

Editorial

Computational Analysis to Study Successive Development of Adaptable Protein Structure and Function during Evolution

Surya P Kilaparty¹ and Nawab Ali^{1*}

¹Department of Applied Science, University of Arkansas at Little Rock, USA

***Corresponding author:** Nawab Ali, Department of Applied Science, University of Arkansas at Little Rock, 2801 South University Avenue, Little Rock, AR 72204, USA, Tel: 501 569 8003; Fax: 501 569 8020; Email: nali@ualr.edu

Received: June 28, 2014; **Accepted:** July 01, 2014;
Published: July 04, 2014

Abstract

Post genomic era poses challenges to both computational scientists and basic biologists equally to assign functions to many proteins for which only the sequence information is available. Predicting protein functions by computational means is an active area of research for many scientists. Understanding functions of unknown proteins will have a huge impact on pharmaceutical industries for drug design and development. Computational methods are also useful for basic biologists particularly in understanding evolutionary aspect of proteins. Some proteins present in lower organisms with fewer or simpler functions are also found in higher organisms with more complex functions. Now, the fundamental question arises how such complexity in protein structure and function is brought in and adapted by higher organisms. The key appears to lie in computational methods for comparative analysis. Therefore, the use of existing as well as development of new computational tools will be an approach in the right direction to understanding evolution and development of proteins.

Keywords: Minpp1; DIPP; Nudix; Inositol phosphate phosphatase; Evolutionary relationship; Computational analysis; Bioinformatics; Motifs; Protein function

Understanding protein function at molecular level has great implications in biomedical and pharmaceutical industry [1-3]. There is a tremendous research currently being done in recognizing functions of uncategorized proteins. Hefty amounts of uncategorized protein sequence data are now available in numerous databases. To analyze these data, a variety of computational algorithms are available. Some algorithms are still being improved while others are in the process of development for predicting functions of proteins from their sequences [4,5]. Obviously, the predicted functions will need to be confirmed by biochemical analysis to define protein function precisely and validate computational results.

Computational tools have traditionally been used to construct phylogenetic trees based on sequence similarity to predict evolutionary relatedness between species. Such analyses have pointed out that there is a constant ongoing evolutionary process and adaptability of changes due to genetic mutations. These mutations originate by way of insertion, deletion or rearrangement in nucleotide bases of which some may reflect changes in amino acids and thus 3D native structures. Variation in a native protein structure between related species might be related to an evolutionary change due to mutation. These changes can be carried along in higher species as a result of natural selection, if the changes are favorable and adaptable. Can computational tools be used to address the basic questions, for example, whether a protein in lower organisms has any homology with the protein in higher organisms? How does a protein change over time in evolutionary context from lower species to higher ones?

We have studied two such proteins namely multiple inositol phosphate phosphatase (Minpp1) and diphosphoinositol

pyrophosphate phosphatase (DIPP) [6,7]. These two proteins belong to histidine super family and Nudix hydrolase family respectively [8,9]. Both of these proteins involve in dephosphorylating higher inositol phosphates (InsP₈, InsP₇, InsP₆) to lower InsPs (InsP₅, InsP₄, InsP₃) [10-13]. Inositol phosphates (InsPs) are ubiquitously present in nature. They are biologically important cellular signaling molecules that participate in calcium mobilization, vesicular trafficking, chromatin remodeling, cell proliferation and differentiation, apoptosis and more recently implicated in diabetes [14-19]. Altered levels of these molecules have also been linked to some diseases or physiologically altered cellular state. Computational studies showed that these two proteins were present in organisms ranging from bacteria to humans [20]; DIPP was also found in viruses [21]. The presence of these proteins in higher organisms can be viewed as the end result of natural selection in evolutionary process for some adaptable functions.

Use of bioinformatics tools not only predicts phylogeny but also provides information on variations in protein structure and their function. The variations found in related species may point to a link for their evolutionary development from lower to higher species. Several tools are available to determine such changes and conservation of motifs in a protein sequence. A Motif is a short amino acid sequence that is conserved and constitutes a signature for a particular function. Analysis of these motifs for any amino acid sequence variation would reflect a variation in its function. These motifs can further be analyzed for any amino acid variations between species. Any variation in motif sequence or an addition of a new motif for a new function in higher organisms relative to lower organisms would be a reflection of complexity in its function.

In order to carry out such evolutionary studies, a comparative analysis of the amino acid sequences is first achieved by blasting the protein of interest against available protein data banks. BLAST search results not only reveal similarities shared but also provide the conserved regions in the protein sequences. However, these search results are not considered highly significant statistically, as these include putative as well as uncharacterized synthetic sequences [22]. To obtain sequences that are highly significant and conserved, multiple sequence alignment (MSA) is carried out using selected sequences. This is an alignment process that compares many sequences and successively considers all possible pair wise alignments. MSA compares the closely related species and any variations, insertions or deletions of amino acids that might have occurred due to evolution. A number of alignment tools based on different computational algorithms are available to perform such sequence analysis. In our studies, MEGA 6, a user-friendly tool, was used [23]. The phylogenetic tree is then constructed from the aligned sequences. The tree is a reflection of an evolutionary relationship between the species. The organisms with dissimilar protein sequences appear far from each other and vice versa. The branches/clades represent the time line of diversification [24]. Additionally, computational tools can be useful to find any isoforms of a protein within a species. In our study (unpublished), four isoforms of human Minpp1 were noted in UniProt protein database analyzed by MEGA 6.

Analysis of secondary and tertiary structures of proteins is crucial for detailed understanding of the function of a protein. Secondary and tertiary structural models of proteins can be constructed by different modelling methods e.g. Homology Modeling, Ab initio Prediction, Sequence-Structure Threading or Docking. A number of meta-servers utilizing these methods are available online. We used I-TASSER meta-server for our analysis [25]. Further details in any variation in protein structure are analyzed by closely comparing the amino acid sequences in the conserved motifs following MSA. The basic motif for inositol phosphate phosphatase that we followed was present in all species in our study, even though slight variations in amino acids were noticed within the motif. Some motifs (e.g. PH domain motif and ER retention motif) are more conserved or only found in higher organisms and not in lower organisms. Such studies reveal that primitive motifs and their functions present in lower organisms are adapted during evolution and evolve to complex or new functions in higher organisms. Additionally, computational tools such as COACH (I-TASSER meta-server) could be used to screen putative ligand-motif interactions [5,26,27] and possibly be confirmed by biological experimentation. Such studies will provide an in-depth understanding about the function of a protein of interest or any uncategorized protein.

References

- Lesk A. Introduction to bioinformatics: Oxford University Press. 2013.
- Xiao X, Lin WZ, Chou KC. Recent advances in predicting protein classification and their applications to drug development. *Curr Top Med Chem*. 2013; 13: 1622-1635.
- Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009; 27: 157-167.
- Kaján L, Yachdav G, Vicedo E, Steinegger M, Mirdita M, Angermüller C, et al. Cloud prediction of protein structure and function with PredictProtein for Debian. *Biomed Res Int*. 2013; 2013: 398968.
- van Westen GJ, Overington JP. A ligand's-eye view of protein similarity. *Nat Methods*. 2013; 10: 116-117.
- Kilaparty SP, Singh A, Ali N. A Bioinformatics Study on Evolutionary Diversification of Multiple Inositol Polyphosphate Phosphatase 1 as an Aid on Understanding its Functional Significance in Mammalian Systems (Abstract). [abstract]. 10th Anniversary Annual MidSouth Computational Biology and Bioinformatics Society Conference. 2013.
- Williams P, Ali N: Evolutionary Relationship of Diphosphoinositol Polyphosphate Phosphatase within NUDIX Family of Proteins. [abstract]. 3rd Annual MidSouth Computational Biology and Bioinformatics Society Conference. 2006.
- Rigden DJ. The histidine phosphatase superfamily: structure and function. *Biochem J*. 2008; 409: 333-348.
- Kraszewska E. The plant Nudix hydrolase family. *Acta Biochim Pol*. 2008; 55: 663-671.
- Ali N, Craxton A, Shears SB. Hepatic Ins(1,3,4,5)P4 3-phosphatase is compartmentalized inside endoplasmic reticulum. *J Biol Chem*. 1993; 268: 6161-6167.
- Chi H, Tiller GE, Dasouki MJ, Romano PR, Wang J, O'keefe RJ, et al. Multiple inositol polyphosphate phosphatase: evolution as a distinct group within the histidine phosphatase family and chromosomal localization of the human and mouse genes to chromosomes 10q23 and 19. *Genomics*. 1999; 56: 324-336.
- Caffrey JJ, Hidaka K, Matsuda M, Hirata M, Shears SB. The human and rat forms of multiple inositol polyphosphate phosphatase: functional homology with a histidine acid phosphatase up-regulated during endochondral ossification. *FEBS Lett*. 1999; 442: 99-104.
- Craxton A, Ali N, Shears SB. Comparison of the activities of a multiple inositol polyphosphate phosphatase obtained from several sources: a search for heterogeneity in this enzyme. *Biochem J*. 1995; 305: 491-498.
- York JD, Odom AR, Murphy R, Ives EB, Wente SR. A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science*. 1999; 285: 96-100.
- Ali N, Craxton A, Sumner M, Shears SB. Effects of aluminium on the hepatic inositol polyphosphate phosphatase. *Biochem J*. 1995; 305: 557-561.
- Shears SB. The versatility of inositol phosphates as cellular signals. *Biochim Biophys Acta*. 1998; 1436: 49-67.
- Kim S, Snyder SH. Nutrient amino acids signal to mTOR via inositol polyphosphate multikinase. *Cell Cycle*. 2011; 10: 1708-1710.
- Mackenzie RW, Elliott BT. Akt/PKB activation and insulin signaling: a novel insulin signaling pathway in the treatment of type 2 diabetes. *Diabetes Metab Syndr Obes*. 2014; 7: 55-64.
- Chakraborty A, Koldobskiy MA, Bello NT, Maxwell M, Potter JJ, Juluri KR, et al. Inositol pyrophosphates inhibit Akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell*. 2010; 143: 897-910.
- Stentz R, Osborne S, Horn N, Li AW, Hautefort I, Bongaerts R, et al. A bacterial homolog of a eukaryotic inositol phosphate signaling enzyme mediates cross-kingdom dialog in the mammalian gut. *Cell Rep*. 2014; 6: 646-656.
- Bessman MJ, Frick DN, O'Handley SF. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J Biol Chem*. 1996; 271: 25059-25062.
- Wolfsberg TG, Madden TL. Sequence similarity searching using the BLAST family of programs. *Current Protocols in Human Genetics* 2001.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30: 2725-2729.
- Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipski A, Kumar S, et al. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A*. 2012; 109: 19333-19338.
- Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated

- protein structure and function prediction. Nat Protoc. 2010; 5: 725-738.
26. Yang J, Roy A, Zhang Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res. 2013; 41: D1096-1103.
27. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics. 2013; 29: 2588-2595.