Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 10907–10908, December 1993

Commentary

What determines where α-helices begin and end?

Lynne Regan

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511

A model for the structure of the α-helix was first proposed by Pauling and Corey (1) more than 40 years ago. By taking into account only the steric and chemical constraints imposed by the polypeptide backbone and assuming that all residues were equivalent, they proposed a helical structure with a 3.6-residue repeat period and a pitch of 5.4 Å.

Since that initial prediction, high-resolution x-ray crystallographic and NMR studies of proteins have confirmed the model and revealed the α-helix to be a ubiquitous element of protein secondary structure. Subsequently, statistical surveys of proteins of known structure have shown that there is a decidedly nonrandom distribution of amino acids, with certain amino acids showing very high probabilities of occurring in α-helical secondary structure, whereas others are much more likely to be found in β-sheets (2).

These later observations led to efforts to develop model systems in which to measure the equilibrium constant for each amino acid to undergo the transition from "random coil" to α-helix (its "propensity" for α-helix formation) and to determine whether this experimentally measurable parameter correlated with the statistically observed preferences.

Measurements of α-helical propensities were originally made in a random copolymer system (3). However, comparisons between the equilibrium constants measured in this complex system and the statistically observed probabilities were inconclusive.

The majority of recent studies have used peptide model systems, motivated by the observation that certain short monomeric peptides display significant α-helix content in aqueous solution at low temperatures (4, 5). The peptide systems are more uniquely defined and allow many aspects of α-helix formation to be studied in greater detail than was possible in the earlier studies.

Several groups have used peptide models to determine the effect of specific amino acid substitutions on helicity and stability (6–9). The results of these studies show good correlation with the results of statistical surveys of proteins of known structure: Amino acids that are frequently found in α-helices in proteins display measurably higher propensities to adopt a helical conformation in small peptides. Nevertheless, the underlying structural and energetic bases for the differences in α-helical propensities are not understood and remain a topic for investigation and debate.

Once it is accepted that certain amino acids are more likely than others to form α-helices, it is natural to pose the question, "What determines where the helices begin and end?" Algorithms that predict protein secondary structure, based on helical propensities, assume that a helix will terminate around the position where the average propensity to be helical drops below a certain value (2). Although this procedure works reasonably well in secondary structure prediction, it is unsatisfying from the perspective of trying to understand the underlying rules that specify protein structure.

Two groups have addressed the issue of helix termination explicitly by studying the distribution of amino acids within α-helices in proteins of known structure (10, 11). Their specific aim was to compare the distribution of residues at the ends of helices with those in the center. Both surveys concluded that there were certain amino acids that were much more likely to be found at the end boundaries of helices than at interior positions. This observation led to the proposal that the N- and C-terminal residues might play a very specific role in ensuring energetically favorable helix termination. These residues were named the N-cap and C-cap, and it was proposed that there might be specific interactions between the side chains of the "capping" residues and the unfulfilled H-bond donors and acceptors that are present at the N and C termini of the helix.

Two papers (12, 13) in this issue of the *Proceedings* present the results of studies that address the issue of helix termination in monomeric peptide models. Chakrabartty *et al.* (12) compare the N- and C-cap preferences in peptide models by systematically substituting each noncharged amino acid and determining the changes in α-helix stability that occur as a consequence of these substitutions. In a related work by Armstrong and Baldwin (13), the results of systematically varying the position in the helix of a charged His residue are described.

The details of the study by Chakrabartty *et al.* (12) are as follows. The reference peptides are based on the sequence $(AAKAA)_n$ (where $n = 2$ or 3) with an additional single Tyr residue at the N terminus to allow accurate determination of peptide concentrations. The amino acid composition was chosen to generate a peptide with relatively high starting helix content that would also be reasonably soluble in aqueous solution. A family of peptides was synthesized with individual substitutions of each noncharged amino acid at the N- and C-cap positions.

When working with peptides, an important decision is how to treat the N and C termini. The N terminus can be left as NH_2 and the C terminus can be left as COOH or, as this is often the choice, the N terminus can be acetylated and the C terminus can be synthesized as $CONH_2$. These latter modifications have the effect of removing the potentially charged end groups, perhaps mimicking more closely the environment of an α-helix within a protein. Initially, Chakrabartty *et al.* (12) worked with peptides in which the N and C termini were blocked in this fashion. They saw essentially no difference in helix stability, regardless of the identity of the N- and C-cap residues. In the current study (12), the N- and C-cap substitutions were repeated, but this time in peptides with free N and C termini. The results are dramatically different from those observed for peptides with blocked termini.

For the N-cap substitutions in the unblocked peptides, experiments were performed at a pH at which the N terminus would be neutral but at which the Lys residues would be charged. Peptide stabilities were ranked on the basis of mean residue ellipticity at 222 nm. A peptide that is 100% helical would be expected to have a mean residue ellipticity of approximately $-30{,}000$ deg·cm^2·dmol^{-1}. Substantial differences in helicity were observed, ranging from a mean residue ellipticity of only -820 deg·cm^2·dmol^{-1} for Gln, the worst N-cap residue, to -6900 deg·cm^2·dmol^{-1} for Asn, the best N-cap residue. Moreover, the rank order of N-terminal capping ability determined in this study correlates well with the N-cap preferences reported in the previous sur-

veys (10, 11). Asn, Gly, Ser, and Thr are clearly the best N-cap residues.

These interesting results prompt the question, "What determines the rank order of N-cap propensities?" There is not an easy answer. The preponderance of small polar side chains is consistent with the hypothesis that the side chains at the N-cap make H-bonds to the main-chain atoms of residues at the N terminus. This cannot be the only explanation, however, because Gly also has a very high N-cap propensity.

It has been proposed by Serrano and coworkers (14, 15) that perhaps the need is for the N-cap side chains to make H-bonds to the peptide backbone themselves or for the N-cap side chains to be small enough to allow access of solvent molecules to H-bond in their place. In this scenario, the explanation for why larger nonpolar residues make poor N-caps would be that their side chains cannot H-bond to the peptide backbone themselves and that their large size prevents the interaction of the polar NH groups with solvent. Further studies are required to unravel these complex effects.

In parallel studies with uncharged C-cap residues, much smaller effects on peptide stability were observed. This does not mean that the identity of the C-terminal residue has no influence on helix stability. In the paper by Armstrong and Baldwin (13), the effect of a charged His residue at various positions in an α-helix was the focus of investigation. In this study, the reference peptides were Ac-Y(AAKAA)$_3$A-NH$_2$ and Ac-Y(AAQAA)$_3$A-NH$_2$ (note the blocked N and C termini). Again the peptides were chosen to be helical and soluble in aqueous solution. In addition, because in this study the test residue is charged, one reference peptide was chosen to be charged and the other was neutral. The effect on peptide stability of introducing a single charged His residue at various locations in the helix was again monitored by determining the ellipticity at 222 nm as a measure of helical content.

In both reference peptides, the authors (13) find that peptide stability is strongly influenced by the location of the charged His residue. The measured helicity varies from 12% when His is at position 6, to 24% when His is at position 3, and to 76% when His is at position 17. The helicity measured when the His residue is at position 17, or at any of the three C-terminal most positions of the α-helix, is unexpectedly high. This result suggests that at these locations a His residue may play a specific role in helix termination. A similar mechanism to N-capping can be invoked, with the His side chain being capable of H-bonding to a main-chain CO group. At the C terminus, where there are unsatisfied H-bonds, this effect is stabilizing. The same effect could explain why His is so destabilizing at interior positions, such as position 6. Here, the His side chain could potentially compete with "normal" helix H-bonding and as a consequence disrupt and destabilize the structure.

In summary, the papers of Chakrabartty *et al.* (12) and of Armstrong and Baldwin (13) represent a significant advance in our understanding of the mechanisms by which the beginnings and ends of α-helices are specified. Clearly, several mechanistic details remain unresolved, but the simple peptide models for helix termination described in these papers provide tractable experimental systems in which such issues can be addressed.

1. Pauling, L. & Corey, R. P. (1950) *J. Am. Chem. Soc.* **72,** 5349–5356.
2. Chou, P. Y. & Fasman, G. (1974) *Biochemistry* **13,** 211–222.
3. Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y. & Scheraga, H. A. (1984) *Macromolecules* **17,** 148–155.
4. Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1987) *Nature (London)* **326,** 563–567.
5. Marqusee, S., Robbins, V. H. & Baldwin, R. L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5286–5290.
6. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M. & Baldwin, R. L. (1990) *Nature (London)* **344,** 268–270.
7. Merkuta, G. & Stellwagen, E. (1990) *Biochemistry* **29,** 894–898.
8. Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. R. (1990) *Science* **250,** 669–673.
9. O'Neil, K. T. & DeGrado, W. F. (1990) *Science* **250,** 646–651.
10. Presta, L. G. & Rose, G. D. (1988) *Science* **240,** 1632–1641.
11. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240,** 1648–1652.
12. Chakrabartty, A., Doig, A. J. & Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11332–11336.
13. Armstrong, K. M. & Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11337–11340.
14. Serrano, L. & Fersht, A. R. (1989) *Nature (London)* **342,** 296–299.
15. Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992) *J. Mol. Biol.* **227,** 544–559.