 8433.
100. A. Kentsis, M. Mezei, T. Gindin and R. Osman, *Proteins: Struct. Funct. Bioinf.*, 2004, **55**, 493.
101. M. Mezei, P. J. Fleming, R. Srinivasan and G. D. Rose, *Proteins: Struct. Funct. Bioinf.*, 2004, **55**, 502.
102. A. L. Rucker, C. T. Pager, M. N. Campbell, J. E. Qualis and T. P. Creamer, *Proteins: Struct. Funct. Genet.*, 2003, **53**, 68.
103. N. H. Andersen and H. Tong, *Prot. Sci.*, 1997, **6**, 1920.
104. B. J. Stapley, C. A. Rohl and A. J. Doig, *Prot. Sci.*, 1995, **4**, 2383.
105. P. Argos and J. Palau, *Int. J. Pept. Prot. Res.*, 1982, **19**, 380.
106. S. Kumar and M. Bansal, *Proteins*, 1998, **31**, 460.
107. W. Shalongo and E. Stellwagen, *Prot. Sci.*, 1995, **4**, 1161.

108. J. K. Sun, S. Penel and A. J. Doig, *Prot. Sci.*, 2000, **9**, 750.
109. J. M. Scholtz, H. Qian, V. H. Robbins and R. L. Baldwin, *Biochemistry*, 1993, **32**, 9668.
110. V. Muñoz and L. Serrano, *Nat. Struct. Biol.*, 1994, **1**, 399.
111. V. Muñoz and L. Serrano, *Biopolymers*, 1997, **41**, 495.
112. J. Fernández-Recio, A. Vásquez, C. Civera, P. Sevilla and J. Sancho, *J. Mol. Biol.*, 1997, **267**, 184.
113. E. Lacroix, A. R. Viguera and L. Serrano, *J. Mol. Biol.*, 1998, **284**, 173.
114. V. Muñoz and L. Serrano, *J. Mol. Biol.*, 1995, **245**, 275.
115. M. Petukhov, V. Muñoz, N. Yumoto, S. Yoshikawa and L. Serrano, *J. Mol. Biol.*, 1998, **278**, 279.
116. A. L. Lomize and H. I. Mosberg, *Biopolymers*, 1997, **42**, 239.
117. J. Wojcik, K. H. Altmann and H. A. Scheraga, *Biopolymers*, 1990, **30**, 121.
118. S. Padmanabhan, E. J. York, L. Gera, J. M. Stewart and R. L. Baldwin, *Biochemistry*, 1994, **33**, 8604.
119. C. N. Pace and J. M. Scholtz, *Biophys. J.*, 1998, **75**, 422.
120. C. A. Rohl, W. Fiori and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 3682.
121. R. J. Kennedy, K.-Y. Tsang and D. S. Kemp, *J. Am. Chem. Soc.*, 2002, **124**, 934.
122. P. C. Lyu, H. X. X. Zhou, N. Jelveh, D. E. Wemmer and N. R. Kallenbach, *J. Am. Chem. Soc.*, 1992, **114**, 6560.
123. M. Petukhov, K. Uegaki, N. Yumoto, S. Yoshikawa and L. Serrano, *Prot. Sci.*, 1999, **8**, 2144.
124. H. X. X. Zhou, P. C. C. Lyu, D. E. Wemmer and N. R. Kallenbach, *J. Am. Chem. Soc.*, 1994, **116**, 1139.
125. M. Petukhov, K. Uegaki, N. Yumoto and L. Serrano, *Prot. Sci.*, 2002, **11**, 766.
126. D. N. Ermolenko, J. M. Richardson and G. I. Makhatadze, *Prot. Sci.*, 2003, **12**, 1169.
127. L. Szalik, J. Moitra, D. Krylov and C. Vinson, *Nat. Struct. Biol.*, 1997, **4**, 112.
128. S. Liehr and H. K. Chenault, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2759.
129. C. D. Andrew, J. Warwicker, G. R. Jones and A. J. Doig, *Biochemistry*, 2002, **41**, 1897.
130. N. Errington and A. J. Doig, *Biochemistry*, 2005, **44**, 7553.
131. A. Horovitz, L. Serrano, B. Avron, M. Bycroft and A. R. Fersht, *J. Mol. Biol.*, 1990, **216**, 1031.
132. G. Merutka and E. Stellwagen, *Biochemistry*, 1991, **30**, 1591.
133. E. Stellwagen, S.-H. Park, W. Shalongo and A. Jain, *Biopolymers*, 1992, **32**, 1193.
134. B. M. Huyghues-Despointes, J. M. Scholtz and R. L. Baldwin, *Prot. Sci.*, 1993, **2**, 80.
135. B. M. Huyghues-Despointes and R. L. Baldwin, *Biochemistry*, 1997, **36**, 1965.

136. B. M. Huyghues-Despointes, T. M. Klingler and R. L. Baldwin, *Biochemistry*, 1995, **34**, 13267.
137. B. J. Stapley and A. J. Doig, *J. Mol. Biol.*, 1997, **272**, 465.
138. S. Padmanabhan and R. L. Baldwin, *Prot. Sci.*, 1994, **3**, 1992.
139. S. Padmanabhan and R. L. Baldwin, *J. Mol. Biol.*, 1994, **241**, 706.
140. A. R. Viguera and L. Serrano, *Biochemistry*, 1995, **34**, 8771.
141. L. K. Tsou, C. D. Tatko and M. L. Waters, *J. Am. Chem. Soc.*, 2002, **124**, 14917.
142. C. D. Andrew, S. Penel, G. R. Jones and A. J. Doig, *Proteins: Struct. Funct. Genet.*, 2001, **45**, 449.
143. L. Mayne, S. W. Englander, R. Qiu, J. X. Yang, Y. X. Gong, E. J. Spek and N. R. Kallenbach, *J. Am. Chem. Soc.*, 1998, **120**, 10643.
144. C. A. Olson, E. J. Spek, Z. S. Shi, A. Vologodskii and N. R. Kallenbach, *Proteins: Struct. Funct. Bioinf.*, 2001, **44**, 123.
145. T. I. Iqbalsyah and A. J. Doig, *J. Am. Chem. Soc.*, 2005, **127**, 5002.
146. Z. Shi, C. A. Olson and N. R. Kallenbach, *J. Am. Chem. Soc.*, 2002, **124**, 3284.
147. Z. Shi, C. A. Olson, A. J. Bell and N. R. Kallenbach, *Biophys. Chem.*, 2002, **101–102**, 267.
148. J. W. Taylor, *Biopolymers*, 2002, **66**, 49.
149. R. M. Campbell, J. Bongers and A. M. Felxi, *Biopolymers*, 1995, **37**, 67.
150. M. Chorev, E. Roubini, R. L. McKee, S. W. Gibbons, M. E. Goldman, M. P. Caulfield and M. Rosenblatt, *Biochemistry*, 1991, **30**, 5968.
151. G. Osapay and J. W. Taylor, *J. Am. Chem. Soc.*, 1990, **112**, 6046.
152. M. Bouvier and J. W. Taylor, *J. Med. Chem.*, 1992, **35**, 1145.
153. A. Kapurniotu and J. W. Taylor, *J. Med. Chem.*, 1995, **38**, 836.
154. C. Bracken, J. Gulyas, J. W. Taylor and J. Baum, *J. Am. Chem. Soc.*, 1994, **116**, 6431.
155. S. T. Chen, H. J. Chen, H. M. Yu and K. T. Wang, *J. Chem. Res. (S)*, 1993, **6**, 228.
156. W. T. Zhang and J. W. Taylor, *Tet. Lett.*, 1996, **37**, 2173.
157. P. Z. Luo, D. T. Braddock, R. M. Subramanian, S. C. Meredith and D. G. Lynn, *Biochemistry*, 1994, **33**, 12367.
158. D. Y. Jackson, D. S. King, J. Chmielewski, S. Singh and P. G. Schultz, *J. Am. Chem. Soc.*, 1991, **113**, 9391.
159. J. R. Kumita, O. S. Smart and G. A. Woolley, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 3803.
160. D. G. Flint, J. R. Kumita, O. S. Smart and G. A. Wolley, *Chem. Biol.*, 2002, **9**, 391.
161. J. R. Kumita, D. G. Flint, O. S. Smart and G. A. Woolley, *Prot. Eng.*, 2002, **15**, 561.
162. M. A. Jiménez, V. Muñoz, M. Rico and L. Serrano, *J. Mol. Biol.*, 1994, **242**, 487.
163. N. R. Kallenbach and Y. X. Gong, *Bioorg. Med. Chem.*, 1999, **7**, 143.
164. M. Petukhov, N. Yumoto, S. Murase, R. Onmura and S. Yoshikawa, *Biochemistry*, 1996, **35**, 387.

165. A. R. Viguera and L. Serrano, *J. Mol. Biol.*, 1995, **251**, 150.
166. K. M. Armstrong and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 11337.
167. J. S. Smith and J. M. Scholtz, *Biochemistry*, 1998, **37**, 33.
168. D. J. Lockhart and P. S. Kim, *Science*, 1993, **260**, 198.
169. 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*, 1991, **88**, 2854.
170. G. Yoder, P. Pancoska and T. A. Keiderling, *Biochemistry*, 1997, **36**, 5123.
171. C. Y. Huang, J. W. Klemke, Z. Getahun, W. F. DeGrado and F. Gai, *J. Am. Chem. Soc.*, 2001, **123**, 9235.
172. S. T. Thomas, V. V. Loladze and G. I. Makhataдзе, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 10670.
173. R. Luo, L. David, H. Hung, J. Devaney and M. K. Gilson, *J. Phys. Chem. B*, 1999, **103**, 366.
174. S. Marqusee and R. T. Sauer, *Prot. Sci.*, 1994, **3**, 2217.
175. P. Luo and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 4930.
176. Z. Shi, C. A. Olson, A. J. Bell and N. R. Kallenbach, *Biopolymers*, 2001, **60**, 366.