Protein Science (1998), 7:21–38. Cambridge University Press. Printed in the USA.

REVIEW

Helix capping[1]

RAJEEV AURORA AND GEORGE D. ROSE

Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205

(RECEIVED June 12, 1997; ACCEPTED July 9, 1997)

Abstract

Helix-capping motifs are specific patterns of hydrogen bonding and hydrophobic interactions found at or near the ends of helices in both proteins and peptides. In an α-helix, the first four >N−H groups and last four >C=O groups necessarily lack intrahelical hydrogen bonds. Instead, such groups are often capped by alternative hydrogen bond partners. This review enlarges our earlier hypothesis (Presta LG, Rose GD. 1988. Helix signals in proteins. *Science 240*:1632–1641) to include hydrophobic capping. A hydrophobic interaction that straddles the helix terminus is always associated with hydrogen-bonded capping. From a global survey among proteins of known structure, seven distinct capping motifs are identified—three at the helix N-terminus and four at the C-terminus. The consensus sequence patterns of these seven motifs, together with results from simple molecular modeling, are used to formulate useful rules of thumb for helix termination. Finally, we examine the role of helix capping as a bridge linking the conformation of secondary structure to supersecondary structure.

Keywords: alpha helix; protein folding; protein secondary structure

The α-helix is characterized by consecutive, main-chain, $i \rightarrow i-4$ hydrogen bonds between each amide hydrogen and a carbonyl oxygen from the adjacent helical turn (Pauling & Corey, 1951). This pattern (Fig. 1) is interrupted at helix termini because, upon termination, no turn of helix follows to provide additional hydrogen bond partners. Such end effects are substantial, encompassing two-thirds of the residues for the protein helix of average length (Presta & Rose, 1988). Further, helix geometry hinders solvent access to amide groups in the first turn of the helix, inhibiting interaction with water and necessitating alternative hydrogen bonds. The term helix "capping" (Richardson & Richardson, 1988a) has been used to describe such alternative hydrogen bond patterns that can satisfy backbone >N−H and >C=O groups in the initial and final turns of the helix (Presta & Rose, 1988).

Many studies involving helix capping have been conducted since publication of our initial hypothesis nine years ago (Presta & Rose, 1988). As we had proposed, amide hydrogens at the helix N-terminus are indeed satisfied predominantly by side-chain H-bond acceptors. In contrast, carbonyl oxygens at the C-terminus are satisfied primarily by backbone >N−H groups from the turn following the helix. Further, these hydrogen-bonding patterns at either helix end are accompanied by a companion hydrophobic interaction between apolar residues in the α-helix and its flanking turn. This hydrophobic component of helix capping was unanticipated.

The main purpose of this review is to enlarge our previous definition of helix capping and to document the common capping motifs. Typically, protein helices terminate in a hydrophobic interaction that straddles the helix termini (i.e., an interaction between two hydrophobic residues close in sequence, one within the helix, the other external to the helix). In this interaction, the polypeptide chain folds back upon itself and, inescapably, occludes polar backbone groups near the helix terminus, thereby inhibiting intermolecular H-bonds with solvent water. Even in the absence of this hydrophobic interaction, helix geometry interferes with solvent access to amide groups at the helix N-terminus. Solvent-shielded polar groups within the initial/final helical turn cannot be satisfied by classical intrahelical hydrogen bonds. Left unsatisfied, they would destabilize the molecule by the equivalent of several hydrogen bonds per helix, an unacceptable energy penalty. These combined constraints at helix ends result in a small number of distinct conformational arrangements that can provide intramolecular H-bond partners while maintaining the hydrophobic interaction. Most common among such structures are the seven motifs described in this review. The capping hydrogen bonds expressed in these motifs must be especially favorable because they are often detected in isolated peptide helices that lack the hydrophobic interaction.

Nomenclature for helices and their flanking residues is as follows:

$$\cdots\text{-N''-N'-Ncap-N1-N2-N3-}\cdots\text{-C3-C2-C1-Ccap-C'-C''-}\cdots$$

Reprint requests to: George D. Rose, Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205; e-mail: rose@grserv.med.jhmi.edu.

[1]Dedicated to Robert L. Baldwin on the occasion of his 70th birthday.

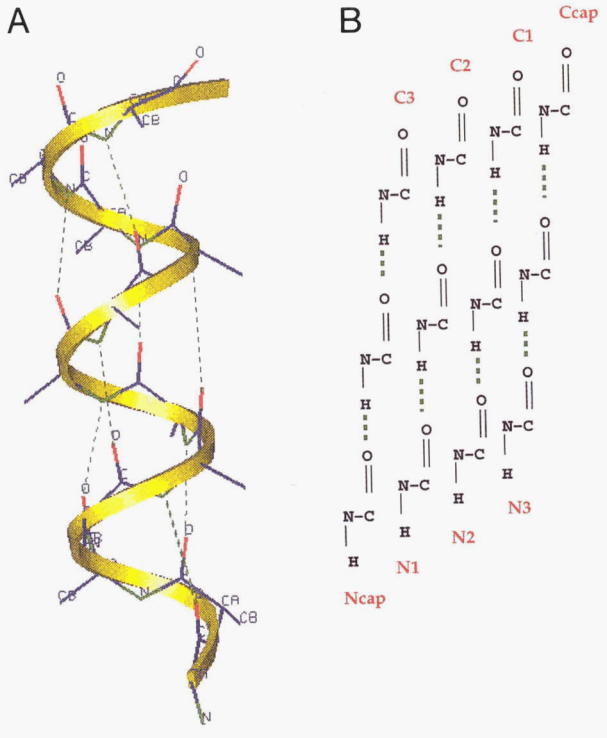


Fig. 1. Two views of the α-helix. **A:** Cartoon with yellow ribbon tracing the helical path of the peptide backbone. N-to-C orientation is from bottom to top; side-chain atoms beyond Cβ are omitted for clarity. Hydrogen bonds ($i \rightarrow i - 4$) between successive amides (green) and carbonyl oxygens (red) are shown as dashed lines. Backbone atoms are labeled in the first and last four residues. Capping by side-chain acceptors is *common* at the N-terminus. There, the C$\alpha \rightarrow$ Cβ vector, which is oriented toward the N-terminus, poises side-chain atoms in capping-competent positions. By definition, Ncap does not have backbone dihedral angles with helical values, further facilitating side-chain capping between N3 and Ncap. Capping by side-chain donors is *uncommon* at the C-terminus. The C$\alpha \rightarrow$ Cβ vector is oriented away from the C-terminus. A side chain within the helix would have to execute a "U-turn" to cap any of the last four carbonyl oxygens. **B:** Helical net representation of the helix in (A). The first four N—H groups (i.e., Ncap through N3) and last four C=O groups (i.e., C3 through Ccap) cannot make intrahelical hydrogen bonds.

where N1 through C1 belong to the helix proper and the primed residues belong to turns that bracket the helix at either end. Ncap and Ccap are bridge residues that belong both to the helix and an adjacent turn. In practice, residues classified as helical have backbone dihedral angles, ϕ and ψ, with mean values of $-64 \pm 7°$ and $-41 \pm 7°$, respectively. Ncap and Ccap make one additional intrahelical hydrogen bond while departing from these means (Presta & Rose, 1988). In this notation, the index of nonhelical residues will be denoted either by iterated prime marks or, when that becomes too cumbersome, by an ordinal number followed by a single prime mark; e.g., either N''' or $N^{3\prime}$ indicates the third residue prior to Ncap.

Understanding molecular architecture requires the ability to parse a protein into its constituent parts. Peptide chain turns link secondary structure into units of supersecondary structure, which, in turn, interact iteratively to form larger modules, leading ultimately to tertiary structure (Rose, 1979). This hierarchic organization is a characteristic feature of globular proteins, and it invites automatic classification.

The automatic classification of residues into categories of secondary structure is not a trivial task. Whereas repetitive secondary structure (i.e., helix and sheet) is conspicuous in protein models, non-repetitive structure (i.e., turns and loops) can be visually ambiguous. Many algorithms have been devised to identify secondary structure from coordinates (Rose & Seltzer, 1977; Richardson, 1981; Kabsch & Sander, 1983; Richards & Kundrot, 1988; Rooman et al., 1990; Fetrow et al., 1997), giving rise to differing, albeit self-consistent, assignments. Related reports describe "standard structures" at the ends of α-helices and β-strands (Edwards et al., 1987; Efimov, 1993; Oliva et al., 1997); see also Cordes et al. (1996). Such structures are peptide chain turns, for which various classification (Rose et al., 1985) and prediction schemes (Wilmot & Thornton, 1988; Hutchinson & Thornton, 1994) have also been developed. These approaches falter at helix ends where an unavoidable structural ambiguity blurs the boundary between the helix terminus and its flanking turn. One useful byproduct of helix capping is that it provides a natural means of classification. The end of a helix is punctuated by a capping motif, just as this sentence is terminated by a period.

Numerous experiments demonstrate that capping stabilizes α-helices in both proteins (Serrano & Fersht, 1989; Bell et al., 1992; Thapar et al., 1996) and peptides (Lyu et al., 1990; Bruch et al., 1991; Chakrabartty et al., 1993; Forood et al., 1993; Yumoto et al., 1993; Zhou & Wemmer, 1994; Odaert et al., 1995; Viguera & Serrano, 1995; Petukhov et al., 1996; Esposito et al., 1997; Reymond et al., 1997; Sukumar & Gierasch, 1997). In addition, studies of the folding kinetics of barnase (Serrano et al., 1992a) and lysozyme (Radford et al., 1992) indicate that helix capping can be an early folding event.

This review presents the results of a global survey of helix capping in proteins of known structure. From the survey, it is apparent that a small repertoire of structural motifs is sufficient to describe many instances of helix termination. Each motif exhibits a characteristic pattern of hydrogen bonds together with a hydrophobic interaction. Some, but not all, of the motifs have been described previously, viz., the capping box (Dasgupta & Bell, 1993; Harper & Rose, 1993; Jimenez et al., 1994; Seale et al., 1994) and the glycine motifs (Schellman, 1980; Bork & Preissner, 1991; Preissner & Bork, 1991; Aurora et al., 1994).

Complementing this global census, simple modeling is used to rationalize the existence of the discovered motifs. Modeling demonstrates that residues at helix termini are highly constrained by local interactions and excluded volume effects. These constraints limit helix termination to a small number of possible structures, and this fact is used to show that our motif definitions are well posed.

In addition to the motifs emphasized in this review, both long-range and intermolecular interactions can cap helices. Long-range capping comprises those cases in which hydrogen bonds and/or hydrophobic interactions are provided by non-local partners (i.e., by groups farther than a few residues in sequence from Ncap/Ccap). Such cases resist classification into identifiable motifs because there are too many conceivable conformations. Also, solvent (e.g., Sundaralingam & Sekharudu, 1989) and other bound molecules (e.g., Quiocho et al., 1987) can participate in capping hydrogen bonds. Here again, both the range and types of interactions are too heterogeneous to invite ready classification. It is also important to emphasize the fact that proteins are complex molecules, and, in some cases, local interactions that foster the observed motifs are superceded by context-dependent singularities.

Capping interactions can be divided into short-range, mid-range and long-range. Short-range interactions are those in which hydrogen bonds and/or hydrophobic interactions for groups in the first/last turn of the helix are provided by partners within an interval of three or four residues in sequence from Ncap/Ccap, and they include the seven motifs described here. Long-range interactions are those in which capping partners are more than seven residues in sequence from Ncap/Ccap. Remaining interactions are classified as mid-range. Most helices are capped by short-and mid-range interactions, as described next.

Capping in 1316 protein helices

The March 1994 pdb_select data set (Hobohm & Sander, 1994) was used in this analysis; it includes 1316 helices, each at least seven residues in length, taken from 274 polypeptide chains in the Protein Data Base (Bernstein et al., 1977). A helix was identified as a series of consecutive residues with backbone dihedral angles near the observed mean values for α-helices ($\phi = -60 \pm 15°$ and $\psi = -40 \pm 15°$). Ncap/Ccap positions were defined as the first/last residue of the series with an N—H(i) $\rightarrow$ C=O(i − 4) backbone hydrogen bond. Hydrogen bonding criteria were established from small-molecule crystal structures, as described previously (Presta & Rose, 1988; Stickle et al., 1992). These data were used to catalog residue preferences in helices, to characterize patterns of helix-terminating hydrogen bonding and hydrophobic interactions, and to define the seven most commonly observed capping motifs.

Residue preferences in helices

Table 1 lists the observed residue preferences in these data as a function of helix position. Results are similar but not identical to those of Richardson and Richardson (1988a); presumably, such differences reflect differing definitions of helix termination. Frequencies in Table 1 have been normalized by dividing the fractional occurrence of each residue at every helical position by the fractional occurrence of that residue in the entire data set. A normalized frequency of unity indicates no preference—that is, the frequency of occurrence of the given residue in that particular position is the same as its frequency at large. Normalized frequencies greater than or less than unity indicate selection for or against the given residue in a particular position. For example, residues

Table 1. *Normalized positional residue frequency at helix termini*[a]

	Helix position																			
AA	N4′	N‴	N″	N′	Nc	N1	N2	N3	N4	N5	C5	C4	C3	C2	C1	Cc	C′	C″	C‴	C4′
G	1.18	1.25	1.26	1.14	0.98	0.55	0.65	0.56	0.37	0.32	0.30	0.29	0.32	0.20	0.33	0.74	*2.74*	1.08	0.97	1.07
A	0.94	0.98	1.05	0.75	0.67	1.10	1.39	1.43	*1.55*	*1.80*	*1.52*	1.49	1.73	1.33	*1.87*	1.19	0.77	0.93	1.09	0.71
V	0.90	0.87	0.62	0.58	0.67	0.76	0.70	1.14	1.18	0.81	1.03	0.94	0.94	1.08	0.51	0.46	0.53	0.74	0.77	1.12
I	1.07	0.88	0.95	0.80	0.78	1.06	0.64	1.18	1.47	1.09	1.25	1.04	1.15	1.58	0.90	0.61	0.64	1.14	0.97	1.05
L	0.95	0.80	0.96	1.01	0.79	0.84	0.91	*1.52*	1.36	1.47	1.26	1.40	*1.80*	*1.63*	1.65	1.36	0.66	1.16	0.87	0.84
F	1.06	1.12	0.95	0.88	0.96	0.90	1.00	1.10	1.39	0.96	1.14	0.86	1.35	1.22	0.67	1.20	1.04	1.05	0.84	0.95
P	1.18	1.31	1.05	*1.33*	1.12	*1.67*	0.94	0.15	0.03	0.15	0.44	0.20	0.07	0.07	0.03	0.10	0.66	1.01	2.01	1.70
M	0.88	1.12	0.99	1.02	0.98	0.90	1.10	1.68	*2.13*	1.64	1.14	*1.84*	*2.21*	1.76	1.35	1.35	0.74	1.11	0.96	0.80
W	0.91	0.90	1.06	0.68	0.94	1.26	1.10	1.68	1.10	0.68	0.68	1.52	1.57	1.00	1.00	0.58	0.58	1.06	0.91	1.25
C	0.60	0.41	0.60	0.66	0.37	0.26	0.52	0.52	0.59	0.55	0.26	0.37	0.63	0.44	0.33	0.44	0.44	0.67	0.26	0.65
S	0.69	1.02	0.96	1.20	1.25	0.81	0.69	0.61	0.44	0.67	0.66	0.68	0.65	0.42	0.71	1.02	0.64	0.71	0.76	0.65
T	0.87	0.80	1.03	1.13	*1.41*	0.77	0.92	0.75	0.65	0.70	0.73	0.79	0.46	0.57	0.50	0.82	0.82	0.84	0.79	0.86
N	0.79	1.05	0.91	1.24	1.28	0.72	0.67	0.55	0.60	0.73	0.58	0.67	0.70	0.64	0.70	*1.33*	1.39	0.82	1.03	0.95
Q	0.94	0.90	0.87	1.08	1.05	1.31	1.60	1.43	1.43	0.97	1.41	1.52	0.88	1.37	1.24	1.43	0.95	1.02	1.08	0.87
Y	1.04	1.12	0.94	0.80	0.82	0.99	0.73	0.65	0.93	0.91	1.04	1.06	1.10	1.02	0.73	1.06	0.93	1.15	0.64	0.85
H	1.15	1.01	1.43	0.96	0.83	0.83	1.36	0.66	0.89	0.46	0.83	0.96	0.76	1.02	0.89	1.55	*1.65*	1.40	0.88	1.13
D	1.19	1.05	1.39	*1.72*	*1.58*	1.14	*1.64*	0.90	0.61	0.90	1.04	0.94	0.68	0.60	0.91	0.72	0.79	1.15	1.17	1.43
E	1.41	1.04	1.11	1.10	0.94	*2.30*	*2.07*	*1.70*	1.34	*1.73*	*1.76*	*1.55*	1.16	1.43	*1.88*	1.27	0.92	1.07	1.31	1.19
K	1.03	1.06	0.97	0.66	0.84	1.08	0.80	0.82	1.27	1.24	1.10	1.17	1.22	*1.71*	1.63	*1.45*	1.19	1.27	1.13	1.10
R	1.15	1.14	0.81	0.90	0.76	1.05	0.95	1.33	1.39	*1.73*	1.49	1.41	1.24	1.39	1.66	*1.45*	1.11	0.96	1.29	1.09

[a]The data set consists of 1,316 helices extracted from 274 polypeptide chains in 263 pdb files (Bernstein et al., 1977), chosen from the March 1994 pdb_select list (Hobohm & Sander, 1994) (http://www.sanders.embl-heidelberg.de/pdbsel). Only protein structures with sequence identity ≤25% and resolution ≤2.5 Å were used; NMR structures and files with only Cα coordinates were excluded. The mean helix length in this data set is 12.3 residues; the shortest helix is seven residues and longest is 51 residues. Each table row contains the normalized positional frequencies of a given residue (in one-letter code) as a function of helix position. The normalized frequency, f, is calculated as

$$f = \frac{\text{fraction } i \text{ in helices}}{\text{fraction } i \text{ in data set}} = \frac{\dfrac{\text{occurrences of residue } i \text{ at helix position } j}{\text{number of helices}}}{\dfrac{\text{occurrences of residue } i \text{ in data set}}{\text{number of residues in data set}}}.$$

For each column (N′–N5 and C5–C′), the highest and next highest normalized frequencies are *italic*, providing the residue in question is not rare (i.e., fraction in data set greater than 5%).

with the highest normalized frequencies of occurrence at Ncap are Asp and Thr, at N1 are Glu and Pro, and at C′ is Gly.

Side-chain capping vs. backbone capping

In proteins, hydrogen bonds are formed preferentially between donors and acceptors that are close in sequence; this overall tendency is especially pronounced at helix ends (Stickle et al., 1992). In our data set of 1316 helices, approximately one-third (31%) have at least one short- or mid-range side-chain–to–backbone hydrogen-bonded capping interaction involving the first four >N—H groups and two-thirds (66%) have at least one short- or mid-range backbone-to-backbone capping interaction involving the last four >C=O groups. These interactions are subdivided into histograms in Figure 2. It is apparent that capping by a side-chain partner from within the helix is common at the helix N-terminus (Fig. 2A), while capping by a backbone donor from the adjacent turn predominates at the C-terminus (Fig. 2C). This difference between capping interactions at N- and C-termini is a plausible consequence of residue chirality. In a right-handed helix of L-amino acids, each $C\alpha \rightarrow C\beta$ vector is oriented toward the helix N-terminus (Fig. 1). Thus, helix geometry predisposes side chains toward upstream backbone positions, facilitating hydrogen bonding between side-chain acceptors within the helix and unsatisfied >N—H groups in the first helical turn. In contrast, side chains within the helix point away from >C=O groups in the last helical turn. Even side-chain donors with multiple degrees of freedom, such as Lys or Arg, must execute a "U-turn" to realize a backbone hydrogen bond because the C → O vector is oriented toward the helix C-terminus. Occasionally, side-chain donors in the turn immediately following the helix (usually from C″ or C‴) cap carbonyl oxygens from the last helical turn (usually C3 or Ccap). In the data set of 1316 helices, such side-chain capping was provided most frequently by Asn (53 occurrences) and Thr (32 occurrences).

Summarizing the histograms (Fig. 2), at the helix N-terminus, capping H-bonds are provided primarily by side-chain partners from residues within the helix. At the helix C-terminus, capping H-bonds are provided primarily by backbone partners from the adjacent peptide chain turn.

Capping involves both hydrogen bonding and hydrophobic interactions

A hydrophobic interaction was associated with every instance of hydrogen-bonded capping in the data set. This hydrophobic component of capping involves two hydrophobic residues that are usually close in sequence, one from the first/last turn of helix, the other external to the helix. The phenomenon of hydrophobic capping was unrecognized until recently (Aurora et al., 1994; Seale et al., 1994; Munoz et al., 1995a).

Table 2 tallies occurrences of hydrophobic capping in the 1316 helices. Almost invariably, both the first and last turn of every protein helix have at least one residue that can participate in a hydrophobic interaction. In the set of 1316 helices, 1309 include such a residue within the first turn, 1313 within the last turn. As seen in the table, almost all hydrophobic capping is between residues close in sequence. Quantifying this observation, at least one short- or mid-range interaction is found at the helix N-terminus in 81% of all cases and at the helix C-terminus in 87% of all cases.

Each table entry lists the number of times a residue from the first/last turn of a helix formed a hydrophobic cap with the designated position external to the helix. These capping interactions were tabulated between the sequentially closest interacting hydrophobic neighbors, one within the helix and the other external to the helix. Predominantly, hydrophobic caps are provided by residues nearby in sequence. Further, specific positional preferences are apparent: at the N-terminus, N3 ~ N4 > N2 > Ncap ~ N1; at the C-terminus, C3 > C2 > Ccap > C1 > C4. At either terminus, the sum of all hydrophobic interactions between residue pairs exceeds the number of helices in the data set. For example, there are 1687 interacting residue pairs among the N-termini of the 1316 helices. Thus, hydrophobic residues often cluster at sites of hydrophobic capping.

Salt bridges between side chains of opposite charge can substitute for hydrophobic capping as observed in our data set and reflected in Table 1 (e.g., a high normalized positional preference for Asp at N′ and Arg or Lys at N4). In some positions, Lys and Arg function as hydrophobic residues because their long, flexible alkyl side-chain moieties can interact with buried apolar surfaces while avoiding concomitant burial of terminal charged groups.

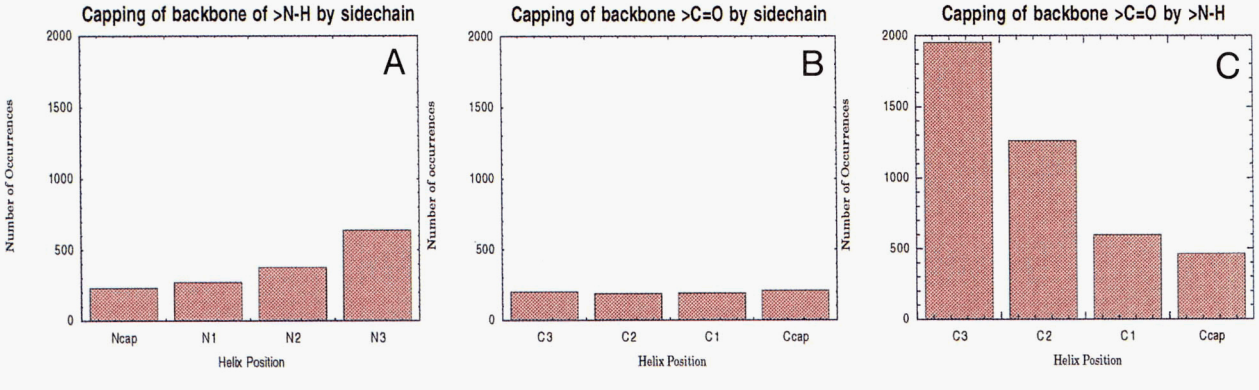


Fig. 2. Histograms of capping hydrogen bonds in the data set of 1,316 helices. **A:** Frequency of capping of the first four N—H groups by a local side chain. **B:** Frequency of capping of the last four C=O groups by a local side chain. Capping by a side chain is infrequent at helix C-termini. **C:** Frequency of capping of the last four C=O groups by a backbone N—H group. At C3, the carbonyl oxygen is often an acceptor for multiple N—H donors.

Table 2. *Survey of hydrophobic capping*[a]

Helix Posn.	Distance from position i														
i	*i*−2	*i*−3	*i*−4	*i*−5	*i*−6	*i*−7	*i*−8	*i*−9	*i*−10	*i*−11	*i*−12	*i*−13	*i*−14	*i*−15	Sum
Nc	43	**68**	49	22	21	18	8	7	8	8	4	6	6	1	269
N1	53	**62**	62	34	20	13	12	6	7	5	3	4	5	4	290
N2	—	**125**	79	34	15	11	12	12	5	6	6	6	6	5	322
N3	—	—	**157**	147	64	27	19	17	19	8	5	5	7	5	480
N4	—	—	—	**143**	47	27	25	21	17	16	10	13	4	3	326
i	*i*+2	*i*+3	*i*+4	*i*+5	*i*+6	*i*+7	*i*+8	*i*+9	*i*+10	*i*+11	*i*+12	*i*+13	*i*+14	*i*+15	Sum
C4	—	—	—	21	**41**	17	22	13	16	29	20	8	13	6	206
C3	—	—	202	**204**	58	47	21	15	11	10	7	3	9	5	592
C2	0	93	**168**	85	65	33	27	12	14	16	11	7	2	1	534
C1	18	**80**	76	35	32	23	22	18	7	8	6	3	3	7	338
Cc	**142**	60	58	29	11	12	16	4	4	4	2	6	4	7	359

[a]Number of occurrences of an interaction between residues at helix positions *i* and $i \pm x$; *x* varies from 2 to 15. Intrahelical interactions are excluded. The cell with the highest frequency is in bold; in case of a tie, the closest interaction is bold. At the N-terminus, 1,060 (80.6%) of the 1,316 helices are capped by at least one hydrophobic interaction. At the C-terminus, 1,147 (87.2%) are capped by at least one hydrophobic interaction.

Soon after algorithms were devised to quantify solvent accessible surface area (Lee & Richards, 1971; Shrake & Rupley, 1973), Chothia (1976) observed that helical residues lose approximately half their accessible surface upon helix formation and the remaining half in the interaction between the helix and the protein. Further, he noted that area loss within the helix is due primarily to buried main-chain polar groups; apolar groups from side chains remain comparatively exposed. This classical picture changes when capping is included. In our data set, the helix proper (i.e., N1 through C1) loses an average of 764 Å^2 upon interaction with the protein, 182 Å^2 in polar area and 582 Å^2 in apolar area. In comparison, the capped helix loses an average of 578 Å^2 upon interaction with the protein, 127 Å^2 in polar area and 451 Å^2 in apolar area. The difference represents helical surface buried by short- and mid-range capping interactions—24% in all, 30% of the polar surface and 23% of the apolar surface. Thus, approximately a quarter of the surface buried between the helix proper and the remainder of the protein is a consequence of local capping.

Based on results presented in this and the preceding section, we propose that the phenomenon of helix capping be redefined to include both hydrogen bonding and hydrophobic interactions. Even in the absence of hydrophobic capping, solvent access to amide groups at the N-terminus is hindered by helix geometry. In detail, the solvent accessible surface area of successive amide nitrogens in an N-acetyl-poly-alanyl-N-methyl-ester α-helix is: N1 = 8.8 Å^2; N2 = 1.6 Å^2; N3 = 0.6 Å^2; N4-C1 = 0.1 Å^2. The hydrophobic interaction that straddles helix termini guides the polypeptide chain back upon itself, burying additional polar surface in the backbone of the first/last helical turn. Shielded from solvent water, these buried >N−H groups at the N-terminus and buried >C=O groups at the C-terminus are satisfied instead by capping hydrogen bonds. Accordingly, the protein must adopt a conformation that solves the problem of satisfying these hydrogen bonds while maintaining the hydrophobic interaction. When interactions are short range, it appears that only a limited number of sterically allowed solutions is possible, as represented by the seven motifs described next.

Capping motifs

The capping motifs described below are summarized in Table 3 and Figure 3. All motifs were first identified in a smaller, 42-protein data set (Stickle et al., 1992; Harper & Rose, 1993; Aurora et al., 1994; Seale et al., 1994; Creamer et al., 1995). Arguably, this 42-protein subset is complete because no previously undiscovered motif was encountered subsequently in the larger 274-protein set used to derive statistics for this review.

Review of previously described motifs

The most conspicuous pattern of capping at helix N-termini is represented by the capping box (Dasgupta & Bell, 1993; Harper & Rose, 1993) that caps two of the initial four backbone amide hydrogen donors of the helix. In particular, the side chain of Ncap forms a hydrogen bond with the backbone of N3 and, reciprocally, the side chain of N3 forms a hydrogen bond with the backbone of Ncap. The capping box is so named because these reciprocal hydrogen bonds appear as unique, box-like patterns in hydrogen bond distance plots (Stickle et al., 1992). The normalized sequence preferences in a capping box are Thr > Ser > Asn at Ncap and Glu > Gln at N3. The definition of the capping box was expanded recently to include the associated hydrophobic interaction between residues N′ and N4. This augmented motif was termed the "expanded capping box" by Seale et al. (1994) and the "hydrophobic staple" by Munoz et al. (1995a).

The big box (Seale et al., 1994) resembles a capping box; it consists of a staggered hydrogen-bonded cycle between the side chain of Ncap and backbone amide of N3 and, reciprocally, the side chain of N3 and backbone amide of N′ (in lieu of Ncap). In a big box, the observed hydrophobic interaction is between apolar side-chain groups in residues N4 and N″ (not N′).

The two primary capping motifs found at helix C-termini are the Schellman and the α_L motifs (Aurora et al., 1994). The Schellman motif is defined by its distinctive, doubly hydrogen-bonded pattern between backbone partners, consisting of 6 →1, 5 →2 hydrogen

Table 3. *Occurrences of capping motif statistics*

No.	Motif	Pattern[a]	Number of occurrences[b]	Number within helix[c]	Number at helix ends[d]	Number with capping interactions[e]
			N-terminal motifs			
Ia	$N' \rightarrow N3$	h-xpxhx	4,039	1,248	154	154
Ib	$N' \rightarrow N4$	h-xpxph	2,631	338	158	146
IIa	$N'' \rightarrow N3$	hp-xpxhx	2,638	316	182	147
IIb	$N'' \rightarrow N4$	hp-xpxph	1,882	244	69	69
IIIa	$N^{3'} \rightarrow N3$	hpp-xpxhx	1,997	245	38	38
IIIb	$N^{3'} \rightarrow N4$	hpp-xpxph	1,504	229	62	62
			C-terminal motifs			
IV	$C'' \rightarrow C3/C'G$	hxpx-Gh	583	44	185 (+ 35)[f]	184 (+ 35)[f]
Va	$C^{3'} \rightarrow C3/C'n$	hxpx-nxh	3,959	803	47	47
Vb	$C^{4'} \rightarrow C3/C'n$	hxpx-nxph	2,942	636	56	56
VIa	$C^{3'} \rightarrow C3/C'G$	hxxx-Gpxh	374	27	42	42
VIb	$C^{4'} \rightarrow C3/C'G$	hxxx-Gpxph	119	8	39	39
VIIa	$C^{4'} \rightarrow C3/C'P$	hxxx-Ppxh	148	0	29	29
VIIb	$C^{5'} \rightarrow C3/C'P$	hxxx-Ppxph	48	0	20	20

[a]Upper case letters denote residues (in one-letter code); lower case letters denote classes. Classes are h (hydrophobic = A, V, I, L, M, F, W, C, neutral H, and within the helix, the alkyl side-chain moieties of K or R); p (polar = G, S, T, N, Q, D, E, K, R, and protonated H); n (non-β-branched = not V, I, T, or P); and indifferent (x). Helix bounds are indicated by a hyphen (-). For N-terminal motifs, Ncap succeeds the hyphen; for C-terminal motifs, Ccap precedes the hyphen.

[b]Number of times the given pattern occurs—regardless of conformation—in the data set.

[c]Number of times the given pattern is found entirely within a helix in the data set.

[d]Number of times the given pattern is found at the terminus of an α-helix of at least seven residues.

[e]Number of times the given pattern expresses the predicted capping motif.

[f]In a $C'' \rightarrow C3/C'G$ motif, when C3 is polar, the hydrophobic contact can be established between C″ and C2 or C4, providing the C″ side chain is sufficiently long (viz., K, R, F, Y, W, or M). The number of such occurrences is given in parentheses.

bonds between the >N−H at C″ and >C=O at C3 and between the >N−H at C′ and >C=O at C2, respectively. The associated hydrophobic interaction is between C3 and C″. If C″ is polar, the alternative α_L motif is observed. The conformational constraints imposed upon the structure by these several interactions can only be satisfied when the C′ residue, typically a glycine, adopts a left-handed conformation (i.e., has a backbone dihedral angle, ϕ, with a positive value). In a Schellman motif, polar or amphipathic residues are highly favored at the C1 position. This tendency is rationalized by the fact that typical protein helices have two distinct facets: a hydrophobic face that is buried against the body of the protein and a hydrophilic face that is exposed to solvent. In a Schellman motif, the $C'' \rightarrow C3$ hydrophobic interaction establishes the hydrophobic face. Consequently, the C1 position, situated half a helical turn away, is relegated to the solvent-exposed face, where polar or amphipathic residues are favored.

The Schellman motif was named in recognition of Charlotte Schellman, who first described it (Schellman, 1980). The hydrogen bond pattern in a Schellman motif has also been noted by others (Milner-White, 1988; Preissner & Bork, 1991; Dasgupta & Bell, 1993).

The α_L motif is defined by a $5 \rightarrow 1$ hydrogen bond between the >N−H at C′ and >C=O at C3. Akin to the Schellman, the C′ residue is typically glycine, which adopts a left-handed conformation (i.e., $\phi > 0$). However, the hydrophobic interaction in an α_L is heterogeneous, occurring between C3 and any of several residues external to the helix (viz., $C^{3'}$, $C^{4'}$, or $C^{5'}$).

There is inherent difference in specificity between the N-and C-terminal motifs. At the N-terminus, helix geometry favors side-chain–to–backbone hydrogen bonding and selects for compatible polar residues (e.g., Thr/Ser at Ncap, Glu at N3 in the capping box). In contrast, at the C-terminus, side-chain–to–backbone hydrogen bonding is disfavored, with backbone hydrogen bonds satisfied instead by post-helical backbone groups (e.g., from C′ and C″ in the Schellman motif). Accordingly, the N-terminus promotes selectivity in all polar positions, especially Ncap and N3, while the C-terminus need only select for C′ residues that can adopt positive values of the backbone dihedral angle ϕ, most notably Gly. These differences in specificity are reflected in the normalized residue preferences listed for each motif in Figure 3. Differing specificities notwithstanding, both termini can realize the same number and type of capping interactions. For example, both the capping box and the Schellman motif include two hydrogen bonds and a hydrophobic interaction, all localized within six consecutive residues that straddle a helix terminus.

Capping nomenclature

Historically, capping motifs have been discovered in piecemeal fashion and christened with proper names at the whim of their discoverers—e.g., the capping box, the Schellman motif, the hydrophobic staple. Although colorful, this idiosyncratic practice has two clear disadvantages. First, it can be confusing when a given motif bears multiple names. For example, the "expanded capping

box" of Seale et al. (1994) is identical to the "hydrophobic staple" of Munoz et al. (1995a). Second, the use of proper names obscures an underlying regularity in the data that becomes apparent with uniform nomenclature.

We propose to name capping motifs systematically, based on hydrophobic capping. This nomenclature is underwritten by the fact that for short-range interactions, a given pairwise hydrophobic contact is sufficient to specify a conformation. For example, all helix caps with interacting hydrophobic residues at N′ and N4 have an identical conformation (viz., the capping box), with backbone dihedral angles for Ncap and N′ that vary by no more than 15°.

Each capping motif is named for the closest pair of interacting hydrophobic residues that straddles the helix terminus. In this nomenclature, a hydrophobic interaction between residues A and B is written as A → B, where the arrow points *from* the hydrophobic residue external to the helix *to* the hydrophobic residue within the helix. For example, N′ → N4 signifies a hydrophobic interaction between residue N′ and N4, the capping box. Several motifs are further qualified by the presence of a particular residue found preferentially at a given position. Such cases are annotated by appending a slash, then the position and one-letter code of the residue. For example, C″ → C3/C′G signifies the Schellman motif, with its characteristic hydrophobic interaction between C″ and C3 and the glycine at C′.

Analysis of the 1316 helices reveals that the hydrogen-bonded capping motifs described in the literature are, in fact, associated with unique hydrophobic patterns. As such, these motifs can be written naturally using the proposed nomenclature. The capping box is an N′ → N4 motif, most often N′ → N4/NcapS,T;N3E. The big box is N″ → N4, and the Schellman is C″ → C3/C′G. The α_L is actually a family, not a unique motif.

The commonly observed motifs can be classified formally into seven cases, at the N-terminus: (I) N′ → N3 and N′ → N4, (II) N″ → N3 and N″ → N4, (III) N‴ → N3 and N‴ → N4; and at the C-terminus: (IV) C″ → C3/C′, (V) $C^{3'}$ → C3/C′n and $C^{4'}$ → C3/C′n, (VI) $C^{3'}$ → C3/C′G and $C^{4'}$ → C3/C′G, and (VII) $C^{4'}$ → C3/C′P and $C^{5'}$ → C3/C′P. Here, "n" designates a non-β-branched residue, viz., any residue except Val, Ile, Thr, or Pro. For each motif, Figure 3 lists the most commonly observed consensus pattern of polar and apolar residues together with ribbon and schematic diagrams of the motif and relevant backbone dihedral angles. Table 3 provides statistics on the frequency of occurrence of these consensus patterns, both in helices and in proteins at large. All motifs are summarized in the paragraphs that follow.

N-terminal capping

The three capping motifs at the helix N-terminus include two that are well described in the literature and one that is not. Motif Ib includes the capping box (see above) (Dasgupta & Bell, 1993; Harper & Rose, 1993; Seale et al., 1994; Munoz et al., 1995a). Motif IIb includes the big box (see above) (Seale et al., 1994). Motif III, which we have dubbed the β-box, consists of a hydrophobic interaction between $N^{3'}$ and N3 or N4 and a hydrogen-bonded β-turn between the >N−H at $N^{3'}$ and O=C< at Ncap. In all three motifs, the backbone dihedral angles for Ncap cluster in the upper left quadrant of a (ϕ,ψ) plot, and side chains also express strong rotamer preferences (Doig et al., 1997).

The normalized frequency of occurrence of proline at N1 is high in all N-terminal motifs (Table 1). The N1 position is well suited to the steric and chemical characteristics of proline (Yun et al., 1991). Proline lacks the flexibility of other residues; its backbone dihedral angle, ϕ, is constrained to a region near −70° (Schulz & Schirmer, 1979), similar to that of an ideal helix. However, when proline adopts a helical conformation, the bulky pyrrolidine ring affects the conformation of the preceding residue, forcing it into a non-helical conformation. Accordingly, proline is sterically compatible with the N1 position because the preceding position, Ncap, has backbone dihedral angles that depart from helical values. Also, proline, an amino acid residue, lacks an >N−H, obviating the need for a hydrogen-bonding acceptor. Finally, it is a little-appreciated fact that proline is the most water soluble of the 20 natural amino acids (Sober, 1977) and is therefore compatible with solvent-exposed positions at helix ends. However, proline can also function as a hydrophobic residue, e.g., at the N‴ position in an N‴ → N3 motif.

C-terminal capping

The four major motifs at the helix C-terminus include two that are well described in the literature and two that are not. Most capping at the C-terminus involves glycines in either of two motifs—motif IV, with the hydrophobic interaction between C″ and C3 (Schellman), and motif VI, with the hydrophobic interaction between either $C^{3'}$ or $C^{4'}$ and C3 (α_L). Both are described above. In either motif, the backbone dihedral angle, ϕ, of the residue at C′ is required to have a positive value. This conformation is unhindered for glycine residues and only moderately disfavored for non-β-branched residues.

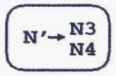
Motif V—$C^{3'}$ or $C^{4'}$ → C3/C′n—resembles C″ → C3 (Schellman) except that, in this case, the C′ position is a non-β-branched residue and the hydrophobic interaction is between C3 and either $C^{3'}$ or $C^{4'}$ in lieu of C″. The backbone dihedral angle, ϕ, of the non-β-branched residue at C′ has a positive value, enabling the characteristic 6 → 1, 5 → 2 hydrogen bonds between the >N−H at C″ and >C=O at C3 and between the >N−H at C′ and >C=O at C2.

Motif VII, a previously undocumented C-terminal capping motif, involves helices terminated by proline residues. Prolines are known to be "helix breakers" (Richardson, 1981) because, when helical, the sequence ···–Xaa–Pro–··· will result in a steric collision between the pyrrolidine ring and Xaa. Consequently, helix-terminating prolines are found at C′, a position in which the $i-1$ residue (i.e., Ccap) can contribute a helical $i \rightarrow i-4$ hydrogen bond while departing from backbone dihedral angles with helical values. Motif VII is defined by a hydrophobic interaction between $C^{4'}$ or $C^{5'}$ and C3 (or C2 if C3 is polar). The motif comprises a turn of 3_{10} helix (C3 through Ccap) and a three-center hydrogen bond linking the carbonyl oxygen at Ccap to amide hydrogens in $C^{3'}$ and $C^{4'}$. In this conformation, both C′ and C″ are solvent exposed and, accordingly, C″ is polar.

Summary of capping motifs

Not all instances of hydrophobic capping are included in these seven motifs. As seen in Table 2, additional interactions are observed between an apolar residue in the first/last helical turn and a hydrophobic cap situated upstream/downstream from $N^{3'}/C^{4'}$. In such cases, a unique motif is neither expected nor found because the number of possible conformations increases exponentially with chain length.

A



(Box)

Ia: h-xpxhx
Ib: h-xxpph

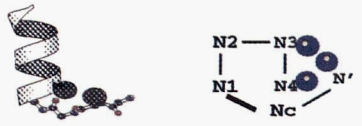
Positional Preferences

N' → N3

N'	Ncap	N1	N2	N3
A(3.0)	D(2.8)	P(2.1)	E(2.2)	M(3.1)
F(2.9)	P(2.3)	K(1.8)	D(2.1)	H(2.9)
I(2.4)	S(1.9)	R(1.5)	A(1.4)	L(2.1)
L(2.1)				

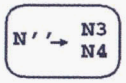
N' → N4

N'	Ncap	N1	N2	N3	N4
M(4.9)	T(4.6)	E(3.8)	E(3.8)	E(7.3)	M(4.1)
L(3.5)	S(2.9)	P(2.6)	D(2.5)	Q(5.0)	F(2.7)
I(3.0)	D(2.4)	R(1.4)	A(1.5)	D(2.5)	V(2.4)
					I(2.4)



	ϕ	ψ
N'	[-102. ± 12.]	140. ± 25.
Ncap	-86. ± 12.	150. ± 25.

B



(Big Box)

IIa: hp-xpxhx
IIb: hp-xxpph

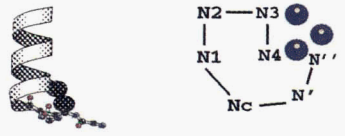
Positional Preferences

N'' → N3

N''	N'	Ncap	N1	N2	N3
A(4.1)	D(3.1)	D(4.4)	P(3.3)	E(2.7)	M(4.4)
F(3.3)	G(2.9)	G(1.8)	E(2.7)	D(1.7)	F(3.6)
Y(2.7)	S(2.3)	T(1.8)	R(2.7)	A(1.7)	A(3.0)

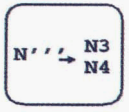
N'' → N4

N''	N'	Ncap	N1	N2	N3	N4
F(5.5)	D(2.9)	T(3.5)	P(4.6)	E(3.6)	Q(6.3)	R(4.0)
I(3.9)	T(2.8)	S(2.7)	A(1.6)	D(2.9)	E(5.8)	M(3.7)
	S(2.1)	N(2.3)		S(1.5)	D(2.9)	L(2.8)
		D(2.0)				



	ϕ	ψ
N''	[-100. ± 15.]	10. ± 15. or 130. ± 15.
N'	-102. ± 14.	10. ± 15. or 170. ± 15.
Ncap	-112. ± 15.	130. ± 24.

C



(β–Box)

IIIa: hpp-xpxhx
IIIb: hpp-xxpph

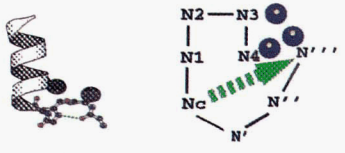
Positional Preferences

N''' → N3

N'''	N''	N'	Ncap	N1	N2	N3
A(3.9)	T(3.1)	D(3.6)	D(4.8)	P(3.0)	E(6.5)	Y(7.6)
I(3.8)	N(2.7)	S(2.7)	E(2.6)	A(2.3)	K(2.3)	F(3.6)
F(3.6)	E(2.6)	N(2.7)	N(2.7)	E(2.6)		R(3.3)

N'' → N4

N'''	N''	N'	Ncap	N1	N2	N3	N4
F(3.7)	E(4.5)	G(3.8)	C(4.4)	K(3.1)	Q(5.0)	E(6.2)	R(4.6)
Y(3.9)	R(3.4)	N(2.7)	D(2.4)	E(1.8)	D(4.9)	Q(3.8)	F(3.7)
I(2.6)	S(3.0)	E(2.7)	T(2.8)	I(1.8)	A(2.1)	N(3.6)	V(2.5)



	ϕ	ψ
N'''	[-82. ± 28.]	-50. ± 12.
N''	-90. ± 14.	85. ± 15.
N'	-95. ± 12.	0. ± 14.
Ncap	-112. 15.	100. 25.

D

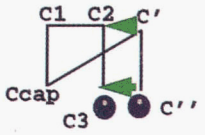
C′′→ C3/C′Gly

(Schellman)

IV: hxpx-Gh

Positional Preference

C3	C2	C1	Ccap	C′	C′′
L(3.1)	Q(2.4)	E(3.0)	H(2.4)	G(12.0)	I(2.2)
A(2.9)	R(2.2)	A(3.0)	K(1.9)		V(1.8)
M(2.2)	K(2.2)	K(2.4)	L(1.6)		K(1.5)



	ϕ	ψ
Ccap	-87. ± 8.	-20. ± 14. or 160. ± 9.
C′	85. ± 9.	20. ± 15.
C′′	-90. ± 11.	[-30. ± 26. or 144. ± 17.]

E

C3′ / C4′ → C3/C′n

(Non-Gly Schellman)

Va: hxpx-nxhx
Vb: hxpx-nxph

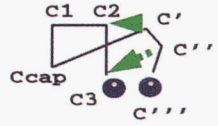
Positional Preference

C′•′ → C3

C3	C2	C1	Ccap	C′	C′′	C′′′
L(2.8)	R(2.8)	R(2.8)	H(3.6)	N(4.6)	L(2.5)	L(2.7)
A(2.8)	Q(2.3)	E(2.8)	N(2.1)	K(3.1)	Y(1.6)	I(2.6)
M(2.2)	I(2.3)	K(2.4)	L(1.9)	D(2.5)		M(2.5)

C4• → C3

C3	C2	C1	Ccap	C′	C′′	C′′′	C4•
L(3.5)	R(3.0)	E(4.7)	F(3.2)	N(8.0)	V(2.5)	D(2.9)	I(5.3)
A(3.4)	Q(2.2)	K(2.1)	L(2.5)	D(3.6)	T(2.0)	T(2.5)	V(3.3)
R(2.0)	I(2.0)	A(1.9)	H(1.5)	K(2.0)		P(2.5)	Y(2.3)



	ϕ	ψ
Ccap	-90. ± 8.	0. ± 15.
C′	60. ± 7.	40. ± 12.
C′′	-90. ± 12.	100. ± 12. or -10. ± 12.
C′′′	-85. ± 14.	-130. ± 14. if C′′ -90., -10.
	-85. ± 15.	-30. ± 12. if C′′ -90., 100.
C′′′′	-90. ± 22.	[120. ± 30.]

F

C3′ / C4′ → C3/C′Gly

(α_L)

VIa: hxxx-Gphx
VIb: hxxx-Gpph

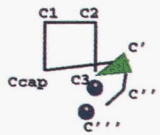
Positional Preference

C′′′ → C3

C3	C2	C1	Ccap	C′	C′′	C′′′
L(2.9)	K(2.8)	R(2.8)	L(2.6)	G(12.0)	G(2.5)	L(2.7)
A(2.8)	L(2.3)	E(2.8)	R(2.1)		N(2.5)	I(2.4)
K(2.2)	N(2.1)	K(2.4)	T(1.9)		T(2.5)	M(2.2)

C4′ → C3

C3	C2	C1	Ccap	C′	C′′	C′′′	C4′
L(3.0)	Q(3.1)	R(3.7)	T(3.2)	G(12.0)	E(4.0)	E(3.9)	V(5.3)
F(3.0)	L(2.8)	E(3.0)	H(2.5)		S(3.0)	S(3.5)	A(4.0)
K(2.9)	R(2.8)	V(2.9)	I(1.5)		D(3.0)	G(2.7)	L(2.4)



	ϕ	ψ
Ccap	-90. ± 9.	-9. ± 7.
C′	85.0 ± 6.	12. ± 9.
C′′	-100.0 ± 13.	127. ± 17.
C′′′	-100.0 12.	[130. 12.]

G

C4′ / C5′ → C3/C′P

(Proline)

VIIa: hxxx-Ppph
VIIb: hxxx-Pppph

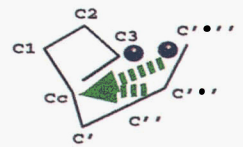
Positional Preference

C4′ → C3

C3	C2	C1	Ccap	C′	C′′	C′′′	C4′
L(3.5)	K(3.5)	K(2.7)	H(4.8)	P(22.0)	E(5.3)	D(2.2)	H(2.8)
Y(3.0)	R(2.8)	E(2.5)	D(2.0)		Q(2.2)	T(2.1)	F(2.6)
Y(3.0)	E(2.2)	Q(2.2)	L(1.8)		N(2.0)	R(1.8)	L(2.0)

C5′ → C3

C3	C2	C1	Ccap	C′	C′′	C′′′	C4′	C5′
K(3.0)	K(3.9)	E(3.0)	Y(4.2)	P(22.0)	E(4.0)	R(4.0)	K(5.3)	V(4.2)
L(2.0)	L(1.9)	H(3.0)	C(2.5)		S(2.0)	Q(2.1)	Q(4.0)	I(3.0)
F(2.0)	R(1.8)	F(2.7)			Q(2.0)	E(2.0)		F(2.1)



	ϕ	ψ
Ccap	-130. ± 12.5	85. ± 12.
C′	- 60. ± 8.5	-22 ± 7. or 150. ± 15.
C′′	- 90. ± 13.5	-22. ± 10. or 30. ± 10.
C′′′	-100. ± 12.5	-27. ± 13. or 122. ± 14.
C′′′′	- 90. ± 14.	[-30. ± 14. or 120. ± 13.]

Fig. 3. *See caption on next page.*

The seven motifs selected for classification were limited intentionally to the most populated categories in our data set, but other minor motifs do exist and might have been included as well. For example, $C^{4'} \rightarrow C3$ or $C^{5'} \rightarrow C3$ is similar to $C'' \rightarrow C3$ (Schellman) and involves a non-β-branched residue at C′. Here, the hydrophobic interaction is between residues C3 and $C^{4'}$ or $C^{5'}$, in lieu of C″. A pyrrolidine ring at $C^{4'}$ or $C^{5'}$ can also adopt this motif, and, accordingly, a peak is observed in the positional frequency for prolines at $C^{4'}$ (Table 1) and $C^{5'}$ (data not shown). In another variant, $C^{5'} \rightarrow C2/C^{3'}G$, a hydrophobic residue at C2 interacts with a hydrophobic residue (or proline) at $C^{5'}$. This interaction requires that the backbone dihedral angle, ϕ, of $C^{3'}$ have a positive value, thereby favoring the selection of glycine or a non-β-branched residue at the $C^{3'}$ position.

Helix capping includes contributions from both hydrogen bonding and hydrophobic interactions. Helix formation per se occludes solvent access to backbone amides at the N-terminus, consistent with Fersht's suggestion that glycine is helix-stabilizing at Ncap due to increased solvent accessibility (Serrano et al., 1992b). This tendency to "dry up" the backbone is further enhanced by the presence of a hydrophobic interaction that straddles the helix terminus and involves two apolar residues, one in the first/last turn, the other external to the helix but usually close in sequence. Capping buries substantial non-polar surface area, an average of 42 Å^2 per hydrophobic cap.

Hydrophobic capping is a commonplace occurrence, more so than hydrogen-bonded capping. In our data set, 81% of the 1316 helices have a short- or mid-range hydrophobic cap at their N-terminus, but only 31% have a hydrogen-bonded cap within this same interval. At the C-terminus, 87% of the helices have a short- or mid-range hydrophobic cap, but only 66% have a backbone-to-backbone hydrogen-bonded cap. Side-chain–to–backbone hydrogen bonds promote residue selectivity at the helix N-terminus, especially at Ncap and N3, but main-chain hydrogen bonds at the C-terminus are less selective, with the exception of motif-specific residues at C′, viz., Gly and Pro.

Helix termination vs. helix continuation

The C-terminal capping motifs can be used to formulate useful rules of thumb that predict when either glycine (Aurora et al., 1994) or proline in a helical sequence would cause helix termination. Notably, these rules depend only on local interactions; they are based entirely on the presence or absence of hydrophobic residues at key sequence locations relative to the position of the Gly or Pro.

Termination in a $C'' \rightarrow C3/C'G$ (Schellman) motif

The key hydrophobic interaction in this motif is between C″ and C3. The interaction is precluded if C3 is polar, in which case the helix continues through the glycine. Lys and Arg can substitute for apolar residues at C″ because their long alkyl side-chain moieties can function as suitable sites for the hydrophobic interaction. At C″, residues with large side chains (e.g., Lys, Arg, Leu, aromatics) can reach C2 or C4, and occasionally, when C3 is polar but either C2 or C4 is hydrophobic, a $C'' \rightarrow C2$ or $C'' \rightarrow C4$ is found in lieu of the usual $C'' \rightarrow C3$.

C1 is also a pivotal position because the $C'' \rightarrow C3$ hydrophobic interaction defines the hydrophobic face of the helix, thereby disposing C1 to the solvent-exposed surface (see Review of previously described motifs, above). Thus, if C1 is hydrophobic, the helix will continue through the glycine. On rare occasion, an apolar residue at C1 is shielded by a longer range interaction; there are two such examples in the data set of 1316 helices. In both, the helix does terminate in a $C'' \rightarrow C3$ hydrophobic cap.

Fig. 3 *(on previous page).* "Wall-chart" summary of the seven capping motifs. Each panel **(A–G)** is organized into five sections that include: (i) (upper left) The motif name given in systematic nomenclature, together with the parenthesized popular name. Below that, h/p/x sequence patterns are shown, where h = hydrophobic, p = polar, and x = indifferent. In detail, hydrophobic = V, I, L, M, F, W, C, neutral H, and within the helix, the alkyl side-chain moieties of K or R; polar = G, S, T, N, Q, D, E, K, R, and protonated H; and x = either h or p. In C-terminal motifs, G indicates glycine and n indicates any non-β-branched residue (i.e., not V, I, T, or P). Proline—the most water soluble of the 20 natural amino acids (Sober, 1977)—can function as either polar or apolar. A hyphen marks the helix boundary, i.e., for N-terminal motifs (panels A–C) N′ precedes the hyphen and Ncap succeeds it; for C-terminal motifs (panels D–G), Ccap precedes the hyphen and C′succeeds it. (ii) (upper right) Residue preferences at each position in the motif, shown as a table. Columns correspond to positions in the motif (e.g., N′, Ncap). In each column, the residues found most frequently at that position are listed (one-letter code) in decreasing order of preference, together with their parenthesized *f*-values. For example, the consensus sequence for $N' \rightarrow N4$ (the capping box) is M-T-E-E-E-M; at Ncap, S is next, and at N3, Q is next. Normalization was performed as in Table 1 except that the numerator is drawn from the motif in question, not from helices at large, i.e., for each residue, i, f = (fraction i in motif)/(fraction i in data set). (iii) (mid-panel, left) Molscript (Kraulis, 1991) representation of the motif with the helix as a ribbon (N-to-C orientation is from bottom to top) and residues flanking the helix shown in ball-and-stick (color code: carbon = gray; nitrogen = blue, and oxygen = red). Interacting hydrophobic side chains are shown as gray spheres and hydrogen bonds as green dashed lines. In N-terminal motifs (panels A–C), the $N4 \rightarrow$ Ncap hydrogen bond is omitted for clarity. (iv) (mid-panel, right) Schematic of capping interactions in the motif. Hydrophobic interactions are depicted as proximate blue spheres and hydrogen bonds as green arrows pointing from donor → acceptor. (v) (bottom) Observed backbone dihedral angles, ϕ,ψ, for each non-helical position in the motif. Dihedral angles at terminal positions, shown in square brackets, are included for completeness but their values do not affect the conformation of the motif. Table 3 lists the frequency of occurrence of each motif in the data set of 1,316 helices. **A:** $N' \rightarrow N3/N4$ Motif described previously as the "capping box" (Harper & Rose, 1993), the "extended box" (Seale et al., 1994), and the "hydrophobic staple" (Munoz et al., 1995a). **B:** $N'' \rightarrow N3/N4$ Described previously as the "big box." **C:** $N''' \rightarrow N3/N4$ The "β-box." The dashed green line indicates the backbone hydrogen bond between Ncap and N‴. **D:** $C'' \rightarrow C3/C'G$ Described previously as the Schellman motif (Aurora et al., 1994). In this case, K or R often function as apolar residues at C″. Dashed green lines indicate the hydrogen bonds from $C' \rightarrow C2$ and $C'' \rightarrow C3$. **E:** $C^{3'}$ or $C^{4'} \rightarrow C3/C'n$ The non-Gly Schellman is similar to (D) except that C′ is a non-β-branched residue in lieu of Gly *and* the hydrophobic interaction is between C3 and $C^{3'}$ or $C^{4'}$, not C″. **F:** $C^{3'}/C^{4'} \rightarrow C3/C'G$ Described previously as α_L (Aurora et al., 1994). The motif includes two related structures—either $C^{3'}$ or $C^{4'}$ is hydrophobic, and C″ is not a proline. **G:** $C^{4'}/C^{5'} \rightarrow C3/C'P$ The proline motif has trans-Pro at C′, a hydrophobic interaction between C3 and $C^{3'}$ or $C^{4'}$, and a three-center hydrogen bond between amide hydrogens at $C^{3'}$ and $C^{4'}$ and the carbonyl oxygen at Ccap.

These same issues apply to repeated glycines (i.e., •••–Gly–Gly–•••). Using DNAse I (1atn) as an example, the helix that spans residues 337–348 includes two glycines in its sequence: YSVWIG^{341}G^{342}ILAS. Termination with Gly^{341} at C′ would situate a hydrophobic residue at C1 (viz., Trp^{339}) and a polar residue at C″ (viz., Gly^{342}), and therefore helix continuation is expected. Similarly, termination with Gly^{342} at C′ would situate a hydrophobic residue at C1 (viz., Ile^{340}); again, helix continuation is expected.

The C″ → C3 interaction can be perturbed by the complicating presence of co-factors, e.g., a heme or NAD (see below). Almost all cases among the 1316 helices in which a suitable C″ → C3 termination sequence is observed but the helix fails to terminate are associated with an exogenous co-factor located within 6.8 Å of the glycine in question. Conversely, no associated co-factor is observed among those helices in which a suitable C″ → C3 sequence results in termination.

These helix termination/continuation rules can be rationalized simply by summing the hydrogen bonding and hydrophobic contributions from the C″ → C3 termination sequence (i.e., $h_1x_1px_2$–Gh_2) in either state. In a continuing helix, this sequence of six residues includes two helical hydrogen bonds and one $i + 4 \rightarrow i$ hydrophobic interaction if x_1 is apolar. Lacking a side chain, the glycine residue makes no additional hydrophobic contribution. In a helix termination motif, this same sequence also forms two helical hydrogen bonds, and the $h_2 \rightarrow h_1$ hydrophobic interaction is assured. Further, even when x_1 is apolar, the surface area buried on helix termination is about 30% larger, on average, than on helix continuation.

Termination in a C$^{4'}$ → C3/C′P (proline) motif

The key hydrophobic interaction in this motif is between C$^{4'}$ and C3. The interaction is precluded if either is polar, in which case the helix continues, providing that the proline is on a solvent-exposed helical face, where the carbonyl oxygen at $i - 4$ can be satisfied by a water molecule. This latter constraint is imposed by the fact that proline, an amino acid, cannot sustain an $i \rightarrow i - 4$ hydrogen bond.

Termination in a C‴ → C3/C′G (α_L) motif

In this motif, the key hydrophobic interaction with C3 varies among C$^{3'}$, C$^{4'}$, or C$^{5'}$. In contrast to the previous two motifs, where helix continuation is favored in the absence of a suitably positioned apolar residue, the competing factors that affect helix continuation in an α_L are more complex due to this heterogeneity in position. For example in Trp repressor (2WRP), the helix spanning residues 45–63 includes the sequence •••R^{48}EAL^{51}G^{52}T^{53}RV^{55}R••• Were the helix to terminate at L^{51}, then G^{52} and T^{53} would be the putative C′ and C″, respectively, and C$^{4'}$ (V^{55}) → C3 (R^{48}) the putative hydrophobic cap. In actuality, a hydrophobic interaction between V^{55} and L^{51} that maintains the helix is observed. In helices with an internal α_L sequence, the most commonly observed competing hydrophobic interaction is between residues at Gly−1 and Gly+3, as illustrated in this example. However, other competing $i \rightarrow i + 4$ interactions are also observed.

Violation of rules of thumb for helix termination

Modeling studies of all motifs show that capping hydrogen bonds are an automatic consequence of the hydrophobic interaction (see Modeling capping motifs, below). With this simplification, the rules of thumb were assessed in detail in our test set of 1,316 helices for both N′ → N4 (box) and C″ → C3/C′G (Schellman) motifs. Assessment is based upon whether or not the occurrence of a box or Schellman sequence in a helix resulted in the corresponding box or Schellman three-dimensional structure. In both motifs, the hydrophobic interaction is usually—but not always—sufficient to foster the predicted structure. It is important to emphasize that the 1,316-helix test set (from 274 polypeptide chains) is not the "learning set"; all motifs were extracted previously by analyzing a smaller, 42-protein data set (Stickle et al., 1992; Harper & Rose, 1993; Aurora et al., 1994; Seale et al., 1994; Creamer et al., 1995).

Table 3 summarizes instances where capping sequences fail to adopt capping structures. For box and Schellman sequences, the simple rules of thumb are quite successful. Almost all failures in these two motifs involve cases where a co-factor is situated at or near the helix terminus (i.e., within three residues of the motif). Tables 4 and 5 list those failures.

Summary of helix-termination rules of thumb

Helix termination by glycine and proline is expected whenever (i) the sequential pattern of apolar residues leads to a hydrophobic interaction that would be present in the motif but absent in the continuing helix, and (ii) the termination motif is not suppressed by the destabilizing presence of apolar residues in solvent-exposed positions.

We comment in passing on the common belief that glycine is favored at helix termini for entropic reasons. It should be emphasized that, relative to the unfolded state, the restriction in conformational freedom imposed on glycine is no greater upon helix continuation than upon helix termination. Even the standard deviations are similar between the backbone dihedral angles in a helix and at the C′ position of a C″ → C3/C′G termination motif. Of course, non-glycyl residues are disfavored in all backbone configurations with $\phi > 0$ for well-understood steric reasons (Richardson, 1981), and C′ is no exception.

Table 4. *Exceptions to the N′ → N4 (box) motif*[a]

PDB file	Helix bounds (Ncap–Ccap)	Box sequence	Co-factor (if any)
1ads	265–272	V-TPERI	NAP
1fnr	212–222	L-YKEEF	FAD
1gox	211–221	L-SWKDV	FMN
1lga	165–174	F-DELEL	Heme
1lts	40–46	I-DLYDH	None
1myp	367–373	F-ASWRV	NAG
1nar	64–74	F-GPERI	None
1pfk	138–160	I-GFFTA	ADP
1rib	185–204	V-SLREL	FEO
2cp4	192–203	M-TFAEA	Heme
2mnr	79–91	L-APVSL	SO_4
2mnr	251–261	L-GPEEM	N″ → N4 (unknown)

[a]N′ → N4 hydrophobic patterns at helix N-termini in which the predicted hydrophobic interaction is not observed (Table 3, column 6 vs. column 7). Most often a co-factor is bound to a residue within the motif.

Table 5. *Exceptions to the C″ → C3/C′Gly (Schellman) motif*[a]

PDB file	Helix bounds (Ncap–Ccap)	Glycine residue number	Co-factor (if any)
1abk	60–74	68	None
1abm	103–117	117	None
1add	26–34	31	1DA (deaza-adenosine)
1add	126–143	136	1DA (deaza-adenosine)
1ads	26–38	38	NAP (NADP)
1apa	146–153	152	PHO (phosphate)
1apm	128–136	136	OCT (mega-8 detergent)
1aoz	149–156	155	Cu ion
1baa	209–226	217	None
1bab	57–75	74	Heme
1bll	152–173	157	Zn ion
1btc	359–373	370	SEO (2-mercaptoethanol)
1cpc	78–101	89	CYC (phycocyanobilin)
1cpt	51–65	64	Heme
1eaf	416–430	421	SO_4 ion
1gla	479–487	485	GOL (glycerol)
1gsr	62–72	71	GTS (glutathioneSO_4)
1gsr	81–106	93	GTS
1ipd	189–205	203	SO_4 ion
1lis	82–96	89	Arg^{87}–Glu^{91} salt bridge
1mio	15–24	23	Fe–S cluster (tetramer interface)
1mgn	58–77	65	Heme
1pda	283–296	287	DPM (dipyrromethane)
1pfk	257–276	268	FBP (fructose 1–6 bisphos)
1pgd	401–415	412	SO_4 (inter-molecular[b] contact)
1prc	259–285	267	MQ7 (menquinone-7-iron)
1tml	126–148	142	SO_4 ion
1utg	31–46	38	None
1vsg	75–84	79	NAG (N-acetylglucosomine)
1vsg	339–360	356	MAN (mannose)
1vsg	339–360	356	E^{353}–K^{357} salt bridge
2aza	56–65	63	Cu ion
2cmd	224–237	235	Citrate
2hpd	140–157	154	Heme
2scp	112–123	117	Heme
3cox	172–183	177	FAD (flavin adenine dinucleotide)
3gbp	15–30	28	Glucose (inter-molecular[b] contact)
3pgk	184–198	195	ATP
4blm	71–83	78	SO_4 ion
4fxn	124–136	132	FMN (flavin mononucleotide)
5fbp	13–24	21	F6P (fructose-6-phosphate)
8abp	18–31	24	G1B (β-D-galactose)
8abp	109–129	129	G1B
8acn	121–134	131	NIC/FS_4 (nitroisocitrate/FeS cluster)

[a]Cases of a C″ → C3/C′G (Schellman) sequence pattern (Table 3, column 5) that fail to terminate the helix. Almost all have a co-factor within the motif.

[b]Cases in which the unit cell includes two monomers. The sidechain of one monomer interacts with a co-factor bound to the other monomer.

Modeling capping motifs

Modeling studies were conducted to demonstrate that the definition of each helix termination motif is well specified. In related earlier studies, it was shown that some motifs are overdetermined because either hydrogen bonding alone or the hydrophobic interaction alone is sufficient to specify the entire motif (Aurora et al., 1994; Creamer et al., 1995). Here, modeling was used merely to show that the motif definitions, given in Figure 3, specify unique conformations.

Every motif was modeled by exhaustive search of the conformational space spanned by a suitably chosen helix termination peptide. That is, each rotatable bond in the peptide was varied systematically through all backbone dihedrals (ϕ, ψ) and/or side-chain torsions (χ). Bond lengths and angles were held rigid with van der Waals radii of all -CH_n groups scaled to 90% of their original values (Bondi, 1964). The approach is illustrated below for two examples, the N′ → N4 and $C^{4'}$ → C3/C′P motifs.

N′ → N4

This motif is defined by an N′ → N4 hydrophobic interaction and by the (N4) N—H···O=C (Ncap) hydrogen bond. The question being addressed is whether these two interactions are sufficient to define the motif uniquely (Fig. 4A).

In detail, all conformations of the blocked hexapeptide

```
Ace-Val-Ala-Ala-Ala-Ala-Val-NCH3
    N'   Nc   N1   N2   N3   N4
```

were generated, subject to the constraint that residues N1 through N4 are maintained in a helical conformation. Remaining torsion angles were varied systematically in 20° increments for backbone torsions and in 120° increments for side-chain torsions (i.e., staggered conformers). The grid search generated 52,488 conformers; all but 1,575 were disallowed due to steric conflict. This remaining set was then filtered using only distance criteria to identify conformations having both an (N4)N—H···O=C (Ncap) hydrogen bond (viz., distance ≤ 3.5 Å) and an N′ → N4 hydrophobic interaction (viz., distance 5.4 Å). Four conformers remained: (1) N′ ψ = +120°, Ncap ϕ = −60°, ψ = +140°; (2) N′ ψ = +140°, Ncap ϕ = −60°, ψ = +140°; (3) N′ ψ = +120°, Ncap ϕ = −60°, ψ = +160°; and (4) N′ ψ = +140°, Ncap ϕ = −40°, ψ = +160°. All are equivalent, with N′ ψ = 130 ± 10° and Ncap ϕ = −55 ± 7°, ψ = −150 ± 10°. This equivalence class has values that resemble the observed means listed in Figure 3 (but differ slightly because the model used idealized geometry). Thus, the two interactions are sufficient to specify the motif uniquely.

$C^{4'}$ → C3/C′P

Using a blocked octamer, all conformations of the peptide

```
Ace-Val-Ala-Ala-Ala-Pro-Ala-Ala-Val-NHCH3
    C3   C2   C1   Cc   C'   C''   C'''  C''''
```

were generated, subject to the constraint that residues C3 through C1 are held in a helical conformation, with remaining torsions sampled systematically (Fig. 4B). For efficiency, backbone torsions were varied using Moult sampling (Moult & James, 1986) while side-chain torsions were sampled in 120° increments, as previously.

The search generated 1.7 million conformers; 90% were disallowed due to steric conflict. Notably, in the absence of the helical residues, an equivalent search generated 1.1 million conformers with only 35% disallowed, an indication of the degree to which the presence of the bulky helix backbone constrains the conformation of the motif. Allowed conformers were then filtered to select those

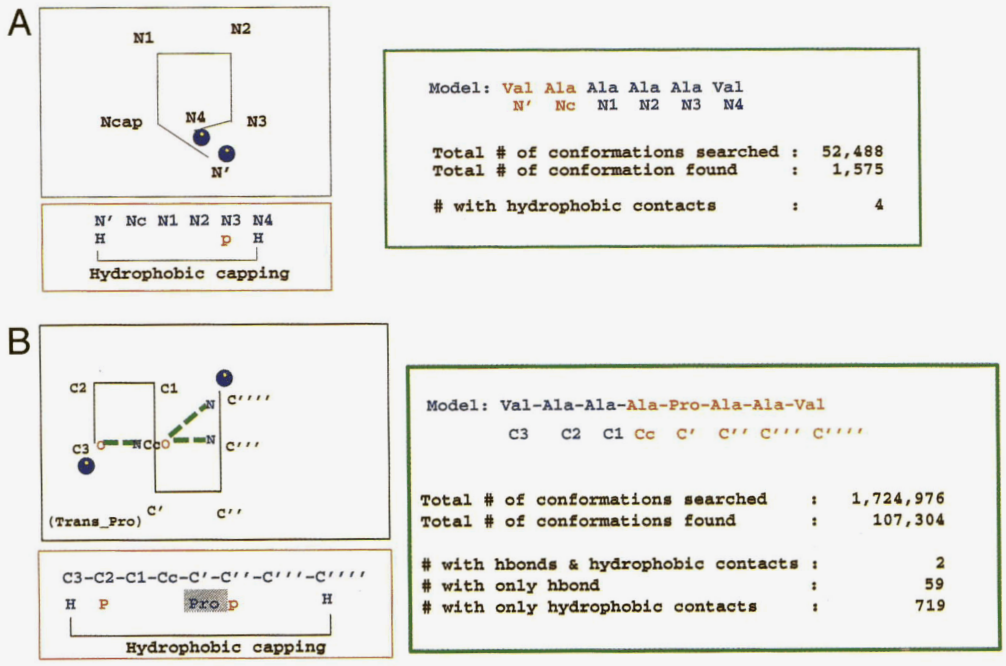


Fig. 4. Summary of modeling. (Left) Schematic of the motif, with the observed pattern of hydrophobic and polar residues indicated. (Right) Exhaustive modeling in a nutshell. Hydrophobic interactions are represented by valine side chains. For backbone dihedral angles, residues in blue were held in an ideal helix; residues in red were varied systematically. For side-chain χ angles, valines were varied in 120° increments (staggered conformers). See text for further detail. **A:** N′ → N4 motif. The sampling regime resulted in 52,488 conformations; all but 1,575 are sterically disallowed. Of these, only four satisfy valine-to-valine distance constraint. **B:** $C^{4'}$ → C3/C′P motif. The sampling regime resulted in 1,724,976 conformations; all but 107,304 are sterically disallowed. Of these, 719 satisfy the valine-to-valine hydrophobic distance constraint, 59 satisfy the $C^{4'}$ → Ccap H-bond distance constraint, and two satisfy both.

having both a C'''' → Ccap hydrogen bond and a C'''' → C3 hydrophobic interaction. Only two equivalent conformers remained, with dihedral angles that resemble the observed means listed in Figure 3. The modeled conformation is, in fact, identical to the proline motif observed at the C-terminus of helix I in carboxypeptidase (5cpa), in which both C3 and C'''' are valines. Thus, the two interactions are sufficient to specify the motif uniquely. Parenthetically, a hydrophobic interaction between C''' and C3 is actually more favorable than one between C'''' and C3, but, in this former case, the carbonyl oxygen at C2 is shielded from solvent access, an energetically disfavored consequence of the conformation. Thus, an implicit constraint at the helix C-terminus is imposed by those carbonyl groups not otherwise satisfied by hydrogen bonds.

Summary of modeling results

Exhaustive modeling confirms that the several requirements of hydrogen bonding, hydrophobic interaction, and excluded volume restrict the polypeptide chain to a small number of allowed conformers at helix termini. This underlying fact accounts for the existence of a recognizable population of familiar capping motifs. Modeling also guarantees that the motif definitions summarized in Figure 3 are not arbitrary but reflect the full set of distinguishable conformational possibilities. Although modeling of only two motifs is described in this review, all have been analyzed in this way (Creamer et al., 1995) (data not shown), with similar conclusions in every case.

Capping studies in peptides

Many groups have used circular dichroism to study capping in peptides. Among the very first were those of Baldwin and coworkers (Fairman et al., 1989), who tested the hypothesis that the customary amide blocking group at the peptide C-terminus might be capping one or more of the otherwise unsatisfied carbonyl oxygens. However, no change in helicity was observed upon changing the amide to the corresponding methyl ester, an unsurprising result in hindsight[2] (see Sidechain capping vs. backbone capping, above).

Another early experiment, directed at the N-terminus, was performed by Gierasch and co-workers (Bruch et al., 1991). Using a peptide derived from a helical segment of carboxypeptidase A, they demonstrated that helicity persisted in the autonomous peptide and was affected by the identity of the initial residue.

Kallenbach and co-workers analyzed capping extensively in *Lyutides* (Lyu et al., 1993), a peptide system invented to mirror the observed positional preferences of residues in protein helices (Richardson & Richardson, 1988a). Some substitutions were observed to have a dramatic effect on helicity. For example, Ala at the Ncap position of a Lyutide reduces helicity by half relative to Ser at that position, whereas, conversely, Ser at a mid-peptide position reduces helicity by half relative to Ala at that position. In general, substitution of hydrogen bond acceptors at the peptide N-terminus increased the helicity of Lyutides (Lyu et al., 1993; Zhou et al., 1994b), while substitution of hydrogen bond donors at the C-terminus had a less pronounced effect (Gong et al., 1995).

The notion that peptide blocking groups might function as an overlooked source of alternative capping (i.e., surrogate capping) was raised by experiments of Nambiar and co-workers (Forood et al., 1993), who showed that helix content is sensitive to the identity of the N-terminal residue in an unblocked duodecapeptide. Yumoto and co-workers (1993) reached similar conclusions using a peptide derived from neuropeptide Y.

Studies were also conducted by Baldwin and co-workers (Chakrabartty et al., 1993) who varied the N-terminal residue systematically in a family of alanine-based peptides. Resultant helix content was found to correlate with the rank order of helix preference for these same residues at the Ncap position in protein helices (Richardson & Richardson, 1988a). In contrast, no change in helix content was observed when analogous substitutions were made at the C-terminal residue of these peptides. The variation in helix content was apparent in peptides with a free amino group but masked in corresponding N-acetylated peptides, confirming the suspicion that the peptide blocking group can function as a capping surrogate.

NMR has also been used to detect capping. The parent Lyutide (Lyu et al., 1993) mentioned above commences with a capping box sequence (M-SEDEL...) that does, in fact, adopt a capping box motif (i.e., N' → N4/NcapS;N3E); NOEs are observed between the Ser–O_γ (Ncap) and the Glu >N−H (N3) and between the Glu–O_ϵ (N3) and the Ser >N−H (Ncap) (Zhou et al., 1994b). Serrano and co-workers also detected a capping box from NOEs (Munoz et al., 1995a). Rohl and Baldwin (1994) used proton exchange to probe the extent of fraying at helix ends, with results that were interpreted using Lifson–Roig helix-coil theory (Lifson & Roig, 1961), modified to take N-capping into account (Doig et al., 1994). Viguera and Serrano (1995) analyzed a designed helical peptide that included both capping box and Schellman sequences. The presence of the capping box was confirmed from the chemical shift, but the Schellman motif could not be detected, although, curiously, the C-terminus of the peptide helix was not found to be frayed. However, another study by Serrano and co-workers (Munoz et al., 1995b) did detect the expected C'' → C3 signature of a Schellman motif in peptide E from Che Y and in peptide D from p21-Ras. Using a peptide with a Schellman motif excised from cellular retinoic acid-binding protein together with numerous controls, Sukumar and Gierasch (1997) demonstrated that the C'' → C3 hydrophobic interaction is both necessary and sufficient for helix termination.

To what degree does capping stabilize peptide helices? The helix-coil transition is inherently a statistical phenomenon that requires a model-dependent statistical-mechanical interpretation (Zimm & Bragg, 1959; Lifson & Roig, 1961; Doig et al., 1994). In this regard, circular dichroism is an appropriate experimental technique because it reports an ensemble-averaged property. Nevertheless, raw values of the mean residue ellipticity (θ) at a suitable wavelength (e.g., 222 nm) are insufficient to compare peptides of differing compositions, lengths, concentrations, or degrees of capping. For example, in comparing two peptides A and B, the fact that $\theta^{A}_{222} > \theta^{B}_{222}$ need not imply that A is more stable than B. A 12-residue peptide with Ncap at residue 2 and Ccap at residue 10 would have at most seven residues in the helix, with maximal helicity less than 60%. Without capping, a 12-residue helix with frayed ends and maximal helicity that exceeds 60% is feasible. In general, capping interactions can stabilize helices by inhibiting fraying, but any residues used to "tie down the ends" in a capping

[2]It had been suggested (Presta & Rose, 1988) that the blocking groups used in peptide studies might be an overlooked source of capping interactions. The specific proposal to use a methyl ester in lieu of the amide was made by one of the authors of this review (G.D.R.), who should have known better.

motif are thereby removed from the helix proper. The removal of even a few residues from a helix can have a proportionately large effect on θ_{222} in short peptides. Thus, the mean residue ellipticity at θ_{222}, though sensitive to helicity in the time-averaged ensemble, does not report on the structure in its entirety. These issues have prompted a renewed interest in helix-coil theory, augmented to include capping (Doig et al., 1994; Shalongo et al., 1994).

A related issue concerns the degree of helix stabilization that can be conferred by capping groups. As an example, the design of a peptide with high helical content was reported recently (Forood et al., 1994). These authors used a series of sulfur-containing blocking groups, each differing in the oxidation state of the sulfur. The higher the oxidative state (i.e., the better the hydrogen bond donor), the higher the helicity. At the extreme, covalently capped peptides can function as though they are permanently prenucleated (Kemp & Curran, 1988; Zhou et al., 1994a; Groebke et al., 1996). These and related studies—e.g., Chakrabartty et al. (1993)—underscore the conclusion that the peptide blocking group can contribute substantially to helix stability.

Capping studies in proteins

Capping has been analyzed in many proteins. Soon after publication of the capping hypothesis (Presta & Rose, 1988; Richardson & Richardson, 1988a), the first such study was conducted in barnase (Serrano & Fersht, 1989). Both helices of barnase have capping motifs at either end. Helix 1 (residues 6–17) begins with an N″ → N4 motif and ends with a $C^{3'}$ → C3/C′His^{18}. Helix 2 (residues 26–33) begins with an N′ → N4 (box) motif and ends with a C″ → C3 (Schellman). Mutations of any interacting residue in these capping motifs decrease protein stability by 1.1 to 3 kcal-mole^{-1} (Serrano et al., 1992a). An analysis based on combinatorial mutagenesis together with quenched flow analysis supports the conclusion that, for both helices, formation of the C-terminus is an early folding event. It is tempting to speculate that C-terminal motif formation per se is the early folding event, though the experiments do not address this point. However, even if capping motif formation were shown to be among the early folding events in barnase, generalization to all helices is unwarranted (Jennings & Wright, 1993).

An elegant study in apocytochrome b_5 used a natural pH-dependent capping switch to make or break a helix (Lecomte & Moore, 1991). In detail, the protonation state of His^{80}, situated at the N-terminus of helix VI, changes upon heme binding. In the holoprotein, His^{80} is neutral, and NOEs affirm the existence of an H-bond between the imidazole and the backbone amide of Arg^{82}. Upon heme removal, the pKa of the imidazole is raised, thereby charging the histidine, disfavoring the H-bonded form, and unraveling the helix. The effect could also be induced in the apoprotein by raising the pH.

Che-Y is a metal-binding (Mg^{2+}) signal transduction protein involved in bacterial chemotaxis. The structure of both apo- and holoproteins is known (Stock et al., 1991). Analogous to heme in cytochrome b_5, magnesium ion in Che Y promotes structural changes. In the apoprotein, helix IV begins at Asn^{94} with an N′ → N4 hydrophobic cap and a side-chain–to–backbone hydrogen bond between Asn^{94} (Ncap) and Ala^{97} (N3). In the presence of magnesium (Belsolell et al., 1994), the helix is extended to Lys^{91}, incarcerating the box-like motif within the helix proper. A peptide corresponding to this region of interest has been synthesized (Munoz et al., 1995b). From NMR, the peptide helix mimics the apoprotein, with Asn at the Ncap position. These examples from cytochrome b_5 and Che Y demonstrate how helix capping can be linked to co-factor binding.

Staphylococcal nuclease (Loll & Lattman, 1989) includes a helix (residues 98–106) that terminates in a Schellman motif. Among the numerous mutations described in this molecule (Shortle, 1992), those that involve any key site in this Schellman motif—viz., C3 ($L^{103}G$ or $L^{103}A$), C′($G^{107}V$), or C″ ($L^{108}G$ or $L^{108}A$)—rank among the most destabilizing observed, reducing protein stability by amounts that range from 4.6 to more than 7.2 kcal-mol^{-1} (Sondek & Shortle, 1990; Green et al., 1992).

T4 phage lysozyme (Remington et al., 1978) has a helix that terminates in an α_L structure, with sequence TFRT-GT^{157}W (C3 through C‴). The predicted α_L interaction between Trp at C‴ and Phe at C2 is observed in the crystal structure. In addition to hydrogen-bond interactions characteristic of the α_L motif, the side chain of Thr^{157} also forms a hydrogen bond with the backbone amide of Asp^{159}. Alber et al. (1987a; 1987b) mutagenized C″ (Thr^{157}) to 13 other residues, and both the X-ray structure and thermal stability of all mutant proteins were determined. As anticipated, substitution of Thr^{157} by uncharged, polar residues has little effect on protein stability because these alternatives can also form an α_L motif with corresponding side-chain hydrogen bonds. In contrast, all substitutions of Thr^{157} by apolar residues destabilized the protein, by amounts ranging from 1.3 to 2.9 kcal-mole^{-1}. Alber et al. (1987a) and Brian Matthews (pers. comm.) point out that diminished stability results from exposure of the apolar side chain to solvent and loss of the side-chain hydrogen bond. In fact, helix termination rules (see Helix termination vs. helix continuation) predict that apolar substitutions at C″ will switch the motif from an α_L to a Schellman, with concomitant burial of the hydrophobic side chain. However, this example is thwarted by the presence of the bulky Trp^{158} at C‴, which inhibits realization of the Schellman structure.

Other examples of capping continue to appear in the current literature. Mutagenesis experiments involving a capping box were conducted in chymotrypsin inhibitor 2 (elMasry & Fersht, 1994) and in human growth hormone (Zhukovsky et al., 1994). In both cases, loss of capping hydrogen bonds diminished protein stability by 1 to 2 kcal-mole^{-1}. Both studies were performed before the hydrophobic contribution to capping was fully appreciated (Seale et al., 1994; Munoz et al., 1995a). Also, mutations were engineered into helix II of RNAaseH1 (Ishikawa et al., 1993). In the natural structure, this helix terminates with the sequence RQG^{77}IT-Q^{80}WI. A Schellman motif was introduced by insertion of a single glycine following Q^{80} (viz., G^{80b}), leading to a net gain in protein stability of 0.4 kcal-mol^{-1}. The mutation G^{77} → Ala decreases stability by 0.9 kcal-mole^{-1}, whereas the double mutant increases stability by 1 kcal-mole^{-1}. In the crystal structure of the double mutant, the helix is extended by a single residue, with Q^{80} as the Ccap residue (in lieu of T^{79}) and a hydrophobic contact between W^{81} (C″) and A^{77} (C3), superceding the W^{81} to R^{75} contact in the insertion-only mutant.

Capping interactions may be conserved across protein families (Starich et al., 1996; Aceto et al., 1997), but not always. In the globins for example, a capping box for helix A appears not to be conserved.

Capping studies in simulations

The nearly parallel alignment of backbone >N−H and >C=O groups in a helix gives rise to a coherent macrodipole. It was

suggested that this macrodipole might stabilize helical conformation under suitable conditions (Hol et al., 1981; Shoemaker et al., 1987). Favorable electrostatic interactions near the "poles" of the macrodipole afforded a plausible explanation for the previously puzzling empirical observation that charged residues in helices are distributed inhomogeneously (Chou & Fasman, 1974; Blagdon & Goodman, 1975), with acidic residues situated preferentially at the helix N-terminus (the positive end of the macrodipole) and basic residues situated preferentially at the C-terminus (the negative end).

Two notable studies using simulations both focused on whether capping effects are due primarily to electrostatic interaction with the macrodipole or to local hydrogen bonding. Simulations are uniquely able to disentangle complex effects because, unlike experiments, specific contributions can be toggled off/on as desired.

The main emphasis in work of Åqvist et al. (1991) was to show that non-specific electrostatic interactions at helix termini are predominantly short-ranged. Specifically, the electrostatic field contributes little to the helix N-/C-terminus apart from interactions localized within the first/last helical turn. Thus, capping is independent of helix length (i.e., charge separation in the macrodipole).

Tidor (1994) also found that electrostatic contributions from the macrodipole decrease markedly beyond the first/last turn, based, in this case, on simulation of a well-characterized mutation in λ Cro (Pakula & Sauer, 1990). By calculating contributions to the free energy from an Ncap mutation (viz., $Y^{26} \rightarrow D$ in the helix that spans residues 26–36), he concluded that the capping effect is due both to specific hydrogen bonding and to electrostatic interactions with spatially proximate groups, some nearby in sequence, some quite distant.

In retrospect, both native and mutant sequences of the λ Cro helix have an N′ → N3/N4 capping motif, with substantial solvent-shielding of the Ncap amide hydrogen. In general, information of this sort is useful in designing not only experiments but also simulations.

Capping links secondary structure to supersecondary structure

Protein monomers can be decomposed neatly into domains, supersecondary structure, and secondary structure (Schulz, 1977). In this descending hierarchy, supersecondary structure is comprised of adjacent, interacting elements of regular secondary structure (i.e., $\alpha\alpha$, $\beta\beta$, or $\beta\alpha\beta$ units). Such higher order units are common in proteins (Levitt & Chothia, 1976; in particular, see Efimov, 1991; Richardson & Richardson, 1988b).

For two adjacent helices separated by a sufficiently short loop (≤7 residues), capping interactions restrict the orientation to a unique, antiparallel conformation. Within this conformation, tolerances have enough play to permit modest flexibility, akin to that seen in globin corners, which undergo natural variation but nevertheless maintain a recognizable fold (Lesk & Chothia, 1980).

The loop between adjacent helices is usually quite short. In the set of 274 proteins, there are 705 adjacent helices, with an average loop size of 4 ± 5 residues. Of these, 80% do fall within the seven-residue threshold.

Capping interactions can impose substantial restriction on the conformation of adjacent helices (Fig. 5). Apolar residues within the connecting loop are poised to satisfy both helices simultaneously, providing C-terminal capping of the first helix and N-terminal capping of the second, and forming a shielded hydrophobic core.

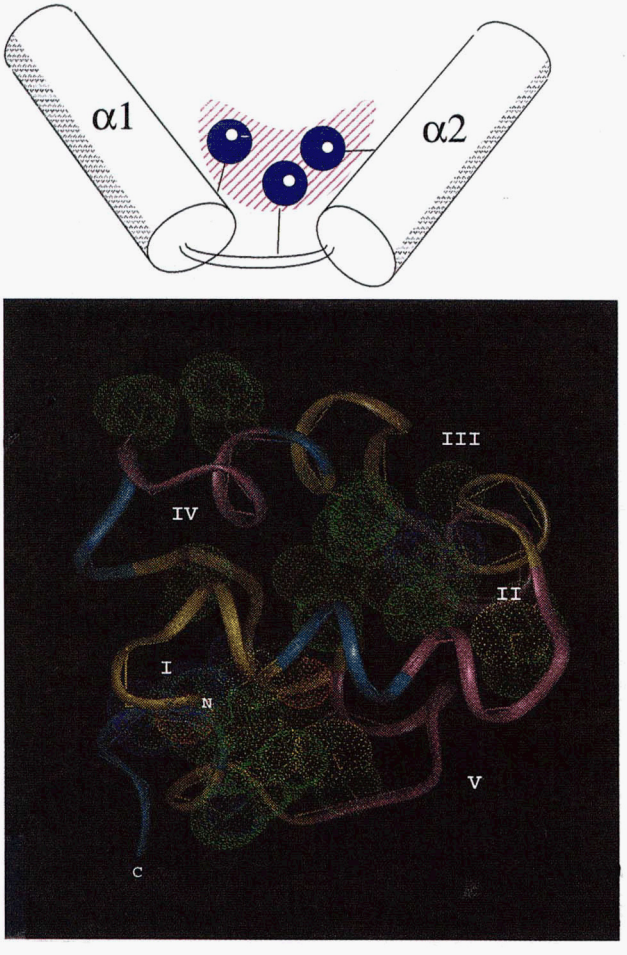


Fig. 5. Capping is a bridge from secondary to supersecondary structure. (Top) For two helices connected by a short (≤7-residue) loop, the constellation of hydrophobic interactions spans the C-cap of one helix, the N-cap of the other, and the intervening non-helical segment. (Bottom) Capping constrains conformation, as illustrated here for 434 Cro (Mondragon et al., 1989), an all-helical protein. In the figure, N-terminal capping motifs are yellow, C-terminal capping motifs are magenta, and other parts of the molecule are cyan. Hydrophobic caps are shown as stippled surface. Capping motifs for the five helices are labeled with Roman numerals: (I) (N-terminus) N″ → N4; (II) $C^{3\prime}$ → C3/C′n followed by N′ → N4; (III) C″ → C3/C′G followed by N′ → N4; (IV) C″ → C3/C′G followed by a three-residue segment, then N″ → N4; and (V) C‴ → C3/C′n followed by N′ → N3. Helix 5 is uncapped at the C-terminus. In all, helices and capping motifs span the entire molecule, with the sole exception of three residues in the helix3–helix4 corner (IV).

As illustrated by Figure 5, capping can play a substantial role in protein conformation, beyond its contribution to the ledger of hydrogen bonds and hydrophobic interactions. For example, in a protein comprised of capped helices interconnected by short loops, the piecewise assembly of these components should be sufficient to establish the overall fold (Aurora & Rose, in prep.). Thus, capping imposes a substantial restriction on the set of allowed conformations, reminiscent of earlier work of Ptitsyn and Rashin (1975) and Richmond and Richards (1978).

Conclusion

The preceding sections review helix capping and summarize evidence that capping interactions can stabilize helices in both pep-

tides and proteins. Hydrophobic capping, in particular, makes a dominant contribution to each motif described in Figure 3. From the perspective of protein folding, the specificity of these seven motifs—i.e., the interactions that distinguish one from another—is largely, though not entirely, a consequence of the pattern of polar/apolar residues surrounding the helix termini. In general, patterns of hydrophobicity are thought to play a significant role in protein folding (Kamtekar et al., 1993; also see Cordes et al., 1996, and references therein).

We close with a caution and a hope. A decade ago, the phenomenon of helix capping ranked as conjecture, but no longer. Yet experiments often neglect these issues still, especially in the design of peptides. Such an oversight manifestly represents an open invitation to an uncontrolled experimental variable. Wishing it otherwise, we hope this review will provide a useful resource.

Acknowledgments

We thank Rajgopal Srinivasan, Trevor Creamer, and Jeffrey Seale for many insightful suggestions, and the NIH (GM29548) for support.

References

Aceto A, Dragani B, Melino S, Allocati N, Masulli M, Di Ilio C, Petruzzelli R. 1997. Identification of an N-capping box that affects the alpha 6-helix propensity in glutathione S-transferase superfamily proteins: A role for an invariant aspartic residue. *Biochem J 322*:229–234.

Alber T, Dao-pin S, Nye JA, Muchmore DC, Matthews BW. 1987a. Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein. *Biochemistry 26*:3754–3758.

Alber T, Dao-pin S, Wilson K, Wozniak JA, Cook SP, Matthews BW. 1987b. Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature 330*:41–46.

Åqvist J, Luecke H, Quiocho FA, Warshel A. 1991. Dipoles localized at helix termini of proteins stabilize charges. *Proc Natl Acad Sci USA 88*:2026–2030.

Aurora R, Srinivasan R, Rose GD. 1994. Rules for alpha-helix termination by glycine. *Science 264*:1126–1130.

Bell JA, Becktel WJ, Sauer U, Baase WA, Matthews BW. 1992. Dissection of helix capping in T4 lysozyme by structural and thermodynamic analysis of six amino acid substitutions at Thr 59. *Biochemistry 31*:3590–3596.

Belsolell L, Prieto J, Serrano L, Coll M. 1994. Magnesium binding to the bacterial chemotaxis protein CheY results in large conformational changes involving its functional surface. *J Mol Biol 238*:489–495.

Bernstein FC, Koetzle TG, Williams G, Meyer E Jr, Brice MD, Rogers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Blagdon DE, Goodman M. 1975. Mechanisms of protein and polypeptide helix initiation. *Biopolymers 14*:241–245.

Bondi A. 1964. Van der Waals volumes and radii. *J Phys Chem 68*:441–451.

Bork P, Preissner R. 1991. On alpha-helices terminated by glycine. 2. Recognition by sequence patterns. *Biochem Biophys Res Commun 180*:666–672.

Bruch MD, Dhingra MM, Gierasch LM. 1991. Sidechain-backbone hydrogen bonding contributes to helix stability in peptides derived from an alpha-helical region of carboxypeptidase A. *Proteins Struct Funct Genet 10*:130–139.

Chakrabartty A, Doig AJ, Baldwin RL. 1993. Helix capping propensities in peptides parallel those in proteins. *Proc Natl Acad Sci USA 90*:11332–11336.

Chothia C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol 105*:1–14.

Chou PY, Fasman GD. 1974. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry 13*:211–222.

Cordes MHJ, Davidson AR, Sauer RT. 1996. Sequence space, folding and protein design. *Curr Opin Struct Biol 6*:3–10.

Creamer TP, Srinivasan R, Rose GD. 1995. Evaluation of interactions between residues in α-helices by exhaustive conformational search. In: Crabb JW, ed. *Techniques in Protein Chemistry*. San Diego: Academic Press. pp 443–450.

Dasgupta S, Bell JA. 1993. Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. *Int J Peptide Protein Res 41*:499–511.

Doig AJ, Chakrabartty A, Klingler TM, Baldwin RL. 1994. Determination of free energies of N-capping in alpha-helices by modification of the Lifson–Roig helix-coil theory to include N-and C-capping. *Biochemistry 33*:3396–3403.

Doig AJ, MacArthur MW, Stapley BJ, Thornton JM. 1997. Structures of N-termini of helices in proteins. *Protein Sci 6*:147–155.

Edwards MS, Sternberg M, Thornton JM. 1987. Structural and sequence patterns in the loops of beta-alpha-beta units. *Protein Eng 1*:173–181.

Efimov AV. 1991. Structure of α-α-hairpins with short connections. *Protein Eng 4*:245–250.

Efimov AV. 1993. Standard structures in proteins. *Prog Biophys Molec Biol 60*:201–239.

elMasry NF, Fersht AR. 1994. Mutational analysis of the N-capping box of the α-helix of chymotrypsin inhibitor 2. *Protein Eng 7*:777–782.

Esposito G, Dhanapal B, Dumy P, Varma V, Mutter M, Bodenhausen G. 1997. Lysine as helix C-capping residue in a synthetic peptide. *Biopolymers 41*: 27–35.

Fairman R, Shoemaker KR, York EJ, Stewart JM, Baldwin RL. 1989. Further studies of the helix dipole model: Effects of a free alpha-NH3+ or alpha-COO-group on helix stability. *Proteins Struct Funct Genet 5*:1–7.

Fetrow JS, Palumbo MJ, Berg G. 1997. Patterns, structures, and amino acid frequencies in structural building blocks, a protein secondary structure classification scheme. *Proteins Struct Funct Genet 27*:249–271.

Forood B, Feliciano EJ, Nambiar KP. 1993. Stabilization of alpha-helical structures in short peptides via end capping. *Proc Natl Acad Sci USA 90*:838–842.

Forood B, Reddy HK, Nambiar KP. 1994. Extraordinary helicity in short peptides via end capping design. *J Am Chem Soc 116*:6935–6936.

Gong Y, Zhou HX, Guo M, Kallenbach NR. 1995. Structural analysis of the N- and C-termini in a peptide with consensus sequence. *Protein Sci 4*:1446–1456.

Green SM, Meeker AK, Shortle D. 1992. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. *Biochemistry 31*:5717–5728.

Groebke K, Renold P, Tsang KY, Allen T, McClure KF, Kemp DS. 1996. Template-nucleated alanine–lysine helices are stabilized by position-dependent interactions between the lysine side chain and the helix barrel. *Proc Natl Acad Sci USA 93*:4025–4029.

Harper ET, Rose GD. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry 32*:7605–7609.

Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci 3*:522–524.

Hol W, Halie LM, Sander C. 1981. Dipoles of the alpha-helix and beta-sheet: Their role in protein folding. *Nature 294*:532–536.

Hutchinson EG, Thornton JM. 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci 3*:2207–2216.

Ishikawa K, Nakamura H, Morikawa K, Kimura S, Kanaya S. 1993. Cooperative stabilization of *Escherichia coli* ribonuclease HI by insertion of Gly-80b and Gly-77 → Ala substitution. *Biochemistry 32*:7136–7142.

Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*:892–896.

Jimenez MA, Munoz V, Rico M, Serrano L. 1994. Helix stop and start signals in peptides and proteins. The capping box does not necessarily prevent helix elongation. *J Mol Biol 242*:487–496.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*:2577–2637.

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science 262*:1680–1685.

Kemp DS, Curran TP. 1988. (2S,5S,8S,11S)-1-acetyl-1,4-diaza-3-keto-5-carboxy-10-thia-tricyclo-[2.8.0^{4,8}]-tridecane, 1 Syntheses of prolyl-proline-derived, peptide-functionalized templates for α-helix formation. *Tetrahedron Lett 29*:4931–4934.

Kraulis PJ. 1991. Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst 24*:946–950.

Lecomte JTJ, Moore CD. 1991. Helix formation in apocytochrome b_5: The role of a neutral histidine at the N-Cap position. *J Am Chem Soc 113*:9663–9665.

Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol 55*:379–400.

Lesk AM, Chothia C. 1980. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J Mol Biol 136*:225–270.

Levitt M, Chothia C. 1976. Structural patterns in globular proteins. *Nature 261*:552–558.

Lifson S, Roig A. 1961. On the theory of helix-coil transition in polypeptides. *J Chem Phys 34*:1963–1974.

Loll PJ, Lattman EE. 1989. The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{2+}, and the inhibitor pdTp, refined at 1.65 Å. *Proteins Struct Funct Genet 5*:183–201.

Lyu PC, Liff MI, Marky LA, Kallenbach NR. 1990. Side chain contributions to the stability of alpha-helical structure in peptides. *Science 250*:669–673.

Lyu PC, Wemmer DE, Zhou HX, Pinker RJ, Kallenbach NR. 1993. Capping interactions in isolated α helices: Position-dependent substitution effects and structure of a serine-capped peptide helix. *Biochemistry 32*:421–425.

Milner-White EJ. 1988. Recurring loop motif in proteins that occurs in right-handed and left-handed forms. Its relationship with alpha-helices and beta-bulge loops. *J Mol Biol 199*:503–511.

Mondragon A, Wolberger C, Harrison SC. 1989. Structure of phage 434 Cro protein at 2.35 Å resolution. *J Mol Biol 205*:179–188.

Moult J, James M. 1986. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins Struct Funct Genet 1*:146–163.

Munoz V, Bianco FJ, Serrano L. 1995a. The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding. *Struct Biol 2*:380–385.

Munoz V, Serrano L, Jimenez MA, Rico M. 1995b. Structural analysis of peptides encompassing all α-helices of three α/β parallel proteins: Che-Y, flavodoxin and P21-Ras: Implications of α-helix stability and the folding of α/β parrellel proteins. *J Mol Biol 247*:648–669.

Odaert B, Baleux F, Huynh-Dinh T, Neumann J-M, Sanson A. 1995. Nonnative capping structure initiates helix folding in an annexin I fragment. A ^{1}H NMR conformational study. *Biochemistry 34*:12820–12829.

Oliva B, Bates PA, Querol E, Aviés FX, Sternberg MJE. 1997. An automated classification of the structure of protein loops. *J Mol Biol 266*:814–830.

Pakula AA, Sauer RT. 1990. Reverse hydrophobic effects relieved by amino acid substitutions at a protein surface. *Nature 344*:363–364.

Pauling L, Corey RB. 1951. The structure of synthetic polypeptides. *Proc Natl Acad Sci USA 37*:241–250.

Petukhov M, Yumoto N, Murase S, Onmura R, Yoshikawa S. 1996. Factors that affect the stabilization of alpha-helices in short peptides by a capping box. *Biochemistry 35*:387–397.

Preissner R, Bork P. 1991. On alpha-helices terminated by glycine. 1. Identification of common structural features. *Biochem Biophys Res Commun 180*:660–665.

Presta LG, Rose GD. 1988. Helix signals in proteins. *Science 240*:1632–1641.

Ptitsyn OB, Rashin AA. 1975. A model of myoglobin self-organization. *Biophys Chem 3*:1–20.

Quiocho FA, Sack JS, Vyas NK. 1987. Stabilization of charges on isolated ionic groups sequestered in proteins by polarized peptide units. *Nature 329*:561–564.

Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature 358*:302–307.

Remington SJ, Anderson WF, Owen J, Ten Eyck LF, Grainger CT, Matthews BW. 1978. Structure of the lysozyme from bacteriophage T4: An electron density map at 2.4 Å resolution. *J Mol Biol 118*:81–98.

Reymond MT, Huo S, Duggan B, Wright PE, Dyson HJ. 1997. Contribution of increased length and intact capping sequences to the conformational preference for helix in a 31-residue peptide from the C terminus of myohemerythrin. *Biochemistry 36*:5234–5244.

Richards FM, Kundrot CE. 1988. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins Struct Funct Genet 3*:71–84.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Prot Chem 34*:168–340.

Richardson JS, Richardson DC. 1988a. Amino acid preferences for specific locations at the ends of alpha-helices. *Science 240*:1648–1652.

Richardson JS, Richardson DC. 1988b. Helix lap-joints as ion-binding sites: DNA-binding motifs and Ca-binding "EF Hands" are related by charge and sequence reversal. *Proteins Struct Funct Genet 4*:229–239.

Richmond TJ, Richards FM. 1978. Packing of alpha-helices: geometrical constraints and contact areas. *J Mol Biol 119*:537–555.

Rohl CA, Baldwin RL. 1994. Exchange kinetics of individual amide protons in ^{15}N-labeled helical peptides measured by isotope-edited NMR. *Biochemistry 33*:7760–7767.

Rooman MJ, Rodriguez J, Wodak SJ. 1990. Automatic definition of recurrent local structure motifs in proteins. *J Mol Biol 213*:327–336.

Rose GD. 1979. Hierarchic organization of domains in globular proteins. *J Mol Biol 134*:447–470.

Rose GD, Gierasch LM, Smith JA. 1985. Turns in peptides and proteins. *Adv Prot Chem 37*:1–109.

Rose GD, Seltzer J. 1977. A new algorithm for finding the peptide chain turns in a globular protein. *J Mol Biol 113*:153–164.

Schellman C. 1980. The alpha-L conformation at the ends of helices. In: Jaenicke R, ed. *Protein folding*. New York: Elsevier/North Holland. pp 53–61.

Schulz GE. 1977. Structural rules for globular proteins. *Angew Chem Int Ed Engl 16*:23–32.

Schulz GE, Schirmer RH. 1979. *Principles of protein structure*. New York: Springer Verlag.

Seale JW, Srinivasan R, Rose GD. 1994. Sequence determinants of the capping box, a stabilizing motif at the N-termini of alpha-helices. *Protein Sci 3*:1741–1745.

Serrano L, Fersht AR. 1989. Capping and alpha-helix stability. *Nature 342*:296–299.

Serrano L, Kellis J Jr, Cann P, Matouschek A, Fersht AR. 1992a. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol 224*:783–804.

Serrano L, Neira JL, Sancho J, Fersht AR. 1992b. Effect of alanine versus glycine in alpha-helices on protein stability. *Nature 356*:453–455.

Shalongo W, Dugad L, Stellwagen E. 1994. Distribution of helicity within the model peptide acetyl(AAQAA)3 amide. *J Am Chem Soc 116*:8288–8293.

Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL. 1987. Tests of the helix dipole model for stabilization of alpha-helices. *Nature 326*:563–567.

Shortle D. 1992. Mutational studies of protein structures and their stabilities. *Quart Rev Biophys 25*:205–250.

Shrake A, Rupley JA. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol 79*:351–371.

Sober HA. 1977. *CRC handbook of biochemistry and molecular biology, 3rd ed.* Cleveland, Ohio: The Chemical Rubber Co.

Sondek J, Shortle D. 1990. Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. *Proteins Struct Funct Genet 7*:299–305.

Starich MR, Sandman K, Reeve JN, Summers MF. 1996. NMR structure of HMfB from the hyperthermophile *Methanothermus fervidus* confirms that this archaeal protein is a histone. *J Mol Biol 255*:187–203.

Stickle DF, Presta LG, Dill KA, Rose GD. 1992. Hydrogen bonding in globular proteins. *J Mol Biol 226*:1143–1159.

Stock JB, Lukat G, Stock AM. 1991. Bacterial chemotaxis and the logic of intracellular signal transduction networks. *Ann Rev Biophys Biophys Chem 20*:109–136.

Sukumar M, Gierasch LM. 1997. Local interactions in a Schellman motif dictate interhelical arrangement in a protein fragment. *Folding & Design 2*:211–222.

Sundaralingam M, Sekharudu YC. 1989. Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science 244*:1333–1337.

Thapar R, Nicholson EM, Rajagopol P, Waygood EB, Scholtz JM, Klevit RE. 1996. Influence of N-Cap mutations on the structure and stability of *Escherichia coli* HPr. *Biochemistry 35*:11268–11277.

Tidor B. 1994. Helix-capping interaction in lambda Cro protein: A free energy simulation analysis. *Proteins Struct Funct Genet 19*:310–323.

Viguera AR, Serrano L. 1995. Experimental analysis of the Schellman motif. *J Mol Biol 251*:150–160.

Wilmot CM, Thornton JM. 1988. Analysis and prediction of the different types of beta-turns in proteins. *J Mol Biol 203*:221–232.

Yumoto N, Murase S, Hattori T, Yumamoto H, Tatsu Y, Yoshikawa S. 1993. Stabilization of α-helix in C-terminal fragments of neuropeptide Y. *Biochem Biophys Res Comm 196*:1490–1495.

Yun RH, Anderson A, Hermans J Jr. 1991. Proline in alpha-helix: Stability and conformation studied by dynamics simulation. *Proteins Struct Funct Genet 10*:219–228.

Zhou HX, Hull LA, Kallenbach NR. 1994a. Quantitative evaluation of stabilizing interactions in a prenucleated alpha helix by hydrogen exchange. *J Am Chem Soc 116*:6482–6483.

Zhou HX, Lyu P, Wemmer DE, Kallenbach NR. 1994b. Alpha-helix capping in synthetic model peptides by reciprocal side-chain–main-chain interactions: Evidence for an N-terminal "capping box." *Proteins Struct Funct Genet 18*:1–7.

Zhou HX, Wemmer DE. 1994. Structure of a C-terminal alpha-helix cap in a synthetic peptide. *J Am Chem Soc 116*:1139–1140.

Zhukovsky EA, Mulkerrin MG, Presta LG. 1994. Contribution to global protein stabilization of the N-capping box in human growth hormone. *Biochemistry 33*:9856–9864.

Zimm BH, Bragg JK. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J Chem Phys 31*:526–535.