Supporting Information

Bryan et al. 10.1073/pnas.1017504108

SI Text

Types of Caps. Several recurrent cap motifs are evident in the known structures of β-helices. Described here are the α-helix cap, subdivided into single and double α-helix motifs, and the visor cap, subdivided into the previous-strand and cross-helix motifs. Together, these cap motifs comprise 83% of the N- and C-terminal ends of the broad sample of β-helical structures listed in Table S1. In addition, one β-helical protein, the trimeric tailspike of *Salmonella* phage P22 (1), caps its three β-helices by interleaving their terminal β-strands to form a second, trimeric β-helix domain. Other structures also include interleaved components, most dramatically the phage T4 cell-puncturing device (1K28), in which the entire β-helix is interleaved. In some cases (1K4Z, 1THJ), the interleaving in crystal structure prevents complete characterization of the cap, as the interleaving may be an artifact of crystallization rather than reflecting native structure.

Extent of Caps. As described in the main paper, the general definition of "cap" used here is a minimum continuous set of residues necessary to fulfill three requirements: (*i*) at least one continuous subset of cap residues maintains van der Waals contact with the hydrophobic core of the β-helix; (*ii*) at least one continuous subset of cap residues maintains van der Waals contact with at least one strand of the terminal rung of the β-helix such that the cap backbone intersects the plane of its β-sheet, and (*iii*) caps do not begin or end within secondary structures as defined by the PDB file annotation. As a consequence of requirement (*iii*), caps only begin and end within or at the ends of loops. Residues forming loop structure were included in the cap if they lie between secondary structure elements of the cap, in order to fulfill the requirement of continuity. In addition, loop residues farthest from the β-helix rungs were included if they were in contact with the exposed edge of the terminal rung of the β-helix. Cap residues not in contact with either the hydrophobic core or an exposed edge of the terminal β-helix rung were included only if they were (*i*) part of a secondary structure which did make such contact or (*ii*) if they were part of a loop connecting such secondary structure elements. In general, our definition restricted the extent of caps to within the area expected to block oligomerization.

α-Helix Cap. α-Helix caps (Fig. 2 *A–C*) all include an α-helix, a cap loop, and an additional element ("AE") of secondary structure. The AE may be a β-strand [cf. 2PEC (2) and 1KK6 (3)], a loop [cf. 1CZF (4)], or an α-helix [cf. 1G95 (5)]. The cap α-helix and AE form an approximate plane, perpendicular to the long axis of the β-helix. The α-helix is slightly offset from the long axis and interacts with the hydrophobic residues in the first rung of the β-helix. Observed examples of α-helix caps (see Table S1) have α-helices of 5 to 21 residues in length. The AE and α-helix lie on the opposite sides of the long axis, with the AE antiparallel to the α-helix and more distant from the long axis. The AE itself may be an α-helix, meaning that α-helix caps may include one or two α-helices.

In a typical N-terminal α-helix cap, the cap elements are the only secondary structures adjacent to the N terminus of the β-helix. The polypeptide chain preceding the cap either interacts with one face of the β-helix or forms unstructured loops adjacent to the cap. There is some evidence, however, that elements distal to the cap may act to immobilize cap elements and hold them in place (6).

The contacts of the cap elements with the exterior side chains of the first rung of the β-helix vary according to the type of β-helix. In right-handed β-helices, the α-helix is often both parallel and proximate to the B2 strand of the first rung (see labels, Fig. 1). For these β-helices, interactions are possible between α-helix residues and outward-facing side chains on β-strand B2. The AE, which is not an α-helix in observed right-handed helices, interacts with the B1 and B3 β-strands of the first rung. In left-handed β-helices, the α-helix often lies perpendicular to one of the three β-strands. For these structures, only one or two α-helix residues are close enough to permit contact with outward-facing β-helix side chains. The AE interacts with a second β-strand of the β-helix.

Visor Caps. Visor caps are a set of motifs usually, though not exclusively, found at the C terminus of β-helices. A representative assortment of visor caps is shown in Fig. 2 *D–I*. Visor caps vary widely in their topology, but share a common feature. In each visor cap motif, the cap loop crosses and interacts with at least one of the three β-strands, preventing hydrogen bond contact with other proteins for the contacted strands. The motifs may be classified according to the location of the crossing as previous-strand or cross-helix motifs: previous-strand motifs interact only with the last strand of the beta-helix, while cross-helix motifs interact with other β-strands (which may also include, but is not confined to, the last β-strand of the β-helix). Unlike helix caps, visor caps may not completely prevent exposure of the core of the β-helix to solvent. For instance, the previous-strand caps of left-handed β-helices (Fig. 2*D*) only interact with two of the four core hydrophobic residues in the last rung of the β-helix.

The smallest visor cap motif is a *previous-strand* motif, where the cap loop is directly connected by a turn to the last strand in the β-helix. The turn element of a previous-stand motif is located below a β-strand, thus making the last strand shorter than those preceding it. Depending on the angle between the last strand and the cap loop at the crossing, this turn is between 180 and 270 degrees. The cap loop crosses under and interacts with the previous strand. The interaction may resemble a hairpin loop involving many amino acids, such as 1KQA (7) (Fig. 2*D*), or may involve only one or two pairs of amino acids, such as 1IDK (8) (Fig. 2*E*).

A larger and opposite motif is a *cross-helix* motif, in which the cap loop crosses at least one of the other strands of the β-helix. Cross-helix motifs are found in both left- [Fig. S3*F*, 1G95 (5)] and right-handed β-helices [Fig. 2*G*, 1DBG (9)]. In contrast to previous-strand motifs, cross-helix motifs tend to originate below turns in the β-helix. An exception is 1HF2 (10), which originates after a shortened β-strand. The angle formed by the turn between the last β-strand of the β-helix and the cap loop is less than 180 degrees. The crossing of a cross-helix motif is close to 90 degrees and creates interactions between only one or two pairs of amino acids.

A few visor cap motifs, such as that for pectate lyase C [Fig. 2*H*, 2PEC (2)], include an α-helix. These motifs are distinguished from the α-helix cap by the relative location of the α-helix. In a "visor + α" motif, the α-helix interacts with one strand of the β-helix, not with the hydrophobic core as in the α-helix cap. The α-helix is instead located between the last strand of the β-helix and the cap loop. Because of the distance between the last β strand and the cap loop, all visor + α motifs are cross-helix motifs (and denoted as such in Table S1).

Common Sequence Features of α-Helix Cap Motifs. Fig. 3*A* depicts a composite sequence alignment of the α-helix caps based on structural considerations (see *Methods*). The alignment displays the structural elements noted by visual observation and structural alignment: an additional element and a consensus α-helix immediately adjacent to the β-helix. The α-helices are broadly amphipathic, with a hydrophobic side facing the β-helix and a polar side facing solvent.

In contrast, visor caps do not display significant similarities in sequence. These caps' sequences vary in length and sequence composition. There is little or no correlation between structural alignment and sequence composition. Instead, some visor cap families share more distant structural elements. For instance, almost half of the members of the pectate lyase superfamily have a secondary structure element after the C-terminal visor cap, aligned parallel to the axis of the β-helix and located beside the B2 β-strand of the β-helix. Another common feature of visor caps is a turn at the beginning of the motif that has a more acute angle than the turns of the β-helix itself.

Correlations Between Cap Location and Motif. Cap location and cap motif appear to be correlated in the sample of deposited structures. N-terminal caps are most often found in the dataset to be α-helix caps, and vice versa. Similarly, C-terminal caps are most often found be to visor caps, and vice versa. However, counterexamples exist for both major types of caps. 1THJ, a carbonic anhydrase (11), bears a C-terminal single α-helix cap, while 1HF2, the MINC protein (10), bears an N-terminal visor cap.

Survey of Genbank for Cap-Like Mechanisms Not Detectable by HELIXCAP. To investigate other cap-like mechanisms, Genbank's nonredundant (nr) dataset was searched with BETAWRAP (15), and the resulting set run on HELIXCAP-HMM. Approximately 1,200 sequences with helix caps were identified by these means. Sequences with BETAWRAP scores above −18 (Fig. S5, a cutoff which includes members of 7 of the 18 families in Table S1) were further analyzed with HELIXCAP-visor, and sequences with BETAWRAP scores above −15 were also analyzed with the more computationally intensive RAPTORX (12). After eliminating hypothetical proteins, proteins known to be beta-helices with caps by families in PFAM (16), proteins shown by RAPTORX to be BETAWRAP false positives (i.e., not beta-helix-like), the remaining proteins were sorted by score (Fig. S4). Those proteins with *Z*-scores below 1.5, a cutoff based on the inflection point of Fig. S4, were examined manually for caps or cap-like structures. Notably, of the three sequences so investigated with RAPTORX results, two (gi|254416357 and gi|254411467) are predicted to have a C-terminal minidomain which fulfills one of the two roles of caps (blockage of hydrogen bonding) by extending one beta-sheet well beyond the other. These structures and the templates they are constructed from (2OMZ and 3CIY) indicate another category of possible solutions to the helix-capping problem. [The third sequence (gi|256423386) cannot be evaluated: the loop at the C terminus of the template structure (2UVE) appears similar to a previous-strand visor, but is too short to clearly form a cap structure.]

Experimental Model. Pertactin (1DAB) was selected for deletion mutant experiments due to the availability of high quality folding data and the independence of its capping mechanism from native state oligomerization. While the folding of another β-helical protein, P22 tailspike (1TSP), is even better characterized, the domain acting as a C-terminal cap is formed only upon trimerization of the protein, and is not stably folded independent of this oligomerization. This requirement for trimerization precludes using tailspike as a convenient model to test mechanisms of β-helix capping. Regularity of the C-terminal rungs of the β-helix was considered and rejected as a requirement for the experiments, as C-terminal irregularity is common amongst β-helices (Fig. S1).

To further verify the identification of the C-terminal visor cap, the more computationally intensive RAPTORX (12) was used to search for additional similar visor caps. Disregarding 1HF2 (identified by HELIXCAP-visor) and hits with high sequence similarity, two structures, 3H09 and 1WXR, additionally displayed visor caps highly similar in topology and orientation to pertactin's (Fig. S2).

Previous experiments suggested the relative stability of the C terminus of pertactin (13). Size exclusion chromatography on the pertactin constructs of this work reveal that the cap-deleted monomer and wild-type pertactin move through the column at the same rate, as do the ΔC-terminal cap dimer peak and the end-to-end covalent dimer. These results suggest that the ΔC-terminal cap of pertactin retains its stability without the cap, forming a shape similar to wild-type pertactin.

The correlation of the ΔC-terminal cap dimer peak and the end-to-end covalent dimer further suggest the formation of an end-to-end dimer. The geometry of beta-helices permits two modes of contact between monomers in a hypothetical aggregative fiber: "head-to-tail," where the C terminus of one beta-helix interacts with the N terminus of another, and "head-to-head, tail-to-tail," where N-termini interact with other N-termini and C-termini with other C-termini. The second mode has been observed in other contexts in prion proteins (14). While irregularities of the C terminus rung structure could be effective in inhibiting "head-to-tail" interactions, "tail-to-tail" interactions would not be hampered, as the interacting geometries are by definition similar. It is notable, and an appropriate subject for future research, that the wild-type N terminus of pertactin seems not to have formed "head-to-head" interactions under our experimental conditions despite the lack of an apparent cap in the crystal structure.

1. Steinbacher S, et al. (1994) Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* 265:383–386.
2. Yoder MD, Keen NT, Jurnak F (1993) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science* 260:1503–1507.
3. Sugantino M, Roderick S (2002) Crystal structure of Vat (D): an acetyltransferase that inactivates streptogramin group A antibiotics. *Biochemistry* 41:2209–2216.
4. van Santen Y, et al. (1999) 1.68-A crystal structure of endopolygalacturonase II from Aspergillus niger and identification of active site residues by site-directed mutagenesis. *Journal of Biological Chemistry* 274:30474–30480.
5. Kostrewa D, D'Arcy A, Takacs B, Kamber M (2001) Crystal structures of Streptococcus pneumoniaeN-acetylglucosamine-1-phosphate uridyltransferase, GlmU, in apo form at 2.33 Å resolution and in complex with UDP-N-Acetylglucosamine and Mg^{2+} at 1.96 Å resolution. *Journal of Molecular Biology* 305:279–289.
6. Simkovsky R, King J (2006) An elongated spine of buried core residues necessary for in vivo folding of the parallel beta-helix of P22 tailspike adhesin. *Proc Nat'l Acad Sci USA* 103:3575–3580.
7. Wang X, Olsen L, Roderick S (2002) Structure of the lac operon galactoside acetyltransferase. *Structure* 10:581–588.
8. Mayans O, et al. (1997) Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure* 5:677–689.
9. Huang W, et al. (1999) Crystal structure of chondroitinase B from Flavobacterium heparinum and its complex with a disaccharide product at 1.7 Å resolution. *Journal of Molecular Biology* 294:1257–1269.
10. Cordell S, Anderson R, Löwe J (2001) Crystal structure of the bacterial cell division inhibitor MinC. *The EMBO journal* 20:2454–2461.
11. Kisker C, Schindelin H, Alber B, Ferry J, Rees D (1996) A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon Methanosarcina thermophila. *The EMBO Journal* 15:2323–2330.
12. Peng J, Xu J (2010) Low-homology protein threading. *Bioinformatics* 26:i294–300.
13. Junker M, et al. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci USA* 103:4918–4923.
14. Krishnan R, Lindquist SL (2005) Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* 435:765–772.
15. Bradley P, Cowen L, Menke M, King J, Berger B (2001) BETAWRAP: successful prediction of parallel beta -helices from primary sequence reveals an association with many microbial pathogens. *Proc Natl Acad Sci USA* 98:14819–14824.
16. Bateman A, et al. (2004) The Pfam protein families database. *Nucleic Acids Res* 32(Database issue):D138–141.

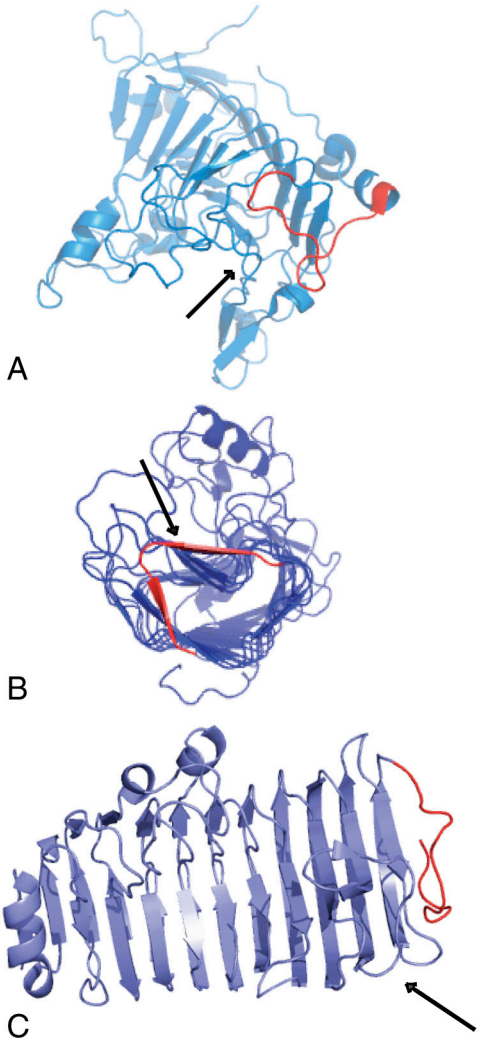

Fig. S1. C-terminal rungs of beta-helices often show irregularity, permitting stability with varied interfaces to the C-terminal visor cap. *A*: 1QJV; *B*: 1BHE; *C*: 1CZF. Arrows denote areas of increased irregularity.

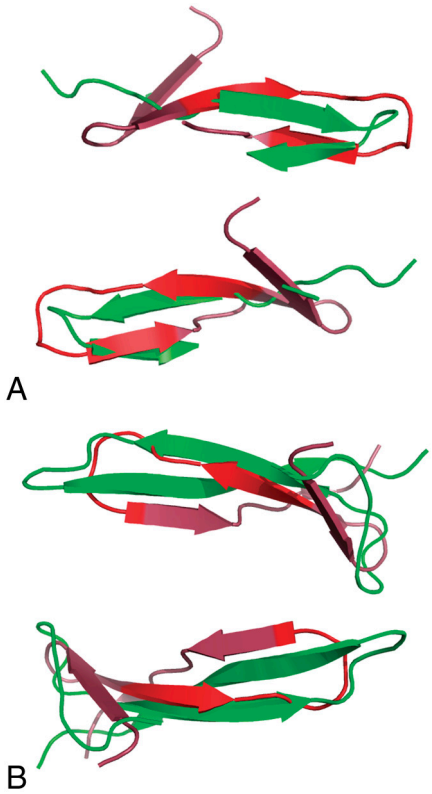

Fig. S2. Additional matches to the C-terminal visor cap of pertactin, as found by RAPTORX (1). *A*: Visor cap of pertactin (red) aligned to visor cap of PDB structure 3H09 (green). *B*: Visor cap of pertactin (red) aligned to visor cap of PDB structure 1WXR (green).

1 Peng J, Xu J (2010) Low-homology protein threading. *Bioinformatics* 26:i294–300.

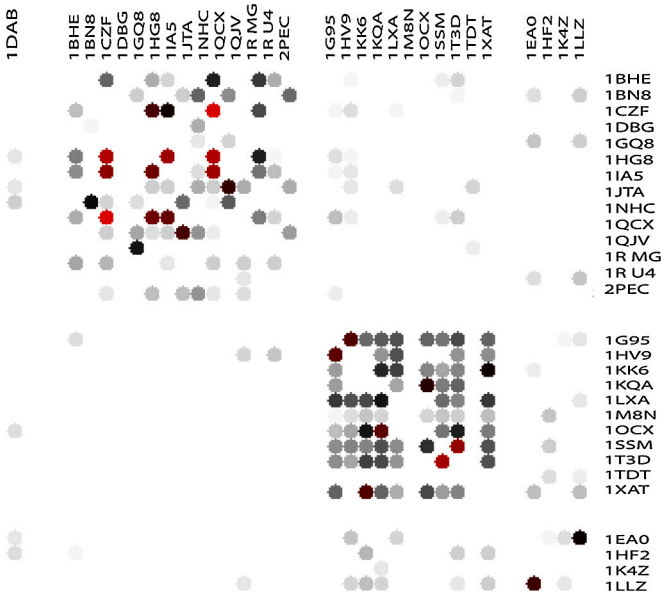

Fig. S3. *Z*-scores of global alignments for all RAPTOR results with C-terminal visor cap-on-cap alignments. Template structures are depicted in rows; query sequences are depicted in columns. Red indicates high *Z*-score, black medium *Z*-score, gray low *Z*-score, and white failure to align.

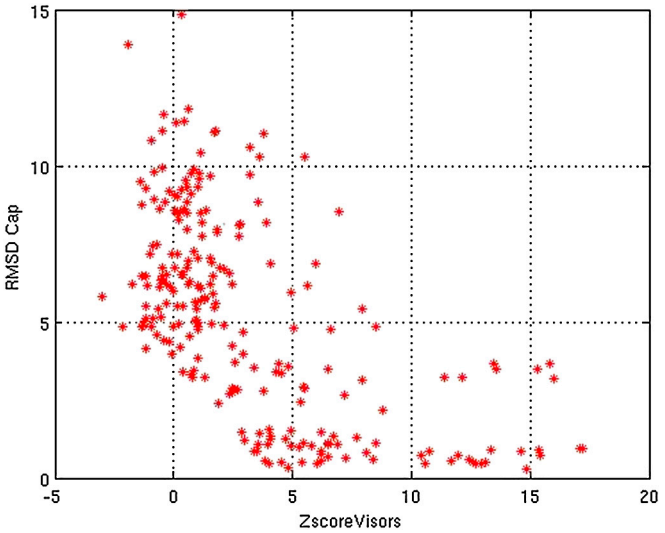

Fig. S4. Plot of *Z*-score against rmsd values for all RAPTOR results with C-terminal visor cap-on-cap alignments.

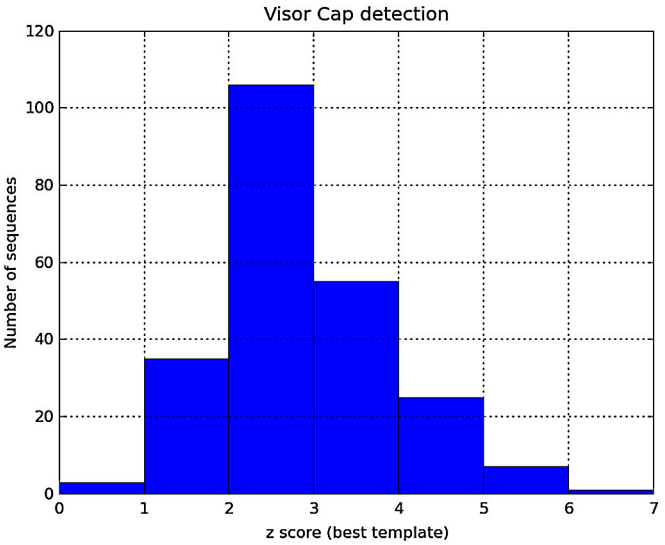

Fig. S5. Histogram of *Z*-scores for Genbank nonredundant sequences with BETAWRAP scores above -18.

Table S1. Classification of β-helix caps by type

PDB ID	Protein name	N-cap*	PDB position†	C-cap*	PDB position†
Pectate lyase superfamily:					
Pectate lyase-like family					
2PEC	pectate lyase C	SH	19-40	CHV	273-314
1JTA	pectin lyase A	SH	34-52	PSV	320-329
1PCL	pectate lyase E	SH	27-44	CHV	280-324
1BN8	pectate lyase	SH	32-48	PSV	354-368
1EE6	pectate lyase	‡		‡	
Pectin lyase family					
1IDK	pectin lyase A	SH	23-38	CHV	317-324
1QCX	pectin lyase B	SH	23-39	CHV	317-328
Galacturonase family					
1RMG	rhamno-galacturonase A	SH	19-51	CHV	343-364
1IA5	polygalacturonase	SH	1-24	CHV	323-339
1K5C	endopoly-galacturonase I	SH	1-17	‡	
1NHC	endopoly-galacturonase I	SH	33-51	CHV	352-368
1CZF	endopoly-galacturonase II	SH	28-45	CHV	347-362
1HG8	polygalacturonase	SH	25-43	CHV	357-373
Pectin methylesterase family					
1GQ8	pectin methylesterase	SH	8-32	PSV	270-286
1QJV	aspartyl esterase	SH	29-56	PSV	330-344
Single-member families					
1DAB	P.69 pertactin virulence factor	‡		PSV	520-528
1BHE	polygalacturonase	SH	19-51	PSV	364-376
1DBG	chondroitinase B	SH	26-40	CHV	421-428
1KTW	iota-carrageenase	SH	46-79	CHV	438-450
1OGM	dextranase	CHV	215-231	PSV	544-574
1RU4	pectate transeliminase	SH	38-72	CHV	349-360
1RWR	filamentous hemagglutinin FhaB	CHV	1-28	‡	
1TSP	P22 tailspike	SH	124-132	Interleaved β-helix§	
Left-handed superfamily:					
Single-member families					
1EWW	spruce budworm antifreeze	PSV	1-18	‡	
1G95	GLMU	SH	239-255	PSV	426-439
1HV9	UDP-N-acetylglucosamine pyrophosphorylase	DH	239-256	PSV	426-440
1KK6	streptogramin A acetyltransferase	SH	16-33	PSV	155-169
1KQA	galactoside O-acetyltransferase	SH	22-54	PSV	172-184
1LXA	UDP-N-acetylglucosamine acyltransferase	‡		PSV	185-197
1M8N	isoform 501, spruce budworm antifreeze	‡		PSV	103-121
1OCX	maltose O-acetyltransferase	DH	19-53	PSV	169-183
1SSM	serine acetyltransferase	SH	102-137	PSV	234-240
1T3D	serine acetyltransferase	SH	105-142	CHV	233-247
1TDT	tetrahydro-diplicolinate N-succinyltransferase	CHV	87-107	CHV	249-256
1THJ	carbonic anhydrase	§	8-13	SH	171-185
1XAT	hexapeptide xenobiotic acetyltransferase	CHV	11-31	PSV	151-164
Non-pectate-lyase superfamily:					
Glutamate synthase family					
1EA0	glutamase synthase	DH	1210-1236	CHV	1400-1415
1LLZ	glutamate synthase	DH	1248-1300	CHV	1436-1451
Single-member families					
1EZG	*tenebrio molitor* antifreeze	‡		‡	
1HF2	MINC	CHV	96-110	CHV	187-202
1K4Z	CAP-1	‡		CHV§	1492-1508
Other families					
1K28	T4 cell-puncturing device	§	389-433	‡	
1WMR	isopullulanase	CHV	200-214	CHV	535-564
2ARA	ARAC	‡		PSV	110-118

*SH, single α-helix cap; DH, double α-helix cap; PSV, previous-strand visor; CHV, cross-helix visor.

†Residue numbers given from start of sequence present in crystal structure.

‡Cap not detected.

§Part of an adjoining domain or interacting with other domains in crystal structure.