Structures of N-termini of helices in proteins

ANDREW J. DOIG,¹ MALCOLM W. MACARTHUR,² BENJAMIN J. STAPLEY,¹
AND JANET M. THORNTON²

¹Department of Biochemistry and Applied Molecular Biology, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom

²Biomolecular Structure and Modelling Unit, Department of Biochemistry, University College London,
Gower Street, London WC1E 6BT, United Kingdom

(RECEIVED August 13, 1996; ACCEPTED October 8, 1996)

Abstract

We have surveyed 393 N-termini of α -helices and 156 N-termini of 3_{10} -helices in 85 high resolution, non-homologous protein crystal structures for N-cap side-chain rotamer preferences, hydrogen bonding patterns, and solvent accessibilities. We find very strong rotamer preferences that are unique to N-cap sites. The following rules are generally observed for N-capping in α -helices: Thr and Ser N-cap side chains adopt the *gauche*– rotamer, hydrogen bond to the N3 NH and have ψ restricted to $164 \pm 8^\circ$. Asp and Asn N-cap side chains either adopt the *gauche*– rotamer and hydrogen bond to the N3 NH with $\psi = 172 \pm 10^\circ$, or adopt the *trans* rotamer and hydrogen bond to both the N2 and N3 NH groups with $\psi = 107 \pm 19^\circ$. With all other N-caps, the side chain is found in the *gauche*+ rotamer so that the side chain does not interact unfavorably with the N-terminus by blocking solvation and ψ is unrestricted. An $i, i + 3$ hydrogen bond from N3 NH to the N-cap backbone C=O is more likely to form at the N-terminus when an unfavorable N-cap is present. In the 3_{10} -helix Asn and Asp remain favorable N-caps as they can hydrogen bond to the N2 NH while in the *trans* rotamer; in contrast, Ser and Thr are disfavored as their preferred hydrogen bonding partner (N3 NH) is inaccessible. This suggests that Ser is the optimum choice of N-cap when α -helix formation is to be encouraged while 3_{10} -helix formation discouraged. The strong energetic and structural preferences found for N-caps, which differ greatly from positions within helix interiors, suggest that N-caps should be treated explicitly in any consideration of helical structure in peptides or proteins.

Keywords: α -helix; hydrogen bond; N-cap; side-chain rotamer; 3_{10} -helix

The residue with non-helical ϕ, ψ angles immediately preceding the N-terminus of an α -helix is called the N-cap. Strong preferences for particular amino acids at the N-cap were first observed in proteins (Argos & Palau, 1982; Richardson & Richardson, 1988), leading to experimental measurements of N-cap preferences in protein (Serrano & Fersht, 1989; Lecomte & Moore, 1991; Bell et al., 1992; Serrano et al., 1992) and peptide (Chakrabarty et al., 1993; Forood et al., 1993; Lyu et al., 1993; Yumoto et al., 1993; Doig et al., 1994; Doig & Baldwin, 1995) systems. The results from surveys of crystal structures and from all experimental systems agree that Asn, Asp, Ser, and Thr are overwhelmingly preferred at the N-cap. The range of helix stability changes encountered when substituting the N-cap is up to 4 kcal·mol^{−1} (Doig & Baldwin, 1995). This range is comparable to the differences measured when substituting residues in interior position of helices (Chakrabarty & Baldwin, 1993), though the order of preferred residues differs greatly. In addition, a number of side chain interactions specific to the helix N-terminus have been identified, including the

capping box (Dasgupta & Bell, 1993; Harper & Rose, 1993; Lyu et al., 1993; Zhou et al., 1994), and hydrophobic staple (Seale et al., 1994; Muñoz et al., 1995).

Despite the recent availability of data on the energetic effects of substituting every natural amino acid at the N-cap, the order of N-cap preferences has not yet been fully rationalized. In this paper we examine the structure of 393 helix N-termini in 85 non-homologous high resolution crystal structures in order to determine the structural origins of the N-cap preferences. We find strong preferences for χ_1 rotamers and hydrogen bond patterns that depend on the identity of the N-cap.

It is useful to divide the amino acids into two groups, namely the good N-cap residues (Asn, Asp, Ser, Thr, and Cys) and the bad N-cap residues (Gly, Ala, Val, Leu, Ile, Phe, Tyr, Trp, Pro, Cystine, Met, Lys, Arg, His, Glu, and Gln), principally based on energetics (Doig & Baldwin, 1995). It is already known that the good N-caps are energetically preferred at the N-cap; in this paper we show that this division can be rationalized in terms of structural preferences observed in crystal structures. In addition, we consider the N-cap structures adopted by residues in 3_{10} -helices. These show some significant differences from α -helices, with implications for protein structure, stability, and design.

Reprint requests to: Andrew J. Doig, Department of Biochemistry and Applied Molecular Biology, UMIST, P.O. Box 88, Manchester M60 1QD, United Kingdom; e-mail: andrew.doig@umist.ac.uk.

Results

χ_1 rotamer preferences of α -Helix N-cap side chains

We used a dataset of 85 structures from the Protein Data Bank (Bernstein et al., 1977) that includes 393 N-cap residues, as defined by the algorithm of Kabsch & Sander (1983). For each amino acid, Table 1 lists the number of occurrences of each N-cap, percentage occurrence, and N-cap propensities. The propensities agree very well with those previously observed (Argos & Palau, 1982; Richardson & Richardson, 1988). We surveyed the χ_1 rotamer distribution of each N-cap side chain, classifying them into the three orientations *gauche*− (*g*−), *trans* (*t*), *gauche*+ (*g*+) with χ_1 clustered around +60°, 180°, −60°, respectively (Table 2). Pro was not considered. Large differences in χ_1 rotamer preferences are apparent between the good and bad N-cap sets; Table 3 summarizes these differences. The bad N-cap side chains overwhelmingly prefer the *g*+ rotamer (80% of bad N-cap side chains are *g*+); the *g*− and *t* rotamers are rare. A potential problem with this conclusion is that a low number of the bad N-caps are found, making their statistical preferences unreliable; nevertheless, a consistent trend is observed across all the bad N-cap side chains.

In contrast, the good N-cap side chains avoid the *g*+ rotamer almost entirely. (Only 6% of good N-cap side chains are *g*+.) Instead, Ser and Thr are found most often in the *g*− rotamer, while Asn and Asp are found in both *g*− and *t* conformations. The N-cap rotamer propensities can be compared to the rotamer propensities measured at interior positions of α -helices and β -strands and in coil structures (Swindells et al., 1995). The propensity values encountered at the N-cap are in general much stronger than those

Table 1. Amino acid propensities for N-caps of α - and 3_{10} -helices drawn from the dataset of 85 high quality non-homologous chains (≤ 2.0 Å resolution, $\leq 20.0\%$ R-factor)

Residue	Total #	α			3_{10}		
		#	%	Propensity	#	%	Propensity
Gly	1,454	57	3.9	1.46	25	1.7	1.9
Ala	1,519	18	1.2	0.44	6	0.4	0.4
Val	1,243	9	0.7	0.27	3	0.2	0.3
Leu	1,369	19	1.4	0.52	7	0.5	0.6
Ile	891	6	0.7	0.25	4	0.4	0.5
Phe	693	7	1.0	0.37	5	0.7	0.8
Tyr	679	6	0.9	0.33	1	0.1	0.2
Trp	243	4	1.6	0.61	2	0.8	0.9
Pro	786	30	3.8	1.42	18	2.3	2.5
Cys	340	2	0.6	0.22	2	0.6	0.7
Met	333	4	1.2	0.45	1	0.3	0.3
Ser	1,169	64	5.5	2.03	8	0.7	0.8
Thr	1,015	59	5.8	2.16	10	1.0	1.1
Lys	1,021	17	1.7	0.62	7	0.7	0.8
Arg	731	17	2.3	0.86	7	1.0	1.1
His	382	14	3.7	1.36	5	1.3	1.5
Asp	992	65	6.6	2.43	26	2.6	2.9
Asn	817	44	5.4	2.00	10	1.2	1.4
Glu	978	16	1.6	0.61	5	0.5	0.6
Gln	607	7	1.2	0.43	4	0.7	0.7
Total	17,262	465	2.7		156	0.9	

Table 2. χ_1 rotamer propensities for α -helix N-caps^a

Residue	<i>g</i> −				<i>t</i>				<i>g</i> +			
	#	%	Propensity	Mean angle	#	%	Propensity	Mean angle	#	%	Propensity	Mean angle
Asn	20	47	1.1	62.2 ± 9.0	19	44	1.5	187.5 ± 10.9	4	9	0.3	−72.7 ± 17.4
Asp	18	28	0.6	64.2 ± 6.7	42	66	2.3	183.7 ± 8.8	4	6	0.2	−65.6 ± 23.8
Cys	0	0	0.0	—	2	100	3.5	169.8 ± 6.9	0	0	0.0	0.00
Ser	41	68	1.6	68.8 ± 10.9	17	28	1.0	179.0 ± 19.2	2	3	0.1	−63.8 ± 10.8
Thr	44	88	2.0	64.5 ± 12.9	2	4	0.1	182.1 ± 10.8	4	8	0.3	−78.1 ± 24.3
Arg	2	14	0.3	67.2 ± 7.4	2	14	0.5	161.5 ± 57.5	10	71	2.6	−50.0 ± 17.8
Gln	0	0	0.0	—	2	33	1.2	175.6 ± 2.3	4	67	2.4	−72.3 ± 12.1
Glu	0	0	0.0	—	0	0	0.0	0.00	7	100	3.6	−57.2 ± 17.0
His	1	10	0.2	43.5	3	30	1.0	191.8 ± 7.0	6	60	2.2	−67.7 ± 6.9
Ile	1	33	0.8	49.1	0	0	0.0	0.00	2	67	2.4	−53.0 ± 24.0
Leu	0	0	0.0	—	1	10	0.4	187.4	9	90	3.3	−64.4 ± 7.2
Lys	0	0	0.0	—	1	8	0.3	198.3	12	92	3.4	−65.1 ± 12.3
Met	1	50	1.1	77.5	0	0	0.0	0.00	1	50	1.8	−45.9
Phe	0	0	0.0	—	0	0	0.0	0.00	4	100	3.6	−50.8 ± 12.4
Trp	0	0	0.0	—	0	0	0.0	0.00	1	100	3.6	−77.9
Tyr	0	0	0.0	—	1	50	1.7	177.7	1	50	1.8	−47.1
Val	1	33	0.8	65.0	0	0	0.0	0.00	2	67	2.4	−61.2 ± 0.6
Sum	129				88				73			
Mean				62.0 ± 14				183.0 ± 13.8				−62.4 ± 15.9

^aIn calculating the overall averages and standard deviations, prolines were excluded. The total is 392, not 393, because the χ_1 of Ser8 in the R chain of 2WRP is not defined.

Table 3. α -Helix N-cap side chain χ_1 rotamer preferences summary

N-Cap residue	<i>g</i> ⁻	<i>t</i>	<i>g</i> ⁺
Good N-cap ^a	56%	37%	6%
Bad N-cap ^b	6%	14%	80%

^aSer, Thr, Asn, Asp, Cys.^bAll others (including Cysteine but excluding Pro).

observed within secondary structure or coil, indicating that the structural constraints at a helix N-cap are greater than at any position yet examined in a protein. These rotamer preferences are rationalized below.

Hydrogen bonding at the helix N-terminus

We surveyed hydrogen bonding patterns of the N3 and N2 amide NH groups and the N-cap side chain. (Successive helical residues are named N-Cap, N1, N2, N3, etc., in the N to C direction.) Table 4 shows the hydrogen bond acceptor partners to the amide N3 NH group as a function of N-cap identity. Table 5 shows the hydrogen bond acceptor partners to the amide N2 NH group. The good N-cap side chains Cys, Ser, and Thr usually accept a hydrogen bond from the N3 NH. As previously noted by Bordo & Argos (1994), Asn and Asp are less likely to hydrogen bond with the N3 NH than Ser and Thr; they often hydrogen bond to the N2 NH or both N2 and N3 simultaneously instead (Table 6).

An N3 NH group that cannot find a hydrogen bonding partner in the N-cap side chain has a number of alternatives. First, it can

hydrogen bond to water. This is rare for the good N-caps, especially Ser and Thr, but is common for the bad N-caps. N3 NH hydrogen bonding to water is more common when the N-cap residue is Gly. This is consistent with the conclusion of Fersht and co-workers that Gly is favored at the N-cap, as it allows better access of water to the free NH groups at the helix N-terminus. (The solvent accessibility data discussed below also supports this observation.) Fersht and co-workers argued that this is because the smaller Gly side chain is less likely to block access to solvent water (Serrano & Fersht, 1989; Serrano et al., 1992; Harpaz et al., 1994). An additional possibility is that the more flexible backbone with Gly as the N-cap allows the adoption of more unusual conformations thus increasing water access.

An alternative hydrogen bonding partner for the N3 NH is the C=O group in the N-cap backbone. The presence of this *i, i + 3* NH-CO hydrogen bond is equivalent to the formation of a single turn of 3_{10} -helix at the N-terminus of an α -helix with a bifurcated hydrogen bond from both N3 and N4 NH to the N-cap C=O. Table 4 shows the number of occurrences of the *i, i + 3* backbone to backbone hydrogen bond as a function of the N-cap identity. The results for N3 NH hydrogen bond partners are summarized in Table 6. There is a clear difference between the good and bad N-cap groups. Good N-cap side chains usually hydrogen bond to the N3 NH. Only a single case of a bad N-cap side chain forming a bond to the N3 NH was found (Gln427 in 1BTC). Bad N-caps are more likely to form an *i, i + 3* bond at the α -helix N-terminus than good N-caps (50 versus 30%). This is because the N3 NH is unable to form a hydrogen bond with the N-cap side chain when a bad N-cap is present. It is therefore more likely to satisfy its hydrogen bonding potential by forming the alternative *i, i + 3* bond hydrogen bond. Nevertheless, *i, i + 3* hydrogen bonds are not rare when a good N-cap is present despite the availability of a

Table 4. α -Helix amide N3 NH hydrogen bond acceptor groups^a

N-Cap	N-Cap C=O (<i>i, i + 3</i>)	N-Cap side chain	None	Water	N3 = Pro	Distant side chain	Other	Total
Asn	14 (33%)	23 (53%)	1 (2%)	5 (12%)	0	0	0	43
Asp	18 (28%)	37 (58%)	5 (8%)	4 (6%)	0	0	0	64
Cys SH	0	1 (100%)	0	0	0	0	0	1
Ser	18 (30%)	44 (72%)	4 (7%)	1 (2%)	0	0	0	61
Thr	16 (32%)	36 (72%)	4 (8%)	1 (2%)	0	0	0	50
Ala	4 (36%)	0	3 (27%)	3 (27%)	0	0	0	11
Arg	5 (36%)	0	4 (29%)	4 (29%)	1 (7%)	0	0	14
Cys SS	0	0	1 (100%)	0	0	0	0	1
Gln	3 (50%)	1 (17%)	0	0	0	0	0	6
Glu	3 (43%)	0	2 (29%)	2 (29%)	0	0	0	7
Gly	22 (36%)	0	8 (13%)	26 (43%)	0	4 (7%)	1 (2%)	61
His	4 (40%)	0	2 (20%)	2 (20%)	1 (10%)	0	0	10
Ile	1 (33%)	0	2 (67%)	0	0	0	0	3
Leu	9 (90%)	0	0	1 (10%)	0	0	0	10
Lys	8 (62%)	0	1 (8%)	4 (31%)	1 (8%)	0	0	13
Met	0	0	1 (50%)	1 (50%)	0	0	0	2
Phe	2 (50%)	0	0	1 (25%)	0	1 (25%)	0	4
Pro	19 (73%)	0	3 (12%)	3 (12%)	1 (4%)	0	0	26
Trp	1 (100%)	0	0	0	0	0	0	1
Tyr	2 (100%)	0	0	0	0	0	0	2
Val	2 (67%)	0	0	1 (33%)	0	0	0	3

^aThe NH is deemed to be hydrogen bonding to water if it is bonded to an explicit water molecule in the crystal structure or has a solvent accessible surface area $>0 \text{ \AA}^2$.

Table 5. α -Helix amide N2 NH hydrogen bond acceptor groups^a

N-Cap	N-Cap side chain	None	Water	N2 = Pro	Distant side chain	Distant backbone	Other	Total
Asn	11 (26%)	5 (12%)	24 (56%)	0	1 (2%)	2 (5%)	1 (2%)	43
Asp	34 (53%)	5 (8%)	26 (41%)	0	0	0	0	64
Cys SH	0	0	1 (100%)	0	0	0	0	1
Ser	2 (3%)	6 (10%)	47 (77%)	0	1 (2%)	3 (5%)	1 (2%)	61
Thr	0	7 (14%)	29 (58%)	0	6 (12%)	3 (6%)	1 (2%)	50
Ala	0	1 (9%)	8 (73%)	0	1 (9%)	0	0	11
Arg	0	1 (7%)	8 (57%)	0	4 (29%)	1 (7%)	0	14
Cys SS	0	1 (100%)	0	0	0	0	0	1
Gln	1 (14%)	1 (14%)	5 (71%)	0	0	0	0	7
Glu	0	0	6 (86%)	0	0	1 (14%)	0	7
Gly	0	5 (8%)	29 (48%)	0	9 (10%)	8 (13%)	9 (15%)	61
His	1	3 (30%)	4 (40%)	0	2 (20%)	0	0	10
Ile	0	0	2 (67%)	0	0	1 (33%)	0	3
Leu	0	1 (10%)	5 (50%)	1 (10%)	2 (20%)	1 (10%)	0	10
Lys	0	3 (23%)	10 (77%)	0	0	0	0	13
Met	0	1 (50%)	0	0	1 (50%)	0	0	2
Phe	0	0	3 (75%)	0	1 (25%)	0	0	4
Pro	0	2 (8%)	17 (65%)	0	7 (27%)	0	0	26
Trp	0	0	1 (100%)	0	0	0	0	1
Tyr	0	1 (50%)	1 (50%)	0	0	0	0	2
Val	0	1 (33%)	2 (67%)	0	0	0	0	3

^aThe NH is deemed to be hydrogen bonding to water if it is bonded to an explicit water molecule in the crystal structure or has a solvent accessible surface area $>0 \text{ \AA}^2$.

well-situated hydrogen bond acceptor in this case. However, many of these N3 NH bonds are bifurcated between the N-cap backbone C=O and side chain. It is likely that these hydrogen bond alternatives are rapidly switching in the native protein. Similarly, Asn and Asp side chains often form bifurcated hydrogen bonds to both the N3 and N2 NH groups. Rapid interconversion and/or bifurcated hydrogen bond formation is likely to be common, especially as N-caps are nearly always on the surface of the protein, where there is more freedom for the side chain to move.

Table 6. α -Helix amide N3 NH hydrogen bond acceptor groups summary

N-Cap residue	N-Cap side chain	N-Cap C=O (<i>i, i + 3</i>)
Good N-cap	141 (64%)	66 (30%)
Bad N-cap	1 (1%)	86 (50%)

Table 7. α -Helix hydrogen bond partners of Asn and Asp N-cap side chains

Rotamer	N2	N2/N3	N3
<i>g</i> ⁺	0	0	0
<i>t</i>	18	10	17
<i>g</i> [−]	2	11	21

Correlation between χ_1 rotamer preferences and hydrogen bonding patterns

Table 7 shows the hydrogen bonding partners of Asn and Asp N-cap side chains as a function of the N-cap side-chain rotamer. If the N-cap is *g*⁺, the side chain is distant from the end of the helix and hydrogen bonding to it is not possible. When the side chain is *g*[−], it usually hydrogen bonds to the N3 NH. When it is in the *t* conformer, it can hydrogen bond to either the N2 and N3 NH groups or form a bifurcated hydrogen bond to both. When the N-cap is Asn or Asp, the 3_{10} -helix hydrogen bond is more likely to form if the N-cap side chain is in the *t* rotamer. [When Asn or Asp are *t* the 3_{10} -helix hydrogen bond is present in 22 out of 61 examples (36%). When Asn or Asp are *g*⁺ or *g*[−], the 3_{10} -helix hydrogen bond is present in 11 out of 46 examples (24%).] In this *t* conformation the hydrogen bonding pattern is N3 NH $\rightarrow$ N-Cap backbone C=O/N2 NH $\rightarrow$ N-Cap side chain C=O. This backbone/side chain hydrogen bonding pattern is reminiscent of the capping box (Harper & Rose, 1993).

Relationship between N-cap ψ and χ_1

The good N-cap residues Ser, Thr, Asp, and Asn show a restricted range of allowed ψ values when at the N-cap. Figure 1 shows a plot of N-cap ψ versus χ_1 for Ser and Thr at the N-cap. χ_1 is usually *g*[−] ($67 \pm 8^\circ$), as noted above, while ψ is confined to the narrow range of $164 \pm 8^\circ$. Figure 2 shows the same plot for Asp and Asn, showing the two alternative conformations adopted by these residues at the N-cap. If χ_1 is *g*[−] ($64 \pm 8^\circ$), ψ is $172 \pm 10^\circ$, similar to Ser and Thr. If χ_1 is *t* ($-176 \pm 9^\circ$), ψ is $107 \pm 19^\circ$. Figure 3 shows the same plot for the bad N-caps. χ_1 is usually *g*⁺ ($-63 \pm 15^\circ$), and ψ adopts the full range of allowed angles for residues in

Fig. 1. Correlation between N-cap ψ and χ_1 when α -helix N-cap is Ser or Thr.

Fig. 2. Correlation between N-cap ψ and χ_1 when α -helix N-cap is Asn or Asp.

the β region of the Ramachandran plot ($56 \pm 71^\circ$). This shows that ψ is restricted when there is a hydrogen bond between the N-cap and helical NH groups. When the N-cap is $g+$ there is no hydrogen bond from the side chain and ψ is unaffected. The narrow clustering of N-cap ψ values was previously noted by Harper & Rose

(1993) in capping box structures. The pairs of angles $\chi_1 = 64^\circ$, $\psi = 172^\circ$ or $\chi_1 = -176^\circ$, $\psi = -107^\circ$ are favored as they position the side chain oxygens close to the N3 NH or to both the N2 and N3 NH groups, respectively, in position for hydrogen bond formation.

Fig. 3. Correlation between N-cap ψ and χ_1 for bad α -helix N-caps.

N-cap Ser is found with χ_1 *t* to an appreciable extent, while with Thr at the N-cap this conformation is almost totally avoided. When Ser is *t*, ψ is $\approx 120^\circ$ (Fig. 1). This is a similar conformation to that often adopted by Asp or Asn; however, unlike Asn and Asp, the Ser side chain oxygen remains closest in space to the N3 NH and is unable to reach the N2 NH. The greater distance between the Ser O and the N3 NH groups in this conformation than when Ser χ_1 is *g*— is perhaps why *g*— is favored. When Thr is in this conformation at the N-cap there is a steric clash between the Thr CH₃ group and the side chain of N3; hence, this conformation is avoided.

Solvent accessible surface areas at the helix N-terminus

Table 8 shows the mean solvent accessible surface areas of the N-Cap, N1, N2, N3, and N4 nitrogen atoms. The total accessibility of the helix N-terminus is the sum of the accessibilities of N1, N2, and N3. A number of trends are apparent from this table: the accessibilities of N4 are very low, as this NH forms an *i*, *i* + 4 hydrogen bond within the helix. Gly has a higher mean accessibility to solvent than average (N1 + N2 + N3 accessibility is 30% greater than mean), consistent with the proposal of Fersht and co-workers, discussed above. The N3 accessibilities of the good N-caps are lower than the mean, as the N3 NH is often blocked from solvent access by the N-cap side chain. The N3 NH has a lower mean accessibility than N1 and N2 as it nearly always partly blocked by the N-cap residue. Helix N-termini are generally highly solvent exposed, as also indicated by the scarcity of tertiary interactions with distant side chains (Tables 4 and 5).

N-cap structures in 3_{10} -helices

3_{10} -Helices with definable N-caps are rarer than similar α -helices, and only 156 examples were found in our dataset of 85 proteins, making it more difficult to draw conclusions on these structures.

3_{10} -helices in proteins are generally shorter than α -helices (Barlow & Thornton, 1988) and common as extensions at the N- and C-termini of an α -helix (Baker & Hubbard, 1984; Barlow & Thornton, 1988).

In accord with a previous study by Karpen et al. (1992), we find that the 3_{10} -capping propensity largely parallels that of α -helix

Table 8. Mean α -helix N-terminal amide nitrogen atom solvent accessible surface areas ($\text{\AA}^2$) for 393 α -helices

N-Cap residue	Number	N-Cap	N1	N2	N3	N4	N1 + N2 + N3
Asn	43	1.35	1.85	1.18	0.25	0.01	3.27
Asp	64	1.43	2.38	0.92	0.26	0.07	3.56
Cys	2	0.00	2.91	0.09	0.00	0.00	3.01
Ser	61	2.04	3.20	1.26	0.26	0.02	4.72
Thr	50	0.25	2.69	0.83	0.20	0.12	3.72
Ala	11	2.74	2.03	1.71	1.11	0.08	4.84
Arg	14	1.60	1.92	1.49	0.56	0.15	3.97
Gln	6	0.64	1.69	1.71	0.38	0.00	3.79
Glu	7	0.89	1.63	1.73	0.55	0.00	3.91
Gly	61	2.84	3.19	1.57	0.44	0.03	5.21
His	10	1.38	1.61	0.85	0.09	0.06	2.55
Ile	3	1.70	2.00	0.18	0.00	0.00	2.19
Leu	10	1.46	1.40	0.58	0.30	0.08	2.28
Lys	13	0.79	3.15	1.52	0.46	0.04	5.14
Met	2	0.00	1.61	0.00	0.08	0.00	1.70
Phe	4	0.26	0.63	1.58	1.32	0.00	3.53
Pro	26	0.13	3.27	0.93	0.15	0.04	4.34
Trp	1	0.00	4.00	4.48	0.00	0.00	8.48
Tyr	2	0.21	1.32	0.68	0.00	0.00	2.00
Val	3	0.00	0.17	1.09	0.07	0.00	1.33
Mean		1.45	2.56	1.17	0.32	0.05	4.06

Table 9. χ_1 rotamer propensities for the 156 3_{10} -helix N-caps^a

Residue	g^-				t				g^+			
	#	%	Propensity	Mean angle	#	%	Propensity	Mean angle	#	%	Propensity	Mean angle
Asn	0	0	0	—	6	60	1.7	-163 ± 20	4	40	0.9	-57 ± 6
Asp	4	15	0.8	66 ± 22	19	73	2.0	-172 ± 8	3	12	0.3	-88 ± 28
Cys	0	0	0	—	1	50	1.4	179	1	50	1.1	-69
Ser	3	38	1.9	72 ± 19	2	25	0.7	172 ± 3	3	38	0.8	-64 ± 7
Thr	5	50	2.6	70 ± 21	2	20	0.6	176 ± 5	3	30	0.7	-56 ± 6
Arg	0	0	0	—	1	14	0.4	-178	6	86	1.9	-60 ± 15
Gln	0	0	0	—	1	33	0.9	-173	2	66	1.5	-63 ± 14
Glu	1	20	1.0	27	1	20	0.6	-138	3	60	1.3	-80 ± 31
His	0	0	0	—	4	80	2.2	-166 ± 6	1	20	0.5	-65
Ile	0	0	0	—	0	0	0	—	4	100	2.2	-62 ± 8
Leu	0	0	0	—	2	29	0.8	-148 ± 26	5	71	1.6	-78 ± 18
Lys	1	14	0.7	61	1	14	0.4	-174	5	71	1.6	-59 ± 11
Met	0	0	0	—	0	0	0	—	1	100	2.2	-51
Phe	0	0	0	—	0	0	0	—	5	100	2.2	-61 ± 8
Trp	1	50	2.6	59	1	50	1.4	-175	0	0	0	—
Tyr	0	0	0	—	1	100	2.8	-177	0	0	0	—
Val	0	0	0	—	2	66	1.9	161 ± 47	1	33	0.8	-66
Sum	15	44	47									
Mean				65 ± 20				-171 ± 13				-65 ± 16

^aIn calculating the overall averages and standard deviations, prolines were excluded.

capping (Table 1). However, there are some differences in propensities and rotamer preferences between the two types of helix. Ser and Thr are relatively less common in 3_{10} -helices than in α -helices, while His and Pro are more common (Karpen et al., 1992). Overall, the propensities for 3_{10} -caps show a similar variability than those of α -helices.

The rotamer preferences for residues in the 3_{10} N-cap position are listed in Table 9. The scarcity of data makes it difficult to infer significant correlations. However, grouping residues into classes according to their propensities, it is seen that Asp, Asn, and His all have a strong preference for the t rotamer and disfavor the g^- state relative to α -helical capping. As with α -helical N-caps, apolar side chains have a propensity to adopt the g^+ rotamer. This allows them to point away from the N-terminal backbone amides, which can then form hydrogen bonds with other parts of the polypeptide or water.

Hydrogen bonds formed by 3_{10} -helix N-cap side chains are listed in Table 10. His is a good N-cap in the 3_{10} -helix (propensity 1.5, Table 1) as it often forms a hydrogen bond to the N2 NH, while it rarely forms interactions in the α -helix (Tables 4 and 5). Ser and

Thr are poorer capping residues in 3_{10} -helix than α -helix because they primarily hydrogen bond with the N3 NH to stabilize the α -helix. In a 3_{10} -helix, the N3 NH is unavailable to the N-cap side chain so Ser and Thr have low 3_{10} -helix N-cap propensities and are g^+ . The longer Asp and Asn side chains can hydrogen bond to both N2 and N3 NH groups in the α -helix; in the 3_{10} -helix they often hydrogen bond to the free N2 NH group (Table 10) when in the t rotamer (Table 9).

Discussion

The structural preferences for hydrogen bonding and rotamer populations discussed above can be fully rationalized. Bad N-cap side chains cannot interact favorably with the N-terminus of the helix so they adopt the g^+ rotamer where the side chain is distant from the helix. Solvation of the free NH groups is improved and the N3 NH is more likely to form a 3_{10} -helix hydrogen bond to the N-cap backbone. Ser, Thr, and Cys at the N-cap nearly always adopt the g^- rotamer where they can hydrogen bond to the N3 NH. Asp and Asn either adopt the g^- rotamer, hydrogen bonding to the N3 NH, or are t when they can form bonds to both the N2 and N3 NH groups. The latter alternative is possible because the oxygen in the Asp and Asn side chains is one more carbon atom distant from the backbone compared to Ser and Thr. The next residues in the series, Gln and Glu, are too long for their oxygens to hydrogen bond to any helical NH groups; instead, they adopt the g^+ rotamer, thus removing the side chain from the vicinity of the helix terminus, in a similar manner to the other bad N-caps.

The results presented here have implications for the modified Lifson-Roig helix/coil model (Doig et al., 1994). At present, an N-cap free energy is applied to any length of consecutive helical residues. However, the fact that the N-cap residue usually interacts with the N3 residue, either via hydrogen bonding to the side chain or by a 3_{10} -helix hydrogen bond, shows that the N-cap weighting

Table 10. Hydrogen bond donors to 3_{10} -helix N-cap side chains

N-Cap residue	N1 amide	N2 amide	N3 amide	Total
Asn	0	2 (20%)	0	10
Asp	1 (4%)	15 (58%)	1 (4%)	26
Ser	0	2 (25%)	0	8
Thr	0	2 (20%)	2 (20%)	10
His	0	4 (80%)	0	5

should only be applied when there are at least three consecutive residues in a helical geometry after the N-cap residue. We are currently making this change.

A number of groups have previously published rotamer libraries for side chains within secondary structure (McGregor et al., 1987; Dunbrack & Karplus, 1994; Swindells et al., 1995). The N-cap rotamer preferences we find here are distinctly different for the preferences observed at interior positions of α -helices or, indeed, in any other structure previously examined. This suggests that N-caps should be treated uniquely in any application of rotamer libraries, such as estimation of side chain conformational entropy (Doig & Sternberg, 1995), structure prediction of proteins, protein/antibody complexes (Totrov & Abagyan, 1994), or protein/inhibitor complexes (Jackson & Sternberg, 1995), and in distinguishing correctly from incorrectly folded protein structures.

The significant differences between α - and 3_{10} -helix N-cap structures have implications for protein design. The choice of Ser or Thr as the N-cap residue will encourage α -helix initiation while discouraging 3_{10} -helix. The choice of Asp, Asn, or Pro may lead to either helix being formed while His may favor 3_{10} -helix formation over α .

The rotamer preferences we find at helix N-termini are greater than those examined in any other protein environment, such as interior positions of helices or sheets. This is because N-termini are solvent exposed and form few tertiary interactions that may move a side chain from its intrinsically preferred conformation. The N-termini of different helices are in similar environments, leading to consistently observed structures dominated by local interactions. The lack of tertiary interactions at N-cap sites in proteins means that studies of N-caps in isolated helical peptides are highly relevant to proteins. The correlation between preferences in peptides and proteins is better for N-caps (Doig & Baldwin, 1995) than interior helix positions (Chakrabarty et al., 1994). This may be because residues in interior positions in helices in proteins are more likely to be involved in tertiary interactions.

Conclusion

The following rules are generally observed for N-capping in α -helices: (1) Thr and Ser N-cap side chains adopt the g^- rotamer and hydrogen bond to the N3 NH. (2) Asp and Asn N-cap side chains either adopt the g^- rotamer and hydrogen bond to the N3 NH or adopt the t rotamer and hydrogen bond to both the N2 and N3 NH groups. (3) With all other N-caps, the side chain is found in the g^+ rotamer so that the side chain does not interact unfavorably with the N-terminus. (4) N-cap ψ angles are restricted to narrow ranges when a good N-cap is present. (5) An $i, i + 3$ N3 NH to N-cap C=O hydrogen bond is more likely to form at the N-terminus when a bad N-cap is present.

In 3_{10} -helices, Asp and Asn remain good N-caps, as they can adopt the t rotamer and hydrogen bond to the N2 NH; Ser and Thr become bad N-caps, as their preferred hydrogen bond donor (N3 NH) is inaccessible.

Materials and methods

The protein structure coordinates were taken from the January 1994 release of the Brookhaven Protein Databank (Bernstein et al., 1977). Only non-homologous chains were considered, defined using the approach of Orengo et al. (1993). From the total of 187 chains thus obtained, a reduced final dataset of 85 was used that excluded

the more poorly resolved structures (Swindells et al., 1995). This involved using the standard rule-of-thumb method for selecting well-resolved, well-refined structures: namely those having a resolution of 2.0 Å or better and an R-factor no greater than 20.0%. Secondary structure and main chain hydrogen bonding assignments were made using the Kabsch and Sander algorithm (Kabsch & Sander, 1983). The N-cap residue is defined as the first residue preceding the helix that is outside the α -right region (core + allowed) of the Ramachandran plot (see Fig. 4). Although this residue is not in the α -helical conformation, its carbonyl oxygen is the first participant in the CO(i) to NH($i + 4$) hydrogen bonding system of the α -helix. 3_{10} -helix N-caps were identified similarly except that the hydrogen bonding network was $i, i + 3$. Only those helices having N-caps that could be defined in this way were considered.

The propensity of a given residue for a particular attribute is defined as the ratio of the actual number of observations to the expected number of observations, where the expected number of observations is given by $(n_{Xaa}/n_{total}) \times n^*$, where

n_{Xaa} = total number of residue type Xaa in the dataset

n_{total} = total number of all residues in the dataset

n^* = total number of residues in the dataset with the attribute of interest (in this case helix N-caps).

Solvent accessibilities were calculated using the method of Lee & Richards (1971), as implemented in the program ACCESS (Hubbard, 1992). The program HBPLUS was used to calculate the potential hydrogen bonds to and from side chain atoms where the recognized donor and acceptor atoms have acceptable geometry (McDonald & Thornton, 1994).

Fig. 4. Ramachandran type map derived from the probability density distribution of ϕ, ψ angles from 462 protein crystal structures. The boundaries are arbitrarily defined according to the density within the $10 \times 10^\circ$ pixels. The most highly favored regions thus defined are shaded dark gray and referred to as the Ramachandran CORE regions. Surrounding the CORE regions are ALLOWED areas of lower population density (shown in a lighter gray).

Acknowledgments

A.J.D. thanks the Nuffield Foundation for financial support. J.M.T. and M.W.M. thank EU Contract BIO2CT-92524 and the BBSRC for financial support. B.J.S. is the grateful recipient of a BBSRC (U.K.) studentship. We thank Simon Hubbard for helpful discussions.

References

- Argos P, Palau J. 1982. Amino-acid distribution in protein secondary structures. *Int J Pept Protein Res* 19:380–393.
- Baker EN, Hubbard RE. 1984. Hydrogen-bonding in globular proteins. *Prog Biophys Mol Biol* 44:97–179.
- Barlow DJ, Thornton JM. 1988. Helix geometry in proteins. *J Mol Biol* 201:601–619.
- Bell JA, Becktel WJ, Sauer U, Baase WA, Matthews BW. 1992. Dissection of helix capping in T4 Lysozyme by structural and thermodynamic analysis of 6 amino-acid substitutions at Thr 59. *Biochemistry* 31:3590–3596.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanovich T, Tasumi M. 1977. The Protein Data Bank: Computer-based archive file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bordo D, Argos P. 1994. The role of side-chain hydrogen-bonds in the formation and stabilization of secondary structure in soluble proteins. *J Mol Biol* 243:504–519.
- Chakrabarty A, Baldwin RL. 1993. Comparison of amino acid helix propensities ("s-values") measured in different experimental systems. In: Cleland J, King J, eds. *Protein Folding: In Vivo and In Vitro*, ACS Symposium Series 526. Washington, DC: American Chemical Society. pp. 166–177.
- Chakrabarty A, Doig AJ, Baldwin RL. 1993. Helix capping propensities in peptides parallel those in proteins. *Proc Natl Acad Sci USA* 90:11332–11336.
- Chakrabarty A, Kortemme T, Baldwin RL. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing interactions. *Protein Sci* 3:843–852.
- Dasgupta S, Bell JA. 1993. Design of helix ends. Amino acid preferences, hydrogen bonding and electrostatic interactions. *Int J Pept Protein Res* 41:499–511.
- Doig AJ, Chakrabarty 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.
- Doig AJ, Baldwin RL. 1995. N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci* 4:1325–1336.
- Doig AJ, Sternberg MJE. 1995. Side chain conformational entropy in protein folding. *Protein Sci* 4:2247–2251.
- Dunbrack R, Karplus M. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein side-chains. *Nat Struct Biol* 1:334–339.
- 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.
- Harpaz Y, ElMasry N, Fersht AR, Henrick K. 1994. Direct observation of better hydration at the N-terminus of the α -helix with glycine rather than alanine as the N-cap residue. *Proc Natl Acad Sci USA* 91:311–315.
- Harper ET, Rose GD. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* 32:7606–7609.
- Hubbard SJ. 1992. Analysis of protein-protein molecular recognition [Ph.D. Thesis]. London: University College.
- Jackson RM, Sternberg MJE. 1995. A continuum model for protein-protein interactions—Application to the docking problem. *J Mol Biol* 250:258–275.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure—Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Karpen ME, DeHaseth PL, Neet KE. 1992. Differences in the amino-acid distributions of 3_{10} -helices and α -helices. *Protein Sci* 1:1333–1342.
- Lecomte JTJ, Moore CD. 1991. Helix formation in Apocytochrome-B5—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.
- 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 helix. *Biochemistry* 32:421–425.
- McDonald IK, Thornton JM. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793.
- McGregor MJ, Islam SA, Sternberg MJE. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol* 198:295–310.
- Muñoz V, Blanco FJ, Serrano L. 1995. The hydrophobic staple motif and a role for loop residues in α -helix stability and protein-folding. *Nat Struct Biol* 2:380–385.
- Orengo CA, Flores TP, Taylor WR, Thornton JM. 1993. Identification and classification of protein fold families. *Protein Eng* 6:485–500.
- Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α -helices. *Science* 240:1648–1652.
- Seale JW, Srinivasan R, Rose GD. 1994. Sequence determinants of the capping box, a stabilizing motif at the N-termini of α -helices. *Protein Sci* 3:1741–1745.
- Serrano L, Fersht AR. 1989. Capping and α -helix stability. *Nature* 342:296–299.
- Serrano L, Neira JL, Sancho J, Fersht AR. 1992. Effect of alanine versus glycine in α -helices on protein stability. *Nature* 356:453–455.
- Swindells MB, MacArthur MW, Thornton JM. 1995. Intrinsic ϕ , ψ propensities of amino acids derived from the coil regions of known structures. *Nat Struct Biol* 2:596–603.
- Totrov M, Abagyan R. 1994. Detailed ab-initio prediction of lysozyme-antibody complex with 1.6-Å accuracy. *Nat Struct Biol* 1:259–263.
- 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.
- Zhou HX, Lyu PC, Wemmer DE, Kallenbach NR. 1994. α -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.