Protein Science (1994), *3*:1741–1745. Cambridge University Press. Printed in the USA.

Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices

JEFFREY W. SEALE,[1] RAJGOPAL SRINIVASAN,[2] AND GEORGE D. ROSE[2]

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

(RECEIVED June 14, 1994; ACCEPTED July 22, 1994)

Abstract

The capping box, a recurrent hydrogen bonded motif at the N-termini of α-helices, caps 2 of the initial 4 backbone amide hydrogen donors of the helix (Harper ET, Rose GD, 1993, *Biochemistry 32*:7605–7609). In detail, the side chain of the first helical residue forms a hydrogen bond with the backbone of the fourth helical residue and, reciprocally, the side chain of the fourth residue forms a hydrogen bond with the backbone of the first residue. We now enlarge the earlier definition of this motif to include an accompanying hydrophobic interaction between residues that bracket the capping box sequence on either side. The expanded box motif—in which 2 hydrogen bonds and a hydrophobic interaction are localized within 6 consecutive residues—resembles a glycine-based capping motif found at helix C-termini (Aurora R, Srinivasan R, Rose GD, 1994, *Science 264*:1126–1130).

Keywords: α-helices; capping box; hydrophobic interactions

The α-helix is characterized by main-chain hydrogen bonds between successive amide hydrogen donors and carbonyl oxygen acceptors situated 4 residues previously in sequence (Pauling et al., 1951). For the helix of average length (i.e., ~12 residues), this pattern results in 8 intrasegment hydrogen bonds. Additional "capping" hydrogen bonds that satisfy the initial 4 amide hydrogens and final 4 carbonyl oxygens (Presta & Rose, 1988; Richardson & Richardson, 1988) may also be present and inhibit fraying of helix ends. Helices and their flanking residues are labeled:

. . . -N″-N′-Ncap-N1-N2-N3- . . . -C3-C2-C1-Ccap-C′-C″- . . . ,

where N1 through C1 have backbone dihedral angles with helical values ($\phi = -64 \pm 7°$; $\psi = -41 \pm 7°$), Ncap and Ccap belong both to the helix and adjacent turn, and the primed residues are in the turns that bracket the helix.

The determinants of helices are insufficiently understood to reliably predict protein helices from sequence alone. Following seminal early studies (Schellman, 1958; Zimm & Bragg, 1959; Sueki et al., 1984), much recent attention has been directed toward understanding the physical basis of helix formation (Presta & Rose, 1988; Richardson & Richardson, 1988; Serrano & Fersht, 1989; Lyu et al., 1990; Merutka et al., 1990; O'Neil & DeGrado, 1990; Padmanabhan et al., 1990; Bruch et al., 1991; Yun & Hermans, 1991; Creamer & Rose, 1992; Horovitz et al., 1992; Blaber et al., 1993; Chakrabartty et al., 1993; Forood et al., 1993; Pickett & Sternberg, 1993; Yumoto et al., 1993; Doig et al., 1994). In particular, the presence of specific capping motifs at helix termini has been noted (Dasgupta & Bell, 1993; Harper & Rose, 1993; Aurora et al., 1994), and their significance in helix initiation/termination is being explored (Lyu et al., 1993; elMasry & Fersht, 1994; Zhou et al., 1994; Zhukovsky et al., 1994).

One such motif, dubbed the "capping box," has been described by several groups (Baker & Hubbard, 1984; Dasgupta & Bell, 1993; Harper & Rose, 1993). The capping box is characterized by a reciprocal hydrogen bonded pattern between the first helical residue, Ncap, and the helical residue situated 3 residues downstream, N3 (Fig. 1 and Kinemage 1). After normalization, the residues observed most frequently in a capping box are serine or threonine at Ncap and glutamate at N3 (Harper & Rose, 1993).

Experimental studies of the capping box motif have been conducted in both peptides and proteins. Kallenbach and coworkers analyzed a 20-residue peptide with a capping box sequence spanning residues 3–6 [-Ser(3)-Glu(4)-Asp(5)-Glu(6)-] (Lyu et al., 1993). Using NMR, they confirmed the presence of a helix commencing at the capping box, with reciprocal hydrogen bonds from Ser(3) $O_\gamma \cdots H-N$ Glu(6) and from Glu(6) $O_\epsilon \cdots H-N$

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; rose@grserv.med.jhu.edu.

[1] Present address: Department of Biochemistry, University of Texas Health Sciences Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284.

[2] Present address: Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205.

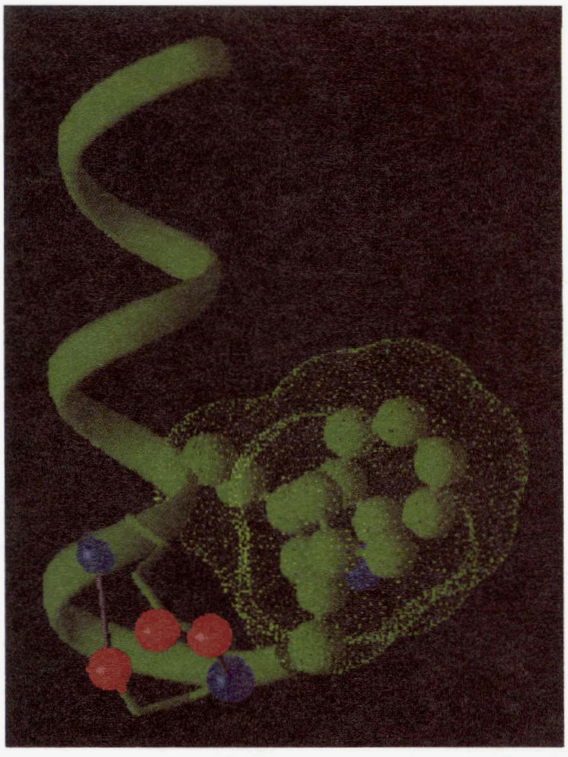

Fig. 1. A capping box from 1MBO, residues 3–18. The helix backbone (N1–N12) is represented by the solid green tube, with amide nitrogens of Ncap and N3 shown as blue spheres. Atoms are colored by type: carbon in green, oxygen in red, nitrogen in blue. Side chains of N′ (leucine), Ncap (serine), N3 (glutamate), and N4 (tryptophan) are drawn explicitly. Side-chain to backbone hydrogen bonds involving residues Ncap and N3 are shown as solid magenta tubes. The van der Waals envelope around the N′ and N4 side chains is represented by the green stippled surface, emphasizing the key hydrophobic interaction between these 2 residues. The figure was generated using *ribbons* (Carson, 1987).

Ser(3) (Zhou et al., 1994). To demonstrate that the capping box initiates the helix, they eliminated the initial 2 residues of the peptide, after which the helix was still observed to commence at the serine (now residue 1), with capping box hydrogen bonds intact (Zhou et al., 1994). Zhukovsky et al. (1994) systematically mutated a Ser-X-X-Glu capping box in helix 2 of human growth hormone and showed that the hydrogen bonds per se contribute to protein stabilization.

We now enlarge the earlier definition of this motif (Harper & Rose, 1993) by documenting the existence of an accompanying hydrophobic interaction between residues immediately adjacent to the box on either side. This hydrophobic interaction, between residues N′ and N4, is observed in almost all capping boxes in our test set of X-ray-elucidated proteins and, as such, appears to be an integral component of the motif.

Upon inclusion of the hydrophobic interaction, the features of the capping box are found to parallel those of the Schellman motif, a glycine-based helix capping motif found exclusively at helix C-termini (Aurora et al., 1994). Both motifs include 2 hydrogen bonds and an accompanying hydrophobic interaction, and both simultaneously stabilize the helix while terminating it. Like the Schellman motif, the capping box appears to be a site at which the observed structure is encoded predominantly by the local sequence, consistent with an underlying stereochemical code for protein folding (Lattman & Rose, 1993).

Results

Capping boxes identified by Harper and Rose (1993) were based upon hydrogen bonding alone. Examination of the residues abutting each box sequence reveals the existence of hydrophobic residues at positions N′ and N4 (Table 1). All but 2 of the 17 boxes have a hydrophobic residue at N′, and all have either a hydrophobic residue or lysine or arginine at N4. Although these basic residues are not typically classified as hydrophobic, their long alkyl side chains provide an apolar surface that can participate in hydrophobic interaction (Aurora et al., 1994; Kinemage 1).

The hydrophobic interaction between N′ and N4, illustrated in Figure 1 and Kinemages 1 and 2, was quantified by calculating the area buried between their side chains, an average of 24 Å^2 (Table 1). Further hydrophobic burial between N′ and N3 is also apparent, an average of 17 Å^2.

Although the capping box is situated at a helix terminus in close proximity to solvent water, capping hydrogen bonds are shielded from solvent access. As shown in Table 1, the amide groups of both Ncap and N3 are generally inaccessible to solvent. Similarly, the Ncap and N3 side chains are partially buried; in contrast, the N1 and N2 side chains, on the opposite side of the helix, are solvent exposed.

In particular, the N2 side chain of helices initiated by capping boxes is highly accessible to solvent (Table 1). Of note, an acidic residue is found preferentially at this position, particularly aspartate. In our data set, 8 of the 17 boxes have aspartate at N2 and 1 has glutamate. Six of the 8 N2 aspartate residues make an additional self-capping hydrogen bond between the side-chain carboxylate and the N2 backbone amide. In 3 cases, this hydrogen bond is made directly, whereas in the remaining 3 cases, it is mediated through a water molecule. Similarly, the glutamate side chain at N2 also hydrogen bonds to its own backbone amide.

In the set of 42 proteins used for this analysis, there are 8 additional helices with a sequence that satisfies our expanded definition of a capping box, but which lack 1 of the 2 hydrogen bonds, typically the one between the N3 side chain and Ncap backbone (Kinemage 3). Zhukovsky et al. (1994) have determined that the other hydrogen bond—i.e., the one between the Ncap side chain and N3 backbone—provides a 2-fold larger contribution to protein stability. These exceptions, listed in Table 2, have characteristics that resemble those of actual capping boxes. By exhaustive conformational searching, we confirmed that these "near" boxes could realize the missing hydrogen bond without rearrangement of the existing backbone structure (data not shown).

A variant of the capping box motif, dubbed the "big" box (Kinemage 4), 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′*). Four instances of this variant, together with 2 additional "near" instances, are listed in Table 3.

The predictive efficacy of the capping box sequence was assessed by searching proteins of known structure for consensus

Table 1. *Capping boxes from 42 high-resolution X-ray crystal structures*[a]

Protein[b]	Helix bounds	Sequence N″″-N6[c]	N′-N4 burial[d]	Normalized solvent accessibilities[e]					
				Ncap NH	N3 NH	Ncap sc	N1 sc	N2 sc	N3 sc
1BP2	89-108	SSEN-NACEAFI	2.5	0.02	0.0	0.47	0.69	0.00	0.25
1MBO	3-18	VL-SEGEWQL	52.6	0.24	0.02	0.61	0.73	—	0.26
2CA2	219-227	PISV-SSEQVLK	33.9	0.0	0.0	0.10	0.48	0.71	0.03
2CTS	37-43	VGQI-TVDMMYG	34.8	0.0	0.0	0.64	0.41	0.91	0.25
2CTS	70-78	FRGY-SIPECQK	13.0	0.0	0.0	0.19	0.03	0.75	0.38
2LHB	12-29	VAPL-SAAEKTK	26.5	0.0	0.0	0.54	1.01	0.80	0.11
2WRP	44-64	NLML-TPDEREA	14.0	0.0	0.0	0.59	0.83	0.93	0.52
2WRP	67-76	RGEM-SQRELKN	27.4	0.28	0.0	0.40	0.74	0.82	0.55
351C	67-80	PNAV-SDDEAQT	51.7	0.23	0.0	0.49	0.69	0.75	0.11
3GRS	383-392	TVGL-TEDEAIH	7.9	0.0	0.0	0.13	0.33	0.75	0.54
3GRS	456-462	KMGA-TKADFDN	24.0	0.0	0.0	0.31	0.06	0.72	0.59
3LZM	59-81	NGVI-TKDEAEK	17.7	0.0	0.0	0.58	0.63	0.89	0.21
3RNT	12-30	SNCY-SSSDVST	14.3	0.25	0.0	0.45	0.71	0.80	0.30
5CPA	14-29	ATYH-TLDEIYD	19.5	0.0	0.0	0.45	0.11	0.77	0.48
5CYT	60-68	GIVW-NNDTLLM	33.8	0.0	0.0	0.45	0.46	0.81	0.19
3TLN	280-297	TPTS-NFSQLRA	0.7	0.0	0.0	0.41	0.01	0.52	0.60
1ECD	93-112	PRGV-THDQLNN	35.6	0.0	0.0	0.53	0.45	0.91	0.10

[a] Three additional capping boxes were added to the list compiled by Harper and Rose (1993). Two of these, 3TLN 280-297 and 1ECD 93-112, involve interchange of the ϵ-carbonyl and ϵ-amino groups of the N3 glutamine. The third, 2WRP 67-76, was accepted, although it fails the hydrogen bond distance criterion between the N3 side chain and Ncap amide by 0.04 Å.

[b] Proteins are specified by their 4-character Brookhaven identifiers (Bernstein et al., 1977).

[c] The subsequence from N″″ through N6. To facilitate alignment, a hyphen precedes the Ncap residue. The key hydrophobic residues, at N′ and N4, are shown in bold.

[d] Area buried between residues N′ and N4, in Å^2, calculated as described in the text.

[e] Fractional solvent accessibility of the respective groups, normalized with respect to the standard state accessibility of Lesser and Rose (1990) as described in the text. No value is reported for glycine.

box sequences and determining how many of these are actual box structures. When normalized for frequency of occurrence, the sequence found most often in a capping box is S/T-X-X-E, where X is any residue (Harper & Rose, 1993). This consensus sequence is found 9 times in Table 1 and once in Table 2. A search of the 42-protein test set identifies 53 consensus sequences, including all 10 actual or near boxes. Inclusion of the adjacent hydrophobic residues in the search template (i.e., hϕ-S/T-X-X-E-hϕ, where hϕ represents C, I, L, M, F, W, V, A, K, R, H, or Y) identifies 23 consensus sequences, again including all 10 actual cases. (Lysine, arginine, and the aromatic residues are classified among the hydrophobics because their side chains contain substantial apolar surface.) Upon inclusion of the adjacent hydrophobic residues, only a single capping box sequence was found in the middle of a helix (viz., 5CPA 298-303), and, in this sequence, the residue corresponding to N2 was apolar. As noted above, the N2 residue in actual capping boxes is highly solvent accessible. Thus, the flanking hydrophobic resi-

Table 2. *"Near" capping boxes from 42 high-resolution X-ray crystal structures*[a]

Protein	Helix bounds	Sequence N″″-N6	N′-N4 burial	Normalized solvent accessibilities					
				Ncap NH	N3 NH	Ncap sc	N1 sc	N2 sc	N3 sc
1GD1	251-265	EKEV-TVEEVNA	18.3	0.0	0.0	0.51	0.84	0.78	0.49
1LZ1	24-37	YRGI-SLANWMC	47.5	0.16	0.0	0.50	0.65	0.91	0.25
3LZM	2-12	M-NIFEMLR	15.0	0.0	0.0	0.55	0.64	0.89	0.51
5CHA	164-172	LPLL-SNTNCKK	19.1	0.84	0.0	0.34	0.88	0.88	0.46
1LZ1	4-15	KVF-ERCELAR	44.4	0.0	0.13	0.86	0.90	0.51	0.32
1ECD	2-17	L-SADQIST	29.4	0.0	0.0	0.69	1.01	0.70	0.28
2CTS[b]	103-118	GQIP-TEEQVSW	25.3	0.01	0.0	0.56	0.94	0.83	0.20
2WRP[c]	8-32	QSPY-SAAMAEQ	29.8	0.02	0.11	—	0.88	0.96	—

[a] Column explanations are given in Table 1.

[b] Missing both side-chain to backbone hydrogen bonds.

[c] Side-chain atoms beyond Cβ are not listed for residues 8 and 11 in Brookhaven file 2WRP.

Table 3. *"Big" capping boxes from 42 high-resolution X-ray crystal structures*[a]

Protein	Helix bounds	Sequence N''''-N6	N'-N4 burial	Normalized solvent accessibilities						
				Ncap NH	N3 NH	N' sc	Ncap sc	N1 sc	N2 sc	N3 sc
1MBO	51-58	KHLK-TEAEMKA	15.0	0.0	0.0	0.83	0.57	0.36	0.89	0.46
1BP2	39-56	GTPV-DDLDRCC	15.0	0.0	0.0	0.16	0.09	0.71	0.01	0.00
2LHB	60-66	KGLT-TADELKK	6.0	0.0	0.0	0.75	0.56	0.95	0.76	0.43
4DFR	77-86	TWVK-SVDEAIA	49.2	0.09	0.01	0.96	0.42	0.88	0.93	0.54
				Near big boxes						
2OVO	33-44	KTYG-NKCNFCN	37.2	0.0	0.0	–	0.01	0.89	0.76	0.47
5PTI	47-56	NNFK-SAEDCMR	19.5	0.02	0.0	1.03	0.34	0.74	0.68	0.51

[a] Column explanations are given in Table 1.

dues are arguably an integral component of the motif because they enhance predictive selectivity substantially, without eliminating any actual boxes.

Discussion

The capping box, a hydrogen bonded capping motif at the N-termini of α-helices, was described recently by several groups (Baker & Hubbard, 1984; Dasgupta & Bell, 1993; Harper & Rose, 1993). In this report, we identified an accompanying hydrophobic interaction that is also an integral component of the motif. Together, the 6 consecutive residues N' through N4 include 2 hydrogen bonds and a hydrophobic interaction.

In a recent study of a capping box in human growth hormone, Zhukovsky et al. (1994) systematically mutated both Ncap and N3 residues, both individually and jointly. Of the 13 mutations tested, they found the consensus sequence to have the largest observed stabilizing effect on the protein. In other recent work, Gronenborn and Clore (1994) showed that the capping box has an identifiable NMR signature. In general, elucidation of specific capping motifs in proteins may be useful in locating helix termini precisely.

After inclusion of the hydrophobic interaction, features of the capping box resemble those of the Schellman motif, a glycine-based capping motif found exclusively at helix C-termini (Aurora et al., 1994). The Schellman motif consists of 2 backbone:backbone 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, together with a hydrophobic interaction between the side chains of residues C'' and C3 (Aurora et al., 1994). Stereochemical details of the 2 motifs differ slightly because the capping box has side-chain to backbone hydrogen bonds, whereas the Schellman motif has backbone to backbone hydrogen bonds.

This minor difference notwithstanding, both motifs have similar hydrophobic interactions and both provide hydrogen bond partners for 2 of the otherwise unsatisfied terminal backbone polar groups in the helix. Moreover, both motifs serve to stabilize the helix while at the same time terminating it and establishing a new trajectory for the polypeptide chain. These capping motifs span the helix termini and include residues that are not within the helix proper. Nevertheless, the motif is clearly an integral part of the observed secondary structure. For this reason, solution studies of peptides with sequences drawn from protein helices would do well to include the associated capping residues.

The capping box and the Schellman motif represent an emerging theme in which specific, pattern-based motifs are found localized at helix termini. In work by Presnell et al. (1992), the sequence pattern designated "NACID" resembles that of a capping box. The existence of such motifs lends support to the more general proposition that protein conformation is specified by an underlying stereochemical code (Lattman & Rose, 1993).

Materials and methods

Helices were identified as appropriately hydrogen bonded sequences with backbone dihedral angles, ϕ and ψ, near their observed means in proteins: $-64 \pm 7°$, $-41 \pm 7°$, respectively (Presta & Rose, 1988). Helix boundaries are delimited by Ncap and Ccap residues, each of which makes 1 additional intrahelical hydrogen bond while departing from helical values of their ϕ, ψ angles (Presta & Rose, 1988).

As previously (Harper & Rose, 1993) reported, 161 α-helices were identified within a set of 42 high-resolution protein structures (resolution ≤ 2.0 Å; *R*-factor ≤ 20) selected from the Protein Data Bank (Bernstein et al., 1977). Helices were identified as sequences of 7 or more consecutive residues including Ncap and Ccap, with backbone dihedral angles of residues N1 through C1 near the observed mean values for α-helices, and with at least 3 $i \rightarrow i-4$ backbone hydrogen bonds. Ncap was defined as the first residue of this sequence with an $i \rightarrow i-4$ backbone:backbone hydrogen bond. Hydrogen bonds were identified using criteria derived from small-molecule crystal studies, as described in Stickle et al. (1992).

Proteins used in this study and in an earlier study (Harper & Rose, 1993), together with their Brookhaven identifiers (Bernstein et al., 1977), are: actinidin (2ACT), α-chymotrypsin A (5CHA), amylase inhibitor (1HOE), avian pancreatic polypeptide (1PPT), azurin (2AZA), carbonic anhydrase (2CA2), carboxypeptidase A α (5CPA), citrate synthase (2CTS), crambin (1CRN), cytochrome *c* (reduced) (5CYT), cytochrome *c*3 (2CDV), cytochrome *c*551 (oxidized) (351C), D-glyceraldehyde-3-P dehydrogenase (1GD1), dihydrofolate reductase (4DFR), erythrocruorin (deoxy) (1ECD), flavodoxin (semiquinone form) (4FXN), γ-II crystallin (1GCR), glutathione peroxidase (1GP1), gluta-

thione reductase (3GRS), hemerythrin (met) (1HMQ), hemoglobin (cyano, met) (2LHB), IG Bence-Jones protein (2RHE), insulin (3INS), lysozyme (human) (1LZ1), lysozyme (phage) (3LZM), myoglobin (oxy) (1MBO), ovomucoid third domain (2OVO), papain (9PAP), penicillopepsin (3APP), pepsin (4PEP), phospholipase A2 (1BP2), plastocyanin (apo) (2PCY), ribonuclease A (7RSA), ribonuclease T1 (3RNT), rubredoxin (1RDG), scorpion neurotoxin (1SN3), staphylococcal nuclease (1SNC), thermolysin (3TLN), trp repressor (2WRP), trypsin (β) (1TPP), trypsin inhibitor (5PTI), and ubiquitin (1UBQ).

Solvent-accessible surface area, as defined by Lee and Richards (1971), was calculated using an algorithm of Richmond (1984). Two residues were considered to be making a hydrophobic contact when the distance between any 2 atoms of either residue was less than or equal to the sum of their van der Waals radii plus the diameter of a water molecule, 2.8 Å. Accessibilities were normalized to a tripeptide standard state using values from Lesser and Rose (1990).

Acknowledgments

We thank Rajeev Aurora, Trevor Creamer, Edwin Harper, Neville Kallenbach, and Leonard Presta for much useful discussion and many helpful suggestions. We also thank Leonard Presta for the preprint of a paper prior to acceptance and Jay Ponder for the area subroutine from his TINKER package. This study was supported by a grant from the NIH (GM 29458).

References

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

Baker EN, Hubbard RE. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol 44*:97–179.

Bernstein FC, Koetzle TG, Williams GJB, Meyer EF 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.

Blaber M, Zhang XJ, Matthews BW. 1993. Structural basis of amino acid alpha helix propensity. *Science 260*:1637–1640.

Bruch MD, Dhingra MM, Gierasch LM. 1991. Side chain–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.

Carson M. 1987. Ribbon models of macromolecules. *J Mol Graphics 5*:103–106.

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

Creamer TP, Rose GD. 1992. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc Natl Acad Sci USA 89*:5937–5941.

Dasgupta S, Bell JA. 1993. Design of helix ends. *Int J Pept Protein Res 41*:499–511.

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

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.

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

Gronenborn AM, Clore GM. 1994. Identification of N-terminal helix capping boxes by means of ^{13}C chemical shifts. *J Biomol NMR 4*:455–458.

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

Horovitz A, Matthews JM, Fersht AR. 1992. α-Helix stability in proteins. II. Factors that influence stability at an internal position. *J Mol Biol 227*:560–568.

Lattman EE, Rose GD. 1993. Protein folding—What's the question? *Proc Natl Acad Sci USA 90*:439–441.

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

Lesser GJ, Rose GD. 1990. Hydrophobicity of amino acid subgroups in proteins. *Proteins Struct Funct Genet 8*:6–13.

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.

Merutka G, Lipton W, Shalongo W, Park SH, Stellwagen E. 1990. Effect of central-residue replacements on the helical stability of a monomeric peptide. *Biochemistry 29*:7511–7515.

O'Neil KT, DeGrado WF. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science 250*:646–651.

Padmanabhan S, Marqusee S, Ridgeway T, Laue TM, Baldwin RL. 1990. Relative helix-forming tendencies of nonpolar amino acids. *Nature (Lond) 344*:268–270.

Pauling L, Corey RB, Branson HR. 1951. The structures of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA 37*:205–210.

Pickett SD, Sternberg MJE. 1993. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol 231*:825–839.

Presnell SR, Cohen BI, Cohen FE. 1992. A segment-based approach to protein secondary structure prediction. *Biochemistry 31*:983–993.

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

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

Richmond TJ. 1984. Solvent accessible surface area and excluded volume in proteins. *J Mol Biol 178*:63–89.

Schellman JA. 1958. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J Phys Chem 62*:1485–1494.

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

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

Sueki M, Lee S, Powers SP, Denton JB, Konishi Y, Scheraga HA. 1984. Helix-coil stability constants for the naturally occurring amino acids in water. XXII. Histidine parameters from poly[(hydroxybutyl) glutamine-*co*-L-histidine]. *Macromolecules 17*:148–155.

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

Yun RH, Hermans J. 1991. Conformational equilibria of valine studied by dynamics simulation. *Protein Eng 4*:761–766.

Zhou HX, Lyu P, Wemmer DE, Kallenbach NR. 1994. 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.

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*:536–535.