

circulation with ISO (Fig. 2C). In a high-resolution Atlantic model, the maximum northward heat transport with ISO is 1 PW (13). Thus, in both models with ISO there is consistency with the empirical estimate at 25°N of 1.1 ± 0.3 PW (20), which does not occur with HOR.

With HOR, convection occurs in many places (Fig. 4A), much more extensively than is thought to occur in nature. In the ACC, the reason for this difference is the extensive upwelling near 60°S (Fig. 2A) that weakens the density stratification and conduces convection. This upwelling is much weaker with ISO (Fig. 2C), and there is consequently much less convection there (Fig. 4B). In general, with ISO deep-water formation occurs only in a few small locations: in the Arctic and far North Atlantic, in the Labrador Sea, and in the Weddell and Ross seas around Antarctica where deep-water formation has been observed.

A new, isopycnally oriented, adiabatic parameterization for mesoscale tracer transport has been tested in a non-eddy-resolving, global ocean model. This parameterization replaces the physically unjustifiable horizontal diffusion that has been the common practice, and it produces substantial improvements: a sharper main thermocline, a cooler abyssal ocean, a meridionally expanded overturning circulation in the North Atlantic that strengthens the poleward and surface heat fluxes, an eddy-induced cancellation of the Deacon cell that greatly diminishes the poleward and surface heat fluxes across the ACC, and a confinement of deep convection to regions where it is known to occur.

REFERENCES

1. W. S. Broecker, *Oceanography* 4, 79 (1991).
2. H. L. Bryden, *J. Mar. Res.* 37, 1 (1979); R. A. de Szoeke and M. D. Levine, *Deep Sea Res.* 28, 1057 (1981).
3. H. L. Bryden and E. C. Brady, *J. Mar. Res.* 47, 55 (1989).
4. J. C. McWilliams *et al.*, in *Eddies in Marine Science*, A. R. Robinson, Ed. (Springer-Verlag, Berlin, 1983), pp. 92–113.
5. K. Bryan and M. D. Cox, *Tellus* 19, 54 (1967); *J. Atmos. Sci.* 25, 945 (1968).
6. F. Bryan, *J. Phys. Oceanogr.* 17, 970 (1987).
7. K. Bryan and L. J. Lewis, *J. Geophys. Res.* 84, 2503 (1979).
8. J. R. Toggweiler, K. Dixon, K. Bryan, *ibid.* 94, 8217 (1989).
9. C. O'D. Iselin, *Eos* 20, 414 (1939); R. B. Montgomery, *Bull. Am. Meteorol. Soc.* 21, 87 (1940).
10. T. J. McDougall and J. A. Church, *J. Phys. Oceanogr.* 16, 196 (1986); M. C. Gregg, *J. Geophys. Res.* 92, 5249 (1987).
11. M. H. Redi, *J. Phys. Oceanogr.* 12, 1154 (1982).
12. M. D. Cox, *Ocean Modelling* 74, 1 (1987).
13. C. Boning, W. R. Holland, F. Bryan, G. Danabasoglu, J. C. McWilliams, unpublished results.
14. P. R. Gent and J. C. McWilliams, *J. Phys. Oceanogr.* 20, 150 (1990).
15. P. R. Gent, J. Willebrand, T. J. McDougall, J. C. McWilliams, unpublished results.
16. D. G. Andrews, J. R. Holton, C. B. Leovy, *Middle Atmosphere Dynamics* (Academic Press, Or-

- lando, FL, 1987); R. A. Plumb and J. D. Mahlman, *J. Atmos. Sci.* 44, 298 (1987).
17. S. Levitus, *Nat. Oceanic Atmos. Adm. Prof. Pap.* 13 (1982), p. 1.
18. K. Trenberth and A. Solomon, *Clim. Dyn.*, in press.
19. A. L. Gordon and H. W. Taylor, in *Numerical*

- Models of Ocean Circulation* (National Academy of Sciences, Washington, DC, 1975), pp. 54–56.
20. H. L. Bryden and M. M. Hall, *Science* 207, 884 (1980).

28 December 1993; accepted 4 March 1994

Rules for α -Helix Termination by Glycine

Rajeev Aurora, Rajgopal Srinivasan, George D. Rose*

A predictive rule for protein folding is presented that involves two recurrent glycine-based motifs that cap the carboxyl termini of α helices. In proteins, helices that terminated in glycine residues were found predominantly in one of these two motifs. These glycine structures had a characteristic pattern of polar and apolar residues. Visual inspection of known helical sequences was sufficient to distinguish the two motifs from each other and from internal glycines that fail to terminate helices. These glycine motifs—in which the local sequence selects between available structures—represent an example of a stereochemical rule for protein folding.

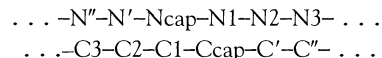
The folded structure of a protein is encrypted in its amino acid sequence (1), written in a code that remains obscure. Is this fold the inevitable outcome of an energetically controlled process, like rolling a ball down a hill? Or do discrete instructions for each structural motif reside somewhere in the sequence? Such alternatives cannot be distinguished by the observation of a folding reaction, any more than a computer algorithm can be deduced by the observation of the way a running program transforms input to output. We have been exploring the hypothesis that protein conformation is specified by a stereochemical code (2), similar in principle to the code for DNA, in which strand complementarity is determined by hydrogen bonds (3). Focusing initially on the α helix, we analyzed proteins of known structure to extract patterns and modeled short fragments to characterize each pattern fully.

The α helix is characterized by consecutive main chain hydrogen bonds between each amide hydrogen and the carbonyl oxygen of the previous helical turn (4). This pattern is truncated at helix ends because, upon termination, no next turn of the helix exists to provide hydrogen bond partners. Such end effects are substantial, encompassing two-thirds of the residues for the helix of average length (5). The term helix “capping” is used to describe those alternative hydrogen bond patterns that can satisfy the backbone N–H and C=O groups in the initial and final turns of the helix, respectively (5, 6). Capping stabilizes α helices in both proteins (7–9) and

peptides (7, 10–12) and inhibits expected fraying (10, 13).

In this report we describe and analyze two Gly-based capping motifs that are prevalent at the COOH-termini of α helices. From the analysis, a set of simple stereochemical rules was developed. With these rules, the visual inspection of known helical sequences was sufficient to differentiate between the two motifs and distinguish them from internal Gly residues that did not terminate helices.

The nomenclature for helices and their flanking residues is as follows:



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 boundary residues that belong to both the helix and the adjacent turn. In practice, residues classified as helical have backbone dihedral angles, ϕ and ψ , with mean values of $-64^\circ \pm 7^\circ$ and $-41^\circ \pm 7^\circ$, respectively; Ncap and Ccap make one additional intrahelical hydrogen bond while departing from these means (5).

All capping interactions were cataloged among the α helices of 42 proteins of known structure (14). Consistent with earlier findings (5), the NH₂-termini of helices were capped predominantly by nearby residue side chains. In contrast, the COOH-termini were usually capped by local backbone interactions, many involving Gly residues.

Approximately one-third of all α helices terminate with a Gly residue at the COOH-terminus. Helices terminating in Gly can be classified by their hydrogen bonding pattern into two primary motifs (Fig. 1). The larger class, which includes 24 of the

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

*To whom correspondence should be addressed.

52 Gly-terminated helices, has a distinctive, doubly hydrogen bonded pattern between backbone partners, consisting of 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, respectively (Table

1). The smaller class, which contains 18 of the remaining Gly-terminated helices (35%), has a 5 → 1 hydrogen bond between the N-H at C' and C=O at C3 (Table 1). Glycine adopts a left-hand conformation in both classes (15). In two

additional cases, Gly at C' is followed consecutively by a Pro residue. Proline, which "kinks" helices, forces a kink here as well, with a 4 → 1 hydrogen bond between the N-H at Ccap and C=O at C3. The eight remaining Gly-terminated structures in our data base involve ligand binding and are probably perturbed variants of the two primary motifs, as discussed below.

Both motifs were described previously (16-18). We refer to the 6 → 1, 5 → 2 hydrogen-bonded structure as the Schellman motif and to the remaining structure with a 5 → 1 hydrogen bond as the α_L motif. In either motif, the requirement that the C' residue adopt a left-hand conformation is fulfilled most easily by Gly (19). Nonetheless, residues with C_β atoms, such as Asn, do occupy this region in proteins (19). Occasionally in our survey, Asn or Lys were observed in lieu of Gly at C' in the Schellman motif. Both motifs are compact, with hydrogen-bonded atoms that are shielded from solvent (20). In contrast, the side chain of the C1 residue and the carbonyl oxygens of C1, Ccap, and C' are solvent-exposed. In the Schellman motif, C1 is invariably either a polar residue or Ala. The solvation of Gly turns has been noted (9).

The existence of hydrophobic interactions between the side chains of C' and C3 was proposed (16), and we have quantified this proposition. Hydrophobic interactions are identified as the juxtaposition of nonpolar atoms with a distance less than the sum of their van der Waals radii plus 2.8 Å (the approximate diameter of a water molecule).

The presence or absence of hydrophobic interactions between C' and C3 distinguishes effectively between the two motifs and thereby constitutes a stereochemical selection rule (Table 2). When C' is apolar, a hydrophobic contact is formed between the side chains of C' and C3; or, to a lesser extent, between C' and C4 or occasionally C2, when C3 is polar. This configuration favors formation of the (C') N-H . . . O=C (C3) hydrogen bond, promoting the Schellman motif. In our data set, at least one of the three residues, C3, C4, or C2, is invariably apolar. On the other hand, when C' is polar, the chain traverses an alternative path between the hydrophobic and hydrophilic interface of the helix. This latter trajectory is incompatible with the Schellman motif, and a (C') N-H . . . O=C (C3) hydrogen bond is formed instead.

When the stereochemical selection rule is applied, Lys and Arg segregate with the apolar residues at C', C3, and either C4 or C2. Although these residues are charged, their long alkyl side chain moieties can also function as suitable sites for hydrophobic interaction (19). Further stabilization arising from a favorable interaction between

Table 1. Partitioning of Gly-terminated helices into two motifs. The sequence of residues C4 through C''' of each helix is shown in one-letter codes (37). Proteins are identified by their PDB identifier (14).

Schellman motif (C' → C3 and C' → C2)			α _L motif (C' → C3)		
PDB code	Helix bounds	Sequence	PDB code	Helix bounds	Sequence
351c	2-10	LFKNK-GCVA	351c	39-50	QRIKN-GSQG
2act	69-80	FIIND-GGIN	2act	24-42	NKITS-GSLI
5cpa	253-261	WSYNQ-GIKY	5cha	164-172	CKKYW-GTKI
2cts	37-43	VDDMMY-GGMR	5cpa	173-186	FVKNH-GNFK
2cts	152-160	RAYAE-GIHR	2cts	89-98	WLLVT-GQIP
2cts	276-292	LQKEV-GKDV*	2cts	208-217	FTNML-GYTD†
2cts	297-311	NTLNS-GRVV	4dfr	77-85	AIAAC-GDVP
2cts	344-364	VLLLEQ-GKAK	1gd1	251-264	KAAAE-GELK‡
2cts	393-415	WSRAL-GFPL	3grs	196-209	ILSAL-GSKT†
4fxn	10-26	GIIES-GKDV	3ins	B8-B19	LYLVC-GERG
4fxn	93-106	RMNGY-GCVV	1lz1	24-36	AKWES-GYNT
1gd1	101-111	KHLEA-GAKK	3lzm	92-106	MVFQM-GETG
3grs	29-42	RAAEL-GARA	3lzm	142-155	TTFRT-GTWD
3grs	227-241	ELENA-GVEV	9pap	24-42	IKIRT-GNLN
3grs	383-391	AIHKY-GIEN*	3rrt	12-29	KLHED-GETV†
3grs	444-453	VAVKM-GATK	3tlh	234-245	YLISQ-GGTH
1lz1	4-15	TLKRL-GMDG	3tlh	280-296	ATDLY-GSTS
3lzm	38-50	LDKAI-GRNC	2wrp	44-63	EELLR-GEMS
1mbo	124-149	KYKEL-GYQG			
9pap	67-78	LVAQY-GIHY*			
4pep	135-143	NLWDQ-GLVS			
4pep	224-235	IQSDI-GASE			
1snc	98-106	ALVRQ-GLAK			
3tlh	300-313	AFDAV-GVK			

*High-temperature factors (B factor > 50) for some atoms in the motif; in these three cases, the 6 → 1 hydrogen bond is not observed, but other criteria for a Schellman motif are satisfied. †Structures with the α_L motif, with an additional hydrogen bond from C' to the C3 acceptor. ‡The hydrogen bond distance criterion is satisfied, but orientation is skewed; however, a high-temperature factor for all Gly atoms is noted (B factor ≈ 47).

Table 2. The stereochemical rules for Gly motifs. Rules for helix continuation or termination are summarized: C' is the key position. If Gly is followed by an apolar residue at C', an interaction with the apolar residue at C3 (or C4 if C3 is polar, or C2 if both C3 and C4 are polar) terminates the helix in a Schellman motif, exposing C1 to solvent. If Gly is followed by a polar residue, the helix terminates in an α_L motif unless a favorable side chain-side chain interaction between C' and C2 prevents helix termination. When C' is Gly (that is, Gly-Gly), the polarity of C''' (in lieu of C') selects between the Schellman and α_L motifs; apolar: Ala, Val, Ile, Leu, Met, Phe, Trp, Cys; polar: Ser, Thr, Asn, Asp, Gln, Glu, Arg, Lys; omitted: His, Tyr, Pro. The long alkyl side chain moieties of Lys and Arg can function as hydrophobic sites (19) and are observed to interact with apolar residues at the indicated positions. Histidine can be polar or apolar, depending on its protonation state, which cannot be determined in x-ray structures. Tyrosine acts ambiguously at C', partitioning to both motifs. Glycine-proline terminates a helix with a (Ccap) N-H . . . O=C (C3) hydrogen bond; unlike the Schellman and α_L motifs, Gly in Gly-Pro structures has a negative value of φ.

C3 > C4 > C2	Residue				Structural motif
	C1	C'	C''	C'''	
Apolar Lys or Arg Polar*	Polar or Ala Apolar*	Gly	Apolar Lys or Arg	Not "bulky"†	Schellman
not Lys not Arg Apolar Lys or Arg preferred	not Ala	Gly	Apolar Lys or Arg		Helix continuation
		Gly only	Polar* not Lys not Arg		α _L

*The helix continues if C1 is apolar or if C3, C4, and C2 are all polar and interaction with C''' is not possible. †In our polyalanine model, Trp was strongly disfavored for steric reasons, although smaller aromatic residues were allowed. More realistic sequences may impose further steric constraints.

the terminal-charged group and the helix macrodipole is also a possibility (21). At C', Tyr, an amphipathic residue, may or may not segregate with the Schellman motif, depending on the microenvironment.

Eight exceptions to this stereochemical selection rule were found in our data set, cases with apolar residues at C'' that did not adopt a Schellman motif as predicted (Table 3, top). Each exception involved either chain termination or interaction with an ion, prosthetic group, disulfide bond, or salt bridge that interfered specifically with the C''-C3 hydrophobic contact. These interactions were probably causal "rule-breakers" because they were present in each exception but absent from all Schellman motifs.

Glycine in a helix has three alternatives: allow helix propagation or cause termination in either a Schellman or an α_L motif. A few simple rules are sufficient to predict whether a helix will propagate through a particular Gly or terminate instead (Table 2). These rules can be rationalized by reference to a helical wheel (22). Glycine followed by an apolar residue at C'' will result in a Schellman motif, providing that C3 (or C4 or C2 in the event that C3 is polar) is also apolar. The interaction between C'' and C3 (or C4 or C2) establishes an extensive hydrophobic patch that partitions with the buried side of the helix. Necessarily, the C1 position is situated on the opposite side, where it is solvent-exposed. Consistent with this logic, either a polar residue or Ala is always found at the C1 position in a Schellman motif (Table 1). Consequently, a polar residue at C3 (or C4 or C2) precludes the hydrophobic interaction with C''. Equivalently, an apolar residue at C1 forces the hydrophobic surface to be solvent-exposed. Either condition is sufficient to spoil the Schellman motif and promote helix propagation.

Alternatively, Gly followed by a polar residue will cause a helix to terminate in an α_L , unless this residue can make a favorable side chain-side chain interaction with the C2 residue. The five exceptions to this termination rule (Table 3, bottom) are similar in kind to the eight Schellman exceptions (Table 3, top) (23).

We define a helix stop signal to be a residue sequence for which it is energetically more favorable to terminate the helix than to continue it. Using exhaustive conformational analysis of suitable peptides (24), we determined that the two motifs are helix stop signals (25). The simple qualitative modeling used here resembles the approach used originally by Pauling and colleagues to predict the existence of the α helix (4). Qualitative modeling relies on stereochemistry to distinguish between allowed and disallowed conformations, without regard to a quantitative estimate of the

particular energies involved.

As a summary of the modeling for the Schellman motif, the 6 \rightarrow 1, 5 \rightarrow 2 hydrogen bonds force the motif, as does the C'' \rightarrow C3

hydrophobic interaction. In addition, termination was energetically favored over continuation of the helix. An automatic consequence of the motif was to position the C' Gly

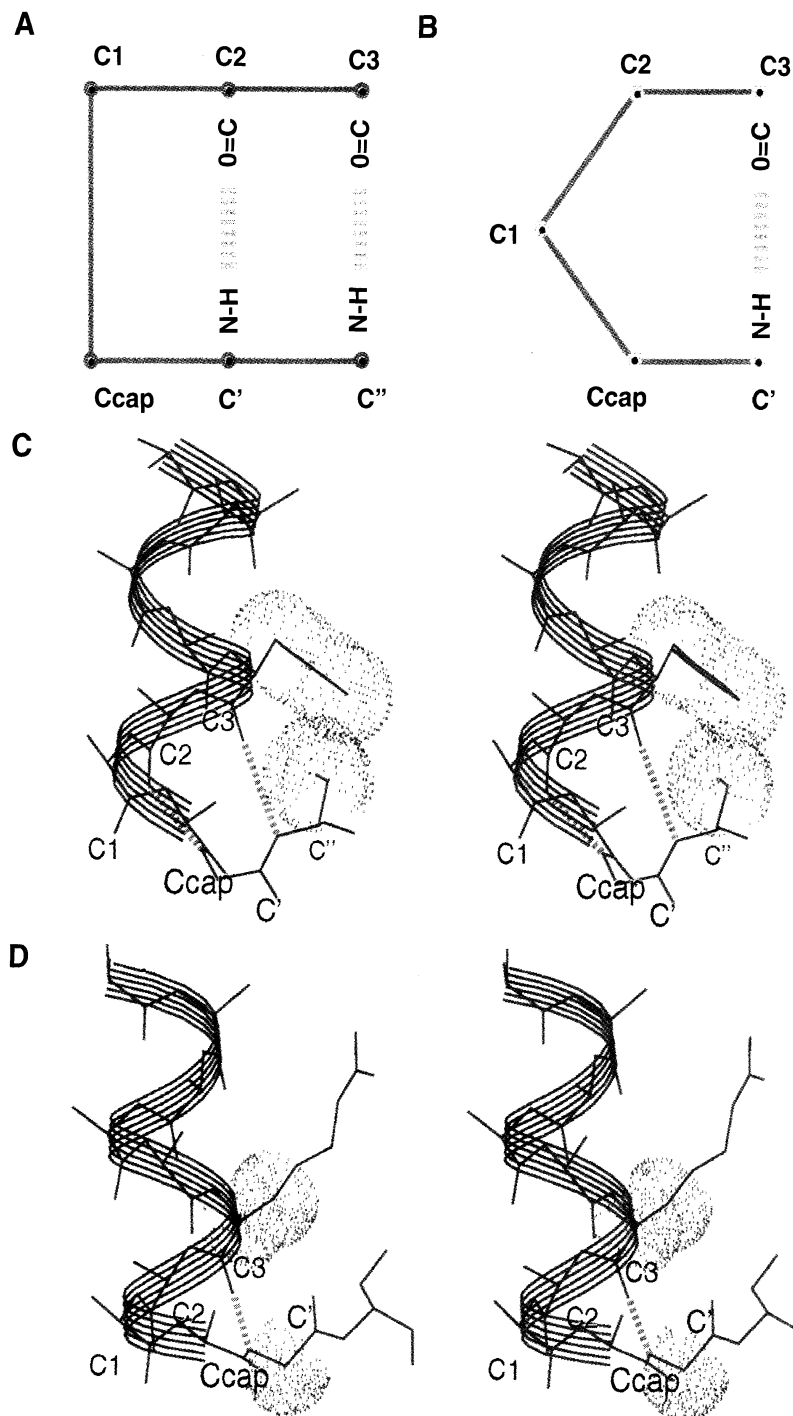


Fig. 1. An example of each Gly motif, from cytochrome C551, designated as 351c in the PDB (14). (A and B) Schematic diagrams of the structures shown in (C) and (D), respectively, in which only the hydrogen bonding pattern is emphasized. Molecular representations (C) and (D) include all backbone (N, C α , C, and O) and C β atoms, with hydrogen bonds represented as green dashed lines. The stipled surface represents the van der Waals radius surrounding selected atoms. (C) Helix from residues 2 (Ncap) to 10 (Ccap), terminating in a Schellman motif. The interaction between side chains of C'' (Cys) and C3 (Phe) and backbone hydrogen bonds from C'' to C3 (i \rightarrow i-5) and C' to C2 (i \rightarrow i-3) is shown. (D) Helix from residues 39 (Ncap) to 50 (Ccap), terminating in an α_L motif. The hydrogen bond from C' to C3 (i \rightarrow i-4) is shown.

in a left-handed turn conformation. For the α_L motif, Gly was the only sterically allowed residue at the C' position. Conclusions from modeling were confirmed experimentally (26). In every instance of an α_L motif, Gly was found at C', unlike the analogous situation for a Schellman motif, in which a non-Gly residue was occasionally used (Fig. 1). At the C' position, electrostatic considerations force the Gly to adopt a left-hand turn conformation. Although it could not be shown conclusively that the termination of a helix in an α_L motif was of lower energy than the extension of the helix through C', this proposition is likely for entropic reasons (27).

These conclusions from modeling were confirmed experimentally by the application of the rules to a second data set. An additional 26 proteins were selected from the Sander set (26). All 26 are high-resolution structures ($\leq 2.0 \text{ \AA}$) with little homology ($\leq 35\%$) to proteins in our learning set and with at least one helix. From

Table 3. Helices in which a Schellman motif is predicted but not found, or where termination of the helix was predicted but not found. Each exception involves chain termination or binding to an external ligand, a disulfide bond, or a salt bridge. Proteins are specified by their PDB identifier (14). Structures are as follows: R3₁₀, right-handed 3₁₀ helix; α_L , the α_L motif; C' → C4, hydrogen bond from Gly at (C') NH to O=C at (C4). Each listed hetero group involves an interaction between at least one residue in the Schellman sequence and atoms in the hetero group, with an interaction distance of $\leq 6.8 \text{ \AA}$. In 1gd1 (312 to 331), the chain terminates at residue C'.

Protein (PDB code)	Observed structure or Gly residue	Helix boundaries (Ncap-Ccap)	Reason for exception
<i>Exceptions to the Schellman selection rule</i>			
2cdv	Gly at Ccap	78-88	Heme
1crn	α_L	22-31	Disulfide at C''
4dfr	α_L	43-50	Methotrexate
1gd1	C' → C4	148-165	SO ₄ ²⁻
1gd1	R3 ₁₀	312-331	No C''
3lzm	α_L	2-12	Salt bridge
5pti	α_L	47-56	Unknown metal
2wrp	α_L	67-76	Trp inducer
<i>Internal glycines associated with hetero groups</i>			
2aza	Gly ⁶³	55-66	Cu ⁺⁺
1mbo	Gly ²⁵	20-36	Unknown*
1mbo	Gly ¹²⁹	124-150	Unknown*
3tlh	Gly ¹⁷³	158-181	Ca ⁺⁺
2wrp	Gly ⁵²	44-63	Trp inducer

*In apomyoglobin (23), neither the B nor the H helix terminates at Gly (25) or Gly (129), respectively. In both helices, Gly is residue N5, with strong NH₂-capping observed. The influence of these NH₂-capping interactions may extend slightly beyond the first helical turn in both cases.

this set, all helices of seven or more residues were identified (15); 219 were found. Of these, 58 terminated in a Gly and 34 had at least one Gly within the helix. All helices with Gly residues were assessed for consistency with the selection rules (Table 2). All but three of the 58 Gly-terminated helices satisfied the rules. Of the three exceptions, two involved additional complicating factors (similar to those listed in Table 3). Thus, only a single exception to the rules was found. Of the 34 internal Gly residues, 12 exceptions to the rules were observed; 10 of these involved direct contact ($< 6.8 \text{ \AA}$ interatomic distance) with external ligands.

In peptides, helix-capping by side chains is known to have a pronounced effect at the NH₂-terminus (10–12) but a questionable effect at the COOH-terminus (10, 12). Nevertheless, the existence of helix stop signals at COOH-termini can be inferred (28). These experimental findings, together with the preceding analysis, prompt us to enlarge our earlier definition (5). Accordingly, we now define capping to include both local side chain and backbone hydrogen bonding interactions that can satisfy any of the otherwise unsatisfied initial four amide hydrogens or final four carbonyl oxygens of the helix.

The two structural motifs described here each have two facets—one helical and the other not. The helical facet rigidifies the final helical turn, by hydrogen bonding to C3, whereas the other facet terminates the helix and directs the polypeptide chain along a new trajectory. An analogous double-faceted structure is found at helix NH₂-termini, where the “capping box” (7) uses side chain–backbone hydrogen bonding to stabilize the initial helical turn, by hydrogen bonding to N3, while simultaneously terminating the helix.

The mutation of key residues in α helices of staphylococcal nuclease and T4 phage lysozyme destabilizes these structures. Such experimental observations can now be explained in the framework of our stereochemical rules. In nuclease (29, 30), mutations at any key site of a Schellman motif—C3 (L103), C' (G107), and C'' (L108)—rank among the most destabilizing observed. In lysozyme (31), C'' (T157) in an α_L structure was mutagenized (32) to 13 other residues. All substitutions at this site by apolar residues were found to be destabilizing. The rules indicate that apolar substitutions at C'' are expected to switch the conformation from an α_L to a Schellman motif. However, in lysozyme the bulky W158 at C'' precludes this possibility (Table 2). We predict that mutation of W158 to a nonaromatic residue will permit formation of a Schellman motif and, in particular, that the I157 mutant (32) will then lose its observed temperature-sensitive phenotype. From studies of the barnase folding pathway, it has been concluded that helix COOH-termini form

early in the folding pathway (33). However, each barnase helix terminates in a Schellman motif, and this fact may limit the generality of earlier conclusions.

Our success in deriving rules by analyzing known helices suggests that the structure of Gly-terminated helices is probably determined by the local sequence. The existence of such rules lends support to the more general proposition that protein conformation is specified by an underlying stereochemical code (2).

REFERENCES AND NOTES

- C. B. Anfinsen, *Science* **181**, 223 (1973).
- E. E. Lattman and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 439 (1993).
- J. D. Watson and F. H. C. Crick, *Nature* **171**, 737 (1953).
- L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205 (1951).
- L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988).
- J. S. Richardson and D. C. Richardson, *ibid.*, p. 1648.
- E. T. Harper and G. D. Rose, *Biochemistry* **32**, 7605 (1993).
- L. Serrano and A. R. Fersht, *Nature* **342**, 296 (1989); J. A. Bell *et al.*, *Biochemistry* **31**, 3590 (1992); J. T. J. Lecomte and C. D. Moore, *J. Am. Chem. Soc.* **113**, 9663 (1991).
- L. Serrano, J. Sancho, A. R. Fersht, *J. Mol. Biol.* **227**, 544 (1992).
- P. Lyu *et al.*, *Biochemistry* **32**, 421 (1993); H. X. Zhou *et al.*, *Proteins Struct. Funct. Genet.* **18**, 1 (1994).
- M. D. Bruch, M. M. Dhingra, L. M. Gierasch, *Proteins Struct. Funct. Genet.* **10**, 130 (1991); P. C. Lyu, H. X. Zhou, N. Jelveh, D. E. Wemmer, N. R. Kallenbach, *J. Am. Chem. Soc.* **114**, 6560 (1992); B. Forood, E. J. Feliciano, K. P. Nambiar, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 838 (1993); N. Yumoto *et al.*, *Biochem. Biophys. Res. Commun.* **196**, 1490 (1993).
- A. Chakraborty, A. J. Doig, R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11332 (1993).
- B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 536 (1959).
- The set of α helices was the same as used previously (7) and consists of all helices identified in 42 representative proteins (resolution $\leq 2.0 \text{ \AA}$; *R* factor ≤ 20) from the Protein Data Bank (PDB): F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977). Proteins, designated by their PDB four-letter identifiers, are as follows: 2ACT, 5CHA, 1HOE, 1PPT, 2AZA, 2CA2, 5CPA, 2CTS, 1CRN, 5CYT, 2CDV, 351C, 1GD1, 4DFR, 1ECD, 4FXN, 1GCR, 1GPI, 3GRS, 1HMQ, 2LHB, 2RHE, 3INS, 1LZ1, 3LZM, 1MBO, 2OVO, 9PAP, 3APP, 4PEP, 1BP2, 7RSA, 3RNT, 1RDG, 1SN3, 1SNC, 3TLN, 2WRP, 1TPP, 5PTI, and 1UBQ.
- Unlike the findings in (5), backbone dihedral angles used in the classification of the C1 residue were allowed to vary by $\pm 20^\circ$ from observed means, near $(-60, -40)$, to compensate for the high degree of flexibility and corresponding uncertainty at Gly-terminated helices. With this adjustment, Gly is found invariably at C' in both hydrogen-bonded motifs. This convention differs from those used in (6) and by G. D. Rose and R. Wolfenden [*Annu. Rev. Biophys. Biomol. Struct.* **22**, 381 (1993)]. In a similar vein, angular constraints between the donor and acceptor were relaxed by 10° from our earlier convention (5). The larger class had mean ϕ and ψ values of $+80^\circ \pm 7^\circ$ and $+23^\circ \pm 16^\circ$, respectively. The smaller class populates either of two distinct conformers; one that is similar to the larger class, with mean ϕ and ψ values of $+82^\circ \pm 14^\circ$ and $+12^\circ \pm 16^\circ$, respectively, and another with mean ϕ and ψ values of $+81^\circ \pm 19^\circ$ and $+180^\circ \pm 31^\circ$, respectively.
- C. Schellman, in *Protein Folding*, R. Jaenicke, Ed. (Elsevier/North-Holland, New York, 1980), pp. 53–61.

17. E. J. Milner-White, *J. Mol. Biol.* **199**, 503 (1988).
18. E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **44**, 97 (1984); R. Preißner and P. Bork, *Biochem. Biophys. Res. Commun.* **180**, 660 (1991); P. Bork and R. Preißner, *ibid.*, p. 666; S. Dasgupta and J. A. Bell, *Int. J. Peptide Protein Res.* **41**, 499 (1993); H. A. Nagarajam, R. Sowdhamini, C. Ramakrishnan, P. Balaram, *FEBS Lett.* **321**, 79 (1993).
19. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
20. In the Schellman motif, the mean accessibility of N and O is 7% in the (C') N-H...O=C (C3) hydrogen bond and 25% in the (C') N-H...O=C (C2) hydrogen bond. Correspondingly, in the α_L motif, the mean accessibility of N and O is 9% in the (C') N-H...O=C (C3) hydrogen bond. Solvent-accessible surface area is defined by B. K. Lee and F. M. Richards [*J. Mol. Biol.* **55**, 379 (1971)] and calculated with ANAREA, an algorithm attributable to T. J. Richmond [*ibid.* **178**, 63 (1984)]. Results were normalized to a percentage with the use of values from G. J. Lesser and G. D. Rose [*Protein Struct. Funct. Genet.* **8**, 6 (1990)].
21. K. R. Shoemaker *et al.*, *Nature* **326**, 563 (1987).
22. M. Schiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967).
23. M. J. Cocco and J. T. J. Lecomte, *Protein Sci.*, in press.
24. Backbone dihedral angles used for C3-C2-C1 were $\phi = -64^\circ$, $\psi = -41^\circ$; van der Waals radii were taken from (34).
25. All conformations of the hexapeptide Leu-Ala-Ala-Gly-Ile were sampled on a 30° grid, with the first three residues held helical. Side chain torsions were allowed to occupy only staggered conformations. The resulting 291,599 allowed conformations constituted the "working set." To assess the role of hydrogen bonds (35) in a Schellman motif, the working set was filtered to select conformations in which hydrogen-bonded partners were within 3.5 Å. Upon energy minimization of the 157 conformations satisfying this criterion, all conformations with hydrogen bonds also had the two hydrophobic residues in contact. Minimizations were performed with the AMBER/OPLS force field with MacroModel (34). In the converse experiment, all conformations with at least one hydrophobic contact between residues 1 and 6 were selected and energy-minimized. The most stable structure was the Schellman motif. Because the Schellman and α helix structures contain the same number of hydrogen bonds, the factor stabilizing the Schellman over the helix is probably the hydrophobic contact. Consistent with this view, energy minimization of the model peptide that was started from a helical conformation resulted in considerable distortion, but the initial structure persisted when started from a Schellman conformation. The number of hydrogen bonds and hydrophobic contacts is identical for an α_L motif and an α helix of the same length, so there is no compelling reason for helix termination over continuation. A possible explanation is the greater conformational freedom available to Gly in the less restrictive α_L conformation. Other residues can substitute for Gly in the Schellman motif but not in the α_L motif, perhaps because the strain imposed by the adoption of a right-handed conformation is counterbalanced by the existence of favorable hydrogen bond and hydrophobic contacts between residues C3 and C' in a Schellman motif. Accordingly, energy minimization showed that a Gly-based α_L is less disfavored than its Ala-based counterpart.
26. U. Hohom, M. Scharf, R. Schneider, C. Sander, *Protein Sci.* **1**, 409 (1992), as applied to the November 1993 release of the PDB (13).
27. T. P. Creamer and G. D. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5937 (1992); S. D. Pickett and M. J. E. Sternberg, *J. Mol. Biol.* **231**, 825 (1993).
28. P. S. Kim and R. E. Baldwin, *Nature* **307**, 329 (1984).
29. P. J. Loll and E. E. Lattman, *Protein Struct. Funct. Genet.* **5**, 183 (1989).
30. D. Shortle *et al.*, *Biochemistry* **29**, 8033 (1990); S. M. Green *et al.*, *ibid.* **31**, 5717 (1992).
31. S. J. Remington *et al.*, *J. Mol. Biol.* **118**, 81 (1978).
32. T. Alber *et al.*, *Nature* **330**, 41 (1987).
33. A. Horowitz *et al.*, *J. Mol. Biol.* **219**, 5 (1991); L. Serrano *et al.*, *ibid.* **224**, 847 (1992).
34. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988); S. J. Weiner *et al.*, *ibid.* **106**, 765 (1984); F. Mohamadi *et al.*, *J. Comput. Chem.* **4**, 440 (1990).
35. D. F. Stickle, L. G. Presta, K. A. Dill, G. D. Rose, *J. Mol. Biol.* **226**, 1143 (1992).
36. Computer-generated drawings were made with Insight II, 2.30; Biosym Technologies, San Diego, CA. Illustrations were provided with SCIAN, 0.852B; E. Pepke, J. Murray, J. Lyons, and T.-Y. Hwo, Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL.
37. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
38. We thank T. Creamer, E. Harper, N. Kallenbach, E. Lattman, and J. Seale for useful discussions; J. Ponder for the area subroutine from his TINKER package; R. Baldwin, N. Kallenbach, and J. Lecomte for preprints of papers; C. Frieden for a critical reading of the manuscript; and K. Henrick and A. Fersht for barnase coordinates. Supported by a grant from the NIH (GM 29458).

7 December 1993; accepted 22 March 1994

DNA Targets for Certain bZIP Proteins Distinguished by an Intrinsic Bend

David N. Paoletta, C. Rodgers Palmer, Alanna Schepartz*

In spite of the large amount of sequence conservation among the DNA binding segments of basic region leucine zipper (bZIP) proteins, these proteins can discriminate differently between target sequences that differ in half-site spacing. Here it is shown that the half-site spacing preferences of bZIP proteins are the result of (i) the differential intrinsic curvature in target binding sites that differ by insertion or deletion of a single base pair and (ii) the ability of some bZIP proteins to overcome this intrinsic curvature through a mechanism dependent on basic segment residues.

The bZIP family of eukaryotic transcription factors contains (i) a short, helical basic segment whose residues participate in DNA contacts, (ii) a zipper segment responsible for protein dimerization (1, 2), and (iii) a six-residue spacer of variable sequence connecting the two (3). X-ray crystallography data from two different bZIP-DNA complexes show that the dimeric bZIP protein contains a pair of uninterrupted α helices that interact with each other along the length of the zipper segment to form a parallel coiled coil (4). These α helices diverge in the vicinity of the nucleic acid and interact with the major groove of the target DNA (5, 6).

Like certain other dimeric DNA binding proteins (7, 8), bZIP proteins discriminate differently between target binding sequences that differ only in half-site spacing. For example, proteins related to Fos and Jun (AP-1 family) bind preferentially to a pseudosymmetric 9-base pair (bp) site, which consists of two ATGA half-sites arranged in an inverted pair separated by a single dC:dG base pair—that is, ATGACTCAT. Proteins related to cyclic adenosine monophosphate response element-binding protein (CREB) and to activating transcription factor-2 (ATF-2) (CREB-ATF family) have higher affinity for

the dyad-symmetric CRE site in which the same inverted pair of half-sites is separated by 2 bp (ATGACGTCAT) (9). In contrast to the AP-1 and CREB-ATF families, the yeast bZIP protein GCN4 binds to both sites with comparable affinity (10). Within the context of B DNA, the additional dG:dC base pair in the CRE site displaces the two ATGA contact surfaces by an axial translation of 3.4 Å and a twist angle of 34° (5). The ability of GCN4 to accept both sequences as specific targets could be the result of the flexibility of an α -helical segment, which permits a structural readjustment that its counterpart in CREB and ATF proteins does not permit (6, 11). Alternatively, GCN4 could bind with the same structure but there could be an induced structural readjustment of one or both of the DNA targets (5). A third possibility is that GCN4 has an equal affinity for the two DNA target sites as a result of an intrinsic deformation of the B-form structure of one target site which compensates for the difference in half-site spacing.

Initially, we used a circular permutation assay (12) to compare the conformation of the CRE and AP-1 target DNA sequences both alone and when bound to the bZIP segments of GCN4 and CRE-BP1, a member of the CREB-ATF family; we have labeled these bZIP peptides ggg and ccc, respectively (Fig. 1). The circular permutation assay is based on the position-dependent effects of DNA distortion on the electrophoretic mobilities of a set of isomeric DNA fragments (12). Oligonucleotides with conformational distortions near

D. N. Paoletta and A. Schepartz, Department of Chemistry, Training Program in Biophysics, Yale University, New Haven, CT 06511, USA.
C. R. Palmer, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA.

*To whom correspondence should be addressed.