Table 1: List of protein analysis tools and databases interfaced by PAT.

Name	**Description**	**Reference**	**Output parser**
Databases			
SWISSPROT	Manually-annotated protein sequences	(1)	
SPTREMBL	SWISSPROT + Computer-analized protein sequences not yet manually-annotated	(1)	
UNIREF100	Non-redundant protein sequences at 100% ident.	(1)	
UNIREF50	Non-redundant protein sequences at 50% ident.	(1)	
PFAM	Protein sequence families	(2)	
Protein Data Bank	Protein 3D structures	(3)	
PDB_seq	PDB sequences in FASTA format	(4)	
Primary sequence analysis			
AACOMPO	Amino acid composition statistics	Local	
PEPTIDE_MASS	Peptide mass computation	(5)	
PEPTIDE_CUTTER	Protease clivage site search	(6)	
SIGNALP	Signal peptide prediction	(7)	+
Cellular localization prediction			
PSORT	Cell location prediction	(8)	+
TARGETP	Cell location and cleavage site predictions	(9)	+
PROTFUN	Prediction of cellular role, enzyme class and GO category	(10,11)	
Non globular structure prediction			
TMAP	Transmembrane segment prediction	(12)	+
TMPRED	Transmembrane segment prediction	(13)	+
TOPPRED	Transmembrane segment prediction	(14)	+
SEG	Compositional bias detection	(15)	+
CAST	Compositional bias detection	(16)	+
NCOILS	Coiled-coil prediction	(17)	+
GLOBULAR	Consensus non globular segment prediction	Local	+
Sequence similarity search			
WUBLAST2	Fast sequence similarity search	(18)	+
PSIBLAST	Fast and iterative sequence similarity search	(19)	+
HMM	Hidden Markov models based similarity search	(20)	+
HMMPFAM	Match a sequence against an HMM database	(20)	+
CDHIT	Exhaustive database homology search	(21)	+
Sequence alignment			
CLUSTALW	Hierarchical multiple sequence alignment	(22)	+
MUSCLE	Multiple sequence alignment	(23)	+
LALIGN	Local pairwise alignments	(24)	+
MVIEW	Multiple sequence alignment viewer	(25)	
Sequence motif search			
PS_SCAN	PROSITE motif search	(26)	

MOTIF	Protein sequence motif search	(27)	
GREP	Regular expression search	Local	
MATRIX	Pairwise scoring matrix from aligned sequences	Local	+
Phylogeny inference and display			
BIONJ	Distance-based phylogeny inference	(28)	+
FASTME	Distance-based phylogeny inference	(29)	+
ATV	Phylogenetic tree applet viewer	(30)	
Secondary structure prediction			
PSIPRED	Secondary structure prediction	(31)	+
PREDATOR	Secondary structure prediction	(32)	+
DSC	Secondary structure prediction	(33)	+
SIMPA96	Secondary structure prediction	(34)	+
PRED2D	Consensus secondary structure prediction	Local	+
Solvent accessibility prediction			
NETASA	Solvent accessibility prediction	(35)	+
Tertiary structure analysis			
DSSP	Secondary structure assignment from PDB files	(36)	+
STRIDE	Structural analysis of PDB files	(37)	+
PDBGEO	PDB file geometrical analysis	Local	+
PDBINTER	Residue surfaces buried by inter-chain contacts	Local	
KNOTER	Standard numbering of knottin structures	(38,39)	
Tertiary structure display			
Jmol	PDB file applet viewer	(40)	
Tertiary structure superpimposition			
CE	Pairwise structural alignment	(41)	+
SHEBA	Pairwise structural alignment	(42)	+
PROFIT	Pairwise structure least square fit	(43)	+
Tertiary structure modelling			
SMD	Combinatorial amino acid side chain placement	(44)	+
SCWRL	Rotamer-based amino acid side chains placement	(45)	+
PDBBLAST	Search and align homologs with known structures	Local	+
Tertiary structure evaluation			
VERIFY3D	Potential-based structure evaluation	(46,47)	+
EVAL23D	Potential-based structure evaluation	(48)	+
TITO	Sequence-structure compatibility evaluation	(49)	
EVDTREE	Potential-based structure evaluation	(50)	
Other tools			
COLOR	Displays colored HTML outputs of aligned sequences, predictions, consensus or evaluations	Local	+
CONSENSUS	Builds sequence and/or prediction consensus	Local	+
SIM2ALI	Creates a multiple alignment from BLAST similarities	Local	+

SORT	Sort protein segments according to name, type, length or position	Local	+

References

1. Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res*, 33 Database Issue, D154-159.
2. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res*, 32, D138-141.
3. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, 28, 235-242.
4. RCSB. Information derived from PDB files: sequences in FASTA format (**http://www.rcsb.org)**.
5. Wilkins, M.R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F. and Appel, R.D. (1997) Detailed peptide characterization using PEPTIDEMASS--a World-Wide-Web-accessible tool. *Electrophoresis*, 18, 403-408.
6. ExPASy. PeptideCutter predicts potential cleavage sites cleaved by proteases or chemicals in a given protein sequence. ***http://caexpasyorg/tools/peptidecutter/***.
7. Bendtsen, J.D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 340, 783-795.
8. Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M. and Brinkman, F.S. (2004) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*.
9. Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 300, 1005-1016.
10. Jensen, L.J., Gupta, R., Blom, N., Devos, D., Tamames, J., Kesmir, C., Nielsen, H., Staerfeldt, H.H., Rapacki, K., Workman, C. *et al.* (2002) Prediction of human protein function from post-translational modifications and localization features. *J Mol Biol*, 319, 1257-1265.
11. Jensen, L.J., Gupta, R., Staerfeldt, H.H. and Brunak, S. (2003) Prediction of human protein function according to Gene Ontology categories. *Bioinformatics*, 19, 635-642.
12. Milpetz, F., Argos, P. and Persson, B. (1995) TMAP: a new email and WWW service for membrane-protein structural predictions. *Trends Biochem Sci*, 20, 204-205.
13. Hofmann, K. and Stoffel, W. (1993) TMBASE - A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler*, 374, 166.
14. Claros, M.G. and von Heijne, G. (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci*, 10, 685-686.
15. Wootton, J.C. (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem*, 18, 269-285.
16. Promponas, V.J., Enright, A.J., Tsoka, S., Kreil, D.P., Leroy, C., Hamodrakas, S., Sander, C. and Ouzounis, C.A. (2000) CAST: an iterative algorithm for the complexity analysis of sequence tracts. Complexity analysis of sequence tracts. *Bioinformatics*, 16, 915-922.
17. Russel, R.B. and Lupas, A.N. (1999) based on Lupas, Van Dyck & Stock (1991) Science 252,1162-1164. *Pfam: coiled-coils*.
18. Gish, W. (1996-2004). ***http://blastwustledu***.
19. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, 3389-3402.
20. Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, 14, 755-763.

21. Li, W., Jaroszewski, L. and Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*, 17, 282-283.
22. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22, 4673-4680.
23. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.
24. Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol*, 132, 185-219.
25. Brown, N.P., Leroy, C. and Sander, C. (1998) MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics*, 14, 380-381.
26. Gattiker, A., Gasteiger, E. and Bairoch, A. (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics*, 1, 107-108.
27. Motif. Sequence motif search. ***http://motifgenomejp/***.
28. Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*, 14, 685-695.
29. Desper, R. and Gascuel, O. (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol*, 9, 687-705.
30. Zmasek, C.M. and Eddy, S.R. (2001) ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics*, 17, 383-384.
31. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, 292, 195-202.
32. Frishman, D. and Argos, P. (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng*, 9, 133-142.
33. King, R.D., Saqi, M., Sayle, R. and Sternberg, M.J. (1997) DSC: public domain protein secondary structure predication. *Comput Appl Biosci*, 13, 473-474.
34. Levin, J.M. (1997) Exploring the limits of nearest neighbour secondary structure prediction. *Protein Eng*, 10, 771-776.
35. Ahmad, S. and Gromiha, M.M. (2002) NETASA: neural network based prediction of solvent accessibility. *Bioinformatics*, 18, 819-824.
36. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22, 2577-2637.
37. Heinig, M. and Frishman, D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res*, 32, W500-502.
38. Gelly, J.C., Gracy, J., Kaas, Q., Le-Nguyen, D., Heitz, A. and Chiche, L. (2004) The KNOTTIN website and database: a new information system dedicated to the knottin scaffold. *Nucleic Acids Res*, 32, D156-159.
39. Chiche, L., Heitz, A., Gelly, J.C., Gracy, J., Chau, P.T., Ha, P.T., Hernandez, J.F. and Le-Nguyen, D. (2004) Squash inhibitors: from structural motifs to macrocyclic knottins. *Curr Protein Pept Sci*, 5, 341-349.
40. SourceForge.net. Jmol is a free, open source Java molecule viewer. ***http://jmolsourceforgenet/***.
41. Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*, 11, 739-747.
42. Jung, J. and Lee, B. (2000) Protein structure alignment using environmental profiles. *Protein Eng*, 13, 535-543.
43. Martin, A.C.R. ProFit: Fitting is performed using the McLachlan algorithm (McLachlan, A.D., 1982 ``Rapid Comparison of Protein Structres", Acta Cryst A38, 871-873). ***http://wwwbioinforguk/software/profit***.
44. Tuffery, P., Etchebest, C., Hazout, S. and Lavery, R. (1991) A new approach to the rapid determination of protein side chain conformations. *J Biomol Struct Dyn*, 8, 1267-1289.

45. Dunbrack, R.L., Jr. and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol*, 230, 543-574.
46. Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, 356, 83-85.
47. Bowie, J.U., Luthy, R. and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253, 164-170.
48. Gracy, J., Chiche, L. and Sallantin, J. (1993) Improved alignment of weakly homologous protein sequences using structural information. *Protein Eng*, 6, 821-829.
49. Labesse, G. and Mornon, J. (1998) Incremental threading optimization (TITO) to help alignment and modelling of remote homologues. *Bioinformatics*, 14, 206-211.
50. Gelly, J.C., Chiche, L. and Gracy, J. (2005) EvDTree: structure-dependent substitution profiles based on decision tree classification of 3D environments. *BMC Bioinformatics*, 6, 4.