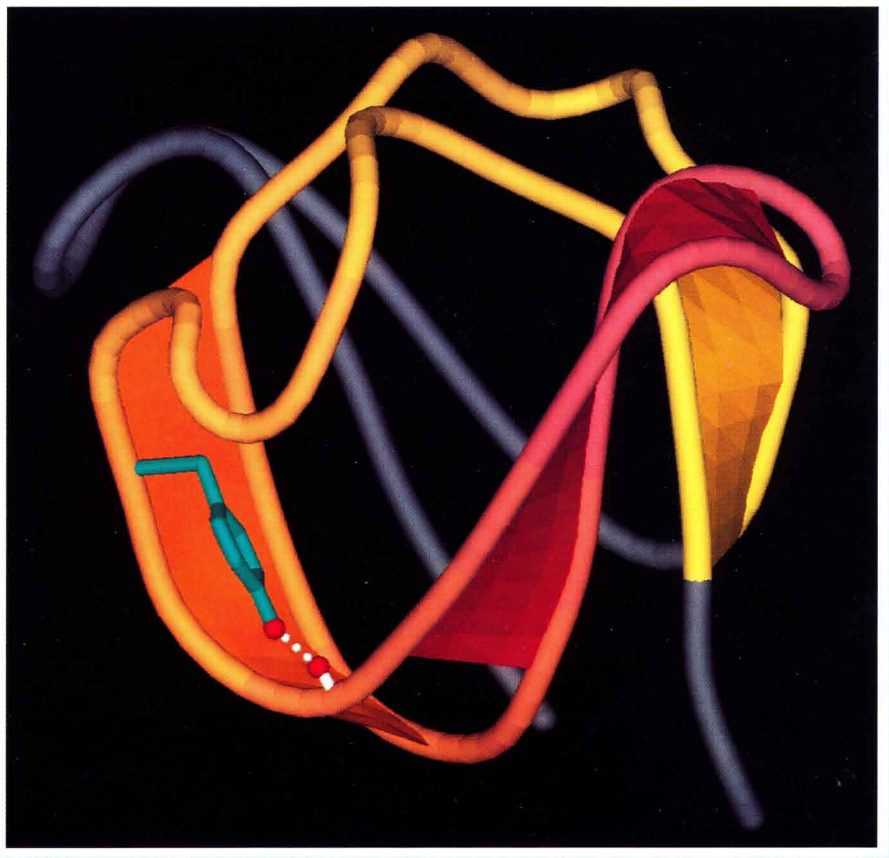
Protein Science

A PUBLICATION OF THE PROTEIN SOCIETY

SUSTAINED IN PART WITH THE SUPPORT OF THE AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY

VOL. 3, NO. 11

NOVEMBER 1994



CAMBRIDGE
UNIVERSITY PRESS

Protein Science (1994), *3*:1927–1937. Cambridge University Press. Printed in the USA.

The tyrosine corner: A feature of most Greek key β-barrel proteins

JENS M. HEMMINGSEN, KIM M. GERNERT, JANE S. RICHARDSON, AND DAVID C. RICHARDSON

Department of Biochemistry, Duke University, Durham, North Carolina 27710

(RECEIVED May 18, 1994; ACCEPTED June 28, 1994)

Abstract

The Tyr corner is a conformation in which a tyrosine (residue "Y") near the beginning or end of an antiparallel β-strand makes an H bond from its side-chain OH group to the backbone NH and/or CO of residue Y − 3, Y − 4, or Y − 5 in the nearby connection. The most common "classic" case is a Δ4 Tyr corner (more than 40 examples listed), in which the H bond is to residue Y − 4 and the Tyr $\chi 1$ is near −60°. Y − 2 is almost always a glycine, whose left-handed β or very extended β conformation helps the backbone curve around the Tyr ring. Residue Y − 3 is in polyproline II conformation (often Pro), and residue Y − 5 is usually a hydrophobic (often Leu) that packs next to the Tyr ring. The consensus sequence, then, is LxPGxY, where the first x (the H-bonding position) is hydrophilic. Residues Y and Y − 2 both form narrow pairs of β-sheet H-bonds with the neighboring strand. Δ5 Tyr corners have a 1-residue insertion between the Gly and the Tyr, forming a β-bulge. One protein family has a Δ4 corner formed by a His rather than a Tyr, and several examples use Trp in place of Tyr. For almost all these cases, the protein or domain is a Greek key β-barrel structure, the Tyr corner ends a Greek key connection, and it is well-conserved in related proteins. Most low-twist Greek key β-barrels have 1 Tyr corner. "Reverse" Δ4 Tyr corners (H bonded to Y + 4) and other variants are described, all less common and less conserved. It seems likely that the more classic Tyr corners (Δ4, Δ5, and Δ3 Tyr, Trp, or His) contribute to the stability of a Greek key connection over a hairpin connection, and also that they may aid in the process of folding up Greek key structures.

Keywords: β-arch; Greek key β-barrel; Greek key connection; protein folding; protein structure; side-chain hydrogen bonding; tyrosine corner

For the process of protein folding and the determination of 3-dimensional structure, it is clear that the long-range pattern of hydrophobic and hydrophilic residues in the sequence is extremely important (Bowie et al., 1991; Kamtekar et al., 1993). There are also specific local patterns of sequence that favor certain secondary structures, for instance, general turn-forming sequences (Chou & Fasman, 1977); residues that strongly favor the N-cap, N + 1, etc., positions at the beginning of an α-helix or the C-cap position at the end (Richardson & Richardson, 1988); and the interaction of charged side chains with the helix dipole (Shoemaker et al., 1987). Even more interestingly, there are some cases known where certain local motifs of sequence and conformation strongly favor a given sort of tertiary structure when they occur in an appropriate context, for example, certain glycine-containing sequences that can adopt the unusual turn conformations needed at tight β-hairpins (Sibanda & Thornton, 1985), and the heptad Leu repeat that favors the coiled-coil or leucine zipper interaction (Landschultz et al., 1988; O'Shea et al., 1991). We here describe another such local motif of sequence and conformation, the "tyrosine corner," which occurs at the end of a β-arch in most Greek key β-barrel proteins, but is found essentially nowhere else.

Reprint requests to: Jane S. Richardson, 211 Nanaline Duke Building, Duke Medical School, Box 3711, Durham, North Carolina 27710; e-mail: jsr@suna.biochem.duke.edu.

Results

After noticing several examples of what we now call the classic Δ4 Tyr corner arrangement, we set out to study its distribution and properties. A large set of proteins were surveyed (see Methods section) for tyrosines on or near a β-strand, with the side-chain OH forming an H bond to the backbone CO and/or NH of a residue close in sequence. Table 1 lists the protein ID code, the local sequence, and the H-bonding for each of the Tyr corners, with related examples grouped. The most common Tyr H bonding is to residue Y − 4 (that is, 4 residues N-terminal to the Tyr, or "Y"), and residue Y − 2 is usually Gly. This H bonding pattern is not the typical one for Tyr; Baker and Hubbard (1984)

Table 1. *Amino acid sequence and H-bonding pattern for various Tyr corners, grouped into classes*[a]

PDB code	Residue	Sequence	Adjacent strand	Comments
Δ4 (classic) Tyr corners				
1GCR	Tyr 62-	L :R: R G D **Y** 62	opp 35 I R V	
1GCR	Tyr 151-	L :R: P G E **Y** 151	opp 124 L N V	
2BB2	Tyr 62-	F :E. K G E **Y** 62	opp 35 V L V	
2BB2	Tyr 151-	L :E. K G R **Y** 151	opp 124 V R V	
1PAZ	Tyr 74-	V T .Q: P G A **Y** 74	opp 90 I A V	
1PCY	Tyr 80-	L S .N: K G E **Y** 80	opp 96 V T V	
1AAN	Tyr 88-	F T .E: A G T **Y** 88	opp 102 V V V	
1TTA	Tyr 69-	F :V: E G I **Y** 69	opp 95 F T A	
1TNF	Tyr 56-	V P :S: E G L **Y** 56	opp 124 F Q L	
1BGH	Tyr 61-	Y :A: P G L **Y** 61	opp 4 V E I	
1CPM	Tyr 17-	Y G: Y G L **Y** 17	opp 91 F D W	
2SNV	Tyr 189-	H P: E G F **Y** 189	opp 196 V Q Y	Between domains
2CD4	Tyr 82-	E D: S D T **Y** 82	opp 93 V Q L	
1CD8	Tyr 92-	E N: E G Y **Y** 92	opp 108 V P V	
1TLK	Tyr 113-	D D: D A K **Y** 113	opp 128 A E L	
1BAF L	Tyr 85-	E D: A A T **Y** 85	opp 102 T K L	
1BAF H	Tyr 94-	E D: T A T **Y** 94	opp 109 T Q V	
1BBD L	Tyr 92-	E D: L A V **Y** 92	opp 108 T K L	= L of 1HIL, 2MCP
1BBD H	Tyr 94-	E D: T A V **Y** 94	opp 112 T S V	= 8FAB H
1DBF L	Tyr 86-	D D: F A T **Y** 86	opp 100 T K V	
1DFB H	Tyr 94-	E D: M A L **Y** 94	opp 120 T M V	
1FAI H	Tyr 94-	E D: A A V **Y** 94	opp 118 T T L	
1HIL B	Tyr 90-	E D: S A M **Y** 90	opp 107 T L V	
1IGF L	Tyr 86-	E D: L G V **Y** 86	opp 102 T K L	= 4FAB L
1IGF H	Tyr 90-	E D: T A I **Y** 90	opp 107 T T L	= H of 1IGM, 2MCP
1IGI L	Tyr 86-	E D: L G I **Y** 86	opp 102 T K L	
1IGI H	Tyr 90-	E D: S A V **Y** 90	opp 107 A S V	
1MAM L	Tyr 86-	E D: M A T **Y** 86	opp 102 T K L	
1MAM H	Tyr 96-	E D: S A T **Y** 94	opp 113 T L V	
1MCW W	Tyr 88-	D D: E A D **Y** 88	opp 105 T K V	
1REI	Tyr 86-	E D: I A T **Y** 86	opp 102 T K L	= L of 1FAI, 1IGM
2FB4 L	Tyr 85-	E D: E T D **Y** 85	opp 103 T K V	= 2BJL L
2FB4 H	Tyr 94-	E D: T G V **Y** 94	opp 112 T P V	
2FBJ L	Tyr 85-	E D: A A I **Y** 85	opp 101 T K L	
2FBJ H	Tyr 94-	E D: T A L **Y** 94	opp 112 T L V	
2HFL L	Tyr 85-	E D: A A E **Y** 85	opp 100 T K L	
2HFL H	Tyr 94-	E D: S G V **Y** 94	opp 110 T T L	
2RHE	Tyr 87-	E D: E A D **Y** 87	opp 105 T K L	= L of 2MCG, 3FAB, 7FAB, 8FAB
3FAB H	Tyr 93-	A D: T A V **Y** 93	opp 111 S L V	= 7FAB H
3HFM L	Tyr 86-	E D: F G M **Y** 86	opp 102 T K L	
3HFM H	Tyr 93-	E D: T G T **Y** 93	opp 107 T L V	
4FAB H	Tyr 96-	E D: M G I **Y** 96	opp 112 T S V	
Δ4 Trp/His/Asn corners				
2CD4	Trp 157-	Q D: S G T **W** 157	opp 172 I D I	
1STP	Trp 21-	G I: T G T **W** 21	opp 29 F I V	
1AVD	Trp 10-	S L: T G K **W** 10	opp 18 M T I	
2SOD	His 41-	L T: E G D **H** 41	opp 85 V T A	
1SPD	His 43-	L T: E G L **H** 43	opp 87 V T A	
1LLA	Asn 498-	L :H: P G Q **N** 498	opp 447 V Q N	Not in 1HCY

(*continued*)

Table 1. *Continued*

PDB code	Residue	Sequence	Adjacent strand	Comments
Δ5 Tyr corners				
1NOA	Tyr 32-	L :Q A G T A **Y** 32	opp 55 V T A	
2AZA	Tyr 108-	L :T P G E A **Y** 108	opp 125 L K L	
1TEN	Tyr 869-	L :K P D T E **Y** 869	opp 889 F T T	
1VIL	Tyr 45-	F Y: E G D C **Y** 45	opp 22 R I	Between β-sheets
Δ3 Tyr corners				
1ACX	Tyr 29-	A :G: E T **Y** 29	opp 51 F T T	
1STP	Tyr 54-	A :E S R **Y** 54	opp 43 Y E S	Not in 1AVD
Δ3 Trp corners (Greek key)				
1GCR	Trp 42-	D S: G C **W** 42	opp 57 L R R	
1GCR	Trp 131-	L E: G S **W** 131	opp 146 L R (P)	
2BB2	Trp 42-	Q A: G P **W** 42	opp 57 F E K	
2BB2	Trp 131-	Q S: G T **W** 131	opp 146 L E K	
2SGA	Trp 66-	I S: A S **W** 66	opp 86 G T R	
Irregular, long, and reverse corners				
Δ6 Tyr corners				
1DFB L	Tyr 190-	Y: E K H K V **Y** 190	opp 207 F N R	
1HIL A	Tyr 192-	Y: E R H N S **Y** 192	opp 209 F N R	= 1BAF, 1BBD, 1FAI, 1IGF, 1IGI, 1MAM, 2MCP, 2FBJ, 2HFL, 3HFM, 4FAB
Δ7 Tyr corners				
7FAB L	Tyr 187-	Q: W K S H K S **Y** 187	opp 202 V A P	
8FAB A	Tyr 191-	Q: W K S H R S **Y** 191	opp 206 V A P	= 2BJL, 2MCG
1TTA	Tyr 105-	:N D S G P R R **Y** 105	opp 12 L P	Parallel strand
Reverse Δ4 Tyr corners				
2PLVvp1	Tyr 185-	**Y** T Y G T: A P A	opp Q V Y 155	
2MEVvp1	Tyr 190-	**Y** N S P L: S V L		Y mc not H-bonded; V188 opp L127
2MEVvp2	Tyr 197-	**Y** V N I A: P T S		Y mc not H-bonded; V195 opp W110
4RHVvp1	Tyr 190-	**Y** V G L A: S A Y		Y mc not H-bonded; V188 opp S126
Reverse Δ4 Trp corner				
1RSP	Trp 50-	**W** P T D W: P V		Not in β
Reverse Δ2,4 Asn corners				
1HGFa	Asn 170-	**N** N :D N F:	opp P G D 241	
1HOE	Asn 25-	**N** G :C A E:	opp P G Q 52	
Reverse Δ3 Tyr, His corner				
1RNB	Tyr 78-	**Y** T S :G:		In loop, not β
1GGI	His 315-	**H** I G P: G:	opp vH 95 E	Peptide/antibody
Reverse Δ5 Tyr corner				
5PTI	Tyr 35-	**Y** G G C R :A	opp A R I 18	
Reverse Δ6 Tyr corner				
2TAA	Tyr 125-	**Y** D G A G S S: V	opp S 172	In parallel β

[a] The classes (Δ4, Δ5, Trp, etc.) are described in the text. Homologous examples are grouped. For each corner, the Brookhaven PDB filecode, chain identifier, residue number, and amino acid sequence are given; note that sequence order is always N- to C-terminal, reading from left to right. The Tyr (or Trp, etc.) of the corner is in boldface and its side-chain H bond(s) are indicated on the amino acid sequence by a colon before the residue for an amide and after it for a carbonyl; a single dot represents a long, or weak, H bond. The sequence of the neighboring β-strand is shown, if present, with a number given for the residue that is opposite the Tyr across a narrow pair of β-sheet H bonds. Y − 5 or Y − 6 hydrophobic residues that pack next to the Tyr ring are shown underlined.

found that only 1 of every 8 side-chain–main-chain H-bonds of a Tyr OH was within ±4 in the sequence.

Conformation and sequence

Figure 1 and Kinemage 1 show a closeup of a typical Δ4 Tyr corner. In order to accommodate the OH H bond to residue Y − 4, the Tyr must have a $\chi 1$ near −60° (the average is −67.2°) and the backbone must curve around the tyrosine ring (Fig. 1A). Y and Y − 1 are in fairly standard β conformation, but Y − 2 is in either left-handed β or very extended β (average ϕ, ψ of +158, −177), and Y − 3 is in polyproline II conformation (average ϕ, ψ of −65°, 135°). Therefore, Gly and Pro work very well in those last 2 positions, but are not required: Y − 2 is Gly about 50% of the time (100% for non-immunoglobulin Δ4 cases) and Y − 3 is Pro about 30% of the time. For reasons we

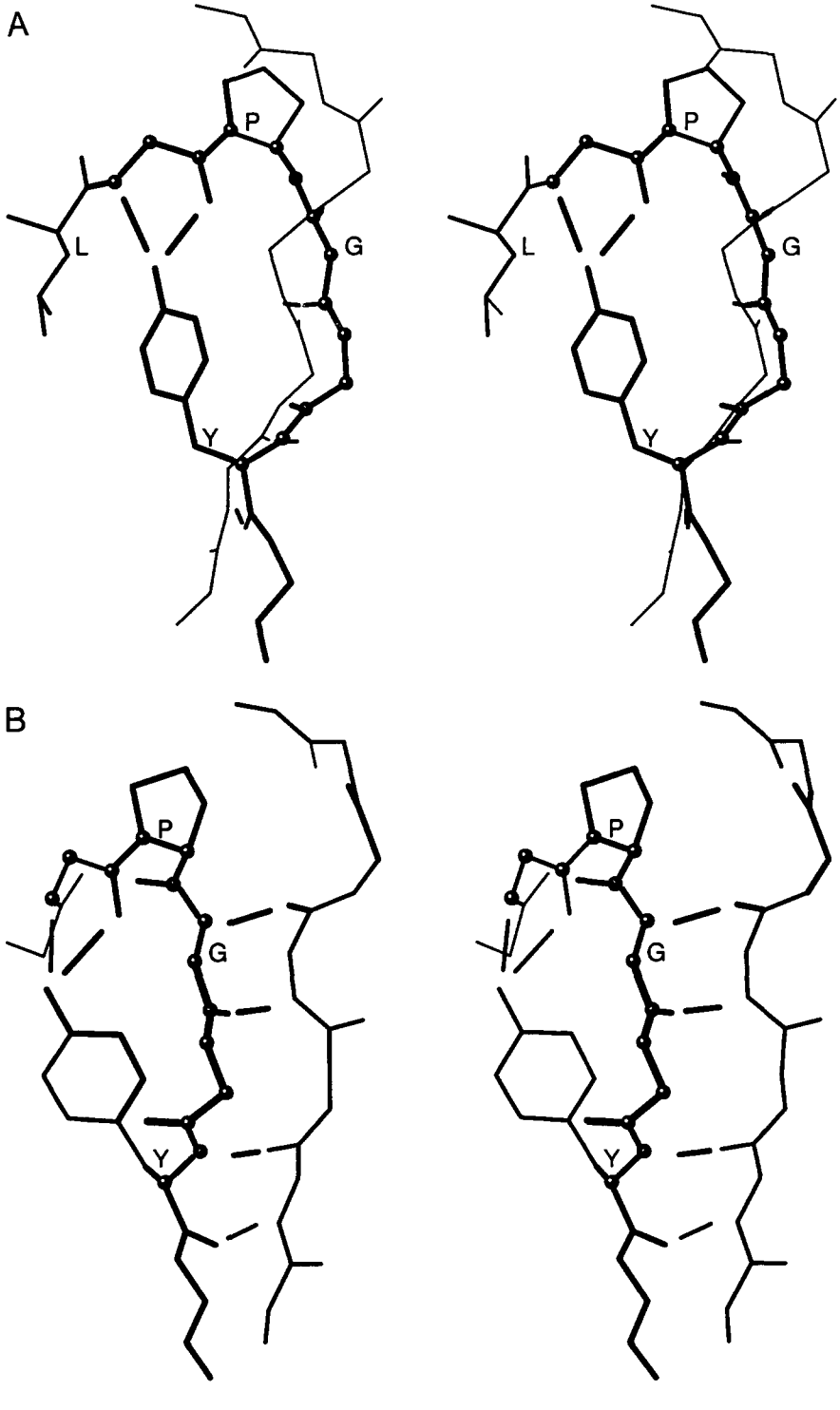


Fig. 1. Stereo diagrams of a classic Δ4 Tyr corner from γ-crystallin (1GCR Tyr 151). **A:** The loop formed by the H bond from the Tyr OH to the backbone at Y − 4. **B:** The β-sheet H bonds to the neighboring strand. Side chains and labels are given for the residues that match the consensus sequence LxPGxY, and their main-chain atoms are shown as balls. (From Kinemage 1.)

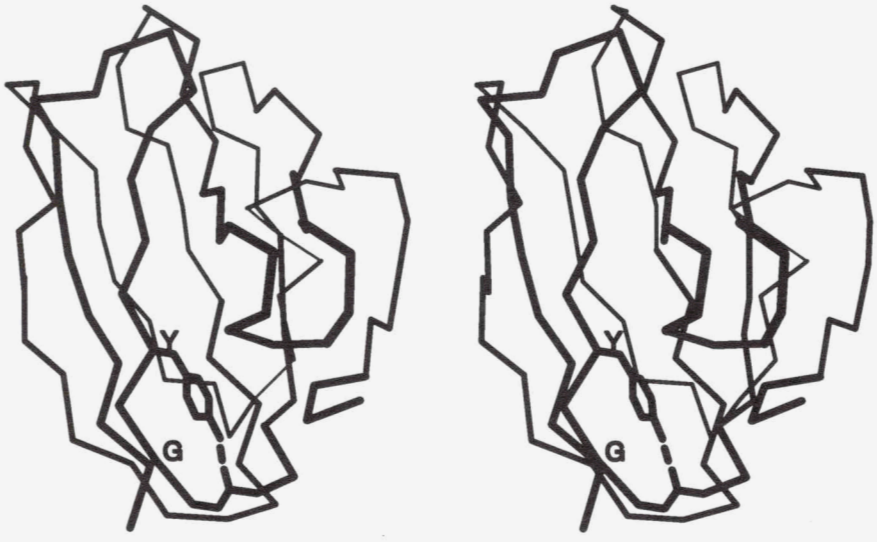


Fig. 2. Stereo diagram showing the tyrosine corner of amicyanin (1AAN Tyr 88) in the context of the overall Cα backbone of the β-barrel. The Tyr side chain is shown, and it and the Y − 2 Gly are labeled. (From Kinemage 2.)

do not understand, Glu is also common (about 30%) in position Y − 3. The H bonding position (Y − 4) is hydrophilic 95% of the time, but any hydrophilic residue works. Usually both the NH and CO of the Y − 4 backbone H bond to the Tyr OH, but sometimes only the CO (as in the immunoglobulins), and sometimes the CO H bond is long (as in the crystallins). Position Y − 5 or Y − 6 is often a hydrophobic (usually Leu) that packs against the Tyr ring as part of the β-barrel interior. The consensus sequence, then, is LxPGxY. Residues Y, Y − 2, and Y + 2 make good antiparallel β-sheet H bonds to the adjacent strand (as seen in Fig. 1B). The β-strand that precedes the β-arch can end as close to the Tyr as residue Y − 4 or as far away as Y − 9.

Position in structure

Figures 2 and 3 and Kinemage 2 show Tyr (or Trp) corners in the context of the proteins or domains in which they occur. They are almost always in antiparallel Greek key β-barrel structures, at the corner that ends a Greek key connection (or "β-arch") and begins a β-strand of the barrel. If there is a definable binding

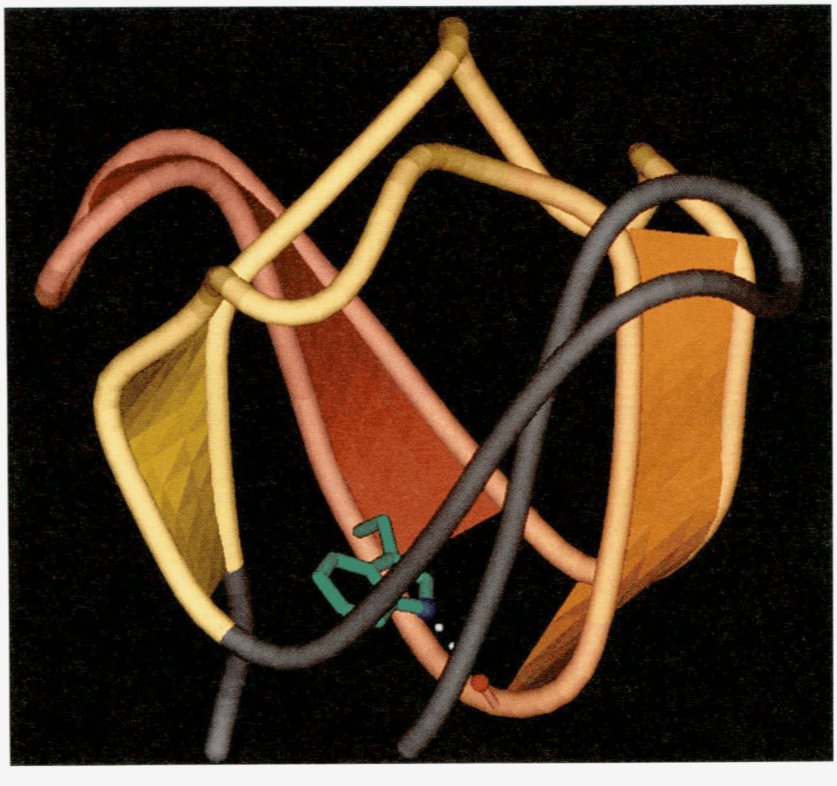


Fig. 3. Ribbon drawing of γ-crystallin domain 1, with the Trp corner side chain in blue-green and dots to mark its H bond to a main-chain carbonyl in the Greek key connection. Colored surfaces emphasize the paired β-strands thought to wrap up together in the folding of the Greek key structure. Color is darkest red near the central hairpin of the paired β-strands and becomes increasingly lighter toward yellow as they wrap around the protein core. The 2 N-terminal strands (in gray) form an independent hairpin not involved in the wrapped strand pairs. Representation made using the VIEW system (Bergman et al., 1993).

or active site, then the Tyr corner is nearly always at the other end of the domain. The difference of shape between the corner formed by a Greek key connection (which arches across the barrel) and a hairpin connection to the adjacent strand is that the former connection is perpendicular to the local plane of the β-sheet, whereas the latter lies in that plane. The H bond and packing interactions with the tyrosine side chain of a Tyr corner presumably stabilize the perpendicular arrangement (that is, a Tyr corner acts as a β-arch support).

Survey of examples

We have already seen a γ-crystallin Tyr corner in Figure 1. There is a classic Δ4 Tyr corner in the equivalent position in each of the 2 domains—Tyr 62 and Tyr 151—in both γ-crystallin (1GCR) and βB2-crystallin (2BB2) (see Kinemage 1 and Table 1). Each has a Gly at residue Y − 2, 1 of the 4 has a Pro at Y − 3, and there are 3 Leu and a Phe at Y − 5. In each case, there is a strong H-bond to the Y − 4 NH and a long one to the Y − 4 CO. Both the Tyr and the Gly form both NH and CO narrow-pair β-sheet H bonds to the appropriate neighboring strand (behind and to the right in the figures). The crystallins are an interesting structure, because their 8-stranded, Greek key domains can be meaningfully described in at least 3 different ways (Richardson et al., 1992). Each domain is a sandwich of 2 4-stranded β-sheets, with no β H bonds between the sheet edges, which best describes the β-sheet geometry in the final structure (Chothia et al., 1977). Each domain has an internal duplication of structure (4-strand Greek key units, with strands 1, 2, and 4 in one sheet and strand 3 in the other sheet) and a low but definite sequence homology, so that the molecule has almost certainly evolved by a 4-fold gene duplication (Wistow et al., 1983). In terms of Greek key folding by wrapped β-strand pairs (Richardson, 1977), the optimal description is that shown in Figure 3, with the first 2 strands an isolated hairpin and the last 6 strands paired correctly to curl up around the interior (see Discussion section). The crystallin Tyr corners (and also the crystallin Trp corners discussed below) occur at the most central foldover in the wrapped strand pairs of each domain; they are not, however, reproduced at the other 2 locations related by gene duplication, which is one piece of circumstantial evidence that Tyr corners may have more to do with the folding of Greek keys than with stability in the final structure.

Kinemage 3 shows 3 classic Δ4 Tyr corners superimposed, from the related blue copper proteins plastocyanin (1PCY), amicyanin (1AAN), and pseudoazurin (1PAZ). (Azurin, the other such protein whose structure is known, has a Δ5 Tyr corner at that position, which will be described below.) The Tyr corners in these blue-copper proteins are extremely similar to the crystallin ones, with Gly in position Y − 2, 1 of 3 Pro in Y − 3, and both Tyr and Gly making a narrow pair of β H bonds to the adjacent strand. The only differences are that the β-sheet below the Tyr is flatter and more regular than in the crystallins, the Tyr OH H bond to the Y − 4 CO is shorter than to the Y − 4 NH, and it is a hydrophobic side chain in position Y − 6 rather than Y − 5 on the connection that packs against the Tyr ring. The blue-copper Tyr corners are all in equivalent locations at a Greek key connection on the end of the barrel opposite the copper site (see Fig. 2), although there is considerable variation in number and placement of edge β-strands among the 4 proteins.

Kinemage 4 shows 5 other unrelated proteins with classic Δ4 Tyr corners—transthyretin (1TTA; or prealbumin), tumor necrosis factor (1TNF), gene 5 protein (1BGH), glucanase (1CPM), and Sindbis virus protein (2SNV). Their local conformation, hydrogen bonding, and sequences are very similar to the examples discussed above. For the first 4 of these, the Tyr corner location is, as usual, at the end of a Greek key connection. The Sindbis virus protein, on the other hand, is an exception in 2 ways: it is a serine protease, which is a high-twist rather than a low-twist β-barrel, and instead of being at one of the Greek key connections, the Tyr corner ends the connection between the 2 domains (see Kinemage 2). However, its local geometry is almost identical to the other classic Δ4 corners, with the connection arching across the barrel and entering the β-strand at a perpendicular to the local β-sheet. Because the Y − 4 residue is Pro in this case, only its CO can H bond to the Tyr OH.

The final group of classic Δ4 Tyr corners is a large related set from the immunoglobulin family (see Kinemage 5). For all VL and VH domains (2RHE, 2FB4, etc.), for the complement proteins CD4 and CD8 (2CD4, 1 CD8), and for telokin (1TLK), this Tyr corner occurs in the equivalent location, at a Greek key connection on the end of the domain away from the antigen-combining site (see Kinemage 2). The sequence and the local conformation from Y to Y − 4 and in relation to the neighboring β-strand is the same as above, except that there is only 25% rather than 100% Gly in position Y − 2, and only the Y − 4 CO H bonds to the Tyr OH. The most striking difference for these immunoglobulin Tyr corners is that instead of coming smoothly across the end of the barrel, the connection makes a sharp jog at residue Y − 4 (ending a turn of 3_{10}-helix). That conformation, as well as the Tyr corner, is conserved across all variable domains in the immunoglobulin family (except for 1 case that uses Trp, discussed below). The equivalent location in constant domains has a longer α-helix in the connection, resulting either in a long, irregular Δ6 or Δ7 Tyr corner, or none at all (see Table 1).

Four unrelated proteins have Δ5 Tyr corners, which in most ways are very like the Δ4 ones (see Fig. 4 and Kinemage 6). In azurin (2AZA), neocarzinostatin (1NOA), and the fibronectin-III-type domain from tenascin (1TEN), there are Δ5 Tyr corners with a β-bulge that puts that puts 2 rather than 1 residue between the Gly and the Tyr. The bulge is a G1 type interlocked with a glycine tight turn (Richardson et al., 1978), so that the Gly at Y − 2 is in left-handed α or left-handed 3_{10} conformation. The residue preceding the Gly is still in polyproline II conformation, and for 2 of the 3 cases, it is a Pro. The Tyr OH H bonds to the backbone NH of residue Y − 5, and all 3 cases have a Leu at Y − 6 that packs against the Tyr ring. The Tyr makes a pair of β H bonds to the neighboring strand, and the bulge adds another 2 H bonds. All 3 of these proteins are low-twist Greek key β-barrels, with the Δ5 Tyr corner at the end of a Greek key connection. Azurin is, of course, related to the other blue-copper proteins discussed above, which have Δ4 Tyr corners in the equivalent location. Neocarzinostatin is related to actinoxanthin (1ACX), which has an unusual Δ3 Tyr corner in the equivalent location. The tenascin domain is a representative of a very large class of "fibronectin III" domains, and sequence comparisons show that the Tyr, the Pro at Y − 4, and the Leu at Y − 6 comprise 3 of the 6 most highly conserved residues in this family (Bork & Doolittle, 1993). The fourth Δ5 Tyr corner occurs in the NMR-determined structure of villin 14T (1VIL;

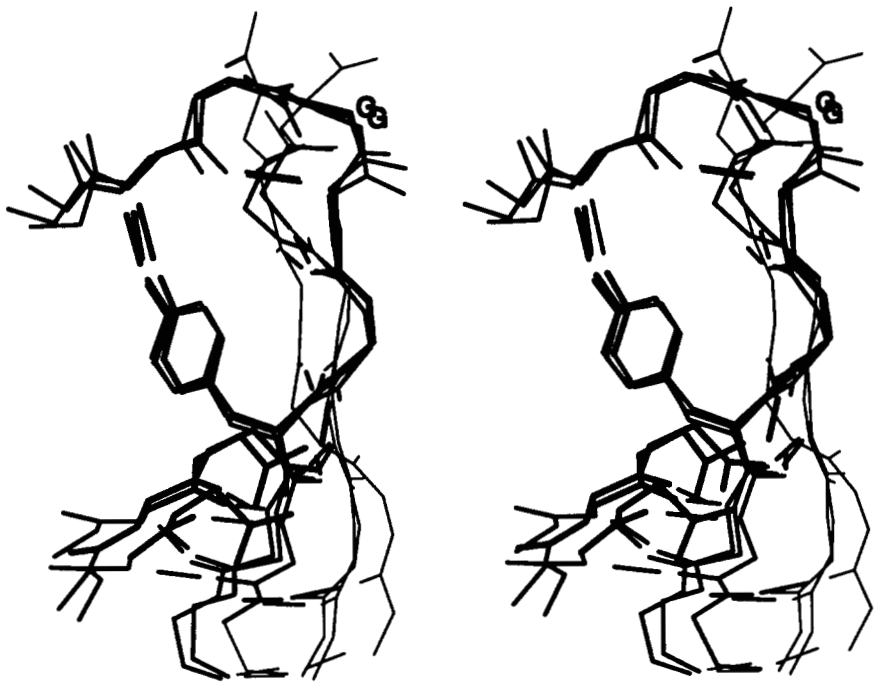


Fig. 4. Stereo diagram of superimposed Δ5 Tyr corners from azurin (2AZA), neocarzinostatin (1NOA), and the fibronectin type III domain from tenascin (1TEN). Note that the extra residue (relative to a classic Δ4 corner) produces both a tight turn and a glycine-type β-bulge. (From Kinemage 6.)

Markus et al., 1994) and has a simpler conformation, without the bulge and turn. Villin is the only α/β protein in which we found a Tyr corner; however, the Tyr corner is in a connection between 2 β-sheets, not in an αβ connection.

Trp, His, or Asn corners

In occasional cases, a side chain other than tyrosine can provide the H bonding and packing interactions for a Tyr corner arrangement (see Kinemage 7). In bovine Cu,Zn superoxide dismutase (2SOD), that side chain is a histidine, whose Nε2 H bonds to the backbone CO of residue H − 4 in the preceding Greek key connection. The conformation and backbone H bonding are the same as for classic Δ4 Tyr corners, and H − 2 is an Lβ Gly and H − 5, a Leu. The Nδ1 of the His H bonds to the backbone of one of the Cu^{2+} ligand residues, contributing to the geometry of the active site. This arrangement is conserved in human SOD (1SPD), but not in yeast SOD (1SDY), nor has it yet been found elsewhere; the necessity for the second H bond may make a His corner more difficult to accommodate than Tyr in barrel interiors.

For the γ- and β-crystallins, in addition to the Δ4 Tyr corner, there is a Δ3 Trp corner on the adjacent strand of the wrapped strand pair (Fig. 3), also at a Greek key connection and also conserved in each domain and in the related proteins. The Nε1 of the Trp ring H bonds to the backbone CO of residue W-3, and residue W-2 is Gly. In the second domain of CD4, the conserved Δ4 Tyr corner of immunoglobulin variable domains is replaced by a homologous one using Trp. There are Δ4 Trp corners on β-strands in both avidin (1AVD) and streptavidin (1STP), one of which is shown in Kinemage 7; however, they follow a helix rather than being in a Greek key connection. *Streptomyces griseus* protease A (2SGA) has a Δ3 Trp corner in the Greek key connection of its first β-barrel; however, in SGPB, that Trp H bonds to a Thr side chain instead, and it is not present at all in the other serine proteases. There are other cases of local Trp H bonds that simply produce a jog in the backbone (such as 2TMV W17 or 1IGI W103), but they are not common, similar to one another, or well conserved.

In the Greek key domain of *Limulus* hemocyanin (1LLA), there is a Δ4 corner made by an asparagine side chain (see Kinemage 7) whose Nδ and Oδ H bond to the backbone CO and NH of residue N − 4. It is quite similar to a Tyr corner in location, sequence, and conformation, but the distance between the Cα of N and N − 4 is only 7.1 Å rather than the 9.3–10.2 Å of a Tyr corner, which does not allow the backbone to curve smoothly out of the Greek key connection. This "Asn" Tyr corner is not present in *Panulirus* hemocyanin (1HCY).

Figure 5 shows a superposition of 1 example each from the 5 groups of Δ4 Tyr and Trp corners described above, to show how well their conformations match. Kinemage 8 also adds a Δ5 Tyr corner example, to show that it matches everywhere but at the β-bulge.

Reverse corners

There are also cases of reverse Δ4 Tyr corners, where the Tyr χ1 is near 180° rather than −60°, and its side-chain OH H bonds to the backbone of residue Y + 4 rather than Y − 4. The Tyr then turns out to be near the end rather than the beginning of a β-strand, and Y + 4 is on the Greek key connection that follows it. We found these reverse Δ4 Tyr corners only in viral coat protein subunits, and only in a rather small subset of those (see Kinemage 9 and Table 1). The 2 cases from mengovirus (2MEV) and the 1 from rhinovirus (4RHV) are in equivalent locations, preceding a long and convoluted Greek key connection, whereas the 1 from poliovirus (2PLV) is in a different location, before a short Greek key connection between sheet edges. β-Sheet H

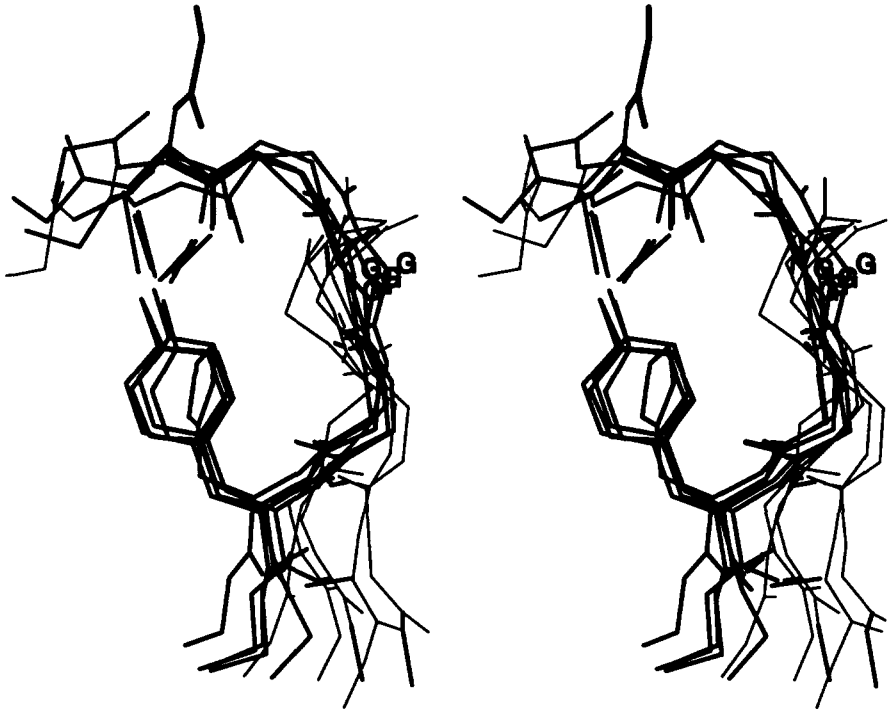


Fig. 5. Stereo drawing of superimposed tyrosine corners, including examples from 4 unrelated protein families: lens crystallins (1GCR Y151), Cu proteins (1PAZ Y74), immunoglobulins (2FB4H Y94), and transthyretin (2PAB Y69). The Δ4 His corner from superoxide dismutase (2SOD) is also shown. (From Kinemage 8.)

bonding to the neighboring strand is minimal, the sequences show little regularity, the connections make different angles to the sheet, and the arrangement is not present in the many other related subunits in the viral coat proteins of known structure. Similarly, reverse Δ3 or reverse Δ5 corners are extremely rare and not well conserved. Therefore, we conclude that reverse Tyr corners are not usually of much importance to the proteins in which they happen to occur.

One interesting example, however, in which a reverse corner may turn out to have importance, is the reverse Δ3 His corner seen in the antigenic peptide from the HIV envelope protein, as complexed to a monoclonal antibody (1GGI; Rini et al., 1993). The conformation of the rest of the envelope protein is not yet known, but the His is on an extended strand, its Nε H bonds to the backbone CO of residues H + 3 and H + 4. Residue H + 2 is an almost completely invariant Gly, whereas the His is the consensus residue at its position, but not always conserved. β-Type H bonds from H and H + 2 are made to a strand in the antibody heavy chain, in parallel orientation.

Another rare type of reverse corner is the reverse Δ2,4 Asn corner, found in the Greek key β-barrels of flu hemagglutinin (1HGFa) and amylase inhibitor (1HOE), as shown in Kinemage 10. The Asn Oδ H bonds to the NH of N + 2 (a quite common pattern for Asn Oδ), but in addition, the Nδ H bonds to the CO of residue N + 4 and there is a β-bulge opposite the Asn. These corners precede Greek key connections, and may help stabilize them over other alternatives, but are not common.

Only 1 example of a Tyr corner was found in α/β proteins and none in αβ connections (see Kinemage 11 and the Δ5 villin example discussed above), even though the geometry of those connections is rather similar to that at the end of a Greek key connection. Some tyrosines near the beginning of β-strands in Greek key β-barrels H bond to backbone across the barrel rather than in the connection (e.g., 3HLA b Tyr 78 to residue Y − 7, or 3GAP Tyr 40 to Y − 11 through a water). Also, not all GxY sequences on β-strands form Tyr corners, even if the Tyr χ1 is near −60°: for example, 5API Tyr 297 H bonds to the backbone of a completely separate β-strand.

Discussion

Tyr corners are a new example of a recognizable feature in the nonrepetitive structure of proteins. They are strongly connected with Greek key β-barrel structures, and several points are suggestive that they might be involved with the folding of those structures. In order to explore that possibility further, let us recapitulate the hypothesis that Greek keys fold by wrapping up from a long, 2-stranded ribbon (Richardson, 1977, 1981; Ptitsyn & Finkelstein, 1980). The pairing of those strands explains the non-near-neighbor connections, and the twist of the original 2-strand ribbon explains the handedness of the Greek key topology. Ray Salemme added another factor to this analysis (Salemme, 1983) by showing that antiparallel β-strand pairs are sided and prefer to curl toward the side that has its side chains between narrow pairs of H bonds; therefore, hydrophobic side chains on the narrow H bond pair side of a β-ribbon will tend to promote its folding and should end up on the inside. Following up on this observation (Richardson & Richardson, 1987), we showed that the wrapped strand pairs of Greek key β-barrels do indeed follow "Ray's Rule." In Δ4 Tyr corners, for the Tyr always and for the Y − 2 Gly almost always, both NH and CO form a narrow pair of β-sheet H bonds to the neighboring strand (listed as "opp" in Table 1). Those 2 strands are the ones related by Ray's Rule, and the Tyr corner is at one of the foldover locations where the strand pair crosses an end of the barrel. H bonding to the β-strand on the other side (the Tyr wide pair side)

Table 2. *Listing of Tyr corners (or lack thereof) in all Greek key β-barrel proteins, grouped according to how many β-strands take part in the Greek key topology (column 2)*[a]

Protein	No. Greek key strands	Corner	Classification
Trypsin, thrombin, tonin, etc.	4	None	Eukaryote
Streptomyces griseus protease A	4	Δ3 Trp	Bacteria
SGPB, SGT, ALP	4	None	Bacteria
Sindbis virus Ser protease	4	Δ4 Tyr	Virus
Staphylococcal nuclease	4	None	Bacteria
Gene 5 protein	4	Δ4 Tyr	Virus
Transthyretin	5	Δ4 Tyr	Eukaryote
Immunoglobulin, VL	(5)	Δ4 Tyr	Eukaryote
Immunoglobulin, VH	(5)	Δ4 Tyr	Eukaryote
Immunoglobulin, CL	5	None	Eukaryote
Immunoglobulin, CH1	5	Δ6 Tyr	Eukaryote
Immunoglobulin, other CHn	5	Δ7 open (pseudo)	Eukaryote
CD4 complement, domain 1	(5)	Δ4 Tyr	Eukaryote
CD4 complement, domain 2	5	Δ4 Trp	Eukaryote
CD8 complement	(5)	Δ4 Tyr	Eukaryote
Telokin	5	Δ4 Tyr	Eukaryote
HLA antigen, subunit b	5	Δ7 open (pseudo)	Eukaryote
Fibronectin III domain	5	Δ5 Tyr	Eukaryote
CuZn SOD, bovine, human	5	Δ4 His	Eukaryote
CuZn SOD, yeast	5	None	Fungus
Hemocyanin, *Limulus*	5	Δ4 Asn	Eukaryote
Hemocyanin, *Panulirus*	5	None	Eukaryote
Plastocyanin	5	Δ4 Tyr	Eukaryote
Azurin	5	Δ5 Tyr	Bacteria
Pseudoazurin, amicyanin	5	Δ4 Tyr	Bacteria
Actinoxanthin	5	Δ3 Tyr	Fungus
Neocarzinostatin	5	Δ5 Tyr	Bacteria
Amylase inhibitor	5	Rev Δ2,4 Asn	Bacteria
γ-, βB2-crystallin, domain 1	6	Δ4 Tyr, Δ4 Trp	Eukaryote
γ-, βB2-crystallin, domain 2	6	Δ4 Tyr, Δ4 Trp	Eukaryote
Taka-amylase, domain 2	6	None	Bacteria
Tumor necrosis factor	7	Δ4 Tyr	Eukaryote
Con A, pea, bean lectins	(8)	None	Eukaryote
Canavalin, phaseolin	8	None	Eukaryote
CAP protein	8	None	Bacteria
Flu hemagglutinin	8	Rev Δ2,4 Asn	Virus
Rhino, mengo, polio virus	8	Rev Δ4 Tyr	Virus
Ten other spherical viruses	8	None	Virus
Phosphocarrier III	12	None	Bacteria

[a] The protein name, type of corner, and phylogenetic classification are also listed. Note that Tyr corners occur in almost all domains with 5, 6, or 7 strands of Greek key structure, but are rare or irregular in those either with just 4 or with 8 or more strands (see Discussion).

is much more limited, seldom continuing past the Tyr; sometimes there is no strand at all on that side. These relationships between Tyr corners and the wrapped β-strand pairs of their Greek keys hold for the Δ4, Δ3, and Δ5 Tyr, Trp, and His corners, but not for the reverse, Asn, or long corners (listed at the end of Table 1).

Not surprisingly, Tyr corners are extremely rare in up-and-down β-barrels (such as the fatty-acid binding proteins and lipo-

calins, or the 3-fold symmetric interleukins and growth factors), because all of their connections are near-neighbor hairpins. The only exception we found is in streptavidin (1STP Tyr 54), and it is not conserved in the closely similar avidin.

We initially expected to find Tyr corners in $\alpha\beta$ connections, where the chain also arches across a hydrophobic core and enters a β-strand approximately perpendicular to the sheet. However, the only such cases we could find were either very dubious examples (as, for instance, in Kinemage 11) or actually occur in an arch connection between 2 β-sheets as in villin (1VIL Tyr 45). Either the difference in positioning between the connection and the sheet is great enough to prevent favorable interaction with the Tyr, or else the primary function of a Tyr corner is to aid folding, which has different requirements in α/β structures than in Greek keys.

Even within the Greek key β-barrels there are some interesting patterns of Tyr corner occurrence, as emphasized by the groupings in Table 2. There is almost never more than 1 Tyr corner per domain, and it lies at the end farthest from the active site. For the low-twist, more sandwich-like proteins with 5, 6, or 7 strands in a Greek key topology (the central section of Table 2), almost every domain has a Tyr corner, they are nearly all classic ones, and they are well conserved among related proteins. Tyr corners are highly unusual, however, in the high-twist Greek key β-barrels (such as the serine proteases), which have only 1 β-arch and only 4 strands that show the Greek key topology; they are listed in the first section of Table 2. At the other extreme, for "jellyroll" domains with 8 or more strands in Greek key topology (the last section of Table 2), only a small percentage have Tyr corners, they are all reverse rather than classic, their locations are less predictable, and they are quite poorly conserved. The lack of Tyr corners in the 8-strand Greek keys appears to correlate with phylogenetic grouping, because most of them are viral proteins; however, that explanation seems unlikely because the phylogenetic correlation is reversed for the 4-strand Greek keys (see Table 2). In general, then, the occurrence of Tyr corners correlates with overall tertiary structure pattern even more strongly than with local conformation, which argues for their involvement in some global process such as folding.

In summary, the Tyr corner is a nearly ubiquitous feature of the low-twist Greek key β-barrel proteins, occurring at the corner between a Greek key connection and the following β-strand. It probably stabilizes that type of connection relative to a hairpin connection. The more classic, common, and well-conserved Tyr corners all occur with similar positioning relative to the wrapped strand pairs that have been hypothesized as a mechanism of Greek key folding, which implies that 1 function of Tyr corners might be to aid such folding. It has long been evident that proteins are better than people at determining unique 3-dimensional structures from sequences, but now we see that they also have the speed and balance to routinely go around corners on one Tyr.

Methods

The initial examples of Tyr corners were found by examining the cases of Gly-Val pairs, found to be one of the most common residue pairs across a narrow pair of hydrogen bonds on the buried sides of antiparallel β-sheets (program by J.M. Hemmingsen). The Gly of such a pair almost always turned out to be the Y − 2 Gly of a Tyr corner. Then a systematic search was undertaken, starting with the 162 proteins of Hobohm et al. (1992), and then adding more α/β proteins and serine proteases, plus an attempted complete set of low-twist Greek key β-barrels up through the end of 1993. The coordinate sets were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977), earlier ones on magnetic tape, and more recent ones from the anonymous FTP directory at 130.199.144.1. Each structure was examined, either with CHAOS (DCR) on an Evans & Sutherland ESV workstation, or with SYBYL (Tripos Assoc.) on an Evans & Sutherland PS390, or with PREKIN and MAGE (DCR; Richardson & Richardson, 1992, 1994) on a Mac II, to find candidates of tyrosines or other H bonding side chains on a β-strand that appeared to interact with nearby main chain. The regions of those candidates were then examined with full main-chain and H bonds. Examples were superimposed using the Calphas of residues Y − 4 to Y + 2 and the 4 residues opposite Y − 2 to Y + 1 on the adjacent strand, plus Cγ and Cz of the Tyr, by the program SUPR (DCR, based on a routine from R. Diamond). Suitable variants of the atom sets were used for the unusual corner types. Kinemages were made of all examples, and distances and conformational angles were measured using MAGE.

Acknowledgments

This work was done with the support of NIH research grant GM-15000, NIH Fellowship F32 GM-15509 to J.M.H., NASA contract NRA-91-OSSA-18 for K.M.G., and using the Macromolecular Graphics Shared Resource of the Duke Comprehensive Cancer Center. Our thanks to Deborah Hill for helping examine the set of buried tryptophans.

References

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

Bergman LD, Richardson JS, Richardson DC, Brooks FP Jr. 1993. VIEW – An exploratory molecular visualization system with user-definable interaction sequences. *Comput Graphics 27*:117–126. (1993 SIGGRAPH Proceedings.)

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

Bork P, Doolittle RS. 1993. Fibronectin type III modules in the receptor phosphate CD45 and tapeworm antigens. *Protein Sci 2*:1185–1187.

Bowie JB, Lüthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*:164–170.

Chothia C, Levitt M, Richardson D. 1977. Structure of proteins: Packing of α-helices and pleated sheets. *Proc Natl Acad Sci USA 74*:4130–4134.

Chou PY, Fasman GD. 1977. β-Turns in proteins. *J Mol Biol 115*:135–175.

Hobohm U, Scharf M, Schneider R, Sander C. 1992. Selection of representative protein data sets. *Protein Sci 1*:409–417.

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

Landschultz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science 240*:1759–1764.

Markus MA, Nakayama T, Matsudaira P, Wagner G. 1994. Solution structure of villin 14T, a domain conserved among actin-severing proteins. *Protein Sci 3*:70–81.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science 254*:539–544.

Ptitsyn OB, Finkelstein AV. 1980. In: Jaenicke J, ed. *Protein folding*. Amsterdam: Elsevier. pp 101–115.

Richardson DC, Richardson JS. 1992. The kinemage: A tool for scientific communication. *Protein Sci 1*:3–9.

Richardson DC, Richardson JS. 1994. Kinemages – Simple macromolecular graphics for interactive teaching and publication. *Trends Biochem Sci 19*:135–138.

Richardson JS. 1977. β Sheet topology and the relatedness of proteins. *Nature 268*:495–500.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–339.

Richardson JS, Getzoff ED, Richardson DC. 1978. The β bulge: A common small unit of nonrepetitive protein structure. *Proc Natl Acad Sci USA 75*:2574–2578.

Richardson JS, Richardson DC. 1987. Some design principles: Betabellin. In: Oxender D, Fox CF, eds. *Protein engineering*. New York: Alan R. Liss. pp 149–163.

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

Richardson JS, Richardson DC, Tweedy NB, Gernert KM, Quinn TP, Hecht MH, Erickson BW, Yang Y, McClain RD, Donlan ME, Surles MC. 1992. Looking at proteins: Representations, folding, packing, and design. *Biophys J 63*:1186–1209.

Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA. 1993. Crystal structure of an HIV-1 neutralizing antibody 50.1 in complex with its V3 loop peptide antigen. *Proc Natl Acad Sci USA 90*:6325–6329.

Salemme FR. 1983. Structural properties of protein β-sheets. *Prog Biophys Mol Biol 42*:95–133.

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

Sibanda BL, Thornton JM. 1985. β-Hairpin families in globular proteins. *Nature 316*:170–174.

Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. 1983. X-ray analysis of the eye lens protein γ-II crystallin at 1.9 Å resolution. *J Mol Biol 170*:175–202.