American-Eurasian J. Agric. & Environ. Sci., 14 (11): 1295-1303, 2014
ISSN 1818-6769

DOI: 10.5829/idosi.aejaes.2014.14.11.12455

Prediction of Structures and Their Characterization of an Unknown Protein from Wheat (*Triticum aestivum* L.)

Sanjay Kumar

Department of Botany, Nagaland University, HQs: Lumami 798627, Nagaland

Abstract: An accepted and uncharacterized protein sequence with accession number Q9LRJ1 was retrieved from the Uni-Prot Knowledgebase. The physico-chemical properties, presence of conserved domains, presence of secondary structures (α-helix, β-sheets or coils), subcellular localization, homology modeling, 3-D structures, Ramachandran plot and Z-scores of the protein sequence and predicted 3-D structure were calculated. The structural and functional analysis of the protein sequence suggested predicting and searching for new functions of many other uncharacterized proteins of plants, animals, bacteria and fungi.

Key words: Uncharacterized protein · Physicochemical properties · Secondary structures · Ramachandran plot · Z-score

INTRODUCTION

The word 'protein' was first introduced to emphasize the importance of protein molecules in biological sciences [1]. The word is imitated from the Greek word 'proteios' which means 'of the first rank' [2]. Proteins are the key apparatus of living organisms and carry out a broad range of fundamental functions in cells [3]. Proteins control metabolic action, catalyze biochemical responses and sustain structural integrity of cells and organisms [4]. Proteins could be listed in a different way of ranges together with their natural functions [5]. A large data with different ranges of proteins could be created using 20 different building blocks of proteins called amino acids [6]. Each of the amino acid has a diverse element arrangement with different properties [7]. An alteration in just one amino acid was able to transform the configuration and function of a protein [8]. Amino acids are the molecules made up of Carbon, Hydrogen, Oxygen and Nitrogen [9]. The amino acids Cysteine and Methionine included Sulfur, Carbon, Hydrogen, Oxygen and Nitrogen [10]. The amino acids comprised of an amino (NH_2) and a carboxyl (COOH) group connected to the same carbon atom called alpha carbon [11]. Amino acids may be different in the side chain and R group which was linked to the alpha carbon [12]. The molecular weight (MW) of proteins might be defined as the mass of one mole of a protein which was usually measured in Daltons unit [13]. One Dalton is the atomic mass of one proton or neutron [14]. Each protein had a distinct and typical solubility in a defined surroundings and any variation in defined environmental conditions (buffer or solvent type, pH, ionic strength, temperature, etc.) might lead the proteins to drop the characteristic of solubility and precipitate out of the liquid medium [15]. The enormous characteristics of these twenty amino acids involved them in formation of many other kinds of protein molecules which were important for the life sustainability.

Similarly, one of the important proteins is Glucose-6-phosphate Dehydrogenase (G6PD). G6PD were observed for its transmissible range [16]. A single nucleotide change in the genetic code prompts to code a different than the normal amino acid and lead to produce as many as variants of G6PD [17]. It could be observed by a wide range of enzyme activity of the variants of G6PD [18]. Enzyme activity might be defined as the measure of the quantity of active enzymes present in a given surroundings and time or Enzyme activity = moles of substrate converted per unit time = rate× reaction volume [19, 20]. The SI unit of the Enzyme activity is katal (1 katal =1 $mols^{-1}$) since the unit is so large, therefore, it not in a practical use. The practical unit of enzyme activity is μmol min^{-1} i.e. 1 enzyme unit (U) = 16.67 nanokatals = 1 μmol min^{-1} [21].

Corresponding Author: Sanjay Kumar, Department of Botany, Nagaland University, HQs: Lumami 798627, Nagaland.

Now a day, G6PD deficiency is extremely frequent and wide-reaching [22]. The deficiency may be competent to stimulate hemolytic anemia by reactions of certain medicines (antibiotics, antipyretics or antimalarials) or incidence of an uncomplicated infection or disease or contamination or food poisoning [23]. The deficiency persuaded in an organism may be hereditary or attained and the diagnosis depended on the source and nature of the disease [24]. The inhibitors of G6PD were in exploration to treat cancers and other diseases for the reason that cell growth and cell proliferations were affected by G6PD [25]. The recognized inhibitor protein for the G6PD is didehydroepiandrosterone (DHEA). The DHEA has been also known as dehydroepiandrosterone or androstenolone or prasterone or 3β-hydroxyandrost-5-en-17-one or 5-androsten-3β-ol-17-one [22]. The main features of the DHEA involved as an endogenous steroid hormone, most circulating steroid, produced in adrenal glands, gonads and brain, functions as metabolic intermediate in the biosynthesis of androgen and estrogen sex steroids [26].

Therefore, the importance of G6PD and its inhibitor protein encouraged to explore the unknown functions and structure of putative uncharacterized protein sequence of G6PD from *Triticum aestivum* L. with accession number Q9LRJ1. Among the food crops, wheat (*Triticum aestivum* L.) is one of the most abundant sources of energy and proteins and its increased production is essential for food security [27]. In the present study describes the physicochemical properties, conserved domains, secondary structures, subcellular localization, homology modeling, 3D structure prediction, ramachandran plot and z-score prediction of the protein sequence.

MATERIALS AND METHODS

An accepted and uncharacterized protein sequence with accession number Q9LRJ1 was retrieved from the Uni-Prot Knowledgebase [28]. The physico-chemical properties of the protein sequence were predicted using ProtParam [29]. The presence of conserved domains within the sequence was predicted from conserved domain databases (CDD) at NCBI [30]. The presence of secondary structures (α-helix, β-sheets or coils) in the protein sequence was predicted by PSIPred [31]. The subcellular localization of the protein sequence was carried out by Plant-mPLoc [32- 35]. The homology modeling of the protein sequence was performed and searched from BLASTp with default parameters [36]. The 3-D structures were predicted of the homology modeling of the protein sequence using Modweb [37]. The Modweb generated 3-D structures were validated with the help of predicted Ramachandran plot from SAVS [38, 39]. The Z-score of the predicted 3-D structure were calculated by ProSA [40].

RESULTS

The physicochemical properties of the uncharacterized protein sequence (Accession number Q9LRJ1) from wheat were presented (Table 1). The conserved domains, secondary structures, subcellular localization, homology modeling, 3D structure prediction, ramachandran plot, z-score prediction and window slide prediction of amino acids were presented (Fig. 1-8).

Table 1: Physicochemical properties of the query protein sequence.

Number of amino acids	509
Total number of atoms	8129
Molecular Weight (dalton)	57747.8
Total number of negatively charged residues (Asp+Glu)	71
Total number of positively charged residues (Arg+Lys)	64
Theoretical pI	5.91
Empirical formula	$C_{2589}H_{4058}N_{698}O_{769}S_{15}$
Instability index	46.66
Aliphatic index	87.88
Grand average of hydropathicity (GRAVY)	-0.391
Extinction coeficients (M-1cm-1 at 280 nm in water)	60530 (abs 0.1%(=1 g/l)) 1.048, assuming all pairs of Cys residues form cystines
	60280 (abs 0.1%(=1 g/l)) 1.044, assuming all pairs of Cys residues are reduced
Estimated half life [N-terminal of the sequence considered is M (Met)]	30 h (mammalian reticulocytes, in vitro)
>20 h (yeast, in vivo)	
>10 h (E. coli, in vivo)	

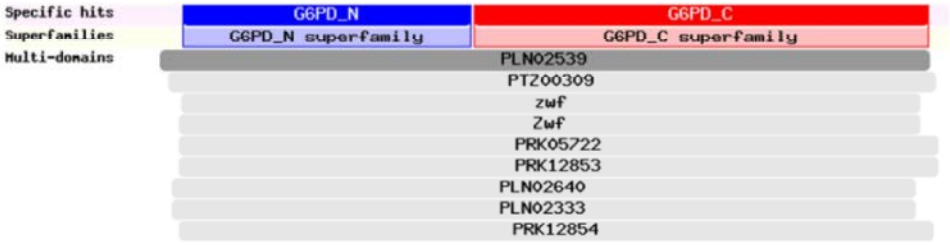


Fig. 1: Conserved domain prediction of protein sequence (G6PD).

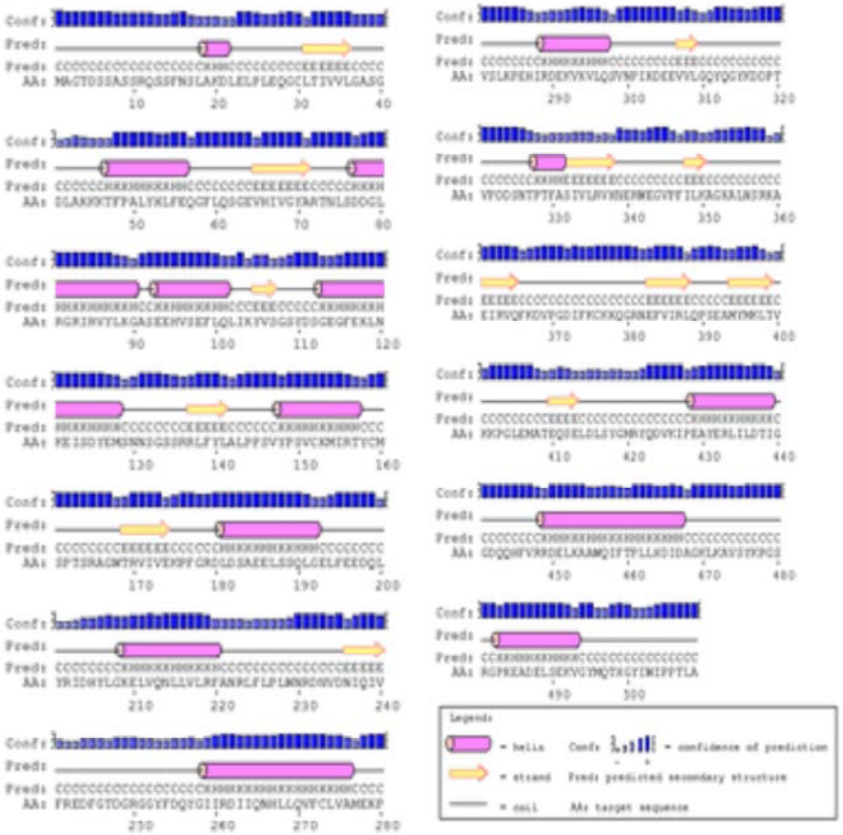
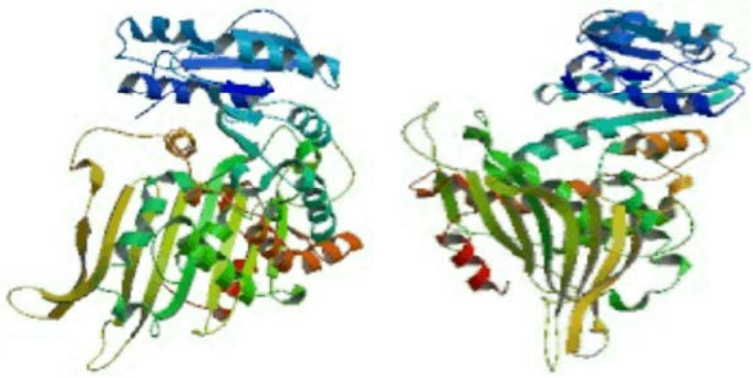


Fig. 2: Secondary structure prediction of the protein sequence by PSIPred.



(a) (b)

Fig. 3: 3D model predicted of the protein sequence. A, 4e9i and B, 2bhl

Fig. 4: Homology of the protein sequence with predicted 3D models.

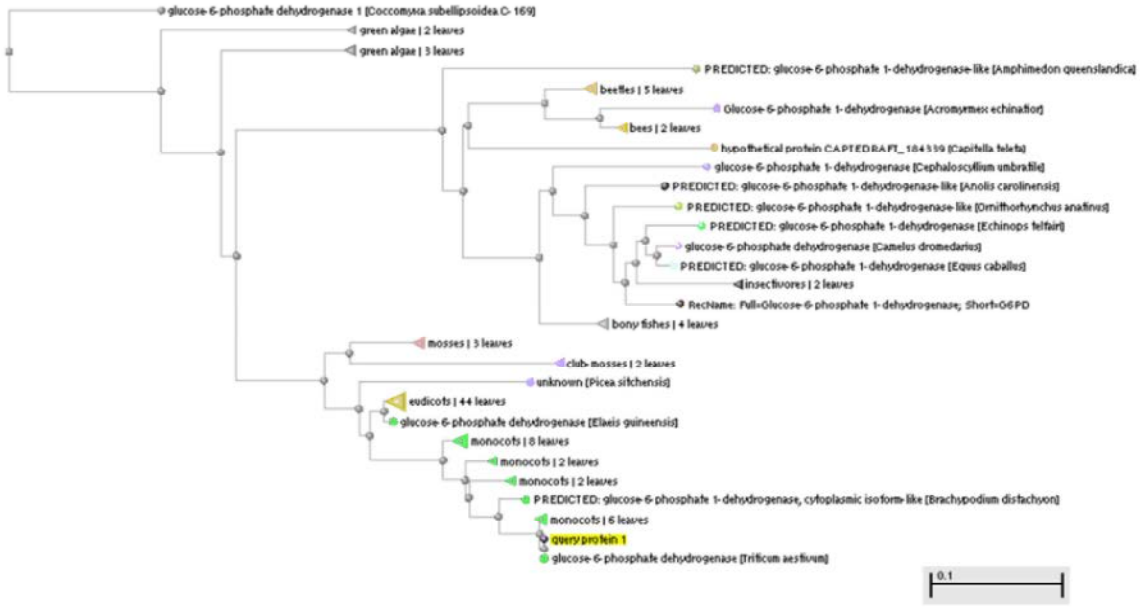


Fig. 5: BLAST prediction of the protein sequence (query protein 1).

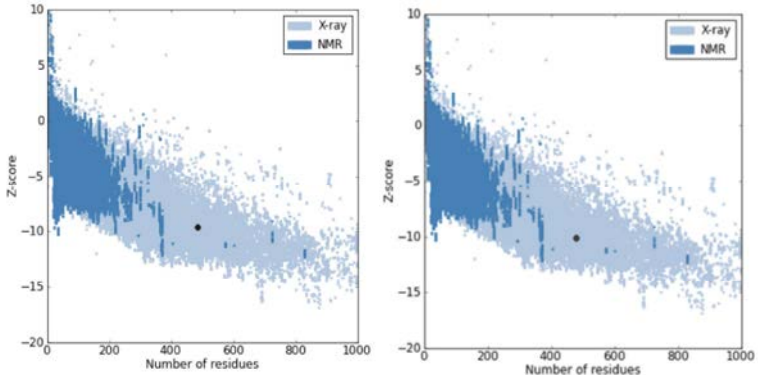


Fig. 6: Z-score prediction. A, 4e9i. B, 2bhl.

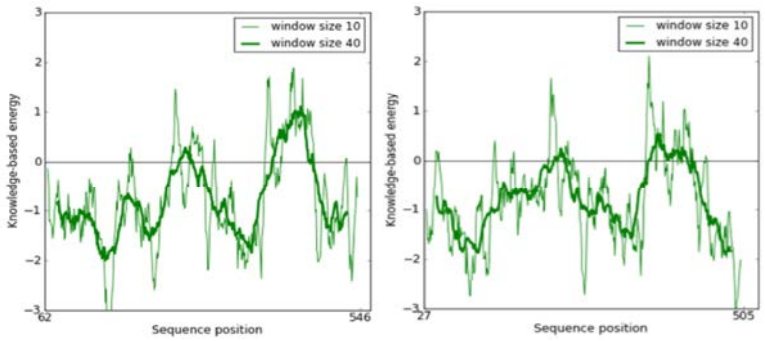


Fig. 7: To predict amino acid sequence position on window. A, 4e9i. B, 2bhl.

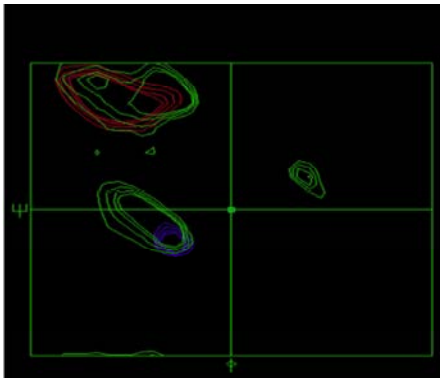


Fig. 8: Predicted Ramachandran plot for both the 3D models. Blue= helix, Red=strand, green=turns and loops.

DISCUSSION

The physicochemical properties of the uncharacterized protein sequence from wheat (*Triticum aestivum* L.) with accession number Q9LRJ1 showed 8129 total number of atoms, 509 amino acids in the sequence with molecular weight of 57747.8 daltons (Table 1). The high molecular weight (57747.8 daltons) of the sequence suggested the bulkiness of the protein in nature [41, 42]. The protein sequence exhibited the 71 negatively charged (Asp and Glu) and 64 positively charged (Arg and Lys) residues of amino acids (Table 1). Each protein has an amino group at one end and a carboxyl group at the other end as well as numerous amino acid side chains, some of which are charged [43]. Therefore, each protein carries a net charge [44]. The net protein charge is strongly influenced by the pH of the solution [45]. The isoelectric value of the protein sequence indicated 5.91 which suggested the sum total of protein as a negatively charged protein (Table 1). A protein carried a pH below their pI indicates a net positive charge on the protein and a pH above their pI indicates a net negative charge on the protein [46, 47]. The empirical formula of the protein was predicted as $C_{2589}H_{4058}N_{698}O_{769}S_{15}$ (Table 1). The instability index provides an estimate of the stability of protein in a test tube [48]. A protein whose instability index is smaller than 40 are predicted as stable and a value above 40 predicts that the protein may be unstable [49]. The aliphatic index of a protein is the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins [50]. The aliphatic and instability index of the desired protein was predicted as 87.88 and 46.66 which suggested the high protein thermal stability because of the presence of aliphatic side chains but unfortunately the protein showed less stability in test tube respectively (Table 1). The grand average of hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence [51]. The GRAVY showed -0.391 which suggested that the protein is hydrophilic and soluble in nature (Table 1). The extinction coefficient indicates how much light a protein absorbs at a certain wavelength [52]. The extinction coefficients of the protein showed 1.048 (all cysteine residues appear as half cystines) and 1.044 (no cysteine appears as half cystines) respectively (Table 1). The molar extinction coefficient of a protein could be estimated with the known amino acid composition. The molar extinction coefficient of tyrosine, tryptophan and cystine at a given wavelength could be used for the estimation of the extinction coefficient of a denatured protein. The reason might be the not appreciable absorbance of the wavelength >260 nm by cysteine than the cystine [53]. The half life is the prediction of time for half of the amount of protein in a cell to disappear after its synthesis in a cell [54]. The estimated half life of the protein sequence showed 30 hours in mammalian reticulocytes, >20 hours in yeast and > 10 hours in *E. coli* assuming N-terminal amino acid of the sequence (Table 1).

The number, type and distribution of nonpolar amino acid residues within the protein determine its hydrophobic character. The distinctive physical, chemical and biological properties associated with an amino acid are the result of the R group. There are 20 major amino acids that differ in their R-group. The R-group can be hydrophobic or polar, aromatic or aliphatic, charged or uncharged. The different R-groups are responsible for amino acids having different polarities, solubilities and chromatographic behaviour. The structure and biological function of a protein depend on its amino acid composition. Proteins are typically characterized by their size (molecular weight) and shape, amino acid composition and sequence, isolelectric point (pI), hydrophobicity and biological affinity. Therefore, the differences in these properties could be used as the basis for separation methods in a purification strategy.

The present study focused on sequence and structural analysis of assumed uncharacterized protein sequence with accession number Q9LRJ1 (g6pd) of *Triticum aestivum* L. The conserved domain analysis of the sequence indicated the presence of NAD type of multi binding domains [55]. The query sequence of 509 amino

acids indicated the two specific hits with superfamilies (G6PD_N and G6PD_C) which included the multidomains (PLN02539, PTZ00309, zwf, PRK05722, PRK12853, PLN02640, PLN02333, PRK12854) (Fig. 1). The secondary structure of the protein sequence was performed [56]. The prediction of secondary structures showed 13 α-helix, 13 β-sheets and mostly coils (Fig. 2). The subcellular localization of the protein sequence was computed by Plant-mPLoc and the predicted location found to be chloroplast and cytoplasm [57]. BLASTp was performed for the protein sequence and searched 3-D models for homology with default parameters against NCBI [58]. The two 3-D homology structures 4e9i and 2bhl were very close to the template protein sequence with 55 and 57 percent similarity respectively (Fig. 3). The sequence similarity of the query protein with the predicted 3D models was almost exactly similar (>=98%) and the two models were very close to each other (Fig. 4). The E-value is the number of different arrangements with a score equal to or superior than S which were anticipated to occur by chance in a data base search [59]. The predicted E-value score become more significant with the lower E-value [31]. The phylogenetic BLAST analysis form NCBI of the protein sequence suggested that the query protein was near to glucose 6 phosphate dehydrogenase of the *Triticum aestivum* and had some different protein sequences (Fig. 5).

The z-scores were calculated from the server and found to be -9.61for 4e9i and -10.13 for 2bhl which suggested the overall model quality respectively (Fig. 6). The result indicates that both the structures are very far from the mean. The z-score values also suggested that standard deviation of the models were below the average. The low value of z-score associated with very small p-values and found in the tails of normal distribution [60]. The z-score of a protein is defined as the energy separation between the native fold and the average of group of misfolds in the units of the standard deviation of the group [61]. The z-score is often used as a way of testing the knowledge-based potentials for their ability to recognize the native fold from other alternatives [62]. The close proximity to zero of the mean z-scores of all the global methods tested means that shuffled alignments score as highly as the original structurally unrelated global alignments [63]. The amino acid positions for both the models were predicted using ProtScale of window size 10 and 40 amino acids respectively (Fig. 7). Window size is the number of amino acids examined at a time to determine a point of hydrophobic character [64]. The window size is the length of the interval of amino acids used for the profile computation [65]. The window size (40) of both the models suggested the hydrophobic nature of the amino acids and makes the core of the protein. The ProtScale permitted computation and representation of the summary formed by any amino acid range on a preferred protein scale [66]. ProtScale might be used for presently 20 amino acids and additional predescribed scales made available on ExPASy. To produce data for a plan, the protein series were scanned with a sliding window of a given size. At each site, the mean scale value of the amino acids within the window was considered and that value was plotted for the midpoint of the window [67]. The predicted 3-D structures (4e9i and 2bhl) were further validated for ramachandran plot (Fig. 8). The plot suggested that the axis of α-helix rotating in the y-plane [68]. The planer peptide bonds of the α-helix rotate about axis and an open centre of the helix visible from the end of the a-helix [69]. The plot suggested the planes of peptide bonds and segments of twisted β-sheets [70]. The α-helix and β-sheets were surrounded by thick turns and loops [71, 72].

Therefore, it might be suggested that several other uncharacterized proteins of plants, animals, bacteria and fungi which have important roles in biological functions may be predicted for their present and future roles and functions and also search for new functions of that proteins with the help of bioinformatics approach.

REFERENCES

1. Wisniak, J., 2000. Jöns Jacob Berzelius A Guide to the Perplexed Chemist. The Chem. Educat. 5(6): 343-350.
2. Ritchie, D.W., 2008. Recent progress and future directions in protein-protein docking. Curr. Prot. Pept. Sci., 9(1): 1-15.
3. Sleator, R.D., 2012. Prediction of protein functions. Methods Mol. Biol., 815: 15-24.
4. Samarin, S. and A. Nusrat, 2009. Regulation of epithelial apical junctional complex by Rho family GTPases. Front. Biosci.: a Journal and Virtual Library, 14(14): 1129-1142.
5. Copland, J.A., M. Sheffield-Moore, N. Koldzic-Zivanovic, S. Gentry, G. Lamprou, F. Tzortzatou-Stathopoulou, V. Zoumpourlis, R.J. Urban and S.A. Vlahopoulos, 2009. Sex steroid receptors in skeletal differentiation and epithelial neoplasia: is tissue-specific intervention possible? BioEssays: news and reviews in mol. cellul. Development. Biol., 31(6): 629-641.

6. Kent, S.B., 2009. Total chemical synthesis of proteins. Chemical Society Reviews, 38(2): 338-351.
7. Joos, T. and J. Bachmann, 2009. Protein microarrays: potentials and limitations. Frontiers in Bioscience, 14(14): 4376-4385.
8. Zhou, Z.H., 2008. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Curr. Opin. Struct. Biol., 18(2): 218-228.
9. Keskin, O., N. Tuncbag and A. Gursoy, 2008. Characterization and prediction of protein interfaces to infer protein-protein interaction networks. Curr. Pharmaceut. Biotech. 9(2): 67-76.
10. Zhang, Y., 2008. Progress and challenges in protein structure prediction. Curr. Opin. Struct. Biol., 18(3): 342-348.
11. Conrotto, P. and S. Souchelnytskyi, 2008. Proteomic approaches in biological and medical sciences: principles and applications. Expt. Oncol., 30(3): 171-180.
12. Mayhew, T.M. and J.M. Lucocq, 2008. Developments in cell biology for quantitative immunoelectron microscopy based on thin sections: a review. Histochem. Cell Biol., 130(2): 299-313.
13. Stepanenko, O.V., V.V. Verkhusha, I.M. Kuznetsova, V.N. Uversky and K.K. Turoverov, 2008. Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes. Curr. Prot. Pept. Sci., 9(4): 338-369.
14. Hey, J., A. Posch, A. Cohen, N. Liu and A. Harbers, 2008. Fractionation of complex protein mixtures by liquid-phase isoelectric focusing. Methods Mol. Biol., 424: 225-239.
15. Standley, D.M., A.R. Kinjo, K. Kinoshita and H. Nakamura, 2008. Protein structure databases with new web services for structural biology and biomedical research. Briefings in Bioinformatics, 9(4): 276-285.
16. Beutler, E., 2008. Glucose-6-phosphate dehydrogenase deficiency: a historical perspective. Blood, 111(1): 16-24.
17. Lozano, R., 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet, 380(9859): 2095-2128.
18. Corpas, F.J., J.B. Barroso, L.M. Sandalio, S. Distefano, J.M. Palma, J.A. Lupiαñez and L.A. Del Río, 1998. A dehydrogenase-mediated recycling system of NADPH in plant peroxisomes. Biochem. J., 330(2): 777-784.
19. Daniel, R.M., M.E. Peterson, M.J. Danson, N.C. Price, S.M. Kelly, C.R. Monk, C.S. Weinberg, M.L. Oudshoorn and C.K. Lee, 2010. The molecular basis of the effect of temperature on enzyme activity. Biochem. J., 425(2): 353-360.
20. NC-IUB, 1979. Nomenclature Committee of the International Union of Biochemistry (NC-IUB) Units of Enzyme Activity. Eur. J. Biochem. 97(2): 319-320.
21. Wienken, C.J., P. Baaske, U. Rothbauer, D. Braun and S. Duhr, 2010. Protein-binding assays in biological liquids using microscale thermophoresis. Nature Communicat. 1(7): 100.
22. Cappellini, M.D. and G. Fiorelli, 2008. Glucose-6-phosphate dehydrogenase deficiency. Lancet, 371(9606): 64-74.
23. Lippi, G., F. Schena, G.L. Salvagno, R. Aloe, G. Banfi and G.C. Guidi, 2012. Foot-strike haemolysis after a 60-km ultramarathon. Blood Transfus. 10(3): 377-383.
24. Baker, W.L., S. Karan and A.M. Kenny, 2011. Effect of dehydroepiandrosterone on muscle strength and physical function in older adults: a systematic review. J. Am. Geriat. Soc., 59(6): 997-1002.
25. Tian, W.N., L.D. Braunstein, J. Pang, K.M. Stuhlmeier, Q.C. Xi, X. Tian and R.C. Stanton, 1998. Importance of glucose-6-phosphate dehydrogenase activity for cell growth. J. Biol. Chem., 273(17): 10609-10617.
26. Lindschau, C., T. Kirsch, U. Klinge, P. Kolkhof, I. Peters and A. Fiebeler, 2011. Dehydroepiandrosterone-induced phosphorylation and translocation of FoxO1 depend on the mineralocorticoid receptor. Hypertension, 58 (3): 471-478.
27. Kumar, S. and A.K. Singh, 2010. A review on herbicide 2, 4-D damage reports in Wheat. J. Chem. Pharmaceut. Res., 2(6): 118-124.
28. Uniprot, C., 2010. Ongoing and future developments at the Universal Protein Resource. Nuc. Acids Res. 39 (Database issue): D214-D219.
29. Shenoy, S.R. and B. Jayaram, 2010. Proteins: Sequence to Structure and Function: Current Status. Curr. Prot. Pept. Sci., 11: 498-514.
30. Marchler-Bauer, A., C. Zheng, F. Chitsaz, M.K. Derbyshire, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, S. Lu, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang and S.H. Bryant, 2012. CDD: conserved domains and protein three-dimensional structure. Nuc. Acids Res. 41 (data base issue): D348-D352.

31. Buchan, D.W.A., S.M. Ward, A.E. Lobley, T.C.O. Nugent, K. Bryson and D.T. Jones, 2010. Protein annotation and modeling servers at university college London. Nuc. Acids Res., 38 (web server issue): W563-W568.
32. Magnus, M., M. Pawlowski and J.M. Bujnicki, 2012. Meta Loc GramN: a meta-Predictor of protein subcellular localization for Gram-negative bacteria. BBA-prot. Proteom. 1824(12): 1425-1433.
33. Goldberg, T., T. Hamp and B. Rost, 2012. LocTree 2 predicts localization for all domains of life. Bioinformatics, 28: i458-i465.
34. Chou, K.C., Z.C. Wu, X. Xiao and Iloc-Euk, 2011. A multi-label classifier for predicting the subcellular localization of singleplex and multiplex eukaryotic protein. Plos ONE, 6: e18258.
35. Mooney, C.m, Y.H. Wang and G. Pollastri, 2011. SCLPred: protein subcellular localization prediction by N to 1 neural networks. Bioinformatics, 27(20): 2812-2819.
36. Oehmen, C.S. and D.J. Baxter, 2013. ScalaBLAST 2.0: Rapid and robust BLAST calculations on multiprocessor systems. Bioinformatics, 29 (6): 797-798.
37. Pieper, U., B.M. Webb, D.T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E.C. Meng, E.F. Pettersen, C.C. Huang, R.S. Datta, P. Sampath Kumar, M.S. Madhusudhan, K. Sjolander, T.E. Ferrin, S.K. Burley and A. Sali, 2011. Modbase, a database of annotated comparative protein structure models and associated resources. Nuc. Acids Res. 39 (database issue): D465-D474.
38. Read, R.J., P.D. Adams, W.B. 3rd Arendall, A.T. Brunger, P. Emsley, R.P. Joosten, G.J. Kleywegt, E.B. Krissinel, T. Lutteke, Z. Otwinowski, A. Perrakis, J.S. Richardson, W.H. Sheffler, J.L. Smith, I.L. Tickle, G. Vriend and P.H. Zwart, 2011. A new generation of crystallographic validation tools for the protein data bank. Structure, 19(10): 1395-1412.
39. Ting, D., G. Wang, R. Mitra, M.I. Jordan and R.L. Dunbrack, 2010. Neighbour-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. PLOS computat.l Biol., 6(4): e1000763.
40. Wiederstein, M. and M.J. Sippl, 2007. ProSA-Web: interactive web service for the recognition of errors in three dimensional structures of proteins. Nuc. Acids Res., 35: W407-W410.
41. Rani, S., P. Sharma, A. Sharma and R. Chatrath, 2013. Comparative computational analysis of ADP Glucose Pyrophosphorylase in Plants. Bioinformation, 9(11): 572-576.
42. Saha, A., Md. J. Ashraf, K.U.M. Alam and A. Mannan, 2013. Computational analysis of bovine alpha-1collagen sequences. Bioinfrmation, 9(1): 42-48.
43. Huang, Yin-Fu and Shu-Ying Chen, 2013. Extracting physicochemical features to predict protein secondary structure. The Scient. World J. article ID 347106, 8 pp, http://dx.doi.org/10.1155/2013/347106
44. Coletta, A., J.W. Pinney, D.Y.W. Solis, J. Marsh, S.R. Pettifer and T.K. Attwood, 2010. Low-complexity regions within protein sequences have position dependent roles. BMC Syst. Biol., 4 : 43.
45. Goel, C., A. Barat, V. Pande and P.K. Sahoo, 2013. Comparative in silico analysis of Mb1 homologues of Teleosts. World J. Fish and Marine Sci., 5(4): 426-429.
46. Fanq, Y. and J. Fanq, 2013. Discrimination of soluble and aggregation-prone proteins based on sequence information. Mol. Biosyst. 9: 806-811.
47. Roy, A., A. Kucukural and Y. Zhang, 2010. I-TASSER: a unified plateform for automated protein structure and function prediction. Nat. Protoc., 5(4): 725-738.
48. Pascual-Garcia, A., D. Abia, R. Mendez, G.S. Nido and U. Bastolla, 2010. Quantifying the evolutionary divergence of protein structures: the role of function change and function conservation. Proteins, 78(1): 181-196.
49. Satpathy, R., R. Behera, S. Ku. Padhi and R.K. Guru, 2013. Computational phylogenetic study and data mining approach to Laccaseenzyme sequences. J. Phylogenet. Evolut. Biol. 1(2): 1000108. Doi:10.4172/2329-9002.1000108.
50. Bakisy, O.H.H. and O.U. Sezerman, 2012. Inferring phylogenies from physicochemical properties of DNA. Am. J. Bioinformatics Res., 2: 1-6.
51. Zha, H.G., T. Liu, J.J. Zhou and H. Sun, 2013. MS-desi, a desiccation-related proteins in the floral nectar of the overgreen velvet bean (Mucuna sempervirens Hems): molecular identification and characterization. Planta, 238(1): 77-89.

52. Sanna, D., K. Varnagy, S. Tiwari, G. Micera and E. Garribba, 2011. VO^{2+} complexion by bioligands showing keto-enol tautomerism: A potentiometric, stereoscopic and computational study. Inorg. Chem. 50: 10328-10341.
53. Gill, S.C. and P.H. von Hippel, 1989. Calculation of Protein Extinction Coeficients from amino acid sequence data. Analyt. Biochem. 182: 319-326.
54. Mbah, A.N. and R.D. Isokpehi, 2013. Identification of functional regulatory residues of the β-lactan inducible penicillin binding protein in methicillin-resistant staphylococcus aureus. Chemother. Res. Pract. article ID 614670, 10pp, http://dx. doi. org/10. 1155/2013/614670.
55. Holm, L. and P. Rosenstrom, 2010. Dali Server: conservation mapping in 3D. Nuc. Acids Res., 38: W545-W549.
56. Pandit, S.B., M. Brylinski, H. Zhou, M. Gao, A.K. Arakaki and J. Skolnick, 2010. PSiFR: an integrated resource for prediction of protein structure and function. Bioinformatics, 26(5): 687-688.
57. Brylinski, M., 2013. Evolver: an optimization engine for evolving protein sequences to stabilize the respective structure. BMC Res. Notes, 6(1): 303.
58. Zhao, K. and X. Chu, 2014. G-BLASTN: accelerating nucleotide alignment by graphics processors. Bioinformatics, 30(10): 1384-1391.
59. Kaczanowski, S. and P. Zielenkiewicz, 2010. Why similar protein sequences encode similar three-dimensional structure? Theor. Chem. Accounts, 125: 543-550.
60. Kryshtafovych, A., K. Fidelis and A. Tramontano, 2011. Evaluation of model quality predictions in CASP9. Proteins, 79(suppl 10): 91-106.
61. Konopka, B.M., J.C. Nebel and M. Kotulska, 2012. Quality assessment of protein model-structures based on structural and functional similarities. BMC Bioinformatics, 13: 242.
62. Zhang, L.I. and J. Scolnick, 1998. What should be the Z-score of native protein structures be? Prot. Sci., 7: 1201-1207.
63. Kalman, M. and N. Ben-Tal, 2010. Quality assessment of protein model structures using evolutionary conservation. Bioinformatics, 26(10): 1299-1307.
64. Brylinski, M. and W.P. Feinstein, 2013. eFindSite: Improved prediction of ligand binding sites in protein models using meta-threading, mavhine learning and auxiliary ligands. J. Comput. Aided Mol. Des., 27(6): 551-567.
65. Brylinski, M., 2013. The utility of artificially evolved sequence in protein threading and fold recognition. J. Theor. Biol., 328:77-88.
66. Benkert, P., M. Biasini and T. Schwede, 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27(3): 343-350.
67. Mishra, A., S. Rao, A. Mittal and B. Jayaram, 2013. Capturing native/ native like structures with phsico-chemical metric (pcSM) in protein folding. BBA-prot. Proteom. 1834(8): 1520-1531.
68. Lovell, S.C., I.W. Davis, W.B. 3rd Arendall, P.I. De Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson, 2003. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins, 50(3): 437-450.
69. Brylinski, M., 2013. Unleashing the power of meta-threading for evolution/structure based function inference of proteins. Front. Genet., 4: 118.
70. Chen, Yen-Ru, Peng Sheng-Lung and Tsay Yu-Wei, 2008. Protein secondary structure prediction based on ramachandran maps. Lectures Notes in Comput. Sci., 5226: 204-211.
71. Zhou, A.Q., C.S. O'Hern and L. Regan, 2011. Revisiting the ramachandran plot from a new angle. Prot. Sci., 20: 1166-1171.
72. Peng, J. and X. Jinbo 2011. A multiple template approach to protein threading. Proteins, 79(6): 1930-1939.