3600–3609 Nucleic Acids Research, 1994, Vol. 22, No. 17

The FSSP database of structurally aligned protein fold families

Liisa Holm* and Chris Sander
Protein Design Group, European Molecular Biology Laboratory, D-69012 Heidelberg, Germany

ABSTRACT

FSSP (families of structurally similar proteins) is a database of structural alignments of proteins in the Protein Data Bank (PDB) [1]. The database currently contains an extended structural family for each of 330 representative protein chains. Each data set contains structural alignments of one search structure with all other structurally significantly similar proteins in the representative set (remote homologs, <30 % sequence identity), as well as all structures in the Protein Data Bank with 70–30 % sequence identity relative to the search structure (medium homologs). Very close homologs (above 70 % sequence identity) are excluded as they rarely have marked structural differences. The alignments of remote homologs are the result of pairwise all-against-all structural comparisons in the set of 330 representative protein chains. All such comparisons are based purely on the 3D co-ordinates of the proteins and are derived by automatic (objective) structure comparison programs. The significance of structural similarity is estimated based on statistical criteria. The FSSP database is available electronically from the EMBL file server and by anonymous ftp (file transfer protocol).

INTRODUCTION

It has been estimated that the biochemistry of all living organisms involves no more than 1,000 divergently related protein families [2]. A majority of newly determined protein sequences can be classified into families by detectable sequence homology. The HSSP database of sequence alignments [3] shows that at least 26% of known sequences deposited in public databases (not counting cDNA fragments) have a relative of known 3D structure. However, protein families are known to retain the shape of the fold even when sequences have diverged below the limit of detection of significant similarities at the sequence level. Structural comparisons merge protein families of known 3D structure into structural classes, the members of which may or may not be evolutionarily related [4–7]. The FSSP database of structural alignments provides a rich source of information for the study of both divergent and convergent aspects of the evolution of protein folds.

The FSSP data sets have a wide field of applications. These include studies to discover remote evolutionary connections in the twilight zone of sequence similarity; to build a multiple alignment of remotely related families for the generation of sequence profiles or sequence patterns that may identify additional remote relatives in sequence databases [8–9]; to classify folds, such as TIM barrels, in order to study their structural principles [10]; to define structural cores for sequence-structure alignment (T. Smith, *pers. comm.*), for modular construction of novel proteins, or for model building by homology [11]; to test the accuracy of sequence alignment methods (B. Rost and R.Schneider, *pers. comm.*); or, to use test sets of remotely homologous pairs for fold recognition (M. Sippl, *pers. comm.*) and to extract representative data sets for statistical structural analyses [12]. Other uses are only limited by your imagination.

FORM AND CONTENT OF THE DATABASE

Structural alignments

For a protein chain in the representative set, with PDB identifier Nxxx (like: 1PPT, 5PCY) and chain identifier Y (omitted if blank), there is an ASCII (text) file Nxxx.FSSP or NxxxY.FSSP which contains a few or tens of proteins structurally similar to the search structure (Z-score above 2 in the pairwise structural comparison, see below), alongside the secondary structure and solvent accessibility extracted from the 3D coordinates of the search structure [13]. The structurally equivalent residues are reported in the form of a multiple alignment and as a list of matching fragments and can be inspected using three-dimensional graphics. The co-ordinates must be retrieved separately from the corresponding PDB data sets, e.g. Nxxx.PDB. Details about the methods used to derive the database are given in [14,15].

Figure 1 shows an example dataset from FSSP, that for the SH3 domain of chicken brain alpha-spectrin (1SHG.FSSP). General information about the structure and notation are given at the top of the dataset. The dataset contains 5 (NALIGN) structurally aligned proteins which are listed in the '# # PROTEINS' section. 1SHF-A is the homologous SH3 domain from *fyn* (PROTEIN column) and is aligned with a positional root mean square deviation of 1.6 Å (RMSD column) over 57 residues (LALI column) and has 33 % sequence identity after structural alignment (%IDE column). The other structural homologs are two more SH3 domains (1HSP is misannotated in the PDB), actinidin, and biotin repressor. Some structural details are given in the '# # ALIGNMENTS' section. Residue W42 (Trp) of 1SHG is in a beta-strand (E) and has a solvent

*To whom correspondence should be addressed

```
FSSP       FAMILIES OF STRUCTURALLY SIMILAR PROTEINS, VERSION 0.3 1994
PDBID      1shg
DATE       file generated on  7-Jun-94
DATABASE   360 chains from the Protein Data Bank with 30 % sequence identity
DATABASE   cutoff, based on PDB-select by Hobohm & al, Protein Science 1, 409-417
METHOD     Dali version 1.0: Holm, L., Sander, C. (1993) J.Mol.Biol. 233,123-138.
PARAMETER  elastic alignment with similarity threshold 0.20
THRESHOLD  This file has been filtered to contain only hits that have similarity
THRESHOLD  scores > two standard deviations above database average.
REFERENCE  Holm, L., Ouzounis, C., Tuparev, G., Vriend, G., Sander, C. (1992)
REFERENCE  A database of protein structure families with common folding motifs.
REFERENCE  Protein Science 1, 1691-1698.
CONTACT    e-mail (Internet) Holm@EMBL-Heidelberg.DE or Sander@EMBL-Heidelberg.DE
CONTACT    phone +49-6221-387361 / fax +49-6221-387306
AVAILABLE  Free academic use. Commercial users must apply for licence.
AVAILABLE  No incorporation into other databases.
HEADER     CYTOSKELETON                              19-MAY-93   1SHG
COMPND     ALPHA SPECTRIN (SH3 DOMAIN)
SOURCE     CHICKEN (GALLUS GALLUS) BRAIN
AUTHOR     M.NOBLE,R.PAUPTIT,A.MUSACCHIO,M.SARASTE,M.SARASTE,
SEQLENGTH   57
NALIGN       5
NCHAIN       1 chain(s) in data set /data/dssp/1shg.dssp
NOTATION:  STRID1/STRID2: PDB identifiers of search protein (STRID1) and structurally
NOTATION:       aligned protein (STRID2) with chain identifier
NOTATION:  RMSD: positional root mean square deviation of superimposed CA atoms in Å
NOTATION:  LALI: total length of the aligned fragments for each pair comparison by
NOTATION:       structural alignment.  The list of alignmentsis sorted by LALI.
NOTATION:  LSEQ2: length of the entire chain of the aligned structure.
NOTATION:  %IDE: percentage of sequence identity over aligned positions
NOTATION:  REVERS: number of fragments matching in reversed chain direction
NOTATION:  PERMUT: number of topological permutations
NOTATION:  NFRAG: total number of aligned fragments
NOTATION:  TOPO: 'S' sequential order of aligned fragments; 'N' non-sequential alignment
NOTATION:  NR: sequential index of structurally aligned pairs
NOTATION:  PROTEIN: COMPND record from the PDB file of the aligned structure
NOTATION:  SeqNo, PDBNo, AA, STRUCTURE, BP1, BP2, ACC: sequential and PDB residue
NOTATION:       numbers, amino acid (lower case = Cys),secondary structure,
NOTATION:       solvent exposure as in DSSP (Kabsch and Sander, Biopolymers 22,
NOTATION:       2577-2637, 1983).  The alignments show the amino acid sequence
NOTATION:       and DSSP code (in lower case) of the aligned fragments.
NOTATION:  NOCC: number of aligned structures spanning this position
NOTATION:  RANGE1/RANGE2: sequential and PDB residue numbers of aligned fragments in
NOTATION:       search structure (RANGE1) and structurally aligned protein (RANGE2);
NOTATION:       topological permutations and matches in reverse chain direction are flagged;
NOTATION:       '<-->' reads 'is equivalent to'; PDB residue numbers in parentheses

## PROTEINS : PDB/chain identifiers and structural alignment statistics
  NR. STRID1 STRID2 RMSD  LALI LSEQ2 %IDE REVERS PERMUT NFRAG TOPO  PROTEIN
   1: 1shg   1shf-A  1.6    57    59   33      0      0     2 S     FYN PROTO-ONCOGENE TYROSINE KINASE
   2: 1shg   1hsp    2.9    56    71   27      0      0     3 S     PHOSPHOLIPASE C$GAMMA (SH2 DOMAIN)
   3: 1shg   1pnj    1.7    53    86   28      0      0     3 S     PHOSPHATIDYLINOSITOL 3-KINASE (P85
   4: 1shg   2act   2.6    47   218    6      0      0     7 S     ACTINIDIN (SULFHYDRYL PROTEINASE)
   5: 1shg   1bia   2.7    44   292    9      0      0     3 S     BIOTIN OPERON REPRESSOR (BIRA) BIO

## ALIGNMENTS     1 -    5
 SeqNo PDBNo AA STRUCTURE BP1 BP2  ACC NOCC  .  .  .  .  :  .  .  .  .  .  :  .  .  .  .  .  :  .  .  .  .  .
    1    6   K              0   0  184    5  V  K  Gs P  Nt
    2    7   E         -    0   0   75    5  T  C  Y  Ve Rs
    3    8   L  E      -AB  27  56A  71    5  Le A  Qe Se Pe
    4    9   V  E      -AB  26  55A   0    5  Fe Vb Ye Ve Ve
    5   10   L  E      -AB  25  54A  49    5  Ve Ke Re Ae Ke
    6   11   A  E      - B   0  53A   2    5  Ae Ae Ae Le Le
    7   12   L         +     0   0   55    5  Ls Ls L  D  Ie
    8   13   Y  S    S-      0   0  126    5  Ys Fs Ys A  Ie
    9   14   D         -     0   0   81    5  D  D  D  A  Gt
   10   15   Y  B      -F   20   0B  18    4  Yb Y  Yb Fh .
   11   16   Q         -     0   0   94    4  E  Kb K  Kh .
   12   17   E         +     0   0   61    4  A  As K  Qh .
   13   18   K         +     0   0  149    4  Rs Qs E  Y  .
   14   19   S  S >  S-      0   0   32    4  Ts Rs Rs A  .
   15   20   P  T 3  S+      0   0  137    4  Es Es Et Ss .
   16   21   R  T 3  S+      0   0  158    4  Ds Dt Et Gs .
   17   22   E  B <   -c    47   0A  10    3  Db Et Db .  .
   18   23   V         -     0   0    2    2  L     I  .  .
   19   24   T         +     0   0   55    3  S  Tb D  .  .
   20   25   M  B      -F   10   0B   3    3  Fb F  Lb .  .
   21   26   K     >   -     0   0  138    3  H  I  H  .  .
   22   27   K  T 3  S+      0   0  137    5  Kt Ks Lt V  Ke
   23   28   G  T 3  S+      0   0   46    5  Gt Ss Gt D  Ee
   24   29   D    <   -      0   0   53    5  E  A  D  He Ie
   25   30   I  E      -A    5   0A  96    5  Ke I  Ie Ae Fe
   26   31   L  E      -A    4   0A   0    5  Fe Ib Le Ie Ge
   27   32   T  E      -AD   3  40A  30    5  Qe Q  Te Ve Ie
   28   33   L  E      + D   0  39A  10    5  Ie Ns V  Ie Se
   29   34   L  E      -     0   0   59    4  Le Vb .  Ve Re
```

```
 30   35   N  E      + D   0  38A  72    4  Ne E  .  Ge Ge
 31   36   S         +     0   0   44    4  S  K  .  Ye Ie
 32   37   T         +     0   0  127    4  Ss Q  .  Ge D
 33   38   N  S     S-     0   0   77    4  Es Es Et .  Kt
 34   39   K  S     S+     0   0  164    4  Gs Gt I  .  Qt
 35   40   D  S     S+     0   0   77    4  Ds Gt G  .  Gs
 36   41   W  E       - E   0  49A  76    4  We W  We .  A
 37   42   W  E       - E   0  48A  39    5  We Wb Le We Le
 38   43   K  E      +DE   30  47A  63    5  Ee R  Ne Ie Le
 39   44   V  E      -DE   28  46A   0    5  Ae Gb Ge Ve Le
 40   45   E  E      -DE   27  45A  57    5  Re Db Ye Ke Ee
 41   46   V  E >    - E   0  44A  16    5  Se Y  Ne N  Qe
 42   47   N  T 3  S-      0   0  149    5  Tt Gs Et Ss Dt
 43   48   D  T 3  S+      0   0  112    5  G  Gs Gt Wb Gt
 44   49   R  E <    - E   0  41A 108    5  E  K  Ee Gt Ie
 45   50   Q  E      + E   0  40A 102    5  Te Kb Re Ye Ie
 46   51   G  E      - E   0  39A   2    5  Ge L  Ge Me Ke
 47   52   F  E      +cE  17  38A  56    5  Ye W  De Re Pe
 48   53   V  E      - E   0  37A   0    5  Ie Fb Fe Ie We
 49   54   P  E >    - E   0  36A  16    4  Pe P  Pe .  Ms
 50   55   A  G >  S+      0   0   15    4  Sg Ss Gs Tg .
 51   56   A  G 3  S+      0   0   76    5  Ng Ns Ts cg Gs
 52   57   Y  G <  S+      0   0   89    5  Yg Ys Ys Gg G
 53   58   V  E <   -B     6  0A  11    5  Ve Ve Ve It Ee
 54   59   K  E     -B     5  0A 107    5  Ae Ee Ee Ms Ie
 55   60   K  E     -B     4  0A  82    5  Pe E  Ye P  Se
 56   61   L  E      B     3  0A  62    5  V  Ms Ie Se Le
 57   62   D                0   0 215    5  D  V  Ge Ye R

## FRAGMENTS: ranges of superimposed residues
  NR. STRID1 STRID2                          RANGE1 <--> RANGE2
   1: 1shg   1shf-A   1 LYS (   6) -  41 VAL (  46) <-->   1 VAL (  84) -  41 SER ( 124)
   1: 1shg   1shf-A  42 ASN (  47) -  57 ASP (  62) <-->  44 THR ( 127) -  59 ASP ( 142)
   2: 1shg   1hsp     1 LYS (   6) -  17 GLU (  22) <-->   6 LYS (   6) -  22 GLU (  22)
   2: 1shg   1hsp    19 THR (  24) -  45 GLN (  50) <-->  24 THR (  24) -  50 LYS (  50)
   2: 1shg   1hsp    46 GLY (  51) -  57 ASP (  62) <-->  52 LEU (  52) -  63 VAL (  63)
   3: 1shg   1pnj     1 LYS (   6) -  28 LEU (  33) <-->   7 GLY (   5) -  34 VAL (  32)
   3: 1shg   1pnj    33 ASN (  38) -  42 ASN (  47) <-->  54 GLU (  52) -  63 GLU (  61)
   3: 1shg   1pnj    43 ASP (  48) -  57 ASP (  62) <-->  66 GLY (  64) -  80 GLY (  78)
   4: 1shg   2act     1 LYS (   6) -   9 ASP (  14) <--> 132 PRO ( 132) - 140 ALA ( 140)
   4: 1shg   2act    10 TYR (  15) -  16 ARG (  21) <--> 144 PHE ( 144) - 150 GLY ( 150)
   4: 1shg   2act    22 LYS (  27) -  32 THR (  37) <--> 160 VAL ( 160) - 170 GLY ( 170)
   4: 1shg   2act    37 TRP (  42) -  43 ASP (  48) <--> 178 TRP ( 178) - 184 TRP ( 184)
   4: 1shg   2act    44 ARG (  49) -  48 VAL (  53) <--> 192 GLY ( 192) - 196 ILE ( 196)
   4: 1shg   2act    50 ALA (  55) -  53 VAL (  58) <--> 205 THR ( 205) - 208 ILE ( 208)
   4: 1shg   2act    54 LYS (  59) -  57 ASP (  62) <--> 211 MET ( 211) - 214 TYR ( 214)
   5: 1shg   1bia     1 LYS (   6) -   9 ASP (  14) <--> 248 ASN ( 273) - 256 GLY ( 281)
   5: 1shg   1bia    22 LYS (  27) -  49 PRO (  54) <--> 258 LYS ( 283) - 285 MET ( 310)
   5: 1shg   1bia    51 ALA (  56) -  57 ASP (  62) <--> 286 GLY ( 311) - 292 ARG ( 317)
```

Figure 1. Format of an FSSP file. One FSSP file contains a structural protein family: the search structure and structurally homologous proteins from the PDB. File organization is line-oriented and strictly formatted. Lines have a maximum length of 132 bytes. The file is divided into four sections, HEADER, PROTEINS, ALIGNMENTS and FRAGMENTS. The sections are separated by double hashes (# #). The HEADER section is mandatory. The HEADER, PROTEINS and ALIGNMENTS sections are similar to those in the HSSP database [3], with obvious modifications of notation that are explained in the HEADER block. The FRAGMENTS section reports the beginning and ending residue numbers of structurally equivalent segments. The residue ranges are given both according to sequential numbering starting from 1 and, in parentheses, according to the numbering in the PDB files.

accessibility (ACC column) of 39 Å^2. W42 has a structurally equivalent residue in 5 (NOCC) of the aligned structures, of which three are tryptophans (W), two are leucines (L), and all five are in beta-strands (b or e in We, Wb, Le, We, Le). Finally, the '# # FRAGMENTS' section says that to superimpose the 3D coordinates of 1SHG with those of 1SHF, residues 6–46 and 47–62 of 1SHG should be equivalenced with residues A84–A124 and A127–A142 of 1SHF.

The default files (Nxxx.FSSP) contain structural alignments generated by the program Dali [15] and are constrained to preserve sequential ordering of the aligned segments. Alignments optimized allowing topological permutations (loop reconnections and chain reversals) are available in files Nxxx__dali.FSSP. Alignments using other methods are available in datasets Nxxx__suppos.FSSP and Nxxx__comp3D.FSSP [14].

Index of protein fold families

To aid navigation in the database, the 330 protein chains contained in the representative set have been clustered into fold families (Table I). A dendrogram of the families was produced by average linkage clustering based on structural similarity scores [15]. Chain length effects were corrected for by transforming the pairwise similarities into statistical significance scores (Z-scores). Families and subfamilies result from truncating the tree at different cut levels of Z-score. The higher the cut, the larger the resulting number of distinct fold families (Figure 2). 142 families resulting from the cut at an average Z-score of 2 are numbered in the first column of Table I. Second and further members of a family are indicated by indentation relative to the first member at the given level of significance. For example, if one decided to derive a

more refined selection of fold families using a Z-score cutoff of 3 instead of 2, then the set of families should be expanded by all subfamilies that are indented by one letter space in Table I, yielding a total of 168 families. The most refined selection possible in the representative set would place each of the 330 chains in a distinct family, but even a cut as high as a Z-score of 10 yields only 255 families (Figure 2).

In comparing proteins with very low sequence identity, there is no direct relationship between the structural Z-score and evolutionary relatedness. To assert descent by common ancestry, the biological function, sequence signatures and architectural detail should be considered. For example, the very distantly related animal/plant lysozymes and T4 lysozyme are classified into two neighbouring families (21 and 22) using the structural Z-score, although they share some structural and biochemical features. As an example of common folding motifs, family 57 in Table I contains six structures with the babbab fold typified by muconolactone isomerase (1MLI).

DISTRIBUTION

Network access

The FSSP data sets can be obtained from the EMBL file server [16]. To get detailed instructions on how to use the service send the messages 'HELP' and 'HELP proteindata' to the network address Netserv@embl-heidelberg.de. If you have access to Internet you can obtain FSSP files by anonymous ftp (file transfer protocol) from ftp.embl-heidelberg.de, directory: /pub/databases/protein_extras \ fssp. Access to the database is also possible over the World Wide Web (WWW), e.g. using the XMosaic interface; the URL address is http://www.embl-heidelberg.de/databases/protein_extras \ fssp. Distribution by the Protein Data Bank (pdb.pdb.bnl.gov) is planned for late 1994.

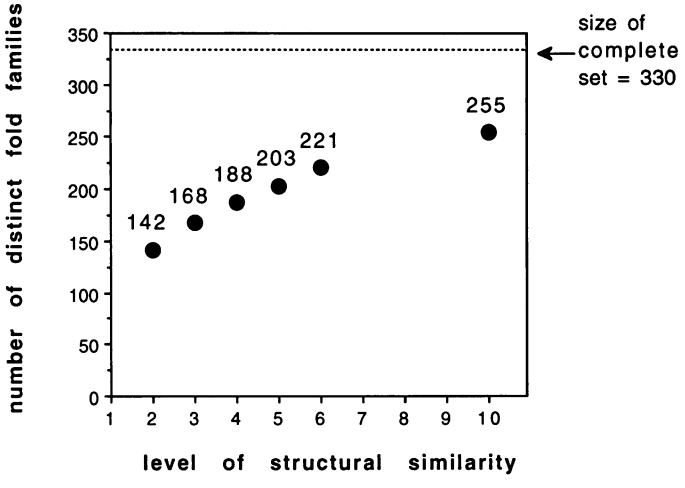

Figure 2. Definition of structural classes. The June 1994 release of the FSSP database is based on a sequence-representative set of 330 protein chains (less than 30 % sequence identity). Average linkage clustering using the similarity scores from an all-against-all structural comparison yielded a tree representation of structural relations in the set (cf. Table 1). Truncating the tree at different levels of structural similarity (horizontal axis, Z-score) defines distinct families, i.e., separated branches of the tree. Cutting at a very low level ($Z << 2$) leads to a collapse into a very few general classes (all-alpha, all-beta). Cutting at a high level increases the number of distinct families, with a gradual approach to one family per protein chain.

The SUPPOS program is available as part of the WHAT IF package (available from G.Vriend, email: vriend@embl-heidelberg.de). The program Dali is currently not available for distribution. Requests for alignments of newly solved crystallographic or solution NMR structures (C^a co-ordinates required) may be sent to L.Holm by email (holm@embl-heidelberg.de).

Conditions

Academic redistribution of single files or of the entire database is permitted. No inclusion in other databases or database services, academic or other, without explicit permission of the authors. All rights reserved. Not to be used for classified research. Users are asked to refer to this paper and ref. 14 in reporting results on use of the database.

Size of the current release

The content and size of the FSSP database is of course tightly coupled to the development of the Protein Data Bank which is currently increasing at the rate of hundreds of datasets every year. The size of the sequence-representative set of PDB files [17], which is used here as a point of departure, has increased from 154 in December 1992 to 204 in October 1993 to 330 in June 1994. The complete set of data files (June 1994) requires about 11 Mb of disk storage. Regular and frequent updates of the database are planned.

Limitations

The structure comparison program Dali [15] defines the extent of the common structural core by maximizing the agreement of *intra*molecular CA−CA distances. The scoring function was deliberately designed to allow inter-domain conformational flexibility; hence, positional root mean square deviations for the corresponding rigid-body superimpositions are often higher than for comparison methods that put an absolute upper limit on *inter*molecular positional deviations. This, however, is only an apparent disadvantage.

The current database contains at most one alignment per pair of full length proteins. In future releases, the significance of alignments will be evaluated at the level of structural domains [18], i.e., parts of structures, and significant suboptimal alignments will be included. PDB data sets are referred to by the PDB code; no provision can be made for asynchronous revisions of the PDB data sets relative to the derived database.

Related data banks and programs

It is often useful to complement the compilation of structure alignments with sequence and variability information by direct reference to the latest version of the HSSP database of sequence-aligned protein families [3]. Users interested in detailed local structural properties of each protein, such as hydrogen bonding patterns, may refer to the DSSP database of secondary structures, derived from PDB files. The HSSP and DSSP databases are available by the same mechanism of network access as FSSP, see above. An X-windows based protein query and 3D inspection system, ProtQuiz 0.7 (Sander & Scharf, unpubl.; test version available via anonymous ftp from ftp.embl-heidelberg.de), can be used for interactive evaluation of pairwise alignments. The FSSP database is cross-referenced with several sequence and other databases in the information retrieval system SRS [19] with access provided on www.embl-heidelberg.de. Kindly report any problems to the authors by electronic mail.

REFERENCES

1. Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M., J. Mol. Biol. 112:535–542 (1977).
2. Chothia C. Nature 357:543–544 (1992).
3. Sander C., Schneider R., Proteins 9:56–68 (1991).
4. Overington J., Johnson M.S., Sali A., Blundell T.L., Proc. R. Soc. Lond. B241:132–145 (1990).
5. Pascarella S., Argos P., Prot. Eng. 5:121–137 (1992).
6. Orengo C.A., Flores T.P., Taylor W.R., Thornton J.M., Prot. Eng. 6:485–500 (1993).
7. Holm L., Sander C., Proteins 19:165–173 (1994).
8. Bashford D., Chothia C., Lesk A.M., J. Mol. Biol. 196:199–216 (1987).
9. Taylor W.R., Prot. Eng. 2:77–86 (1988).
10. Wilmanns M., Hyde C.C., Davies D.R., Kirschner K., Jansonius J.N., Biochemistry 30:9161–9169 (1991).
11. Sutcliffe M.J., Haneef I., Carney D., Blundell T.L., Prot. Eng. 1:377–384 (1987).
12. Maiorov V.N., Crippen G.M., J.Mol.Biol. 235:625–634 (1994).
13. Kabsch W., Sander C., Biopolymers 22:2577–2637 (1983).
14. Holm L., Ouzounis C., Sander C., Tuparev G., Vriend G., Protein Science 1:1691–1698 (1992).
15. Holm L., Sander C., J. Mol. Biol. 233:123–138 (1993).
16. Stoehr P.J., Omond R.A., Nucleic Acids Res. 17:6763–6764 (1989).
17. Hobohm U., Scharf M., Schneider R., Sander C., Protein Science 3:409–417 (1992).
18. Holm L., Sander C., Proteins 19:256–268 (1994).
19. Etzold T., Argos P., CABIOS 9:49–57 (1993).

Table I. Protein fold families

family	PDB code	protein
1	1acx	ACTINOXANTHIN
1	_____1cobB	SUPEROXIDE DISMUTASE (*CO SUBSTITUTED)
2	1ten	TENASCIN (THE THIRD FIBRONECTIN TYPE III REPEAT)
2	_____2hhrB	HUMAN GROWTH HORMONE COMPLEX WITH ITS RECEPTOR
2	____2hlaA	HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN AW 68.1
2	_____4fabL	4-4-20 (IG*G2A=KAPPA=) FAB FRAGMENT - FLUORESCEIN (DIANION)
2	_____3hlaB	HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN A2.1
2	_____1fc2D	IMMUNOGLOBULIN FC AND FRAGMENT B OF PROTEIN A COMPLEX
2	_____1cid	CD4 (DOMAINS 3 AND 4)
2	_____1tlk	TELOKIN
2	_____1cd8	T CELL C0-RECEPTOR CD8
2	_____1cd4	/CD4$ (1 - 183 PLUS ASP - THR) (/D1D2$) (N-TERMINAL
2	_____1cdb	CD2 (T LYMPHOCYTE GLYCOPROTEIN, ADHESION DOMAIN)
3	1ltsC	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN, AB5 TOXIN
4	2hmgB	HEMAGGLUTININ (/G146(A)D$) (BROMELAIN DIGESTED) (MUTANT
5	1sh1	NEUROTOXIN I (SH I) (ENERGY MINIMIZED AVERAGE STRUCTURE)
6	1dfnB	DEFENSIN /HNP$-3
7	1prcM	PHOTOSYNTHETIC REACTION CENTER
8	2bp2	PROPHOSPHOLIPASE A=2=
9	1mat	METHIONINE AMINOPEPTIDASE (E.C.3.4.11.18)
10	2hhmA	HUMAN INOSITOL MONOPHOSPHATASE DIMER COMPLEXED WITH
10	_____3fbpB	FRUCTOSE-1,6-BISPHOSPHATASE (D-FRUCTOSE-1,6-BISPHOSPHATE
11	1cseI	SUBTILISIN CARLSBERG (E.C.3.4.21.14) (COMMERCIAL PRODUCT
12	2rveB	ECO RV ENDONUCLEASE COMPLEX WITH DNA
13	1lgaA	LIGNIN PEROXIDASE (LIP) (E.C.1.11.1.-) (FERRIC)
13	_____3ccp	YEAST CYTOCHROME $C PEROXIDASE (E.C.1.11.1.5) MUTANT WITH
14	1lfb	TRANSCRIPTION FACTOR LFB1 (HOMEODOMAIN)
14	_____1hddC	ENGRAILED HOMEODOMAIN COMPLEX WITH /DNA$
15	1vsgB	VARIANT SURFACE GLYCOPROTEIN (N-TERMINAL DOMAIN)
15	____1hlhA	HELIX-LOOP-HELIX DOMAIN (ONLY) FROM THE E47 PROTEIN PRODUCT
15	_____1ropA	ROP: COL*E1 REPRESSOR OF PRIMER
15	_____1fha	FERRITIN (H-CHAIN) MUTANT (LYS 86 REPLACED BY GLN) (K86Q)
15	__2tmvP	INTACT TOBACCO MOSAIC VIRUS (FIBER DIFFRACTION STUDY)
15	____1lpe	APOLIPOPROTEIN-*E3 (/LDL$ RECEPTOR BINDING DOMAIN)
15	_____1bbhB	CYTOCHROME C (PRIME)
15	256bA	CYTOCHROME $B562 (OXIDIZED)
15	_____2ccyB	CYTOCHROME $C(PRIME)
15	____2hmzA	HEMERYTHRIN (ADIZOMET)
15	_____1brd	BACTERIORHODOPSIN
16	1mrrA	MANGANESE SUBSTITUTED PROTEIN R2 OF
16	__2ztaA	/GCN4$ LEUCINE ZIPPER
16	__3inkC	INTERLEUKIN-2 MUTANT WITH CYS 125 REPLACED BY ALA (C125A)
16	_____1bgc	GRANULOCYTE COLONY STIMULATING FACTOR (RBG-CSF)
16	_____1rcb	INTERLEUKIN-4
16	_____1ifa	INTERFERON BETA (MURINE)
16	_____1gmfA	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR
17	1dsbA	DSBA (DISULFIDE BOND FORMATION PROTEIN)
17	_____1gp1A	GLUTATHIONE PEROXIDASE (E.C.1.11.1.9)
18	3trx	THIOREDOXIN (REDUCED FORM)
18	_____1gstB	ISOENZYME 3-3 OF GLUTATHIONE S-TRANSFERASE (2.5.1.18)
18	_____1ego	OXIDIZED GLUTAREDOXIN
18	_____1aba	GLUTAREDOXIN MUTANT WITH VAL 15 REPLACED BY GLY AND TYR 16
19	1gal	GLUCOSE OXIDASE EC 1.1.3.4
19	_____1trb	THIOREDOXIN REDUCTASE (E.C.1.6.4.5) MUTANT WITH CYS 138
19	_____1npx	NADH PEROXIDASE (E.C.1.11.1.1) NON-ACTIVE FORM WITH
19	_____1lpfA	LIPOAMIDE DEHYDROGENASE (E.C.1.8.1.6)
19	_____1phh	$P-*HYDROXYBENZOATE HYDROXYLASE (/PHBH$) (E.C.1.14.13.2) -
19	_____3grs	GLUTATHIONE REDUCTASE (E.C.1.6.4.2), OXIDIZED FORM (E)
20	1ipd	3-ISOPROPYLMALATE DEHYDROGENASE (E.C.1.1.1.85)
20	_____6icd	ISOCITRATE DEHYDROGENASE (E.C.1.1.1.42) (MUTANT WITH SER 113
20	__1glt	GLUTATHIONE SYNTHASE
20	____1grcB	GLYCINAMIDE RIBONUCLEOTIDE TRANSFORMYLASE (EC 2.1.2.2)
20	_____1gd1R	$HOLO-*D-*GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE
20	_____5ldh	LACTATE DEHYDROGENASE H=4= AND S-$LAC-/NAD$==+== COMPLEX
20	_____4mdhB	CYTOPLASMIC MALATE DEHYDROGENASE (E.C.1.1.1.37)
20	_____1dhr	DIHYDROPTERIDINE REDUCTASE (DHPR) (E.C.1.6.99.10) COMPLEX
20	_____1hsdA	3ALPHA,20BETA-*HYDROXYSTEROID DEHYDROGENASE (HOLO FORM)
20	_____1udpA	URIDINE DIPHOSPHOGALACTOSE 4-EPIMERASE (E.C.5.1.3.2)
20	_____1pgd	6-PHOSPHOGLUCONATE DEHYDROGENASE (6-PGDH)
20	_____8adh	APO-LIVER ALCOHOL DEHYDROGENASE (E.C.1.1.99.8)
20	_____1hmy	HHAI DNA (CYTOSINE-C5-)-METHYLTRANSFERASE (E.C.2.1.1.37)

Table I. (*cont.*)

family	PDB code	protein
20	_____1pfkA	PHOSPHOFRUCTOKINASE (E.C.2.7.1.11) (R-STATE) COMPLEX WITH
20	_____2atcA	ASPARTATE CARBAMOYLTRANSFERASE (ASPARTATE TRANSCARBAMYLASE)
20	_____1gdhA	D-GLYCERATE DEHYDROGENASE (APO FORM) (E.C.1.1.1.29)
20	_____8abp	L-*ARABINOSE-BINDING PROTEIN (MUTANT WITH MET 108 REPLACED
20	_____1dri	D-RIBOSE-BINDING PROTEIN
20	_____2gbp	D-*GALACTOSE/D-*GLUCOSE BINDING PROTEIN (/GGBP$)
20	_____2liv	LEUCINE(SLASH)*ISOLEUCINE(SLASH)*VALINE-BINDING PROTEIN
20	_____3chy	CHE*Y
20	_____2fcr	FLAVODOXIN
20	_____1fx1	FLAVODOXIN
20	_____1nipA	NITROGENASE IRON PROTEIN
20	_____1etu	ELONGATION FACTOR TU (DOMAIN I) - *GUANOSINE DIPHOSPHATE
20	_____5p21	$C-*H-RAS $P21 PROTEIN (AMINO ACIDS 1 - 166) COMPLEX WITH
20	____1minA	NITROGENASE MOLYBDENUM-IRON PROTEIN
20	_____1minB	NITROGENASE MOLYBDENUM-IRON PROTEIN
20	__1tgl	TRIACYLGLYCEROL ACYLHYDROLASE (E.C.3.1.1.3)
20	_____2sc2	SERINE CARBOXYPEPTIDASE II (E.C.3.4.16.1) (CPDW-II)
20	_____2had	HALOALKANE DEHALOGENASE ($P*H 6.2)
20	_____1ace	ACETYLCHOLINESTERASE (E.C.3.1.1.7)
20	__1tpt	THYMIDINE PHOSPHOYLASE (E.C.2.4.2.4)
20	____1ulb	PURINE NUCLEOSIDE PHOSPHORYLASE (E.C.2.4.2.1) COMPLEX WITH
20	_____3cpa	CARBOXYPEPTIDASE A=ALPHA= (COX) (E.C.3.4.17.1) COMPLEX WITH
20	_____1lap	LEUCINE AMINOPEPTIDASE (E.C.3.4.11.1)
20	____3adk	ADENYLATE KINASE (E.C.2.7.4.3)
20	_____1gky	GUANYLATE KINASE (E.C.2.7.4.8) COMPLEX WITH
21	1l84	LYSOZYME (E.C.3.2.1.17) MUTANT WITH CYS 54 REPLACED BY THR,
22	1baa	BARLEY ENDOCHITINASE (26 KD)
22	_____1lhm	LYSOZYME (E.C.3.2.1.17) (MUTANT WITH CYS 77 REPLACED BY ALA
23	4fisB	FIS PROTEIN (FACTOR FOR INVERSION STIMULATION) MUTANT
24	1wrpR	$TRP REPRESSOR (TRIGONAL FORM)
25	1snc	STAPHYLOCOCCAL NUCLEASE (E.C.3.1.31.1) COMPLEX WITH
26	1bovA	VEROTOXIN-1 (B-OLIGOMER), ALSO CALLED SHIGA-LIKE TOXIN-1
26	_____1ltsD	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN, AB5 TOXIN
27	1phs	PHASEOLIN
28	2dpv	CANINE PARVOVIRUS, STRAIN D, VIRAL PROTEIN 2
28	__2bpa1	BACTERIOPHAGE PHIX174 CAPID PROTEINS GPF, GPG, GPJ AND
28	_____1tnfA	TUMOR NECROSIS FACTOR-ALPHA (CACHECTIN)
28	____1bmv2	BEAN POD MOTTLE VIRUS (MIDDLE COMPONENT)
28	_____1bbt1	FOOT AND MOUTH DISEASE VIRUS O=1=BFS 1860 (FMDVO=1=BFS)
28	_____1r093	RHINOVIRUS 14 (/HRV$14) COMPLEX WITH ANTIVIRAL AGENT
28	_____2mev1	MENGO ENCEPHALOMYOCARDITIS VIRUS COAT PROTEIN
28	_____2tbvB	TOMATO BUSHY STUNT VIRUS
28	_____1r092	RHINOVIRUS 14 (/HRV$14) COMPLEX WITH ANTIVIRAL AGENT
28	_____2mev3	MENGO ENCEPHALOMYOCARDITIS VIRUS COAT PROTEIN
28	_____4sbvA	SOUTHERN BEAN MOSAIC VIRUS COAT PROTEIN
28	_____1rmu1	RHINOVIRUS 14 (/HRV$14) (MUTANT WITH CYS 1 199 REPLACED BY
28	_____1bmv1	BEAN POD MOTTLE VIRUS (MIDDLE COMPONENT)
28	_____2bpa2	BACTERIOPHAGE PHIX174 CAPID PROTEINS GPF, GPG, GPJ AND
28	_____2stv	SATELLITE TOBACCO NECROSIS VIRUS
28	__1lte	LECTIN COMPLEX WITH LACTOSE
28	_____1ayh	HYBRID(1−3,1−4)-BETA-D-GLUCAN-4-GLUCANOHYDROLASE H(A16-M)
28	__3hmgE	HEMAGGLUTININ (/L226(A)Q$) (BROMELAIN DIGESTED) (MUTANT
29	1gsgP	GLUTAMINYL-T/RNA$ SYNTHETASE (GLN/RS$) COMPLEX WITH
29	_____3ts1	TYROSYL-TRANSFER /RNA$ SYNTHETASE (E.C.6.1.1.1) COMPLEXED
30	8catA	CATALASE (E.C.1.11.1.6)
30	____1ifb	INTESTINAL FATTY ACID BINDING PROTEIN (APO FORM 1)
30	_____1mup	MAJOR URINARY PROTEIN COMPLEX WITH 2-(SEC-BUTYL)
30	_____1rbp	RETINOL BINDING PROTEIN
30	_____1bbpA	BILIN BINDING PROTEIN (/BBP$)
31	1ms2A	/MS$2 VIRUS (BACTERIOPHAGE)
31	__3bcl	BACTERIOCHLOROPHYLL-A PROTEIN
31	_____1omf	MATRIX PORIN (OMPF)
31	_____2por	PORIN (CRYSTAL FORM B)
32	3dpa	PAP*D
32	__4ait	TENDAMISTAT (ENERGY MINIMIZED MODEL USING '/FANTOM$')
33	1thi	THAUMATIN I
34	1pcdA	PROTOCATECHUATE 3,4-*DIOXYGENASE (E.C.1.13.11.3)
34	_____2pabB	PREALBUMIN (HUMAN PLASMA)
35	1higA	INTERFERON-GAMMA
36	1c5a	DES-ARG==74==-COMPLEMENT C5A
37	1ppt	AVIAN PANCREATIC POLYPEPTIDE

Table I. (*cont.*)

family	PDB code	protein
38	1pafA	POKEWEED ANTIVIRAL PROTEIN
38	_____1aaiA	RICIN
39	2il8A	INTERLEUKIN 8 (IL-8) (NEUTROPHIL ACTIVATION PROTEIN) /NAP$
40	1dpi	/DNA$ POLYMERASE I (KLENOW FRAGMENT) (E.C.2.7.7.7) D/CMP
41	1powA	PYRUVATE OXIDASE (E.C.1.2.3.3) (WILD TYPE)
41	____3pgk	PHOSPHOGLYCERATE KINASE (E.C.2.7.2.3) COMPLEX WITH ATP,
42	1tfi	TRANSCRIPTIONAL ELONGATION FACTOR SII (TFIIS, NUCLEIC-ACID
43	1rnt	RIBONUCLEASE T=1=(E.C.3.1.27.3) ISOZYME-2(PRIME)-GUANYLIC
43	____1rnbA	BARNASE (G SPECIFIC ENDONUCLEASE) (E.C.3.4.21.15) COMPLEX
43	_____1sarA	RIBONUCLEASE SA (E.C.3.1.4.8)
44	3fxc	FERREDOXIN
44	____2pia	PHTHALATE DIOXYGENASE REDUCTASE (E.C.1.18.1.)
44	_____1fnr	FERREDOXIN:/NADP=+=$ OXIDOREDUCTASE (FERREDOXIN REDUCTASE)
45	1ubq	UBIQUITIN
45	__2gb1	PROTEIN G (B1 DOMAIN) (/NMR$, RESTRAINED MINIMIZED AVERAGED
46	3gf1	INSULIN-LIKE GROWTH FACTOR (NMR, 10 STRUCTURES)
47	9insB	INSULIN
48	1lab	LIPOYLATED DOMAIN (RESIDUES 1-80) OF THE LIPOAMIDE
49	1f3g	PHOSPHOCARRIER III=GLC=FAST=
50	1pda	PORPHOBILINOGEN DEAMINASE (HYDROXYMETHYL BILANE
50	_____1abg	SULFATE-BINDING PROTEIN WITH SULFATE
50	_____1abh	PHOSPHATE-BINDING PROTEIN COMPLEX WITH PHOSPHATE
50	_____1omp	D-MALTODEXTRIN-BINDING PROTEIN
51	4dfrB	DIHYDROFOLATE REDUCTASE (E.C.1.5.1.3) COMPLEX WITH
51	_____3dfr	DIHYDROFOLATE REDUCTASE (E.C.1.5.1.3) COMPLEX WITH NADPH AND
51	__2reb	REC*A PROTEIN
51	____3aat	ASPARTATE AMINOTRANSFERASE (E.C.2.6.1.1) (MUTANT WITH ARG
52	1lis	LYSIN
53	3cp4	CYTOCHROME P450CAM (CAMPHOR MONOOXYGENASE) (E.C.1.14.15.1)
54	2utgA	UTEROGLOBIN
55	4tms	THYMIDYLATE SYNTHASE (E.C.2.1.1.45)
56	1pba	PROCARBOXYPEPTIDASE *B (E.C.3.4.17.2) (ACTIVATION DOMAIN)
57	1mli	MUCONOLACTONE ISOMERASE (E.C.5.3.3.4)
57	__1nrcA	PROTEIN FROM U1 SMALL NUCLEAR RIBONUCLEOPROTEIN (SNRNP U1)
57	____1aps	ACYLPHOSPHATASE (E.C.3.6.1.7) (NMR, 5 STRUCTURES)
57	____1ndk	NUCLEOSIDE DIPHOSPHATE KINASE (E.C.2.7.4.6) MUTANT WITH
57	__1tbpA	TATA-BINDING PROTEIN (TBP, C-TERMINAL 179 AMINO ACIDS)
57	____2glsA	GLUTAMINE SYNTHETASE (E.C.6.3.1.2)
58	1pgi	D-GLUCOSE 6-PHOSPHATE ISOMERASE (E.C.5.3.1.9)
59	1msbA	MANNOSE BINDING PROTEIN *A (LECTIN DOMAIN) COMPLEX WITH
60	2crd	CHARYBDOTOXIN (NMR, 12 STRUCTURES)
60	__1gps	GAMMA-1-P THIONIN (NMR, 8 MODELS)
61	3csc	CITRATE SYNTHASE (E.C.4.1.3.7)- L-MALATE - ACETYL
62	1abd	ACYL-COENZYME A BINDING PROTEIN (ACBP)
62	__1fc2C	IMMUNOGLOBULIN FC AND FRAGMENT B OF PROTEIN A COMPLEX
63	1aak	UBIQUITIN CONJUGATING ENZYME
64	1shaA	V-SRC TYROSINE KINASE TRANSFORMING PROTEIN (PHOSPHOTYROSINE
64	__1sryA	SERYL-TRNA SYNTHETASE (E.C.6.1.1.11)
64	_____1bia	BIOTIN OPERON REPRESSOR (BIRA) BIOTIN HOLOENZYME SYNTHETASE
65	1cpcA	C-PHYCOCYANIN
65	_____1cpcB	C-PHYCOCYANIN
65	_____1lh3	LEGHEMOGLOBIN (CYANO,MET)
65	_____1ecd	HEMOGLOBIN (ERYTHROCRUORIN, DEOXY)
65	_____2mba	MYOGLOBIN
65	_____1cohB	ALPHA-FERROUS-CARBONMONOXY, BETA-COBALTOUS-DEOXY HEMOGLOBIN
65	_____1mbn	MYOGLOBIN (FERRIC IRON - METMYOGLOBIN)
65	_____1colA	COLICIN *A (C-TERMINAL DOMAIN) (PORE-FORMING DOMAIN)
66	1gly	GLUCOAMYLASE (GLUCAN 1,4-ALPHA-GLUCOSIDASE)
67	1glaG	GLYCEROL KINASE (ATP:GLYCEROL PHOSPHOTRANSFERASE
67	_____1atnA	DEOXYRIBONUCLEASE I COMPLEX WITH ACTIN
67	_____2yhx	YEAST HEXOKINASE B (E.C.2.7.1.1) COMPLEX WITH
67	_____1hsc	44 K /ATP$ASE FRAGMENT (N-TERMINAL) OF 70K HEAT-SHOCK COGNATE
68	1rnh	SELENOMETHIONYL RIBONUCLEASE H (E.C.3.1.26.4)
68	__1hmi	HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 REVERSE TRANSCRIPTASE
69	1wsyB	TRYPTOPHAN SYNTHASE (E.C.4.2.1.20)
69	__3pgm	PHOSPHOGLYCERATE MUTASE (E.C.2.7.5.3) DE-PHOSPHO ENZYME
70	8atcB	ASPARTATE CARBAMOYLTRANSFERASE (ASPARTATE TRANSCARBAMYLASE)
71	1mypC	MYELOPEROXIDASE (E.C.1.11.1.7)
72	1tml	ENDO-1,4-BETA-D-GLUCANASE (E.C.3.2.1.4)
72	_____3cbh	CELLOBIOHYDROLASE /II$ CORE PROTEIN (E.C.3.2.1.91) (/CBHII$)
72	_____1pyk	PYRUVATE KINASE (E.C.2.7.1.40)

Table I. (*cont.*)

family	PDB code	protein
72	_____1cgt	CYCLODEXTRIN GLYCOSYLTRANSFERASE (E.C.2.4.1.19)
72	_____1btc	BETA-AMYLASE COMPLEXED WITH ALPHA-CYCLODEXTRIN, (ALPHA-1,
72	_____1ads	ALDOSE REDUCTASE (E.C.1.1.1.21) COMPLEX WITH NADPH
72	_____1ada	ADENOSINE DAMINASE (E.C.3.5.4.4) COMPLEX WITH
72	_____1gox	GLYCOLATE OXIDASE (E.C.1.1.3.1)
72	_____1pii	N-(5'PHOSPORIBOSYL)ANTHRANILATE ISOMERASE (E.C.5.3.1.6):
72	_____4rubB	RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE(SLASH)OXYGENASE
72	_____1ypiA	TRIOSE PHOSPHATE ISOMERASE (/TIM$) (E.C.5.3.1.1)
72	_____2taaA	TAKA-*AMYLASE A (E.C.3.2.1.1)
72	_____2rusA	RUBISCO (RIBULOSE-1,5-BISPHOSPHATE
72	_____1wsyA	TRYPTOPHAN SYNTHASE (E.C.4.2.1.20)
72	_____1ald	ALDOLASE *A (E.C.4.1.2.13)
72	_____1mle	MUCONATE LACTONIZING ENZYME (CIS,CIS MUCONATE
72	_____5xiaB	D-*XYLOSE ISOMERASE (E.C.5.3.1.5), XYLITOL COMPLEX
73	9apiA	MODIFIED ALPHA=1=-*ANTITRYPSIN
74	2act	ACTINIDIN (SULFHYDRYL PROTEINASE) (E.C.NUMBER NOT ASSIGNED)
75	1hsp	PHOSPHOLIPASE C$GAMMA (SH2 DOMAIN) (/NMR$, MINIMIZED MEAN
75	__1pnj	PHOSPHATIDYLINOSITOL 3-KINASE (P85-ALPHA SUBUNIT,
75	_____1shg	ALPHA SPECTRIN (SH3 DOMAIN)
76	1hid	HISTIDINE-CONTAINING PHOSPHOCARRIER PROTEIN HPR (NMR)
76	__1eaa	CATALYTIC DOMAIN (RESIDUES 384–637) OF DIHYDROLIPOYL
76	_____2cla	CHLORAMPHENICOL ACETYLTRANSFERASE (E.C.2.3.1.28)
77	1hstA	HISTONE H5 (GLOBULAR DOMAIN)
77	____3gapA	CATABOLITE GENE ACTIVATOR PROTEIN - CYCLIC /AMP$ COMPLEX
78	1arrA	ARC REPRESSOR
79	1cmcB	E.COLI MET HOLOREPRESSOR (METJ)
80	1prcC	PHOTOSYNTHETIC REACTION CENTER
81	8rubS	RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE(SLASH)OXYGENASE
82	1cbp	CUCUMBER BASIC PROTEIN
82	____1paz	PSEUDOAZURIN (OXIDIZED CU ++ AT $P*H 6.8)
82	_____6pcy	PLASTOCYANIN (CU1+,$P*H 3.8)
82	_____1nrd	NITRITE REDUCTASE (E.C.1.7.99.3)
82	_____2azaA	AZURIN (OXIDIZED)
83	1sil	SIALIDASE (E.C.3.2.1.18) COMPLEX WITH 2-DEOXY-2,3-
83	_____6nn9	NEURAMINIDASE N9 (E.C.3.2.1.18) (SIALIDASE) (MUTANT WITH
83	_____1nsbB	NEURAMINIDASE SIALIDASE (E.C.3.2.1.18)
84	4tln	THERMOLYSIN (E.C.3.4.24.4) COMPLEX WITH
85	1lccA	LAC REPRESSOR ('HEADPIECE') COMPLEX WITH AN 11 BASE-PAIR
85	__3croL	434 CRO PROTEIN COMPLEX WITH 20 BASE PAIR PIECE OF /DNA$
85	_____1lmbB	$LAMBDA REPRESSOR-OPERATOR COMPLEX
86	1prcH	PHOTOSYNTHETIC REACTION CENTER
87	1atnD	DEOXYRIBONUCLEASE I COMPLEX WITH ACTIN
88	2ssi	STREPTOMYCES SUBTILISIN INHIBITOR
89	2fxb	FERREDOXIN
90	1crn	CRAMBIN
91	2ca2	CARBONIC ANHYDRASE /II$ (CARBONATE DEHYDRATASE) (/HCA II$)
92	1omb	OMEGA-AGA-IVB (NMR, MINIMIZED AVERAGE STRUCTURE)
93	1cbh	C-TERMINAL DOMAIN OF CELLOBIOHYDROLASE I (/CT-CBH$ I)
94	2kaiA	KALLIKREIN A (E.C.3.4.21.8) COMPLEX WITH BOVINE PANCREATIC
94	_____4ptp	BETA TRYPSIN, DIISOPROPYLPHOSPHORYL INHIBITED
94	_____1arb	ACHROMOBACTER PROTEASE I
94	_____1p04A	ALPHA-LYTIC PROTEASE (E.C.3.4.21.12) COMPLEX WITH
94	_____1sgt	TRYPSIN (/SGT$) (E.C.3.4.21.4)
95	1croA	CRO REPRESSOR
96	1tpm	TISSUE-TYPE PLASMINOGEN ACTIVATOR (TYPE 1 FIBRIN-BINDING
97	1ixa	EGF-LIKE MODULE OF HUMAN FACTOR IX
98	4tgf	DES-VAL==1==,VAL==2==-TRANSFORMING GROWTH FACTOR ALPHA
98	__1epi	EPIDERMAL GROWTH FACTOR (EGF) IN PH 6.8 SOLUTION (/NMR$,
99	1hc6	ARTHROPODAN HEMOCYANIN (DEOXYGENATED) SUBUNIT 6 REFINED
100	1end	ENDONUCLEASE V
101	2pmgB	PHOSPHOGLUCOMUTASE (E.C.2.7.5.1)
102	1prf	PROFILIN 1A
102	__3blm	BETA-*LACTAMASE (E.C.3.5.2.6)
103	1rhd	RHODANESE (E.C.2.8.1.1)
104	1pec	PECTATE LYASE C (PLC) (E.C.4.2.2.2)
105	1avr	ANNEXIN V (RHOMBOHEDRAL)
106	5acn	ACONITASE (E.C.4.2.1.3) (INACTIVE (3FE-4S) CLUSTER FORM)
107	1rec	RECOVERIN (CALCIUM SENSOR IN VISION)
107	_____2scpA	SARCOPLASMIC CALCIUM BINDING PROTEIN
107	_____5cpv	CALCIUM-BINDING PARVALBUMIN B
107	_____5tnc	TROPONIN-*C

Table I. (*cont.*)

family	PDB code	protein
107	_____3icb	CALCIUM-BINDING PROTEIN (VITAMIN D-DEPENDENT, MINOR A FORM)
108	1fas	FASCICULIN 1
108	_____1cdtA	CARDIOTOXIN V=4===/II$== (TOXIN /III$)
108	_____3ebx	ERABUTOXIN $B
109	1cms	CHYMOSIN B (FORMERLY KNOWN AS RENNIN) (E.C.3.4.23.4)
109	_____3er3E	ENDOTHIA ASPARTIC PROTEINASE (ENDOTHIAPEPSIN)
109	____1mvpA	MYELOBLASTOSIS ASSOCIATED VIRAL PROTEASE (E.C.3.4.23)
109	_____3hvp	(/ABA$==67,95==)-/HIV$-1 PROTEASE (/SF2$ ISOLATE)
110	3sdpB	IRON SUPEROXIDE DISMUTASE (E.C.1.15.1.1)
111	1bds	/BDS-I$ (/NMR$, MINIMIZED MEAN STRUCTURE)
112	2ila	INTERLEUKIN-1*ALPHA (/IL$-1*ALPHA)
112	_____1aaiB	RICIN
112	_____1tie	ERYTHRINA TRYPSIN INHIBITOR (KUNITZ) DE-3
112	_____4i1b	INTERLEUKIN-1*BETA (/IL$-1*BETA)
113	2polA	BETA SUBUNIT OF POL III (E.C.2.7.7.7)
114	1abk	ENDONUCLEASE III (E.C.3.1.25.1) (ACS REG 60184-90-9)
115	1rbbB	RIBONUCLEASE B (E.C.3.1.27.5)
116	1pyaB	PYRUVOYL-DEPENDENT HISTIDINE DECARBOXYLASE (L-HISTIDINE
117	1bw3	BARWIN, BASIC BARLEY SEED PROTEIN, HOMOLOGOUS TO THE
118	1mon	MONELLIN
119	1pyp	INORGANIC PYROPHOSPHATASE (E.C.3.6.1.1)
120	1bbl	E3-BINDING DOMAIN OF THE DIHYDROLIPOAMIDE
121	1hleB	HORSE LEUCOCYTE ELASTASE INHIBITOR (HLEI)
121	__8apiB	MODIFIED ALPHA=1=-*ANTITRYPSIN
122	1pyaA	PYRUVOYL-DEPENDENT HISTIDINE DECARBOXYLASE (L-HISTIDINE
123	2gn5	GENE 5 /DNA$ BINDING PROTEIN
124	1tgsI	TRYPSINOGEN COMPLEX WITH PORCINE PANCREATIC SECRETORY
124	_____1choI	ALPHA-CHYMOTRYPSIN (E.C.3.4.21.1) COMPLEX WITH TURKEY
125	1eps	5-ENOL-PYRUVYL-3-PHOSPHATE SYNTHASE (E.C.2.5.1.9)
126	1fkf	/FK506$ BINDING PROTEIN (/FKBP$) COMPLEX WITH
127	1erp	PHEROMONE ER-10 (NMR, 20 MODELS)
128	1pi2	BOWMAN-*BIRK PROTEINASE INHIBITOR /PI-II$
129	1cpl	CYCLOPHILIN
130	1c2rA	CYTOCHROME $C=2=
130	_____1ycc	CYTOCHROME C (ISOZYME 1) (REDUCED)
130	____1c53	CYTOCHROME C553
130	_____451c	CYTOCHROME $C=551= (REDUCED)
130	_____1cc5	CYTOCHROME C=5= (OXIDIZED)
131	1ltsA	HEAT-LABILE ENTEROTOXIN (LT); CHOLERA-LIKE TOXIN, AB5 TOXIN
132	1cy3	CYTOCHROME $C=3=
133	1hcc	16TH COMPLEMENT CONTROL PROTEIN (/CCP$) OF FACTOR H
134	1mhu	CD-7 METALLOTHIONEIN-2 (ALPHA DOMAIN) (/NMR$)
135	2mrt	CD-7 METALLOTHIONEIN-2 (BETA DOMAIN) (/NMR$)
136	4sgbI	SERINE PROTEINASE B COMPLEX WITH THE POTATO INHIBITOR
137	8pti	BOVINE PANCREATIC TRYPSIN INHIBITOR (/BPTI$) MUTANT (TYR 35
138	2mev4	MENGO ENCEPHALOMYOCARDITIS VIRUS COAT PROTEIN
139	6hir	HIRUDIN (MUTANT WITH LYS 47 REPLACED BY GLU) (/K47E$)
140	7wgaB	WHEAT GERM AGGLUTININ (ISOLECTIN 1)
141	4cpaI	CARBOXYPEPTIDASE A=ALPHA= (COX) (E.C.3.4.17.1) COMPLEX WITH
142	1isuA	HIGH-POTENTIAL IRON-SULFUR PROTEIN (HIPIP)
142	_____2hipB	HIGH POTENTIAL IRON SULFUR PROTEIN (HI/PIP$)

Structural classification of protein chains in the database of three-dimensional structures (PDB). The sequential index of the fold family is followed by PDB and chain identifiers and protein names of a family member. Family 1 has 2 members (1acx, 1cobB), family 2 has 11 members (1ten, 2hhrB, ...) and so on. Indentation in the 'PDB code' column means that a protein belongs to the same family/subfamily as the protein above. The families are defined by cutting an average linkage clustering tree at a similarity level of 2 standard deviations above expected (Z = 2). Subfamilies are defined by cuts at similarity levels of Z = 3, 4, 5, 6 and 10; more refined family divisions can be made at each level of similarity. For example, 3dpa and 4ait of family 32 are split in two separate families if the cut is made at Z = 3 rather than at Z = 2; 1acx and 1cobB (family 1) end up in different families if a cut is made at Z=5; 2hhmA and 3fbpB (family 10) stay together even at Z = 10. Only chains in the sequence-representative set (maximally 30% sequence identity) are reported here; higher than 30% sequence identity between homologous proteins implies, in general, structural similarity that would be far off the scale to the right.