

PROSITE: a dictionary of sites and patterns in proteins

BAIROCH, Amos Marc

Abstract

PROSITE is a compilation of sites and patterns found in protein sequences. The use of protein sequence patterns (or motifs) to determine the function of proteins is becoming very rapidly one of the essential tools of sequence analysis. This reality has been recognized by many authors. While there have been a number of recent reports that review published patterns, no attempt had been made until very recently [5,6] to systematically collect biologically significant patterns or to discover new ones. It is for these reasons that we have developed, since 1988, a dictionary of sites and patterns which we call PROSITE. Some of the patterns compiled in PROSITE have been published in the literature, but the majority have been developed, in the last two years, by the author.

Reference

BAIROCH, Amos Marc. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic acids research*, 1992, vol. 20 Suppl, p. 2013-8

DOI : 10.1093/nar/20.suppl.2013

PMID : 1598232

Available at:

<http://archive-ouverte.unige.ch/unige:36259>

Disclaimer: layout of this document may differ from the published version.



UNIVERSITÉ
DE GENÈVE

PROSITE: a dictionary of sites and patterns in proteins

Amos Bairoch

Department of Medical Biochemistry, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

INTRODUCTION

PROSITE is a compilation of sites and patterns found in protein sequences. The use of protein sequence patterns (or motifs) to determine the function of proteins is becoming very rapidly one of the essential tools of sequence analysis. This reality has been recognized by many authors [1,2]. While there have been a number of recent reports [3,4] that review published patterns, no attempt had been made until very recently [5,6] to systematically collect biologically significant patterns or to discover new ones. It is for these reasons that we have developed, since 1988, a dictionary of sites and patterns which we call PROSITE.

Some of the patterns compiled in PROSITE have been published in the literature, but the majority have been developed, in the last two years, by the author.

FORMAT

The PROSITE database is composed of two ASCII (text) files. The first file (PROSITE.DAT) is a computer-readable file that contains all the information necessary to programs that make use of PROSITE to scan sequence(s) with pattern(s). This file also includes, for each of the patterns described, statistics on the number of hits obtained while scanning for that pattern in the SWISS-PROT protein sequence data bank [7]. Cross-references to the corresponding SWISS-PROT entries are also present in that file. The second file (PROSITE.DOC), which we call the textbook, contains textual information that documents each pattern. A user manual (PROSUSER.TXT) is distributed with the database, it fully describes the format of both files. A sample textbook entry is shown in Figure 1 with the corresponding data from the pattern file.

LEADING CONCEPTS

The design of PROSITE follows four leading concepts:

Completeness. For such a compilation to be helpful in the determination of protein function, it is important that it contains a significant number of biologically meaningful patterns.

High specificity of the patterns. In the majority of cases we have chosen patterns that are specific enough not to detect too many unrelated sequences, yet that detect most if not all sequences that clearly belong to the set in consideration.

Documentation. Each of the patterns is fully documented; the documentation includes a concise description of the family of protein that it is supposed to detect as well as an explanation on the reasons that led to the selection of the particular pattern.

Periodic reviewing. It is important that each pattern be periodically reviewed, so as to insure that it is still relevant.

CONTENT OF THE CURRENT RELEASE

Release 8.10 of PROSITE (March 1992) contains 530 documentation entries describing 605 different patterns. The list of these entries is provided in Appendix 1.

DISTRIBUTION

PROSITE is distributed on magnetic tape and on CD-ROM by the EMBL Data Library. For all enquiries regarding the subscription and distribution of PROSITE one should contact:

EMBL Data Library
European Molecular Biology Laboratory
Postfach 10.2209, Meyerhofstrasse 1
6900 Heidelberg, Germany
Telephone: (+49 6221) 387 258
Telefax : (+49 6221) 387 519 or 387 306
Electronic network address: datalib@EMBL-heidelberg.de

PROSITE can be obtained from the EMBL File Server [8]. Detailed instructions on how to make best use of this service, and in particular on how to obtain PROSITE, can be obtained by sending to the network address netserv@EMBL-heidelberg.de the following message:

HELP
HELP PROSITE

If you have access to a computer system linked to the Internet you can obtain PROSITE using FTP (File Transfer Protocol), from the following file servers:

GenBank On-line Service [9]
Internet address: genbank.bio.net (134.172.1.160)
NCBI (National Library of Medicine, NIH, Washington D.C., U.S.A.)
Internet address: ncbi.nlm.nih.gov (130.14.20.1)
Swiss EMBnet FTP server (Biozentrum, Basel, Switzerland)
Internet address: bioftp.unibas.ch (131.152.1.7)
ExPASy (Expert Protein Analysis System server, University of Geneva, Switzerland)
Internet address: expasy.hcuge.ch (129.195.254.61)

The present distribution frequency is four releases per year. No restrictions are placed on use or redistribution of the data.

REFERENCES

1. Doolittle R.F. (In) Of URFs and ORFs: a primer on how to analyze derived amino acid sequences., University Science Books, Mill Valley, California, (1986).
2. Lesk A.M. (In) Computational Molecular Biology, Lesk A.M., Ed., pp17–26, Oxford University Press, Oxford (1988).
3. Barker W.C., Hunt T.L., George D.G. Protein Seq. Data Anal. 1:363–373(1988).
4. Hodgman T.C. CABIOS 5:1–13(1989).
5. Bork P. FEBS Lett. 257:191–195(1989).
6. Smith H.O., Annau T.M., Chandrasegaran S. Proc. Natl. Acad. Sci. USA 87:826–830(1990).
7. Bairoch A., Boeckmann B. Nucleic Acids Res. 20:2019–2022(1992).
8. Stoehr P.J., Omund R.A. Nucleic Acids Res. 17:6763–6764(1989).
9. Benton D. Nucleic Acids Res. 18:1517–1520(1990).

a

```

(PDOC00107)
(PS00116; DNA_POLYMERASE_B)
(BEGIN)
*****
* DNA polymerase family B signature *
*****

```

Replicative DNA polymerases (EC 2.7.7.7) are the key enzymes catalyzing the accurate replication of DNA. They require either a small RNA molecule or a protein as a primer for the de novo synthesis of a DNA chain. On the basis of sequence similarities a number of DNA polymerases have been grouped together [1 to 6] under the designation of DNA polymerase family B. The polymerases that belong to this family are:

- Human polymerase alpha.
- Yeast polymerase I/alpha (gene POL1), polymerase II/epsilon (gene POL2), polymerase III/delta (gene POL3), and polymerase REV3.
- Escherichia coli polymerase III (gene *dinA* or *polB*).
- Polymerases of viruses from the herpesviridae family (Herpes type I and II, Epstein-Barr, Cytomegalovirus, and Varicella-zoster).
- Polymerases from Adenoviruses.
- Polymerases from Baculoviruses.
- Polymerases from Poxviruses (Vaccinia virus and Fowlpox virus).
- Bacteriophage T4 polymerase.
- Podoviridae bacteriophages Phi-29, M2, and PZA polymerase.
- Tectiviridae bacteriophage PRD1 polymerase.
- Polymerases encoded on linear DNA plasmids: Kluyveromyces lactis pGKL1 and pGKL2, Ascobolus immersus pAI2, and Claviceps purpurea pCLK1.
- Putative polymerase from the maize mitochondrial plasmid-like S1 DNA.

Six regions of similarity (numbered from I to VI) are found in all or a subset of the above polymerases. The most conserved region (I) includes a perfectly conserved tetrapeptide which contains two aspartate residues. The function of this conserved region is not yet known, however it has been suggested [3] that it may be involved in binding a magnesium ion. We use this conserved region as a signature for this family of DNA polymerases.

- Consensus pattern: [YA]-x-D-T-D-S-[LIVMT]
- Sequences known to belong to this class detected by the pattern: ALL, except for yeast polymerase II/epsilon which has Glu instead of Tyr/Ala and has Gly instead of Ser, and for Ascobolus immersus plasmid pAI2 which also has Gly instead of Ser
- Other sequence(s) detected in SWISS-PROT: chicken vitellogenin 2.
- Note: the residue in position 1 is Tyr in every family B polymerases, except in phage T4, where it is Ala.
- Last update: December 1991 / Text revised.

- [1] Jung G., Leavitt M.C., Hsieh J.-C., Ito J. Proc. Natl. Acad. Sci. U.S.A. 84:8287-8291(1987).
- [2] Bernad A., Zaballos A., Salas M., Blanco L. EMBO J. 6:4219-4225(1987).
- [3] Argos P. Nucleic Acids Res. 16:9909-9916(1988).
- [4] Wang T.S.-F., Wong S.W., Korn D. FASEB J. 3:14-21(1989).
- [5] Delarue M., Poch O., Todro N., Moras D., Argos P. Protein Engineering 3:461-467(1990).
- [6] Ito J., Braithwaite D.K. Nucleic Acids Res. 19:4045-4057(1991).

(END)

b

```

ID  DNA POLYMERASE_B; PATTERN.
AC  PS00116;
DT  APR-1990 (CREATED); NOV-1990 (DATA UPDATE); DEC-1991 (INFO UPDATE).
DE  DNA polymerase family B signature.
PA  [YA]-x-D-T-D-S-[LIVMT].
NR  /RELEASE=20,22654;
NR  /TOTAL=31(31); /POSITIVE=30(30); /UNKNOWN=0(0); /FALSE_POS=1(1); /FALSE_NEG=2(2);
CC  /TAXO-RANGE=?BEPV; /MAX-REPEAT=1;
DR  P09884, DPOA_HUMAN, T; P13382, DPOA_YEAST, T; P15436, DPOD_YEAST, T;
DR  P14284, DPOX_YEAST, T; P21189, DPO2_ECOLI, T; P03261, DPOL_ADE02, T;
DR  P04495, DPOL_ADE05, T; P05664, DPOL_ADE07, T; P06538, DPOL_ADE12, T;
DR  P08546, DPOL_HCMVA, T; P03198, DPOL_EBV, T; P04293, DPOL_HSV11, T;
DR  P07917, DPOL_HSV1A, T; P04292, DPOL_HSV1K, T; P09854, DPOL_HSV1S, T;
DR  P07918, DPOL_HSV21, T; P09252, DPOL_VZVD, T; P09804, DPOL_KLULA, T;
DR  P05468, DPO2_KLULA, T; P22373, DPOL_CLAPU, T; P10582, DPOL_MAIZE, T;
DR  P18131, DPOL_NPVAC, T; P20509, DPOL_VACCC, T; P06856, DPOL_VACCV, T;
DR  P21402, DPOL_FQWPV, T; P03680, DPOL_BPPH2, T; P06950, DPOL_BPPZA, T;
DR  P19894, DPOL_BPM2, T; P10479, DPOL_BPPRD, T; P04415, DPOL_BPT4, T;
DR  P22374, DPOL_ASCIM, N; P21951, DPOE_YEAST, N;
DR  P02845, VIT2_CHICK, F;
DO  PDOC00107;
//

```

Figure 1. Sample data from PROSITE. a. A documentation (textbook) entry. b. The corresponding entry in the pattern file.

Appendix 1. List of patterns documentation entries in release 8.10 of PROSITE

Post-translational modifications

N-glycosylation site
 Glycosaminoglycan attachment site
 Tyrosine sulfation site
 cAMP- and cGMP-dependent protein kinase phosphorylation site
 Protein kinase C phosphorylation site
 Casein kinase II phosphorylation site
 Tyrosine kinase phosphorylation site
 N-myristoylation site
 Amidation site
 Aspartic acid and asparagine hydroxylation site
 Vitamin K-dependent carboxylation domain
 Phosphopantetheine attachment site
 Prokaryotic membrane lipoprotein lipid attachment site
 Farnesyl group binding site (CAAX box)

Domains

Endoplasmic reticulum targeting sequence
 Peroxisomal targeting sequence
 Gram-positive cocci surface proteins 'anchoring' hexapeptide
 Nuclear targeting sequence
 Cell attachment sequence
 ATP/GTP-binding site motif A (P-loop)
 EF-hand calcium-binding domain
 Actinin-type actin-binding domain signatures
 Cofilin/tropomyosin-type actin-binding domain
 Apple domain
 Kringle domain signature
 EGF-like domain cysteine pattern signature
 Fibrinogen beta and gamma chains C-terminal domain signature
 Type II fibronectin collagen-binding domain
 Hemopexin domain signature
 Somatomedin B domain signature
 Thyroglobulin type-I repeat signature
 'Trefoil' domain signature
 Cellulose-binding domain, bacterial type
 Cellulose-binding domain, fungal type
 Chitin recognition or binding domain signature
 WAP-type 'four-disulfide core' domain signature
 Phorbol esters / diacylglycerol binding domain
 C2 domain signature

DNA or RNA associated proteins

'Homeobox' domain signature
 'Homeobox' antennapedia-type protein signature
 'Homeobox' engrailed-type protein signature
 'Paired box' domain signature
 'POU' domain signatures
 Zinc finger, C2H2 type, domain
 Zinc finger, C3HC4 type, signature
 Nuclear hormones receptors DNA-binding region signature
 GATA-type zinc finger domain
 Poly(ADP-ribose) polymerase zinc finger domain
 Fungal Zn(2)-Cys(6) binuclear cluster domain
 Leucine zipper pattern
 Fos/jun DNA-binding basic domain signature
 Myb DNA-binding domain repeat signatures
 Myc-type, 'helix-loop-helix' putative DNA-binding domain signature
 p53 tumor antigen signature
 'Cold-shock' DNA-binding domain signature
 CTF/NF-I signature
 Ets-domain signatures
 HSF-type DNA-binding domain signature
 IRF family signature
 LIM domain signature
 SRF-type transcription factors DNA-binding and dimerization domain
 TEA domain signature
 Transcription factor TFIID repeat signature
 TFIIS cysteine-rich domain signature
 DEAD-box family ATP-dependent helicases signature
 Eukaryotic putative RNA-binding region RNP-I signature
 Fibrillarin signature

Bacterial regulatory proteins, araC family signature
 Bacterial regulatory proteins, asnC family signature
 Bacterial regulatory proteins, crp family signature
 Bacterial regulatory proteins, gntR family signature
 Bacterial regulatory proteins, lacI family signature
 Bacterial regulatory proteins, lysR family signature
 Bacterial regulatory proteins, merR family signature
 Bacterial histone-like DNA-binding proteins signature
 Histone H2A signature
 Histone H2B signature
 Histone H3 signature
 Histone H4 signature
 HMG1/2 signature
 HMG-I and HMG-Y DNA-binding domain (A+T-hook)
 HMG14 and HMG17 signature
 Chromo domain
 Protamine P1 signature
 Nuclear transition protein 1 signature
 Ribosomal protein L2 signature
 Ribosomal protein L3 signature
 Ribosomal protein L5 signature
 Ribosomal protein L6 signature
 Ribosomal protein L11 signature
 Ribosomal protein L14 signature
 Ribosomal protein L15 signature
 Ribosomal protein L16 signature
 Ribosomal protein L22 signature
 Ribosomal protein L23 signature
 Ribosomal protein L29 signature
 Ribosomal protein L33 signature
 Ribosomal protein L19e signature
 Ribosomal protein L32e signature
 Ribosomal protein L46e signature
 Ribosomal protein S3 signature
 Ribosomal protein S5 signature
 Ribosomal protein S7 signature
 Ribosomal protein S8 signature
 Ribosomal protein S9 signature
 Ribosomal protein S10 signature
 Ribosomal protein S11 signature
 Ribosomal protein S12 signature
 Ribosomal protein S14 signature
 Ribosomal protein S15 signature
 Ribosomal protein S17 signature
 Ribosomal protein S18 signature
 Ribosomal protein S19 signature
 Ribosomal protein S4e signature
 Ribosomal protein S6e signature
 Ribosomal protein S24e signature
 DNA mismatch repair proteins mutL / hexB / PMS1 signature
 DNA mismatch repair proteins mutS / hexA / Duc1 / Rep1 signature

Enzymes**Oxidoreductases**

Zinc-containing alcohol dehydrogenases signature
 Iron-containing alcohol dehydrogenases signature
 Short-chain alcohol dehydrogenase family signature
 Aldo/keto reductase family signatures
 L-lactate dehydrogenase active site
 Glycerate-type 2-hydroxyacid dehydrogenases signature
 Hydroxymethylglutaryl-coenzyme A reductases signatures
 3-hydroxyacyl-CoA dehydrogenase signature
 Malate dehydrogenase active site signature
 Malic enzymes signature
 Isocitrate and isopropylmalate dehydrogenases signature
 6-phosphogluconate dehydrogenase signature
 Glucose-6-phosphate dehydrogenase active site
 IMP dehydrogenase / GMP reductase signature
 Bacterial quinoprotein dehydrogenases signatures
 FMN-dependent alpha-hydroxy acid dehydrogenases active site
 Eukaryotic molybdopterin oxidoreductases signature
 Prokaryotic molybdopterin oxidoreductases signatures

Aldehyde dehydrogenases active site
 Glyceraldehyde 3-phosphate dehydrogenase active site
 Fumarate reductase / succinate dehydrogenase FAD-binding site
 Acyl-CoA dehydrogenases signatures
 Glutamate / Leucine / Phenylalanine dehydrogenases active site
 Delta 1-pyrroline-5-carboxylate reductase signature
 Dihydrofolate reductase signature
 Pyridine nucleotide-disulphide oxidoreductases class-I active site
 Pyridine nucleotide-disulphide oxidoreductases class-II active site
 Respiratory chain NADH dehydrogenase 30 Kd subunit signature
 Respiratory chain NADH dehydrogenase 49 Kd subunit signature
 Nitrite reductases and sulfite reductase putative siroheme-binding sites
 Uricase signature
 Cytochrome c oxidase subunit I, copper B binding region signature
 Cytochrome c oxidase subunit II, copper A binding region signature
 Multicopper oxidases signatures
 Peroxidases signatures
 Catalase signatures
 Glutathione peroxidase selenocysteine active site
 Lipoygenases, putative iron-binding region signature
 Extradiol ring-cleavage dioxygenases signature
 Intradiol ring-cleavage dioxygenases signature
 Bacterial ring hydroxylating dioxygenases alpha-subunit signature
 Bacterial luciferase subunits signature
 Bioppterin-dependent aromatic amino acid hydroxylases signature
 Copper type II, ascorbate-dependent monooxygenases signatures
 Tyrosinase signatures
 Fatty acid desaturases signatures
 Cytochrome P450 cysteine heme-iron ligand signature
 Heme oxygenase signature
 Copper/Zinc superoxide dismutase signatures
 Manganese and iron superoxide dismutases signature
 Ribonucleotide reductase large subunit signature
 Ribonucleotide reductase small subunit signature
 Nitrogenases component 1 alpha and beta subunits signature
 Nickel-dependent hydrogenases large subunit signatures

Transferases

Thymidylate synthase active site
 Methylated-DNA--protein-cysteine methyltransferase active site
 N-6 Adenine-specific DNA methylases signature
 N-4 cytosine-specific DNA methylases signature
 C-5 cytosine-specific DNA methylases signatures
 Serine hydroxymethyltransferase pyridoxal-phosphate attachment site
 Phosphoribosylglycinamide formyltransferase active site
 Aspartate and ornithine carbamoyltransferases signature
 Acyltransferases ChoActase / COT / CPT-II family signatures
 Thiolases signatures
 Chloramphenicol acetyltransferase active site
 cysE / lacA / nifP / nodL acetyltransferases signature
 Beta-ketoacyl synthases active site
 Chalcone and resveratrol synthases active site
 Gamma-glutamyltranspeptidase signature
 Transglutaminases active site
 Phosphorylase pyridoxal-phosphate attachment site
 UDP-glucuronosyl and UDP-glucosyl transferases signature
 Purine/pyrimidine phosphoribosyl transferases signature
 Glutamine amidotransferases class-I active site
 Glutamine amidotransferases class-II active site
 S-Adenosylmethionine synthetase signatures
 Polyphenyl synthetases signature
 EPSP synthase active site
 Aspartate aminotransferases pyridoxal-phosphate attachment site
 Aminotransferases class-II pyridoxal-phosphate attachment site
 Aminotransferases class-III pyridoxal-phosphate attachment site
 Phosphoserine aminotransferase signature
 Hexokinases signature
 Galactokinase signature
 Phosphofructokinase signature
 pfkB family prokaryotic carbohydrate kinases signatures
 Phosphoribulokinase signature
 Thymidine kinase cellular-type signature
 Prokaryotic carbohydrate kinases signature
 Protein kinases signatures
 Pyruvate kinase active site signature

Phosphoglycerate kinase signature
 Aspartokinase signature
 ATP:guanido phosphotransferases active site
 PTS Hpr component phosphorylation sites signatures
 PTS permeases phosphorylation sites signatures
 Adenylate kinase signature
 Nucleoside diphosphate kinases active site
 Phosphoribosyl pyrophosphate synthetase signature
 Bacteriophage-type RNA polymerase family active site signature
 Eukaryotic RNA polymerase II heptapeptide repeat
 Eukaryotic RNA polymerases 30 to 40 Kd subunits signature
 DNA polymerase family A signature
 DNA polymerase family B signature
 DNA polymerase family X signature
 Galactose-1-phosphate uridyl transferase active site signature
 CDP-alcohol phosphatidyltransferases signature
 PEP-utilizing enzymes phosphorylation site signature
 Rhodanese active site
 Hydrolases
 Phospholipase A2 active sites signatures
 Lipases, serine active site
 Colipase signature
 Carboxylesterases type-B active site
 Pectinesterase signature
 Alkaline phosphatase active site
 Fructose-1-6-bisphosphatase active site
 Serine/threonine specific protein phosphatases signature
 Tyrosine specific protein phosphatases active site
 Prokaryotic zinc-dependent phospholipase C signature
 3'-5'-cyclic nucleotide phosphodiesterases signature
 cAMP phosphodiesterases class-II signature
 Sulfatases signatures
 Ribonuclease III family signature
 Ribonuclease T2 family histidine active sites
 Pancreatic ribonuclease family signature
 Beta-amylase signature
 Polygalacturonase signature
 Clostridium cellulases repeated domain signature
 Alpha-lactalbumin / lysozyme C signature
 Lysosomal alpha-glucosidase / sucrase-isomaltase active site
 Alpha-galactosidase signature
 Alpha-L-fucosidase putative active site
 Glycosyl hydrolases family 1 active site
 Glycosyl hydrolases family 9 active site
 Glycosyl hydrolases family 10 active site
 Glycosyl hydrolases family 17 signature
 Alkylbase DNA glycosidases alkA family signature
 Uracil-DNA glycosylase signature
 Aminopeptidase P and proline dipeptidase signature
 Serine carboxypeptidases, active sites
 Zinc carboxypeptidases, zinc-binding regions signatures
 Serine proteases, trypsin family, active sites
 Serine proteases, subtilase family, active sites
 ClpP proteases active sites
 Eukaryotic thiol (cysteine) proteases active site
 Ubiquitin carboxyl-terminal hydrolase, putative active-site signature
 Eukaryotic aspartyl proteases active site
 Neutral zinc metallopeptidases, zinc-binding region signature
 Matrixins cysteine switch
 Insulinase family signature
 recA signature
 Proteasome subunits signature
 Signal peptidase complex SPC21/SPC18 subunits signature
 Amidases signature
 Asparaginase / glutaminase active site
 Urease active site
 Dihydroorotase signatures
 Beta-lactamases classes -A, -C, and -D active site
 Arginase and agmatinase signatures
 Adenosine and AMP deaminase signature
 Inorganic pyrophosphatase signature
 Acylphosphatase signatures
 ATP synthase alpha and beta subunits signature
 ATP synthase gamma subunit signature
 ATP synthase delta (OSCP) subunit signature
 ATP synthase a subunit signature

ATP synthase c subunit signature
E1-E2 ATPases phosphorylation site
Sodium and potassium ATPases beta subunits signatures
Cutinase, serine active site

Lyases

DDC / GAD / HDC pyridoxal-phosphate attachment site
Orotidine 5'-phosphate decarboxylase signature
Phosphoenolpyruvate carboxylase active site
Phosphoenolpyruvate carboxykinase (GTP) signature
Phosphoenolpyruvate carboxykinase (ATP) signature
Ribulose biphosphate carboxylase large chain active site
Fructose-biphosphate aldolase class-I active site
Fructose-biphosphate aldolase class-II signature
Malate synthase signature
Citrate synthase signature
KDPG and KHG aldolases active site signatures
Isocitrate lyase signature
DNA photolyases signature
Carbonic anhydrases signature
Fumarate lyases signature
Aconitase family signature
Enolase signature
Serine/threonine dehydratases pyridoxal-phosphate attachment site
Enoyl-CoA hydratase/isomerase signature
Tryptophan synthase alpha chain signature
Tryptophan synthase beta chain pyridoxal-phosphate attachment site
Delta-aminolevulinic acid dehydratase active site
Phenylalanine and histidine ammonia-lyases signature
Porphobilinogen deaminase cofactor-binding site
Guanylate cyclases signature
Ferrochelatase signature

Isomerases

Alanine racemase pyridoxal-phosphate attachment site
Aldose 1-epimerase putative active site
Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature
FKBP-type peptidyl-prolyl cis-trans isomerase signatures
Triosephosphate isomerase active site
Xylose isomerase signatures
Phosphoglucose isomerase signature
Phosphoglycerate mutase family phosphohistidine signature
Methylmalonyl-CoA mutase signature
Eukaryotic DNA topoisomerase I active site
Prokaryotic DNA topoisomerase I active site
DNA topoisomerase II signature

Ligases

Aminoacyl-transfer RNA synthetases class-I signature
Aminoacyl-transfer RNA synthetases class-II signatures
ATP-citrate lyase and succinyl-CoA ligases active site
Glutamine synthetase signatures
Ubiquitin-activating enzyme signature
Ubiquitin-conjugating enzymes active site
Adenylosuccinate synthetase active site
Argininosuccinate synthase signatures
Phosphoribosylglycinamide synthetase signature
ATP-dependent DNA ligase putative active site

Others

Isopenicillin N synthetase signatures
Site-specific recombinases signatures
Thiamine pyrophosphate enzymes signature
Biotin-requiring enzymes attachment site
2-oxo acid dehydrogenases acyltransferase component lipoyl binding site
Putative AMP-binding domain signature

Electron transport proteins

Cytochrome c family heme-binding site signature
Cytochrome b5 family, heme-binding domain signature
Cytochrome b/b6 signatures
Cytochrome b559 subunits heme-binding site signature
Thioredoxin family active site
Glutaredoxin active site

Type-1 copper (blue) proteins signature
2Fe-2S ferredoxins, iron-sulfur binding region signature
4Fe-4S ferredoxins, iron-sulfur binding region signature
High potential iron-sulfur proteins signature
Rieske iron-sulfur protein signatures
Flavodoxin signature
Rubredoxin signature

Other transport proteins

Class I metallothioneins signature
Ferritin iron-binding regions signatures
Bacterioferritin signature
Transferrins signatures
Plant hemoglobins signature
Hemerythrins signature
Arthropod hemocyanins / insect LSPs signatures
ATP-binding proteins 'active transport' family signature
Binding-protein-dependent transport systems inner membrane component signature
Serum albumin family signature
Avidin / Streptavidin family signature
Eukaryotic cobalamin-binding proteins signature
Lipocalin signature
Cytosolic fatty-acid binding proteins signature
LBP / BPI / CETP family signature
Plant lipid transfer proteins signature
Uteroglobin family signatures
Mitochondrial energy transfer proteins signature
Sugar transport proteins signatures
Sodium symporters signatures
Prokaryotic sulfate-binding proteins signature
Amino acid permeases signature
Aromatic amino acids permeases signature
Anion exchangers family signatures
MIP family signature
General diffusion gram-negative porins signature
Eukaryotic porin signature
Insulin-like growth factor binding proteins signature

Structural proteins

43 Kd postsynaptic protein signature
Actins signatures
Annexins repeated domain signature
Clathrin light chains signatures
Clusterin signatures
Connexins signatures
Crystallins beta and gamma 'Greek key' motif signature
Dynamin family signature
Intermediate filaments signature
Kinesin motor domain signature
Myelin basic protein signature
Myelin P0 protein signature
Myelin proteolipid protein signature
Neuromodulin (GAP-43) signatures
Profilin signature
Surfactant associated polypeptide SP-C palmytoylation sites
Synapsins signatures
Synaptobrevin signature
Synaptophysin / synaptoporin signature
Tropomyosins signature
Tubulin subunits alpha, beta, and gamma signature
Tubulin-beta mRNA autoregulation signal
Tau and MAP2 proteins repeated region signature
Neuraxin and MAPIB proteins repeated region signature
F-actin capping protein beta subunit signature
Amyloidogenic glycoprotein signatures
Cadherins extracellular repeated domain signature
Insect flexible cuticle proteins signature
Gas vesicles protein GVPa signature
Gas vesicles protein GVPC repeated domain signature
Flagella basal body rod proteins signature
Type-4 pili methylation site
Plant viruses icosahedral capsid proteins 'S' region signature
Potexviruses and carlaviruses coat protein signature

Receptors

Neurotransmitter-gated ion-channels signature
 G-protein coupled receptors signature
 Visual pigments (opsins) retinal binding site
 Bacterial rhodopsins retinal binding site
 Receptor tyrosine kinase class II signature
 Receptor tyrosine kinase class III signature
 Growth factor and cytokines receptors family signatures
 Integrins alpha chain signature
 Integrins beta chain cysteine-rich domain signature
 Natriuretic peptides receptors signature
 Photosynthetic reaction center proteins signature
 Photosystem I psaA and psaB proteins signature
 Phytochrome chromophore attachment site
 Speract receptor repeated domain signature
 TonB-dependent receptor proteins signature
 Type-II membrane antigens family signature
 Bacterial chemotaxis sensory transducers signature

Cytokines and growth factors

HBGF/FGF family signature
 Nerve growth factor family signature
 Platelet-derived growth factor (PDGF) family signature
 Small cytokines signatures
 TGF-beta family signature
 TNF family signature
 Wnt-1 family signature
 Interferon alpha and beta family signature
 Interleukin-1 signature
 Interleukin-2 signature
 Interleukin-6 / G-CSF / MGF family signature
 Interleukin-7 signature
 Interleukin-10 signature
 LIF / OSM family signature

Hormones and active peptides

Adipokinetic hormone family signature
 Bombesin-like peptides family signature
 Calcitonin / CGRP / IAPP family signature
 Corticotropin-releasing factor family signature
 Granins signatures
 Gastrin / cholecystokinin family signature
 Glucagon / GIP / secretin / VIP family signature
 Glycoprotein hormones beta chain signature
 Gonadotropin-releasing hormones signature
 Insulin family signature
 Natriuretic peptides signature
 Neurohypophysial hormones signature
 Pancreatic hormone family signature
 Parathyroid hormone family signature
 Pyrokinins signature
 Somatotropin, prolactin and related hormones signatures
 Tachykinin family signature
 Thymosin beta-4 family signature
 Cecropin family signature
 Mammalian defensins signature
 Insect defensins signature
 Endothelins / sarafotoxins signature

Toxins

Plant thionins signature
 Snake toxins signature
 Myotoxins signature

Heat-stable enterotoxins signature
 Aerolysin type toxins signature
 Shiga/ricin ribosomal inactivating toxins active site signature
 Channel forming colicins signature
 Hok/gef family cell toxic proteins signature
 Staphylococcal enterotoxins / Streptococcal pyrogenic exotoxins signatures
 Thiol-activated cytotoxins signature
 Membrane attack complex components / perforin signature

Inhibitors

Pancreatic trypsin inhibitor (Kunitz) family signature
 Bowman-Birk serine protease inhibitors family signature
 Kazal serine protease inhibitors family signature
 Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature
 Serpins signature
 Potato inhibitor I family signature
 Squash family of serine protease inhibitors signature
 Cysteine proteases inhibitors signature
 Tissue inhibitors of metalloproteinases signature
 Cereal trypsin/alpha-amylase inhibitors family signature
 Alpha-2-macroglobulin family thiolester region signature
 Disintegrins signature
 Lambdoid phages regulatory protein CIII signature

Others

Pentraxin family signature
 Immunoglobulins and major histocompatibility complex proteins signature
 Prion protein signature
 Cyclins signature
 Proliferating cell nuclear antigen signature
 Arrestins signature
 Chaperonins signature
 Heat shock hsp70 proteins family signatures
 Heat shock hsp90 proteins family signature
 Ubiquitin signature
 GTPase-activating proteins signature
 Stathmin family signature
 SRP54-type proteins GTP-binding domain signature
 GTP-binding elongation factors signature
 Eukaryotic initiation factor 5A hypusine signature
 S-100/ICaBP type calcium binding protein signature
 Hemolysin-type putative calcium-binding region signature
 HlyD family secretion proteins signature
 P-II protein uridylation site
 Small, acid-soluble spore proteins, alpha/beta type, signature
 Caseins alpha/beta signature
 Legume lectins signatures
 Vertebrate galactoside-binding lectin signature
 Lysosome-associated membrane glycoproteins signatures
 Glycophorin A signature
 Seminal vesicle protein I repeats signature
 Seminal vesicle protein II repeats signature
 HCP repeats signature
 Bacterial ice-nucleation proteins octamer repeat
 Cell cycle proteins ftsW / rodA / spoVE signature
 Staphylocoagulase repeat signature
 11-S plant seed storage proteins signature
 Dehydrins signature
 Small hydrophilic plant seed proteins signature
 Pathogenesis-related proteins BetvI family signature
 Thaumatin family signature