*In Silico* Biology 5, 0034 (2005); ©2005, Bioinformation Systems e.V.

# A new domain family in the superfamily of alkaline phosphatases

**Rana Bhadra[1], Narayanaswamy Srinivasan[1]* and Shashi B. Pandit[2]**

[1] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India
[2] Present address:
Center of Excellence in Bioinformatics, University at Buffalo, New York, USA

* Corresponding author

Phone: +91-80-2293 2837; Fax: Fax: +91-80-2360 0535
Email: ns@mbu.iisc.ernet.in

Edited by E. Wingender; received March 06, 2005; revised May 13, 2005; accepted May 15, 2004; published May 21, 2005

---

## Abstract

During the course of our large-scale genome analysis a conserved domain, currently detectable only in the genomes of *Drosophila melanogaster*, *Caenorhabditis elegans* and *Anopheles gambiae*, has been identified. The function of this domain is currently unknown and no function annotation is provided for this domain in the publicly available genomic, protein family and sequence databases. The search for the homologues of this domain in the non-redundant sequence database using PSI-BLAST, resulted in identification of distant relationship between this family and the alkaline phosphatase-like superfamily, which includes families of aryl sulfatase, N-acetylgalactosomine-4-sulfatase, alkaline phosphatase and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (iPGM). The fold recognition procedures showed that this new domain could adopt a similar 3-D fold as for this supefamily. Most of the phosphatases and sulfatases of this superfamily are characterized by functional residues Ser and Cys respectively in the topologically equivalent positions. This functionally important site aligns with Ser/Thr in the members of the new family. Additionally, set of residues responsible for a metal binding site in phosphatases and sulphtases are conserved in the new family. The in-depth analysis suggests that the new family could possess phosphatase activity.

*Keywords*: functional domains, phosphatases, protein domains, sequence analysis, sulfatases

## Introduction

The completion of genome sequencing for a number of organisms offers us an opportunity to understand the molecular basis of their physiology, metabolism, regulation and evolution [Aravind *et al.*, 2003; Pearl *et al.*, 2002; Pawlowski *et al.*, 1999]. This is essentially inferred from the functional characterization of the gene products encoded in the genomes. The clues about the functions of newly discovered proteins can be obtained by their similarity to experimentally well characterized proteins whose sequences are available in the sequence databases. However, there are a large number of proteins encoded in genomes that do not exhibit obvious sequence similarity to proteins of known function. Moreover, most of these proteins, referred to as hypothetical proteins, have not been experimentally explored for their possible functions. The hints about the functions of these hypothetical proteins can be obtained by exploring their relationships to proteins of known function by means of sophisticated sequence analysis and fold recognition procedures.

During the course of our large-scale genome analysis [Pandit *et al.*, 2004; Namboori *et al.*, 2004] a set of homologous hypothetical proteins, currently detectable only in the genomes of *Drosophila melanogaster*, *Caenorhabditis elegans* and *Anopheles gambiae*, has been identified. The function of this family of proteins is currently unknown and no function annotation is present in the publicly available genomic sequence databases. Moreover in a sequence domain database, Pfam [Bateman *et al.*, 2002], this family is classified as "Domain of unknown function" (DUF229). Using the promoter prediction tools such as NNPP [Reese, 2001] and Promoter2.0 [Knudsen, 1999] it is observed that these putative proteins contain promoter regions. In order to associate function to these hypothetical proteins we have explored their relationship to proteins of known funcion.

We have employed sensitive profile-based sequence similarity search tools such as PSI-BLAST [Altschul *et al.*, 1997] and 3-D fold recognition to explore the relatedness of this hypothetical protein family. Using these procedures, we could establish relationship between the hypothetical protein family and the members of alkaline phosphatase-like superfamily. The detailed investigation of the key functional residues suggests probable phosphatase functionality for these hypothetical proteins. Based on the occurrence of this hypothetical protein family in nematode and fly genome and associated putative phosphatase function we refer this family as "Nematode Fly putative phosphatase" or "NFPP" family.

---

## Methods

### Databases

The sequences of members belonging to various families were retrieved from various publicly available sequence databases and domain databases. The non-redundant sequence database, NRDB, has been obtained from National Center for Biotechnology Information (ftp://ftp.ncbi.nlm.nih.gov/blast/db).

### Sequence analysis

The PSI-BLAST [Altschul *et al.*, 1997] was used to search against NRDB to retrieve the homologues. The RPS-BLAST searchable profiles for PALI [Balaji *et al.*, 2001; Gowri *et al.*, 2003] families have been generated as described in methods in Pandit *et al.*, 2002. The sensitive profile matching method RPS-BLAST was used to match each sequence against PALI profiles with $E$-value cut-off of $10^{-5}$. Multiple sequence alignment was done using T-Coffee [Notredame *et al.*, 2000] and ClustalW [Thompson *et al.*, 1994].

---

## Results and discussion

### Recognition of conserved residues in NFPP family

Sequences of the members of the new family were retrieved by searching of non-redundant sequence database (NRDB) at NCBI, using gapped BLAST [Altschul *et al.*, 1997] program with one of the members of this new family as a query. Both $E$-value and $h$-value cut-off was set to $10^{-3}$. The homologues identified, with sequence identity greater than 30%, mostly in the first round, are considered as the member of the family. The homologous members of this new family were aligned using the multiple sequence alignment tool ClustalW [Thompson *et al.*, 1994]. A representative set of sequences of the new family is shown aligned in Figure 1. As evident from the multiple sequence alignment of the family members, there is conservation of Asp, His and Ser/Thr in specific positions in the alignment apart from other conserved hydrophobic and polar residues.

**Figure 1:** Multiple sequence alignment of representative sequences belonging to the NFPP domain family. These sequences were aligned using ClustalW multiple alignment tool. The ranges of sequence identity and sequence similarity of aligned sequences are 31%-61% and 47%-75% respectively. The proposed functionally conserved (metal binding or catalytic) residues are indicated in bold and blue in colour.Alignment of these metal binding residues are also indicated by * at the top of the alignment block. The other conserved residues within the family are shown in bold and brown color and these positions are indicated by #. The proposed functionally conserved residue positions in the first and second sequences (gi23172514 and gi21430616) are shown in superscript. The sequences are associated with Gene identification (GI) number from GenBank [Benson *et al.*, 2003]. The sequences considered for alignment are: gi23172514, gi21430616 from *Drosophila melanogaster*, gi17543326, gi17559486 from *C. elegans* and gi21296810, gi21300280 from *Anopheles gambiae*. The other members of this family and not present in alignment are: *D. melanogaster* proteins (gi7297472, gi22945998, gi7298655, gi21430366, gi21430910, gi23092768, gi21356013, gi7300665, gi7295100, gi7297470, gi7297815, gi22945479, gi21064045, gi21483202, gi7299531, gi7295901, gi21064005, gi7299997, gi21357449, gi23171351, gi7302136, gi20129693, gi19527787, gi17569219), gi21299542 from *A. gambiae* and gi17559746, gi17559316, gi17542258, gi17544218 are from *C. elegans*..

### Relationship between NFPP family with alkaline phosphatase-like superfamily

We explored the possibility of this hypothetical protein family being distantly related to any of the other protein families, using profile-based search procedures and 3-D fold recognition methods. When we used these hypothetical protein family members as queries in PSI-BLAST searches against NRDB, in the second, third and fourth rounds of iterations we identified homologues of eukaryotic sulfatases with *E*-

values from 2 x $10^{-5}$ to 6 x $10^{-4}$. From fifth and sixth rounds of iteration, in addition to eukaryotic and prokaryotic sulfatases, we identified members of 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (iPGM) and nucleotide pyrophosphatase (NPP) with *E*-values from 6 x $10^{-5}$ to 1 x $10^{-4}$. Furthermore, in subsequent iterations (round nine onwards) we could identify homologues of alkaline phosphatase. Interestingly, sulfatases, NPP, iPGM and alkaline phosphatase members have been shown to be related by superfamily relationship [Galperin *et al.*, 1998; Galperin and Jedrzejas 2001; Gijsbers *et al.*, 2001] and are classified in alkaline phosphatase-like superfamily. In the 3-dimensional protein structural classification database, SCOP, alkaline phosphatase-like superfamily includes arylsulfatase (AS-A), N-acetylgalactosomine-4-sulfatase (AS-B), alkaline phosphatase (AP) and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (iPGM) members. Using RPS-BLAST, searches have been made with hypothetical protein sequences as queries against structure-based sequence profile database of protein domain families of known structure available in PALI database [Balaji *et al.*, 2001; Gowri *et al.*, 2003]. These hypothetical protein members could be related to sulfatase with an *E*-value of 3 x $10^{-4}$. We have made an entry of this predicted relationship in our SUPFAM database [Pandit *et al.*, 2004] against the entry for the family DUF229 and alkaline phosphatase-like superfamily.

The region of alignment of hypothetical protein members with the members of the alkaline phosphatase-like super families is typically from positions 170 to 580. However, the length of homologous hypothetical proteins varies in the range of 500 to 750 residues. Hence, the occurrence of alkaline phosphatase-like superfamily in the hypothetical proteins can be considered to correspond to a domain. We could not identify any other known domains in the sequences of these hypothetical proteins. The best pairwise sequence identities between a member of this new domain and sulfatases and phosphatases are 19% and 15% respectively suggesting a remote relationship. The best pairwise sequence identity of this domain with iPGM and NPP is 19% and 13% respectively. This indicates that this domain family is distantly related to any member of alkaline phosphtase-like superfamily. However, fold recognition for the sequences of the hypothetical protein family showed that this domain sequences could be comfortably accommodated into the fold of alkaline phosphatase-like fold. This was done by using fold recognition methods, GenThreader [Jones, 1999] (http://bioinf.cs.ucl.ac.uk/psiform.html), 3D-PSSM [Kelley *et al.*, 2000] (http://www.sbg.bio.ic.ac.uk/~3dpssm) and FUGUE [Shi *et al.*, 2001] (http://www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html), with sequences of hypothetical protein domain family as queries. All the methods mentioned above reliably associated the hypothetical protein domain family with the 3-D fold of alkaline phosphatase. The *E*-values for the various queries from hypothetical protein family all resulting in the top most hit of alkaline phosphatase-like fold in GenThreader and 3D-PSSM are less than 0.01 and 0.005 respectively. The Z-scores in FUGUE are greater than 6. The best sequence alignment we could obtain between hypothetical protein and alkaline phosphatase-like superfamily members is shown in Figure 2. The predicted secondary structures of the hypothetical proteins obtained using PHD [Rost, 1996] (http://www.cubic.bioc.columbia.edu/predictprotein) and SSPRO [Pollastri *et al.*, 2002] (http://www.igb.uci.edu/tools/scratch/) are very similar. The predicted $\alpha$-helical and $\beta$-strand regions (shown at the top of the alignment blocks in Figure 2) of the members of DUF229 aligned very well with the observed helical and strand regions, respectively (shown at the bottom of the alignment blocks in Figure 2), reported in the protein databank files corresponding to the distant homologues of known 3-D structure (Figure 2).

**Figure 2:** Alignment of NFPP family members with APs, iPGMs, AS-A and AS-Bs that belong to alkaline phosphatase-like superfamily showing structural similarities in the catalytic domains. The NFPP family members were aligned with members of alkaline phosphatse-like superfamily using T-coffee and considering structure-based alignment for superfamily members taken from PASS2 database [Bhaduri *et al.*, 2004] The secondary structure prediction of the catalytic domain of *D. melanogaster* NFPP family (gi23172514 and gi21430616) was obtained using PHD and SSPRO secondary structure prediction tools. The pdb codes 1ED8 , 1EQJ , 1AUK and 1FSU corresponds to alkaline phosphatase (AP), phosphoglycerate mutase (iPGM), aryl sulfatase (AS-A) and N-acetyl galactosamine-4-sulfatase (AS-B) respectively. The cylinder and arrow in green color at the top of the alignment blocks represent the predicted $\alpha$-helices and $\beta$-stands of NFPP sequences respectively. The orange color cylinder and arrows at the bottom of the alignment blocks represent the conserved $\alpha$-helices and $\beta$-stands, in the members of alkaline phosphatase-like superfamily with known 3-D structure. The numbers in parentheses represent the numbers between two aligned regions and the superscript numbers represent the position of functionally important residues. The catalytic residues are shown in red and their alignment positions are indicated by # at the top of the alignment. The blue color residues indicated by * at the top of the alignment are metal coordinating residues (coordinates $Mg^{2+}$ in AS-A, $Zn^{2+}$(2) in AP, $Mn^{2+}$(2) in iPGM and $Ca^{2+}$ in AS-B) conserved in all the members of this superfamily. The pink color residues (indicated by single underline) are other metal coordinating residues (coordinates $Zn^{2+}$(1) in AP and $Mn^{2+}$(1) in iPGM) conserved only in AP and iPGM families and residues in violet (indicated by double underline) are metal coordinating residues in AS-A and AS-B. The residues conserved in NFPP family and are proximal in 3-D space with a putative role in metal binding are shown in green and indicated by wavy underline.

### Conserved residues between new family and alkaline phosphatase-like superfamily

The alkaline phosphatase-like superfamily members are known to occur ubiquitously in several prokaryotes and eukaryotes [Coleman, 1992; Galperin *et al.*, 1998]. This superfamily includes various enzymatic functions such as isomerases, hydrolases and putative lyase. The substrates of these enzymes vary in their size and chemical nature [Galperin *et al.*, 1998]. However, they share similar metal binding residues and similar mechanism [Galperin *et al.*, 1998; Galperin and Jedrzejas, 2001]. We explored the possibility of a hypothetical protein family to possess phosphatase or sulfatase enzymatic activity as suggested by the sequence similarity. Most of the phosphatases and sulfatases are characterised by functional residues Ser [Coleman, 1992; Jedrzejas *et al.*, 2000; Kim *et al.*, 1991; Stec *et al.*, 2000] and Cys [Schmidt *et al.*, 1995; Lukatela *et al.*, 1998; Bond *et al.*, 1997] respectively. The alignment between NFPP members and alkaline phosphatase-like superfamily was obtained using T-coffee [Notredame *et al.*, 2000]. Moreover, the alignment was adjusted manually in correspondence with structure-based alignment of superfamily members taken from PASS2 database [Bhaduri *et al.*, 2004] as shown in Figure 2. This alignment revealed a remarkable conservation in length and position of $\alpha$-helices and $\beta$-strands. As evident from the alignment (Figure 2), catalytic Ser of phosphatases is aligned with Ser/Thr of hypothetical protein and with catalytic Cys of sulfatases. The Cys residue has been shown to be critical for sulfatase activity and any substitution of this residue abolishes the activity [Brooks *et al.*, 1995; von Bülow *et al.*, 2001]. However, in some bacterial sulfatases such as from *Klebsiella pneumoniae*, *E. coli* Cys is substituted with Ser [Murooka *et al.*, 1990; Daniels *et al.*,1992]. Furthermore, all eukaryotic and prokaryotic sulfatases have a conserved motif [CS]XPXRXXXLTG [Dierks *et al.*, 1998] which is not present in this new family. From this alignment (Figure 2), it has also been noticed that residues coordinating with one of the metals (indicated by * at the top of the alignment block) which are conserved in alkaline phosphatase-like superfamily, are also conserved in this new family. The residues coordinating with the other metal (indicated by single underline) conserved in only AP and iPGM, are not conserved in this new NFPP family.

However, structural modelling of one of the new family members showed that another set of conserved Asp and His (indicated by wavy underline in Figure 2) in the new family are proximal in 3-D space and lie near catalytic site (Figure 3). This suggests that these residues may coordinate additional metal ions. The eukaryotic and prokaryotic phosphatases (alkaline phosphatases, nucleotide pyrophosphatases and 2,3-bisphosphoglycerate-independent phosphoglycerate mutase) have Ser or Thr residue required for the activity. These observations indicate a probable phosphatase activity for the new family. However, the possibility of this new family to act as sulfatases can not be completely ruled out since both sulfatases and phosphatases are essentially involved in cleaving similar functional groups viz. sulphate and phosphates.

**Figure 3:** Structure of catalytic core of one of the members of NFPP (gi23172514) that is modelled using one of the known structures as a template structure. The residue numbers shown corresponds to the positions in gi23172514. The catalytic residue is shown in red color. The metal coordinating residues conserved in the superfamily is shown in blue. The residues that are proximal to the catalytic site and could have a putative metal binding activity in NFPP are shown in green. This figure has been produced using Setor software [Evans, 1993].

Interestingly, homologues of NFPP family are identified from *D. melanogaster*, *C. elegans* and *A. gambiae* only. Some of the currently known members of alkaline phsophatase-like superfamily are already known to occur in these organisms. This suggests that NFPP family have diverged extensively to meet the functional requirement in the fly andnematode.

---

## Conclusions

The in-depth analysis showed that family of hypothetical protein has a conserved domain that is remotely related to alkaline phosphatase-like superfamily and has putative catalytic residue Ser/Thr and metal binding residues in one of the metal binding sites. The phosphatases require Ser or Thr residue at a specific location in the sequence for their activity. Mainly based on this observation we propose that this new domain would belong to this superfamily and may exhibit phosphatase activity. We report identification of homologues of this family in *Anopheles gambiae*. It is interesting that the members of this new family are identified only in fruitfly, worm and mosquito. All these organisms are already known to have members of the alkaline phosphatase-like superfamily, which are very distantly related to the hypothetical proteins of current investigation. By means of detailed comparative analysis in relation to distant homologues of known structure we suggest residues which are potentially involved in metal binding and catalysis. With the availability of more genomic data in the future a clear evolutionary relationship of this new family with other families in the superfamily of alkaline phosphatase may be established.

---


## Acknowledgements

R.B is supported by the NMITLI project sponsored by CSIR, New Delhi and by the Wellcome Trust, London. S.B.P is supported by Council of Scientific and Industrial Research, New Delhi. This research is supported by the award of International Senior Fellowship to N.S. by the Wellcome Trust, London and by the computational genomics project funded by the Department of Biotechnology, New Delhi.

## References


- 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.
- Aravind, L., Iyer, L. M., Wellems, T. E. and Miller, L. H. (2003). *Plasmodium* Biology: Genomic Gleanings. Cell 115, 771-785.
- Balaji, S., Sujatha, S., Kumar, S. S. C. and Srinivasan, N. (2001). PALI-a database of Phylogeny and ALIgnment of homologous protein structures. Nucleic Acids Res. 29, 61-65.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. (2002). The Pfam protein families database. Nucleic Acids Res. 30, 276-280.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. (2003). GenBank. Nucleic Acids Res. 31, 23-27.
- Bhaduri, A., Pugalenthi, G. and Sowdhamini, R. (2004). PASS2: an automated database of protein alignments organised as structural superfamilies. BMC Bioinformatics 5, 35.
- Bond, C. S., Clements, P. R., Ashby, S. J., Collyer, C. A., Harrop, S. J., Hopwood, J. J. and Guss, J. M. (1997). Structure of a human lysosomal sulfatase. Structure 5, 277-289.
- Brooks, D. A., Robertson, D. A., Bindloss, C., Litjens, T., Anson, D. S., Peters, C., Morris, C. P. and Hopwood, J. J. (1995). Two site-directed mutations abrogate enzyme activity but have different effects on the conformation and cellular content of the N-acetylgalactosamine 4-

sulfatase protein. Biochem. J. 307, 457-463.

- Coleman, J. E. (1992). Structure and mechanism of alkaline phosphatase. Annu. Rev. Biophys. Biomol. Struct. 21, 441-483.
- Daniels, D. L., Plunkett, G. 3rd, Burland, V. and Blattner, F. R. (1992). Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science 257, 771-778.
- Dierks, T., Miech, C., Hummerjohann, J., Schmidt, B., Kertesz, M. A. and von Figura, K. (1998). Posttranslational formation of formylglycine in prokaryotic sulfatases by modification of either cysteine or serine. J. Biol. Chem. 273, 25560-25564.
- Evans, S. V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. J. Mol. Graph. 11, 134-138.
- Galperin, M. Y. and Jedrzejas, M. J. (2001). Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes. Proteins 45, 318-324.
- Galperin, M. Y., Bairoch, A. and Koonin, E. V. (1998). A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases.Protein Sci. 7, 1829-1835.
- Gijsbers, R., Ceulemans, H., Stalmans, W. and Bollen, M. (2001). Structural and catalytic similarities between nucleotide pyrophosphatases/phosphodiesterases and alkaline phosphatases. J. Biol. Chem. 276, 1361-1268.
- Gowri, V.S., Pandit, S.B., Karthik, P.S., Srinivasan, N. and Balaji, S. (2003). Integration of related sequences with protein three-dimensional structural families in an updated version of PALI database. Nucleic Acids Res. 31, 486-488.
- Jedrzejas, M. J., Chander, M., Setlow, P. and Krishnasamy, G. (2000). Structure and mechanism of action of a novel phosphoglycerate mutase from *Bacillus stearothermophilus*. EMBO J. 19, 1419-1431.
- Jones, D. T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method

for genomic sequences. J. Mol. Biol. 287, 797-815.

- Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 299, 499-520.
- Kim, E. E. and Wyckoff, H. W. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. J. Mol. Biol. 218, 449-464.
- Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics 15, 356-361.
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K. and Saenger, W. (1998). Crystal structure of human arylsulfatase A the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. Biochemistry 37, 3654-3664.
- Murooka, Y., Ishibashi, K., Yasumoto, M., Sasaki, M., Sugino, H., Azakami, H. and Yamashita, M. (1990). A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the arylsulfatase (*atsA*) gene and the *atsB* gene. J. Bacteriol. 172, 2131-2140.
- Namboori, S., Srinivasan, N. and Pandit, S.B. (2004). Recognition of remotely related structural homologues using sequence profiles of aligned homologous protein structures. In Silico Biol. 4, 0037.
- Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205-217
- Pandit, S.B., Gosar, D., Abhiman, S., Sujatha, S., Dixit, S.S., Mhatre, N.S., Sowdhamini, R. and Srinivasan, N. (2002). SUPFAM - a database of potential protein superfamily relationships derived by comparing sequence-based and structure-based families: implications for structural genomics and function annotation in genomes. Nucleic Acids Res. 30, 289-293.
- Pandit, S. B., Bhadra, R., Gowri, V. S., Balaji, S., Anand, B. and Srinivasan, N. (2004). SUPFAM: A database of sequence superfamilies of protein domains. BMC Bioinformatics 5, 28.
- Pawlowski, K., Zhang, B., Rychlewski, L. and Godzik, A. (1999). The *Helicobacter pylori* genome from sequence analysis to structural and functional predictions. Proteins 36, 20-30.

- Pearl, F. M., Lee, D., Bray, J. E., Buchan, D. W., Shepherd, A. J. and Orengo, C. A. (2002). The CATH extended protein-family database providing structural annotations for genome sequences. Protein Sci. 11, 233-244.

---

- Pollastri, G., Przybylski, D., Rost B. and Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 47, 228-235.

---

- Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput. Chem. 26, 51-56.

---

- Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. 266, 525-539.

---

- Schmidt, B., Selmer, T., Ingendoh, A. and von Figura, K. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. Cell 82, 271-278

---

- Shi, J., Blundell, T. L. and Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J. Mol. Biol. 310, 243-257.

---

- Stec, B., Holtz, K. M. and Kantrowitz, E. R. (2000). A Revised Mechanism for the alkaline phosphatase Reaction Involving Three Metal Ions. J. Mol. Biol. 299, 1303-1311.

---

- 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.

---

- von Bülow, R., Schmidt, B., Dierks, T., von Figura, K. and Usón, I. (2001). Crystal structure of an enzyme-substrate complex provides insight into the interaction between human arylsulfatase A and its substrates during catalysis. J. Mol. Biol. 305, 269-277.