

Richard C. Garratt
Christine A. Orengo

The Protein Chart

The "periodic table" of proteins illustrates the beauty, diversity, and complexity of proteins in one place. It is a remarkable teaching and learning tool.

(Donald Voet, author of "Biochemistry")

The ideal tool to understand and to teach the principles of protein structure, and beautifully designed. I wholeheartedly recommend it.

(Robert Huber, Nobel Prize in Chemistry 1988)



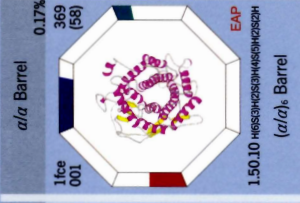
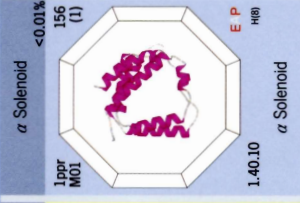
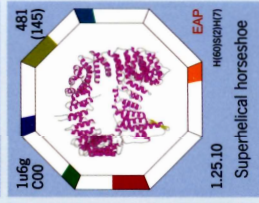
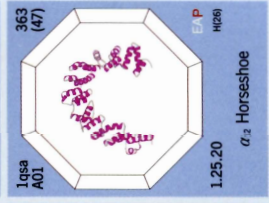
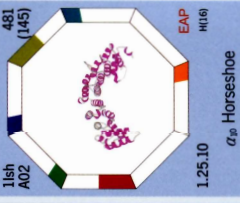
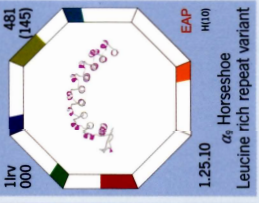
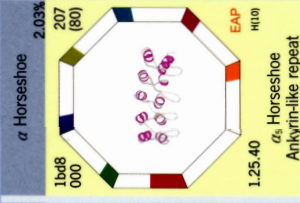
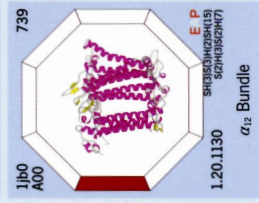
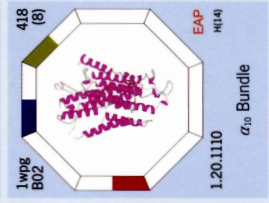
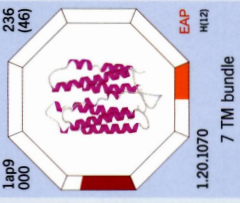
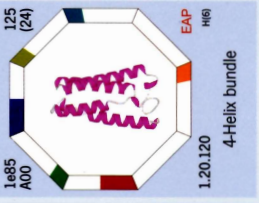
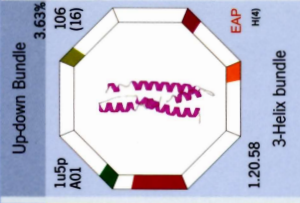
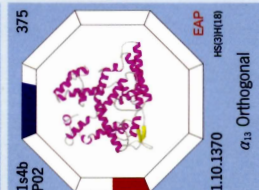
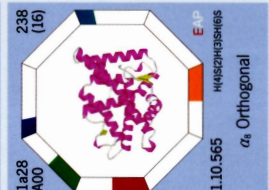
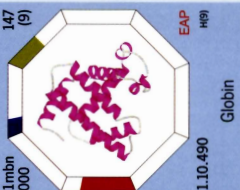
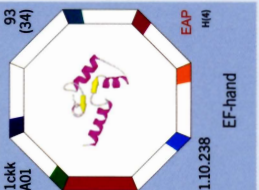
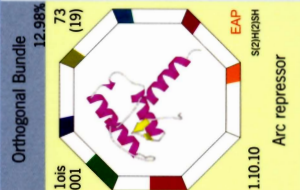
WILEY-VCH

The columns of the table are based on domain architectures as defined by the CATH hierarchical classification. Each cell provides information on an interesting fold group within that architecture and highlights a particular structural domain from that group. The first row of each column typically contains the most basic fold group for that architecture followed by fold groups with more complex structures. The population given as a percentage for each architecture is calculated from the 527 genomes present in Gene3D version 6.0.

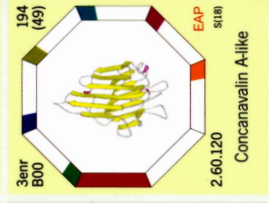
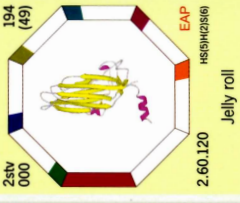
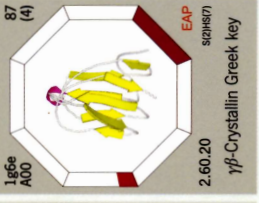
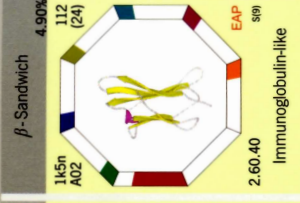
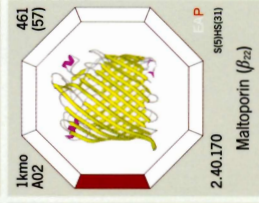
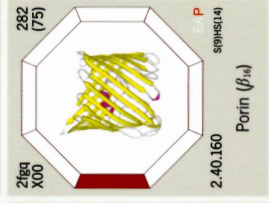
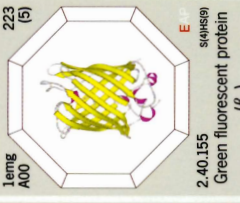
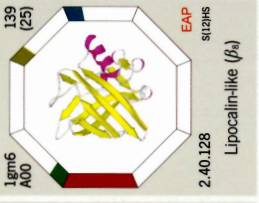
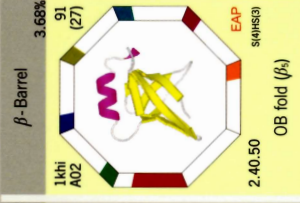
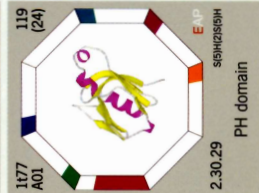
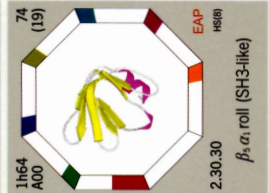
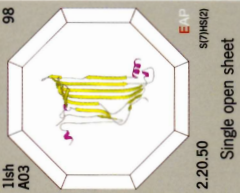
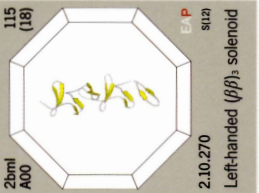
Known functions have been automatically assigned to one of eight categories in the Gene Ontology (GO) molecular function classification (see legend). These categories are represented as a coloured octagon around the structure and are based on a classification scheme devised by Christos Ouzounis. For each fold group, the GO categories are those identified for all structures within that fold group, excluding electronically inferred annotations, as well as all annotated sequence homologues to those structures (at 60% sequence identity, 80% overlap of the larger domain) in Gene3D. Functions are assigned based on the whole structure to which the domain belongs and may therefore not always represent a specific functional attribute of that domain.

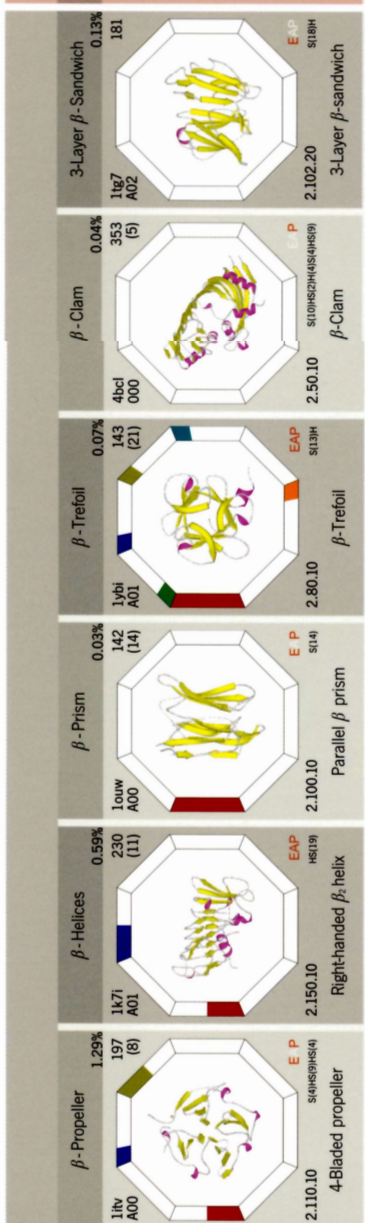
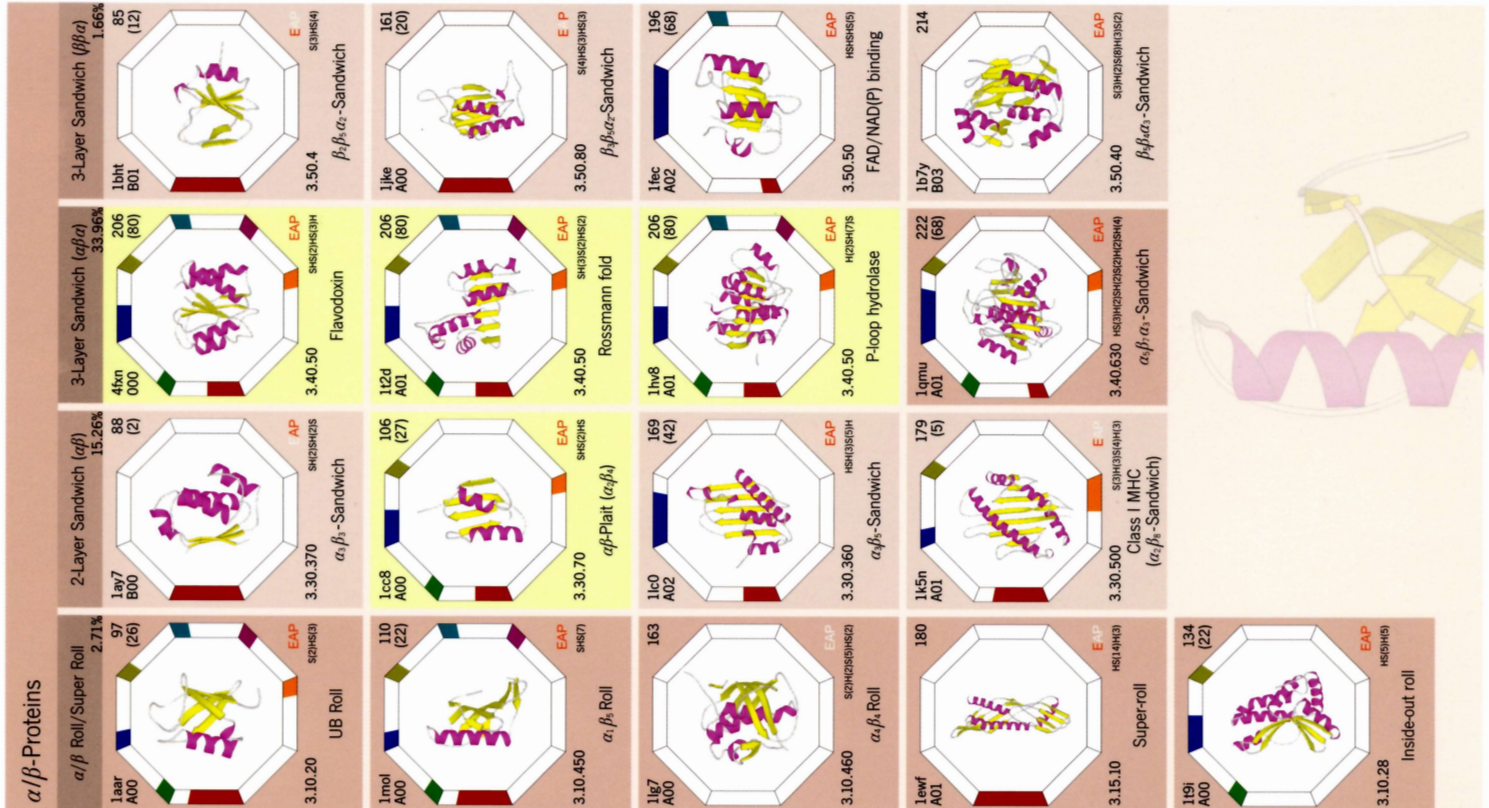
A white octagon tile means that no proteins in that fold group have that function. The incremental filling of the tile (by 1/4, 1/2, 3/4, 1/1) indicates the presence of the respective functions in the fold group and their relative importance (i.e. up to 25 / 50 / 75 / 100% of all proteins in that fold group have that function). For the fibrous proteins the functional mapping is simply that of the particular structure shown and its 60% sequence identity homologues. A completely blank octagon reflects the fact that currently no function can be automatically mapped to that fold, but not necessarily that no known function exists.

α -Proteins



β -Proteins





Legend
 Domain ID for reference structure (PDB code, chain ID, domain number)

1ybi A01 (21)

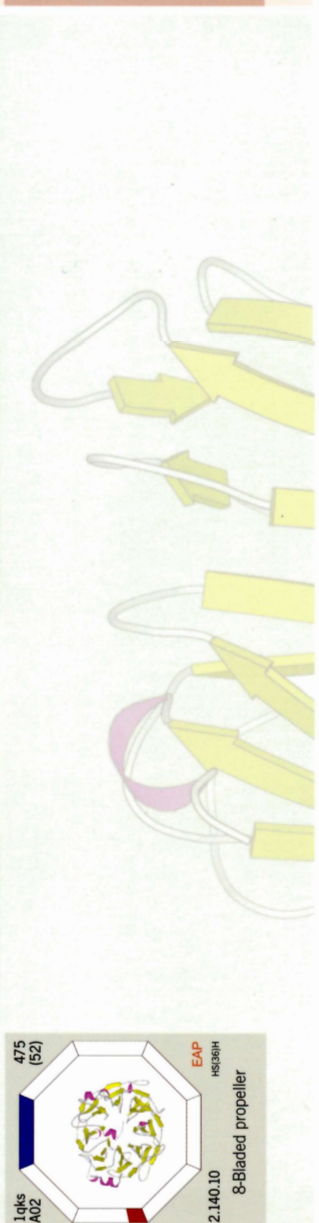
Average length (in residues) and standard deviation for the fold group.

Taxonomy (E=eukarya, A=archaea, P=bacteria) Red indicates the presence of the domain in that lineage.

Fold name
Secondary structure string (H=helix, S=strand)

CATH code for fold group

β -Trefoil
 2.80.10
 S(13)H
 EAP



yellow background indicates top ten most highly populated fold groups (superfolds)

Compiled by
Richard C. Garratt
 (University of São Paulo, Brazil)
 and
Christine A. Orengo
 (University College London, UK)

Designed by
Adam Design
 (Weinheim, Germany)

The collaboration of Alison Cuff,
 Ian Sillitoe, Humberto Pereira and
 Janet Thornton is gratefully acknow-
 ledged.
 With a few exceptions, all protein
 structures have been taken from the
 CATH database (www.cathdb.info).

All rights reserved, including those
 of translation into other languages.
 Reproduction (by photoprinting,
 microfilm, or any other means) and
 translation into machine language
 not permitted.

© 2008 WILEY-VCH Verlag
 GmbH & Co. KGaA,
 Weinheim



Knots & Fibres

Residue A00	Description	Residue EAP
278 299c	α -Helical Fibrous Left-handed coiled coil	H16
1726 1s1j	Spectrin-like 3-helix bundle	H88
448 1qu7	Antiparallel 4-helix bundle	H19
804 1av1	Antiparallel 4-helix closed bundle	H17
160 1mx1	3. Trefoil knot	S162 H121
217 1xd3	4. (Figure-of-eight) knot	S162 H121
152 1ye	5 ₂ Knot	H121
88 1cag	Other Fibrous Collagen-like	S162 H121
540 113w	Cadherin-like	S162 H121

Miscellaneous

Residue A00	Description	Residue EAP
258 1c6q	α/β Box	S162 H121
210 1g6s	α/β Prism	S162 H121
313 7cdc	α/β Barrel TIM barrel	H121
246 117e	Tubby fold	H121

Layer Sandwich ($\alpha\beta\beta\alpha$)

Residue A00	Description	Residue EAP
251 1b90	$\alpha_1\beta_1\alpha_2\beta_2$ -Sandwich	S162 H121
209 1b25	$\alpha_1\beta_1\beta_2\alpha_2$ -Sandwich	S162 H121
225 1xf1	$\alpha_1\beta_1\alpha_2$ -Sandwich	S162 H121

Layer Sandwich ($\beta\beta\beta$)

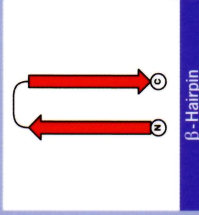
Residue A00	Description	Residue EAP
251 1b90	$\beta_1\alpha_1\beta_2\alpha_2$ -Sandwich	S162 H121
209 1b25	$\alpha_1\beta_1\beta_2\alpha_2$ -Sandwich	S162 H121
225 1xf1	$\alpha_1\beta_1\alpha_2$ -Sandwich	S162 H121

Layer Sandwich ($\beta\alpha\beta$)

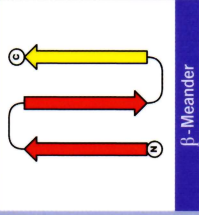
Residue A00	Description	Residue EAP
115 1d15	$\beta\alpha\beta$ -Sandwich	H121
128 1jw3	$\beta\alpha\beta$ -Sandwich	S162 H121
202 1hw7	Hsp33 domain ($\beta\alpha\beta$)	S162 H121
467 1k0e	Orthogonal sandwich	S162 H121



Basic topologies of secondary structure

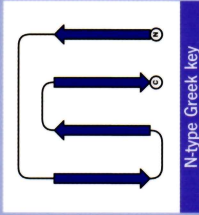


β -Hairpin

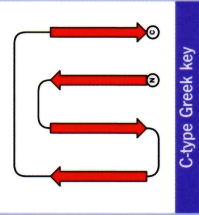


β -Meander

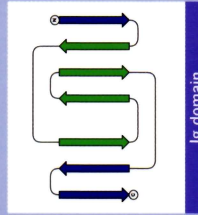
The **β -hairpin** and the **β -meander** are examples of the simplest type of up-and-down antiparallel β -sheet topologies. The latter has an additional strand (yellow) with respect to the former. They can be observed in innumerable examples of β -structures shown in the main table.



N-type Greek key

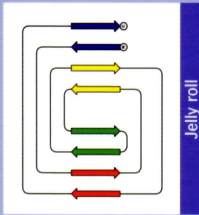


C-type Greek key

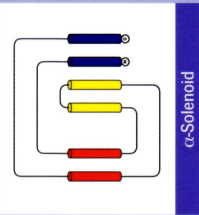


Ig domain

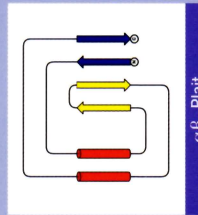
The **Ig fold** is one of the most well-known of all protein structures. An Ig constant domain, as shown here, has an **N-type Greek key** embedded within it (shown in green).



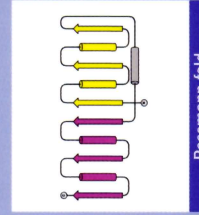
Jelly roll



α -Solenoid



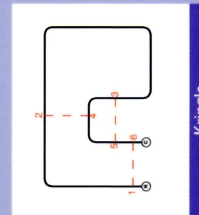
$\alpha\beta$ -Pleit



Rossmann fold

The **Rossmann fold** was first described by Michael Rossmann as a motif observed in lactate dehydrogenase. It is common in α/β proteins with open β -sheets and particularly

in nucleotide binding proteins. It is composed of two topologically identical and pseudo-symmetrically related substructures, which are shown in different colours. The term Rossmann fold is also occasionally used to refer to such substructures. Several examples can be seen in the main table.

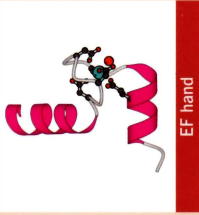


Krigle

The **kringle** takes its name from the topology imposed by its three disulphide bridges, which in two dimensions resembles a Danish pastry of the same name. It is a common modular element in proteins of the coagulation pathways.



HTH motif



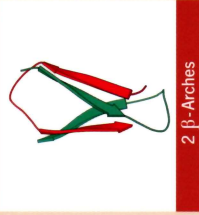
EF hand

The **helix-turn-helix** and **EF-hand** motifs are both characterized by two orthogonal α -helices. The former is a specific example of an α/α corner and is found in DNA binding proteins, where the second (recognition) helix inserts into the major groove. The EF-hand is observed in Ca^{2+} binding proteins, where the Ca^{2+} is bound by a loop between the two helices.

Important structural motifs



Greek key

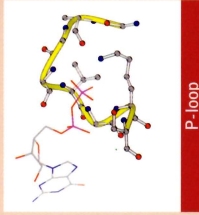


2 β -Arches

The **Greek key** and the interdigitated **β -arches** are two of the most commonly observed motifs (or substructures) in β -proteins. The former is predominant, observed at the edges of antiparallel β -sheets where the motif is often divided between two such sheets. The latter has been described as the most common sub-structure observed in β -sandwiches.

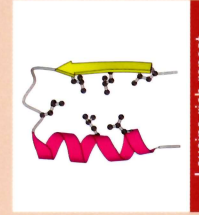


Zinc finger



P-loop

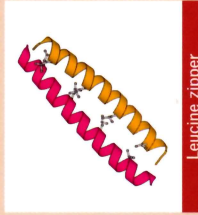
Zinc fingers are metal binding motifs involved in DNA recognition. They differ in their Zn^{2+} ligands, 3D structures and DNA binding modes. The example shown is a 'classical' Zinc finger involving two His and two Cys ligands. The **P-loop** is a glycine rich motif involved in nucleotide binding, where it interacts directly with the α and β phosphate moieties.



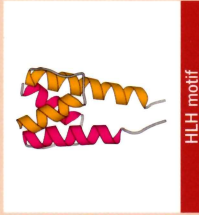
Leucine rich repeat

The **Leucine rich repeat** (LRR) is characterized by a sequence motif which typically contains 6 leucines. They form a structural motif of a β -strand, α -helix and connecting loop. Several examples of proteins, containing different numbers of repeats, can be seen in the $\alpha\beta$ -horsehoes of the main table.

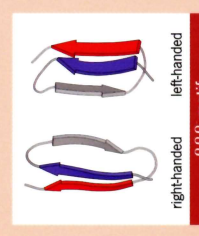
Both the **leucine zipper** and the **helix-loop-helix** are dimerization motifs. In the former case this occurs via the formation of a classical left-handed coiled coil, with leucines at every 7th position (the d position of the coiled coil). In the case of the HLH, the two helices of the motif come together with those of the second monomer to form a 4-helix bundle.



Leucine zipper

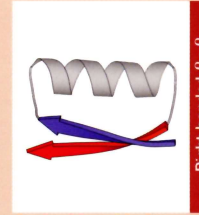


HLH motif



$\beta\beta$ motif
right-handed
left-handed

Most connections between parallel β -strands are right-handed, but exceptions are to be seen in the left-handed β -helices of the main table. The **$\beta\alpha\beta\beta$ motif** includes an additional intervening antiparallel strand.



Right handed $\beta\alpha\beta$



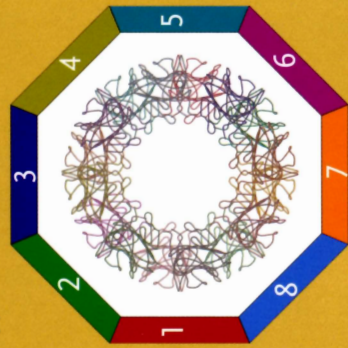
$\beta\alpha\beta\beta$ motif

Legend

PDB code for reference structure

Number of subunits

1qtj



D₈ 822

SAP from horseshoe crab

▲ Protein name
 ▲ Point group symmetry (Schoenflies nomenclature)
 ▲ Point group symmetry (International nomenclature)

- 1 Multiple sites, cross-linking, membrane association
- 2 Cooperativity/allostery
- 3 Cavities, channels and pores
- 4 Functional (active) site formation
- 5 Size and stability
- 6 Economy of genetic material
- 7 "Rulers" (exact separation between binding sites)
- 8 Multiple functions (in hetero-oligomers)

Yellow background indicates dihedral symmetry

Gray background indicates cubic/icosahedral symmetry



WILEY-VCH

Oligomeric Proteins

Highest Order Rotation Axis

1	2	3	4	5	6	7	> 7
1a0d C ₁ HLA class II	1wrp C ₂ Trp repressor	1rtm C ₃ C-type mannose binding protein	1ob8 C ₄ Potassium channel	1lts C ₅ Heat labile enterotoxin (B subunit)	1do0 C ₆ HslU ATPase	1l8f C ₇ SmAP	16 C ₈ LHC II Rhodospirillum rubrum
3hw C ₂ HIV reverse transcriptase	2pol C ₂ Bacterial polymerase III β subunit (E. coli)	1cd5 D ₆ Glucosamine 6-P deaminase	1cuk C ₄ RUVA (DNA recombination protein)	1msl C ₅ Mechanosensitive channel	1gby C ₆ Helicase RepA of plasmid R57 1010	7ah C ₇ Alpha-hemolysin	18 C ₈ LHC II Rhodospirillum rubrum
1hz C ₄ IgG	3phv C ₂ HIV protease	1raa D ₁₂ Aspartate transcarbamoylase	1dhn D ₈ 7, 8-Dihydroneopterin aldolase	1gtp D ₁₀ GTP cyclohydrolase I	1y12 C ₆ Protein secretion apparatus (HCP1)	1gr D ₁₄ GroEL	11 C ₁₁ TRP RNA binding attenuation protein (TRAP)
2aai C ₂ Ricin	4hb C ₄ Human haemoglobin	1dps T ₁₂ DNA binding protein Dps	1a6d D ₁₆ Thermosome	1rvw I ₆₀ Icosahedral lumazine synthase (B. subtilis)	1f52 D ₁₂ Glutamine synthase	1pma D ₂₈ Proteasome	16 D ₁₆ SAP from horseshoe crab
	4pk D ₄ Phosphofructokinase	3pcg T ₂₄ Protocatechuate 3,4-dioxygenase	2fha O ₂₄ Ferritin	1am I ₆₀ Satellite panicum mosaic virus	1g3k D ₂₄ HslUV ATP dependent protease		

ISBN 978-3-527-31963-3



9 783527 319633