km on Lake Tanganyika (Fig. 2C). Single airguns or multigun arrays in the size range of 40 to 140 cubic inches (0.65 to 2.3 liters) were used for the seismic source. Incoming seismic signals were received on a 960-m long, 48-channel cable and then recorded on a Texas Instruments DFS V acquisition system aboard the Project's research vessel, *Nyanja*. TRANSAT satellites coupled to ship's velocity logs and supplemented by radar provided most navigational control. Global positioning system satellite navigation was used during about 20% of the Lake Malawi data acquisition. All Lake Malawi data have been stacked and a few lines have been migrated to date. All Lake Tanganyika dip lines have been migrated.
2. Lake Malawi is about 570 by 75 km and has a maximum depth of about 700 m (Fig. 2B). The lake is permanently anoxic below depths of about 200 m (*3*). Lake Tanganyika is about 650 by 70 km and has a maximum depth of about 1500 m (Fig. 2D). It is permanently anoxic below depths of about 150 m (*4*). Both lakes currently have outlets, although both have been closed in historical times (*5*).
3. D. H. Eccles, *Limnol. Oceanogr.* **19**, 730 (1974).
4. E. T. Degens, R. P. Von Herzen, H. K. Wong, *Naturwissenschaften* **58**, 229 (1971).
5. L. C. Beadle, *The Inland Waters of Tropical Africa* (Longman, London, ed. 2, 1981).
6. B. R. Rosendahl, *Annu. Rev. Earth Planet. Sci.* **15**, 445 (1987); C. J. Ebinger, B. R. Rosendahl, D. J. Reynolds, *Tectonophysics* **141**, 215 (1987).
7. T. Specht and B. R. Rosendahl, *J. Afr. Earth Sci.*, in press; J. Flannery, thesis, Duke University (1988). Some of the section imaged by these reflection seismic data may be of Cretaceous or Permo-Triassic (Karroo) age, although none of this section has been drilled. Preliminary velocity analyses show low values (<3500 m/s) for most of the stratigraphic section, which suggests that it is probably Cenozoic in age (C. A. Scholz, unpublished data).
8. C. F. Burgess *et al.*, in *Rifting and the Opening of the Atlantic Ocean*, W. Manspeizer, Ed. (Elsevier, Amsterdam, in press).
9. B. R. Rosendahl, J. W. Versfelt, C. A. Scholz, L. D. Woods, Eds., *Seismic Atlas of Lake Tanganyika* (Project PROBE, Duke University, Durham, NC, 1988). The Lake Tanganyika rift sediments also have never been drilled. The thickest section of sediment in this rift shows no major depositional hiatus in the deepest parts of the basin. We interpret these rift sediments to be Cenozoic in age.
10. Sequence boundaries are the unconformities and their correlative conformities that define depositional sequences. A review of the concepts of sequence stratigraphy is presented in C. E. Payton, Ed., *Am. Assoc. Pet. Geol. Mem. 26* (1977).
11. T. J. Dunkelman, J. A. Karson, B. R. Rosendahl, *Geology* **16**, 258 (1988); T. J. Dunkelman, thesis, Duke University (1986); C. J. Ebinger *et al.*, *Nature* **308**, 627 (1983).
12. T. C. Johnson, J. D. Halfman, B. R. Rosendahl, G. S. Lister, *Geol. Soc. Am. Bull.* **98**, 439 (1987).
13. A. S. Cohen *et al.*, in *Sedimentation in the African Rifts*, L. E. Frostick, R. W. Renault, I. Reid, J. J. Tiercelin, Eds. (Spec. Publ. 25, Geological Society of London, London, 1986), pp. 127–139.
14. K. W. Butzer, G. L. Isaac, J. L. Richardson, C. Washbourn-Kamau, *Nature* **175**, 1069 (1972).
15. R. L. Kendall, *Ecol. Monogr.* **39**, 121 (1969).
16. R. E. Hecky and E. T. Degens, *Woods Hole Oceanogr. Inst. Rep. 73-28* (1973).
17. M. E. Meadows, *Paleoecol. Afr.* **16**, 353 (1984).
18. B. R. Rosendahl and D. A. Livingstone, *Episodes: J. Int. Union Geol. Sci.* **1983**, 14 (1983).
19. T. C. Johnson, unpublished data; K. A. Haberyan, thesis, Duke University (1988).
20. C. H. Pilskaln and T. C. Johnson, in preparation. The laminated diatomaceous muds and silts consist of couplets of white diatom-rich layers 1 mm thick and dark layers consisting of diatoms, mineral grains, and organic debris. These components are interpreted to represent seasonal flux events.
21. E. M. Van Zinderen Bakker and J. H. Mercer, *Paleogeogr. Paleoclimatol. Paleoecol.* **56**, 217 (1986).
22. P. E. Olsen, *Science* **234**, 842 (1986).
23. D. Harkin, *Geol. Surv. Tanganyika Mem. II* (1960).
24. K. A. Haberyan and R. E. Hecky, *Paleogeogr. Paleoclimatol, Paleoecol.* **61**, 169 (1987). Carbon-dated sediments from a core from southern Lake Tanganyika yielded sedimentation rates on the order of 1 mm/year. Sediment geochemistry and fossil diatom analysis indicate Lake Tanganyika was closed before about 10,000 years ago.
25. B. R. Rosendahl *et al.*, in *Sedimentation in the African Rifts*, L. E. Frostick, R. W. Renault, I. Reid, J. J. Tiercelin, Eds. (Spec. Publ. 25, Geological Society of London, London, 1986), pp. 29–43. Accommodation zones are defined as those regions where half-grabens of opposing polarity link along the strike of the rift.
26. D. A. Livingstone, *Limnol. Oceanogr.* **10**, 607 (1965); R. E. Hecky, *Pol. Arch. Hydrobiol.* **25**, 159 (1978).
27. D. A. Livingstone, *Annu. Rev. Ecol. Syst.* **6**, 249 (1975).
28. F. A. Street and A. T. Grove, *Nature* **261**, 385 (1976); A. T. Grove, *ibid.* **234**, 403 (1971).
29. G. Fryer and T. D. Iles, *The Cichlid Fishes of the Great Lakes of Africa, Their Biology and Evolution* (Oliver & Boyd, Edinburgh, 1972).
30. R. H. Lowe-McConnell, *Ecological Studies in Tropical Fish Communities* (Cambridge Univ. Press, Cambridge, 1987).
31. J. E. Brooks, *Q. Rev. Biol.* **25**, 131 (1950); E. B. Worthington, *Nature* **173**, 1064 (1954).
32. We thank the governments of Burundi, Malawi, Mozambique, Tanzania, and Zaire for permission to conduct this research. Numerous individuals at Project PROBE helped in the acquisition and processing of the data. We thank J. McGill, captain of the R.V. *Nyanja*, for his efforts during data acquisition and L. D. Woods for drafting the figures. D. A. Livingstone and T. C. Johnson reviewed earlier versions of the manuscript. We thank T. C. Johnson for permission to examine his unpublished data. This work has been supported by Agip, Amoco, Arco, Conoco, Eason, Exxon-Esso, Marathon, Mobil, Pecten, Pennzoil, Petrofina, Placid, Shell International Petroleum, Texaco, and the World Bank.

13 November 1987; accepted 22 April 1988

Amino Acid Preferences for Specific Locations at the Ends of α Helices

JANE S. RICHARDSON AND DAVID C. RICHARDSON

A definition based on α-carbon positions and a sample of 215 α helices from 45 different globular protein structures were used to tabulate amino acid preferences for 16 individual positions relative to the helix ends. The interface residue, which is half in and half out of the helix, is called the N-cap or C-cap, whichever is appropriate. The results confirm earlier observations, such as asymmetrical charge distributions in the first and last helical turn, but several new, sharp preferences are found as well. The most striking of these are a 3.5:1 preference for Asn at the N-cap position, and a preference of 2.6:1 for Pro at N-cap + 1. The C-cap position is overwhelmingly dominated by Gly, which ends 34 percent of the helices. Hydrophobic residues peak at positions N-cap + 4 and C-cap − 4.

THE EMPIRICAL PREFERENCES OF each amino acid for the different types of secondary structure have formed a central theme of the efforts to predict three-dimensional protein structure from amino acid sequence. In almost all cases (*1–4*), those amino acid preferences have been smoothed over windows of 3 to 20 residues along the sequence, which gives statistically solid but relatively weak preference values. The only secondary structures that have been analyzed for unaveraged, single-position preferences are specific types of turns or connections; those have shown preferences as high as the 3:1 value for proline in tight turn position two (*5*), and sometimes even higher for glycine in very specifically defined connection types (*6*). Position-specific preferences have not been compiled for helices or β strands, presumably because the database was not large enough, and also because of the difficulty of settling on an unambiguous definition for placing segment ends. The database of well-determined three-dimensional protein structures is now large enough to provide statistically meaningful single-position helix preferences for all but the rarer amino acids. We use a definition of helix ends that optimizes correlation with the amino acid sequence, in order to compile those preferences here.

The sample of 215 α helices was taken from 42 proteins in the Brookhaven Data Bank (*7*), plus three revised coordinate sets: bacteriochlorophyll protein (*8*), cro repressor (*8*), and cytochrome b562 (*9*). From families of homologous proteins, only examples with low sequence similarity to each other were included. We examined the structures on an Evans and Sutherland PS 330 with a Tektronix stereo window, using the display program CHAOS (*10*), which runs entirely inside the PS 330 processor. The utility programs that calculate vectors, ribbons, H bonds, and dot surfaces for CHAOS run on a MassComp workstation under UNIX. The backbone dihedral angles ϕ and ψ were calculated by means of the DIHDRL program (*11*), modified for the UNIX MassComp. The superpositions in Fig. 2 were done with the INSIGHT pro-

Department of Biochemistry, Duke University, Durham, NC 27710.

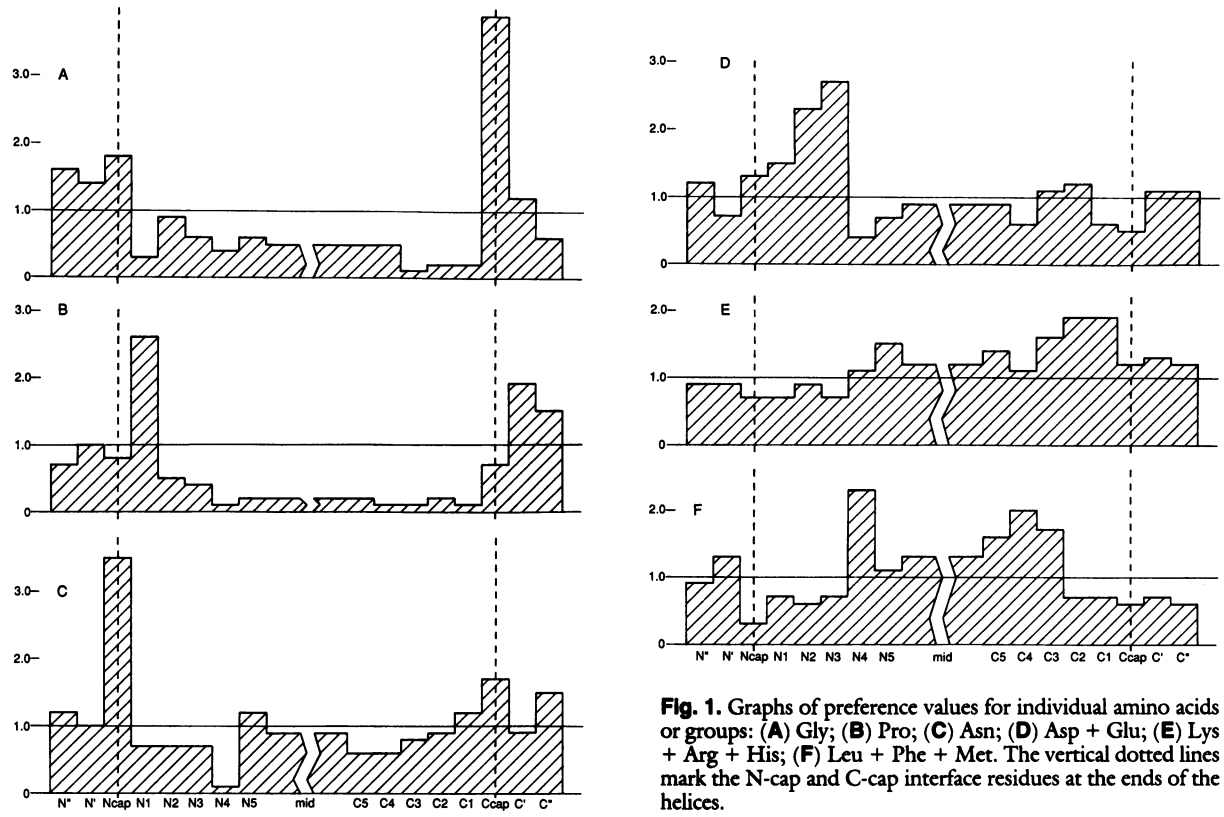

Fig. 1. Graphs of preference values for individual amino acids or groups: (**A**) Gly; (**B**) Pro; (**C**) Asn; (**D**) Asp + Glu; (**E**) Lys + Arg + His; (**F**) Leu + Phe + Met. The vertical dotted lines mark the N-cap and C-cap interface residues at the ends of the helices.

gram from Biosym.

Helix boundaries can be defined in terms of ϕ and ψ angles (*12*), hydrogen bonding (*12, 13*), or α-carbon positions (*14*); a tolerance margin must be chosen and it is necessary to decide whether an interface residue with only one correct H bond or one correct conformational angle is counted as inside or outside the helix. Fortunately, for compiling position-specific preferences the last decision is irrelevant, since the interface residue (here called the N-cap or C-cap) is treated separately. A subset of the sample was examined to test whether H bonding [based on the definitions from (*9*)], ϕ and ψ angles (values from DIHDRL), or α-carbon positions (with the use of visual examination on computer graphics, as described below) would give the strongest and most position-specific amino acid preferences. The α-carbon definition performed best, for two separate reasons. One is that it accentuated known residue preferences such as Gly at the C-cap (*15*) and Ser at the N-cap (*16*) and produced new preferences in addition, more sharply defined than for either of the other two definitions. The second reason is that there are rather frequent cases where a single peptide in a helix is rotated far enough to destroy reasonable ϕ and ψ values or H bonding but leaves the α carbons unchanged. These local distortions show amino acid preferences characteristic of helix interior rather than ends.

Therefore, we adopted an α-carbon definition of helix ends. Specifically, the N-cap (or C-cap) is the first (or last) residue whose α carbon lies approximately in the cylinder formed by the helix backbone and approximately along the helical spiral path. One N-cap and three C-caps were ambiguous by this definition and were not included in the database. No helices were included that were shorter than eight residues including caps. The 215 α helices are listed in (*17*). Preferences were normalized by the number of occurrences expected, on the basis of the overall percentage of each amino acid found in a sample of 135 proteins of known structure (*18*). Preferences are tabulated for the caps at each end, two positions out from the ends, and five positions in from the end; remaining helical residues are grouped together as "middle."

The results are presented in the form of amino acid preferences for specific positions along the α helices (Table 1); the values either up or down from random expectation by more than 3 standard deviations are indicated. These strong preferences include some that have been previously observed, such as Gly at the C-caps (*15*), here seen for fully one-third of the helices (Fig. 1A), lack of Pro in helix interiors (Fig. 1B), and the asymmetrical preference for negatively charged side chains in the first turn and positively charged ones in the last turn (Fig. 1D) (*19, 20*). However, there are also some strong and highly position-specific preferences that have not been seen before. The most notable of these are for Asn in the N-cap position and for Pro at position N1 (the first residue after the N-cap).

The preference distribution for Pro (in Fig. 1B) indicates its near absence from helix interiors, where it breaks at least two adjacent H bonds and where its ring pushes the preceding turn of backbone away by about an angstrom. Although Pro has the lowest helix preference of any residue, it actually occurs in the interiors of 8 percent of the helices in this sample (*21*). Pro is fairly common, but not spectacularly so, at the cap positions and beyond the ends, and its average preference value for the three positions in the first turn is about 1.0, as has been found before (*1*). However, the detailed tabulation shows that Pro residues in the first turn are almost exclusively in the N1 position, which has a normalized preference value of 2.6 as opposed to 0.4 to 0.5 for N2 and N3. In this one location, at least, it seems that Pro is perhaps better described as a helix-initiator rather than a helix-breaker.

Pro in the first turn of a helix fits well in terms of its own backbone conformation, and it should have some stabilizing influence because with only one rotatable angle it loses less entropy than other amino acids in forming a regular structure. In addition, Pro in the first turn should help block continuation of the helix in the NH_2-terminal direc-

tion, since its ring produces steric interference with the necessary backbone positioning and it also prevents H bonding both from its own nitrogen (which is covalently bonded in the ring) and, indirectly, from at least one neighboring NH. Both of these effects should apply in positions N1, N2, or N3, but might be expected to show most strongly in N3, which is immediately under the next potential residue that could add to the helix. Overall, these considerations do not seem to account at all adequately for the 5:1 preference of Pro for N1 over N2 or N3 (Table 1 or Fig. 1B). The work of Presta and Rose (*22*) now suggests a possible explanation. In our sample of native-protein helices, side-chain H bonds are frequently seen to the backbone NH of N2 and especially N3, but almost never to N1. If a principal contribution of Pro is simply to tie up one of those otherwise exposed NH groups, then it should be most valuable in position N1.

Fig. 2. Stereo image of 16 superimposed examples of Asn N-caps, all of which make side-chain H bonds with the backbone NH of helical residue N3 or N2. Side chains are shown only for the N-cap Asn residues and for Pro residues in position N1.

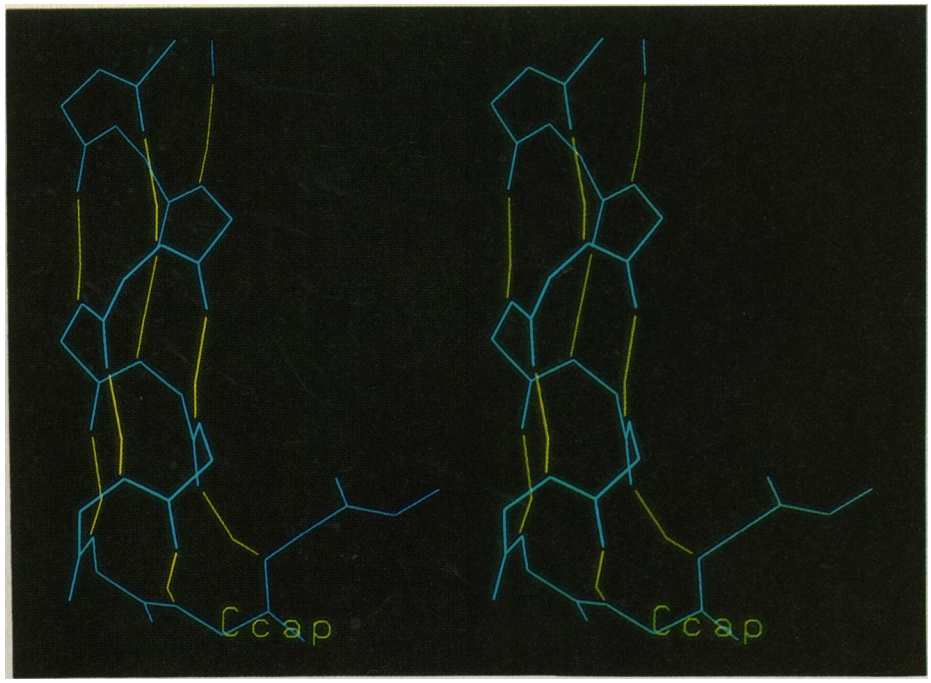

Fig. 3. Stereo picture of a Gly C-cap (1TIM 120), showing how the peptide NH groups on either side of the Gly form H bonds to the C4 and C3 carbonyl groups in reverse order (*15*), while making further propagation of the helix impossible. This type of C-cap conformation is $L3_{10}$ (near $\phi = 70°$, $\psi = 20°$), so that most cases are Gly; in this helix sample there are 60 Gly, 12 Lys, 7 Asn, and 12 other C-cap residues with positive ϕ angles. The H bonds are shown bent at the calculated H position.

The asparagine preferences are shown in Fig. 1C, where the height and sharpness of the peak at the N-cap position can be appreciated. Of the 34 Asn N-caps, 23 form H bonds from Oδ to the backbone NH of residue N3 (or sometimes N2) exposed in the first turn of the helix. This type of H bonding has often been noted for Ser, which is indeed just as common as Asn but has a significantly lower preference value since Asn is a rarer amino acid overall. Asp and Thr can also fill this role, but their preferences are a bit lower and less position-specific. In contrast, Glu and Gln do not have the right geometry for this interaction and are quite rare as N-caps. In helices and helix ends, Asn and Gln are in fact very nonconservative replacements for one another in spite of their apparent similarity. Eighty-five of the 214 helix NH_2-termini have H bonds from the N-cap side chain to N3 or N2 as follows: 26 from Ser, 23 Asn, 17 Thr, 16 Asp, 1 Glu, and 1 His. A set of 16 Asn N-caps are superimposed in Fig. 2, showing that their conformations vary only slightly. All are within the broad β region, but one cluster has ψ near 100° and can H bond to either N3 or N2, while the other cluster has ψ near 170° and H bonds to N3. The Ser residues that form this H bond have more variable conformations than the Asn residues, but most are within the polyproline region (approximately β, but with $\phi < -90°$), which makes the carbonyl groups surrounding them point in perpendicular directions.

For those amino acid preferences noted here, the dominant theme is the importance of possible hydrogen bonding patterns in determining the ends of α helices. The sharpest preferences are at the C-cap, where a Gly in $L\alpha$ or $L3_{10}$ conformation (*15*) can satisfy two successive backbone CO groups while turning the chain in a direction that prevents continuation of the helix (Fig. 3), and at the N-cap, where the side chains of Asn, Ser, Asp, or Thr can H bond to the open backbone NH's. Asn is especially suitable in this position because not only does its side chain mimic a peptide, but the entire residue actually mimics a dipeptide: following backward along the chain from the CO of the succeeding peptide, through the Asn Cα, and out the side chain to Oδ (which plays the role of the second peptide CO).

Asn would help to specify the location of the helix NH_2-terminus, because it can simultaneously stabilize the first helical turn by providing an additional interaction equal to that of a residue, and also discourage further helix propagation in the NH_2-terminal direction by competing with the backbone to provide that interaction. In an odd sense, Asn is a less "handed" amino acid residue than any except Gly, because two of the substituents at its α carbon are approximately equivalent: the preceding peptide out to the CO, and the side chain out to O δ. Good N-cap residues are relatively rare in the interior of α helices, perhaps because they compete with the helical H bonds. When Ser does occur in α helix, its OH often forms a second H bond to a backbone CO on the previous helical turn; this presumably destabilizing arrangement often involves a slightly lengthened distance between the two turns.

Moving in from the helix ends, one sees a preference for negative charges in the first turn and positive charges in the last turn (Fig. 1, D and E) confirming what has been noted previously (*19*). Those distributions are fairly uniform over the first or the last helical turn except for the relative rarity of Asp in position N1, which may be because it makes itself the N-cap when it is that close. As N-cap, it can both interact favorably with the helix dipole (*20*) and H bond to one of the exposed backbone NH groups.

Still further in from the helix ends, there is a peak of preference for hydrophobic amino acids in position N4 and in C3 or C4 (Fig. 1F); these peaks are quite strong for Leu, Phe, and Met and are somewhat less pronounced for Val and Ile. Positions N4 and C4 are underneath the polypeptide chain leaving the helix, and also usually on its interior face, since the chain at each end must connect to the rest of the protein. This hydrophobic preference pattern shows once again the classic inside versus outside distribution on helices (*23*, *24*). A new feature is the demonstration that this periodicity has a well-defined relation to the N-cap and C-cap positions; this has been noted before for specific cases (*25*, *26*) but not as a general

Table 1. Position-specific amino acid preferences in helices. The upper entry is the observed number of occurrences out of a total number listed at the bottom. The lower entry is the relative preference value: the ratio of observed occurrences to the expected number based on average percentage composition [column 1; adapted from reference (*18*)]. Boldface values are higher, and underscored values lower, than expectation by more than 3 standard deviations. Out of 215 helices, one NH_2-terminus and three COOH-termini were omitted because their location could not be specified unambiguously.

%		N″	N′	N-cap	N1	N2	N3	N4	N5	Mid	C5	C4	C3	C2	C1	C-cap	C″	C′
4.5	Pro	6	9	8	**25**	5	4	1	2	10	2	0	0	2	0	7	17	13
		0.7	1.0	0.8	**2.6**	0.5	0.4	0.1	0.3	0.3	0.3	0	0	0.2	0	0.7	1.9	1.5
8.7	Gly	29	25	**33**	6	16	12	6	8	29	6	8	2	3	4	**72**	21	10
		1.6	1.4	**1.8**	0.3	0.9	0.6	0.4	0.6	0.5	0.5	0.5	0.1	0.2	0.2	**3.9**	1.2	0.6
6.9	Ser	22	17	**34**	11	12	6	6	12	27	5	6	8	23	25	11	9	12
		1.6	1.2	**2.3**	0.7	0.8	0.4	0.4	1.1	0.6	0.5	0.5	0.6	1.6	1.7	0.8	0.7	0.9
6.1	Thr	4	10	21	11	9	13	6	6	43	4	6	9	12	13	4	10	**25**
		0.3	0.8	1.6	0.8	0.7	1.0	0.5	0.6	1.0	0.5	0.6	0.7	0.9	1.0	0.3	0.8	**2.1**
4.5	Asn	11	9	**34**	7	7	7	0	9	28	4	5	8	9	11	15	8	13
		1.2	1.0	**3.5**	0.7	0.7	0.7	0	1.2	0.9	0.6	0.6	0.8	.9	1.2	1.6	0.9	1.5
3.6	Gln	7	7	3	5	6	12	9	8	30	5	1	10	12	**16**	7	10	8
		1.0	0.9	0.4	0.7	0.8	1.5	1.3	1.4	1.3	1.0	0.2	1.3	1.6	**2.1**	0.9	1.4	1.1
6.0	Asp	17	10	**27**	10	**33**	**28**	4	6	38	6	5	7	9	5	9	17	17
		1.4	0.8	**2.1**	0.8	**2.6**	**2.2**	0.3	0.6	1.0	0.7	0.5	0.6	0.7	0.4	0.7	1.4	1.4
5.6	Glu	11	8	5	**27**	**24**	**40**	6	8	30	9	7	19	20	10	4	9	8
		1.0	0.7	0.4	**2.2**	**2.0**	**3.3**	0.5	0.9	0.8	1.1	0.7	1.6	1.7	0.8	0.3	0.8	0.7
6.7	Lys	14	16	10	9	14	12	11	**21**	49	13	12	**31**	**27**	**28**	18	16	17
		1.0	1.1	0.7	0.6	1.0	0.8	0.8	**1.9**	1.1	1.4	1.0	**2.2**	**1.9**	**2.0**	1.3	1.2	1.3
3.8	Arg	3	4	3	6	7	3	11	7	33	5	5	6	**17**	8	7	11	8
		0.4	0.5	0.4	0.7	0.9	0.4	1.5	1.2	1.3	1.0	0.8	0.8	**2.1**	1.0	0.9	1.4	1.1
2.1	His	5	4	5	3	3	3	6	3	14	7	7	5	8	**15**	6	5	4
		1.2	0.9	1.1	0.7	0.7	0.7	1.5	0.9	1.0	2.4	1.9	1.1	1.8	**3.4**	1.3	1.2	1.0
9.0	Ala	13	16	10	24	31	19	20	20	**106**	**23**	21	13	27	21	15	18	12
		0.7	0.8	0.5	1.2	1.6	1.0	1.1	1.4	**1.8**	**1.8**	1.3	0.7	1.4	1.1	0.8	1.0	0.7
7.7	Leu	15	20	3	15	5	10	**39**	13	63	13	19	**31**	13	11	11	14	8
		0.9	1.2	0.2	0.9	0.3	0.6	**2.6**	1.1	1.2	1.2	1.4	**1.9**	0.8	0.7	0.7	0.9	0.5
7.1	Val	10	15	1	17	9	16	21	9	56	4	17	19	10	10	3	9	14
		0.7	1.0	0.1	1.1	0.6	1.1	1.5	0.8	1.2	0.4	1.4	1.3	0.7	0.7	0.2	0.6	1.0
5.1	Ile	9	14	2	10	8	4	11	7	41	9	14	15	4	8	8	11	7
		0.9	1.3	0.2	0.9	0.7	0.4	1.1	0.9	1.2	1.3	1.6	1.4	0.4	0.7	0.7	1.1	0.7
3.8	Phe	9	9	2	4	7	5	14	6	34	10	**19**	14	2	6	4	1	9
		1.2	1.1	0.2	0.5	0.9	0.6	1.9	1.0	1.3	1.9	**2.9**	1.8	0.3	0.7	0.5	0.1	1.2
3.4	Tyr	13	3	6	13	3	9	4	1	19	6	5	8	2	9	6	6	3
		1.9	0.4	0.8	1.8	0.4	1.2	0.6	0.2	0.8	1.3	0.8	1.1	0.3	1.2	0.8	0.9	0.5
1.8	Met	1	7	3	1	4	4	6	5	18	7	**9**	4	5	4	3	3	0
		0.3	1.8	0.8	0.3	1.0	1.0	1.7	1.7	1.5	2.7	**2.8**	1.0	1.3	1.0	0.8	0.8	0
1.3	Trp	3	2	1	6	5	4	**8**	3	14	2	5	1	1	0	0	1	7
		1.1	0.7	0.3	2.1	1.7	1.4	**3.1**	1.4	1.5	1.1	2.1	0.4	0.4	0	0	0.4	2.7
2.4	Cys	3	4	3	4	6	3	5	6	11	0	3	1	6	8	2	4	2
		0.6	0.8	0.6	0.8	1.2	0.6	1.1	1.6	0.7	0	0.7	0.2	1.2	1.6	0.4	0.8	0.4
Total		205	209	214	214	214	214	194	160	650	140	174	211	212	212	212	200	197
Chain-termini		9	5														12	15

regularity.

Some amino acid residues, notably Ala, show relatively smooth and favorable distributions along the helix but are rarer at the caps. Ala has visible but slight preferences for outside over inside and middle over ends. Several of the amino acid residues that form side-chain H bonds (Ser, Thr, and Gln) show a smoothly increasing occurrence for C3, C2, and C1 but drop abruptly at the C-cap.

Preferences are relatively weak for the positions outside the caps, presumably because those residues have a variety of other roles and constraints. The charge asymmetry is essentially gone by then (Fig. 1, D and E). There is a significant but weak preference for H bonding residues (up by 2 sigma) over hydrophobic residues (down by 2.5 sigma) in positions N'', C', and C'', while position N' somewhat prefers the hydrophobics. The only individual preference in this region that is up by more than 3 standard deviations is for Thr in C''. Most of those C'' Thr's are on the solvent side of a β strand or at the beginning of the next α helix, but six of them show a specific interaction where they are in position 4 of a tight turn and both Oγ and NH of the Thr are hydrogen-bonded to the carbonyl O of residue C1.

It is interesting that a simple definition of helix ends in terms of Cα position gives stronger and more position-specific amino acid preferences than more precise definitions based on hydrogen bonding or ϕ, ψ angles. Some of this effect may come from local errors in the less well-determined crystal structures, but in looking at the structures it is hard to avoid the conviction that many of the local irregularities represent late folding adjustments of the protein to its detailed environment in the final tertiary structure. Interactions important for secondary-structure formation early in the folding process may persist only approximately in the native structure, as is also strongly suggested by the work of Presta and Rose (*22*). Presumably the definition that correlates most strongly with amino acid sequence will be best for use in structure predictions.

It remains to be seen whether these individual position amino acid preferences can be integrated into an improved helix prediction algorithm, and whether substitution of strongly preferred residues in appropriate positions can help to stabilize isolated helices or entire proteins.

REFERENCES AND NOTES

1. P. Y. Chou and G. Fasman, *Biochemistry* **13**, 211 (1974).
2. J. Garnier, D. J. Osguthorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978).
3. J. Kyte and R. Doolittle, *ibid.* **157**, 105 (1982).
4. G. Rose and J. Seltzer, *ibid.* **113**, 153 (1977).
5. P. Y. Chou and G. Fasman, *ibid.* **115**, 135 (1977).
6. B. L. Sibanda and J. M. Thornton, *Nature* **316**, 170 (1985).
7. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
8. From B. W. Matthews.
9. From F. S. Mathews.
10. D. C. Richardson and M. E. Zalis developed this program.
11. Enrique Abola; available on Brookhaven Data Bank distribution tapes.
12. IUPAC-IUB Commission on Biochemical Nomenclature, *J. Biol. Chem.* **245**, 6489 (1970).
13. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
14. M. Levitt and J. Greer, *J. Mol. Biol.* **114**, 181 (1977).
15. C. Schellman, in *Protein Folding*, R. Jaenicke, Ed. (Elsevier/North-Holland, New York, 1980), p. 53.
16. J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. E. Dickerson, *Nature* **190**, 666 (1961).
17. List of helices (The abbreviations are those used in the Brookhaven Data Bank.): 1HMQ: 18–38, 40–66, 69–86, 90–110; 1ECA: 2–18, 19–31, 52-73, 76–91, 93–112, 117–136; 2ACT: 24–43, 49–58, 69–81, 99–106, 120–131; 2CYP: 15–33, 42–55, 84–99, 103–120, 150–162, 164–177, 200–209, 232–241, 241–254, 254–272; 3TLN: 64–89, 136–154, 159–181, 233–247, 259–275, 280–297, 300–314; 2MBN: 3–20, 20–37, 36–43, 51–58, 58–79, 82–98, 100–116, 124–150; 1CPV: 7–19, 25–34, 39–51, 59–?, 78–89, 98–108; 1CCY: 4–32, 39–63, 78–103, 103–?; 1LZM: 2–12, 38–51, 59–81, 81–91, 92–107, 107–114, 114–124, 125–136, 136–143, 142–156; 1MLT: 1–26; b562: 2–20, 22–43, 59–82, 83–106; 7LYZ: 4–16, 24–37, 88–102, 108–116; 1LHB: 12–30, 29–46, 45–52, 60–67, 67–88, 91–106, 111–127; 3CAT: 9–19, 53–67, 157–166, 177–189, 188–?, 258–271, 323–331, 347–366, 437–450, 451–469, 470–485, 485–500; 4DFR: 24–36, 43–51, 77–86, 96–107; 5CPA: 14–30, 72–90, 93–103, 112–122, 173–187, 215–234, 253–262, 285–307; 4FXN: 10–27, 39–46, 62–75, 93–107, 124–137; 2CAB: 154–167, 219–229; 2GRS: 29–43, 62–80, 95–122, 196–210, 227–241, 338–356, 383–391, 439–454, 456–463, 469–477; 2ADK: 1–8, 20–33, 38–50, 51–64, 68–84, ?–109, 121–137, 142–168, 178–194; 4LDH: 30–44, 55–70, 108–131, 141–154, 165–181, 224–244, 247–264, 308–326; 4ADH: 46–55, 170–188, 201–215, 225–236, 249–259, 271–282, 304–311, 323–338, 354–365; 1SBT: 5–12, 12–20, 63–73, 103–118, 132–146, 219–239, 242–254, 269–275; 1RHD: 11–22, 42–50, 76–88, 106–119, 129–137, 163–174, 224–236, 251–264, 274–282; 2AZA: 52–67; CRO: 6–15, 15–24, 26–37; 2STV: 12–24, 116–123; 2SNS: 54–69, 98–107, 121–136; BCL: 121–128, 150–164, 169–180, 181–191, 283–294, 334–347; 2SSI: 45–57, 99–107; 5RSA: 3–12, 24–34, 50–60; 2PAB: 74–83; 2SGA: 55–64, 230–238; 2APP: 139–148, 222–233; 2PTN: 164–172, 234–245; PSTI: 33–44; 5PTI: 47–56; LOVO: 33–45; 1SN3: 22–31; 3BP2: 1–13, 39–57, 89–109; 1INS: a1–a9, a12–a20, b8–b20; 2CDV: 64–71, 78–88, 90–99; 3CYT: 2–18, 49–57, 60–70, 70–76, 87–103; 1PPT: 13–32; 2B5C: 8–16, 42–50, 54–62, 64–75, 80–87.
18. H. Nakashima, K. Nishikawa, T. Ooi, *J. Biochem. Tokyo* **99**, 153 (1986).
19. O. B. Ptitsyn, *J. Mol. Biol.* **42**, 501 (1969).
20. K. R. Shoemaker, P. S. Kim, E. J. York, R. L. Baldwin, *Nature* **326**, 563 (1987).
21. J. Richardson and D. Richardson, in *Prediction of Protein Structure and Principles of Protein Conformation*, G. Fasman, Ed. (Plenum, New York, in press).
22. L. Presta and G. Rose, *Science* **240**, 1591 (1988).
23. W. Schiffer and A. Edmundson, *Biophys. J.* **7**, 121 (1967).
24. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1982).
25. M. S. Edwards, M. J. E. Sternberg, J. M. Thornton, *Prot. Eng.* **1**, 173 (1987).
26. A. V. Efimov, *FEBS Lett.* **166**, 33 (1984).

27. Supported by NIH grant GM-15000.

29 February 1988; accepted 28 April 1988

Effect of Forskolin on Voltage-Gated K^+ Channels Is Independent of Adenylate Cyclase Activation

TOSHINORI HOSHI, SARAH S. GARBER, RICHARD W. ALDRICH

Forskolin is commonly used to stimulate adenylate cyclase in the study of modulation of ion channels and other proteins by adenosine 3',5'-monophosphate (cAMP)–dependent second messenger systems. In addition to its action on adenylate cyclase, forskolin directly alters the gating of a single class of voltage-dependent potassium channels from a clonal pheochromocytoma (PC12) cell line. This alteration occurred in isolated cell-free patches independent of soluble cytoplasmic enzymes. The effect of forskolin was distinct from those of other agents that raise intracellular cAMP levels. The 1,9-dideoxy derivative of forskolin, which is unable to activate the cyclase, was also effective in altering the potassium channel activity. This direct action of forskolin can lead to misinterpretation of results in experiments in which forskolin is assumed to selectively activate adenylate cyclase.

NEURONAL EXCITABILITY CAN BE modulated by phosphorylation of ion channel proteins by protein kinases that are activated by specific second messengers, such as cAMP (*1*). Because of its high affinity for adenylate cyclase, forskolin (FSK) has been used to increase intracellular cAMP levels, leading to the presumed phosphorylation of ion channels by cAMP-dependent protein kinases (*1–9*). Recent reports, however, have suggested that some of the actions of FSK on ion channels and the glucose transporter may be inconsistent with a selective activation of adenylate cyclase and subsequent phosphorylation (*9–*

Department of Neurobiology, Stanford University School of Medicine, Stanford, CA 94305.