ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Computational Sequence Analysis and Structure Prediction of Jack Bean Urease

Shashank Gupta[1], Atul Kathait[1], Vineet Sharma[1*]

School of Biosciences, Apeejay Stya University, Sohna-Palwal Road, Gurgaon 122103, Haryana, India

Manuscript Info

Manuscript History:

Received: 11 December 2014
Final Accepted: 22 January 2015
Published Online: February 2015

Key words:

Jack Bean Urease, Homology Modelling, Computational biology, Secondary Structure

Corresponding Author

Vineet Sharma

Abstract

Ureases are ubiquitous in nature having similar catalytic function yet sometimes different protein structures. Catalytic hydrolysis of urea results in the production of ammonia causing abrupt pH increase of the surrounding medium. This increase in pH has has been shown to have adverse effects on human health and agriculture. The efficiency of soil nitrogen fertilization with urea gets decreased due to activity of soil ureases. Large amount of urease present in ureolytic bacteria has been also shown as a major cause of human diseases. Both bacterial and plant ureases have been shown to posses high sequence similarity, suggesting that they may have similar three-dimensional structures and a conserved catalytic mechanism. In the present work, homology modelling was performed to construct 3D structural model of jack bean urease. Homology modelling of jack bean urease can be used to study the mechanism of substrate/inhibitor binding with its active site. Computational studies based on these homology models, may provide informations for structure based drug designing against bacterial ureases.

INTRODUCTION

Ureases (urea amidohydrolases, EC 3.5.1.5) are a group of enzymes found in nature among plants, bacteria, algae, fungi and invertebrates [Mobley and Hausinger 1989], with different protein structures, but proposed to have common catalytic function. Urease catalyzes hydrolysis of urea and produce ammonia and carbonic acid. The carbamate produced during this reaction spontaneously decomposes, at physiological pH, to give a second molecule of ammonia and bicarbonate [Mobley *et al*., 1995]. The hydrolysis of the reaction products causes an abrupt overall pH elevation, the major cause for the negative side effects of the action of urease both for human and animal health, and as well for agriculture also. Urease serves as a virulence factor in human and animal infections of the urinary and gastrointestinal tracts. They are also involved in kidney stone formation, pyelonephritis, catheter encrustation, ammonia encephalopathy, hepatic coma, and urinary tract infections [Mobley *et al*., 1995; Collins and D'Orazio, 1993]. The ureolytic activity of *Helicobacter pylorii* is also the major cause of pathologies (including cancer) induced by gastro-duodenal infections by such microorganisms. In agriculture context, urea used as a soil fertilizer, representing a fundamental source of nitrogen for plant nutrition, worldwide. This is made possible by the large amount of urease in soil present both in living ureolytic bacteria [Severo *et al*., 2006] and as extracellular urease. The latter is found aggregated with clays and humic substances, which prevent its degradation by extracellular proteolytic enzymes and microorganisms. The efficiency of soil nitrogen fertilization with urea is severely decreased by this widespread urease activity, which causes the release of large amounts of ammonia in the atmosphere and further induces plant damage by ammonia toxicity and soil pH increase, thereby posing significant environmental and economic problems [Ciurli *et al.*, 1999]. The bacterial and plant ureases have been found to have sequence

similarity [Follmer, 2008] but mostly different in terms of number of subunits. Among the plant ureases, only the crystallization and preliminary X-ray analysis of JBU were reported earlier [Jabri *et al*., 1992, Sheridan *et al*., 2002]. Computational approaches can provide homology models, which can be used in molecular dynamic simulations, and automatic docking in order to demonstrate the function of proteins and to illustrate the mode of substrate binding [Grundy, 1998]. These types of methods can be used successfully in enzyme–substrate systems and can provide useful information for future studies. The aim of this study was to perform computational sequence analysis and 3D structural modelling of a jack bean urease.

METHODS

Sequence sources

Jack bean urease protein sequence (P07374.3) was retrieved from the National Centre for Biotechnology Information (NCBI) database.

Analysis of proteins sequence

The amino acid composition of the urease sequences was computed using the PEPSTATS analysis tool [Rice *et al*., 2000]. The physicochemical parameters such as the molecular weight, isoelectric point (pI), aliphatic index, amino acid property, extinction coefficient, half-life instability index, and grand average hydropathy (GRAVY) were calculated using the ProtParam tool of the ExPASy proteomics server [Gasteiger *et al*., 2003]. N-glycosylation sites of the proteins sequence was evaluated by using the **NetNglyc 4.0** [Gupta and Brunak, 2002]. Sub-cellular localization of any protein is important understanding protein function. Prediction of sub-cellular localization of protein was carried out by **CELLO v.2.5** (subCELlular LOcalization predictor [*Yu et al.,* 2004]. Prediction of transmembrane regions and signal peptides in protein sequences were analyzed with **TMHMM program** [Krogh *et al*., 2001] to decide whether the enzymes contain transmembrane sequences. The presence of signal sequences and their possible cleavage sites were predicted with the **SIGNALP V4.1** program [Nielsen *et al.,* 1997] using hidden Markov model method.

Secondary structure prediction

The sequence of jack bean urease was used for the secondary structure prediction using **PDBsum tool** [Laskowski, 2001]. PDBsum tool was employed for calculating the secondary structural features of the protein sequence considered for this study.

Three-Dimensional structure prediction

Protein 3D structure is very important in understanding the protein interactions, functions and their localization. The three dimensional structure of protein was predicted with the **I-TASSER** server [Zhang, 2008]. 3D models were built from multiple alignments of protein sequences with known structure and function [Zhao *et al*., 2010]. Model evaluation was performed by PROCHECK server [Laskowski *et al.,* 1993] for Ramachandran plot, Verify3D [Eisenberg *et al*., 1997], ERRAT [Colovos and Yeates, 1993] is performed for validation. PyMOL (Schrödinger Inc.) was used to generate publishable images of the jack bean urease model.

Function annotation of the protein

To functionally annotate the jack bean urease protein, ProFunc was used and to find the conserved domains in protein to identify its family, it was searchedagainst close orthologous family members. NCBI Conserved Domain Database (NCBI CDD) [Bauer *et al*., 2007] was used to find the conserved domains or ancient domains in the protein sequence.

RESULTS AND DISCUSSION

Analysis of Protein sequence

The primary structure analysis of jack bean urease protein showed that the most abundant amino acid is glycine which accounts for 9.4% and Alanine 8.8% of the enzyme's primary structure respectively. The least common amino acids observed were tryptophan (0.47%) and cysteine (1.78%). The computed isoelectric point (pI) of urease was 6.05; indicates that enzyme is likely to precipitate in acidic buffers. The extinction coefficient (EC) of urease was 53290. The instability indices (Ii) of urease was 31.75, which was below 40, this classifies the protein as stable. Urease has negative GRAVY scores attesting to their solubility in hydrophilic solvents. The aliphatic index (Ai) which evaluates the relative volume occupied by the side chains of hydrophobic amino acids was generally higher in urease. A high aliphatic index indicates that a protein may remain stable over a wide range of temperatures. Glycosylation is the most abundant and diverse posttranslational modification of proteins [Steentoft *et al*., 2013]. The NetOglyc server produces neural network predictions of mucin type GalNAco-glycosylation sites in mammalian proteins [Steentoft *et al*., 2013]. NetOGlyc predicts that 16 sites of urease protein are glycosylated at 114, 120, 127, 246, 258, 262, 438, 441, 442, 444, 446, 448, 570, 571, 658 and 788 positions. Cellular functions are

often localized in specific compartments; therefore, predicting the subcellular localization of unknown proteins can give information about their functions and can also help in understanding disease mechanisms and developing drugs. The subcellular localization prediction using CELLO predicted that our protein is a cytoplasmic protein. Transmembrane regions and signal peptides were analysed in urease protein sequences, there are no transmembrane regions and signal peptides was found.

Secondary structure prediction

The predicted secondary structure composition of jack bean urease protein was determined using the PDBsum tool. The secondary structure prediction server revealed that 25.5% of amino acids resided in α-helices, while 19.3% of residues were in β-sheets and 3.5% of residues were in 3_{10} helix **(Fig 1).** The rest of the amino acids were found in other conformations such as β-hairpins and β-turns.

Three-Dimensional structure prediction

The 3D structure of jack bean urease was generated using I-TESSAR server. In this method the target sequences are first threaded using a representative PDB structure library to search for the possible folds by Profile- Profile Alignment (PPA), PSI-BLAST profiles, Hidden Markov Model, Needleman-Wunch and Smith-Waterman alignment algorithms. The top 10 alignments obtained are from the following threading programs i.e. MUSTER, dPPAS, Neff-PPAS, wdPPAS, PPAS, SPARKS-X, SP3, PROSPECT2, HHSEARCH2F and FAS03. The PDB ID: 3LA4 had the best Z-score using all the ten algorithms and was used for modelling urease structure **(Table 1)**. I-TASSER server predicted 5 models from which the model with best C-Score of 2.0 was selected with estimated accuracy of 0.99±0.03 (TM-Score) and 3.3±2.3 Å (RMSD) C-score is a confidence score for estimating the quality of predicted models by I-TASSER server. It is calculated on the basis of the significance of threading template alignments and the convergence parameters of the structure assembly simulations. Validation of the model including the geometric properties of the backbone conformations, were analyzed using various structure evaluation programs. Ramachandran plot calculations were calculated with PROCHECK program. Ramachandran plot of the model was shown in **Fig 2**. 80.1 % residues are in most favoured regions, 17.9% residues are in additionally allowed region, there are no residues in generously allowed region and 0.4% residues are in disallowed region. These results revealed that the majority of the amino acids are in a phi-psi distribution that is consistent with a right-handed α-helix and the model is reliable and stable. ERRAT **(Fig 3)** and Verify3D **(Fig 4)** confirm the quality of predicted 3D structure by I-TASSER as more reliable and within an accepted range **(Table 2)**. The final model predicted by I-TASSER was visualized by pymol software shown in **Fig 5**.

Function annotation of the protein

To hypothetically annotate the function of the urease protein ProFunc was used. It was discovered that protein is involved in three biological processes: metabolic process, cellular process and nitrogen metabolic process. The biochemical function of the protein is catalytic activity and hydrolase activity acting on carbon\nitrogen bonds. To investigate functional family of protein NCBI Conserved Domain Database (NCBI CDD) was used. The result showed that the urease protein is nickel-dependent metalloenzyme belongs to metallo-dependent hydrolyses superfamily. The enzyme consists of 3 subunits, alpha, beta and gamma, which can be fused and present on a single protein chain and which in turn, forms multimers, mainly trimers. The large alpha subunit is the catalytic domain containing an active site with a bi-nickel center complexed by a carbamylated lysine. The beta and gamma subunits play a role in subunit association to form the higher order trimers.

CONCLUSION

Urease serves as a virulence factor in human and animal infections of the urinary and gastrointestinal tracts. [Mobley *et al*., 1995 Collins and D'Orazio, 1993]. Urease is considered as a significant drug target for various diseases. In the present work, a homology based 3D model of jack bean urease is constructed using I-Tasser server. I-Tasser predict 5 models with different Z-score. The best models produced by server were further assessed by Procheck, Errat and Verify3D. I-Tasser server produced satisfactory Ramachandran plot statistics, Errat plot quality factor. Based on the ProFunc server, jack bean urease is nickel-dependent metalloenzyme belongs to metallo-dependent hydrolases superfamily. Since, structurally plant urease is similar to the bacterial urease; we can use homology modelling technique to develop a model for bacterial urease which can be further use for drug discovery.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

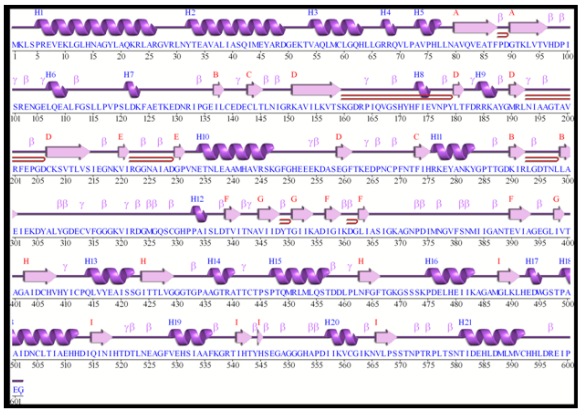

Figure 1: Schematic showing the secondary structural elements in the urease protein. α-helices are labeled with the letter "H", and β-strands are lettered in uppercase. β, γ, and hairpin turns are also labeled. (The secondary motif map and topology diagram were calculated using the PDBsum tool.

Figure 1: Schematic showing the secondary structural elements in the urease protein. α-helices are labeled with the letter "H", and β-strands are lettered in uppercase. β, γ, and hairpin turns are also labeled. (The secondary motif map and topology diagram were calculated using the PDBsum tool.

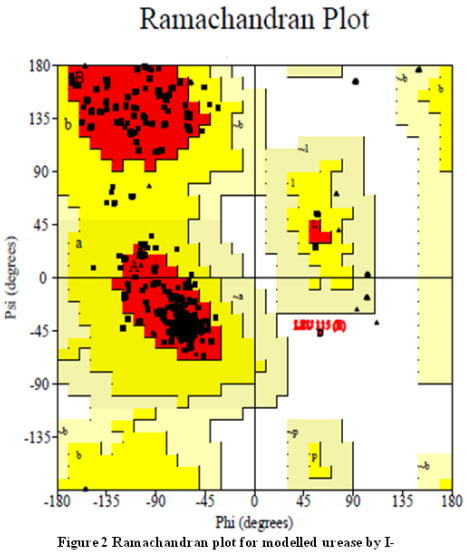

Figure 2 Ramachandran plot for modelled urease by I-TASSER.

Figure 2: Ramachandran plot for modelled urease by I-TASSER.

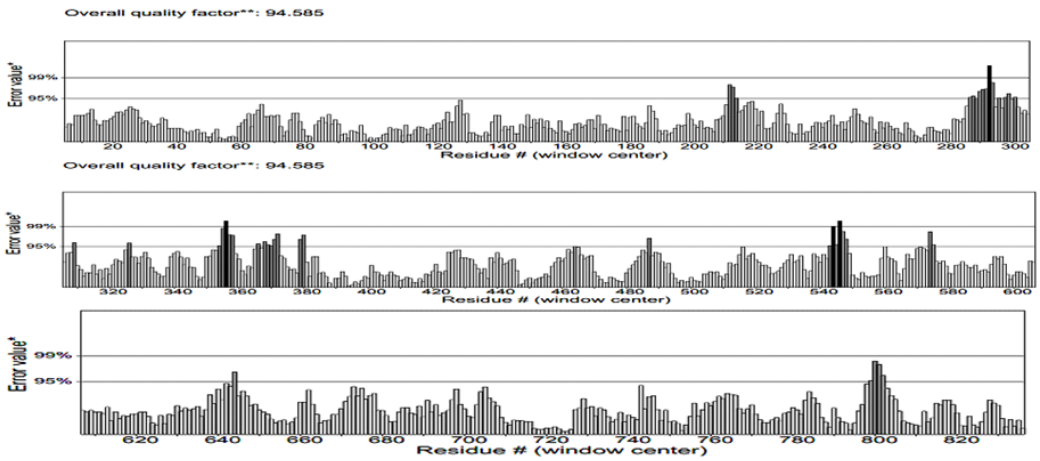

Figure 3: Errat results for Predicted Jack Bean Urease.

Figure 3: Errat results for Predicted Jack Bean Urease.

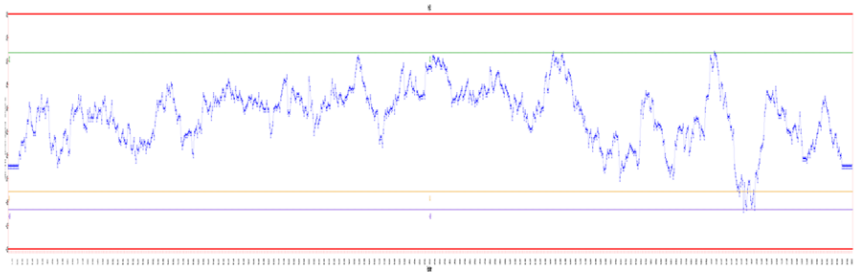

Figure 4: Verify 3D for Predicted Jack Bean Urease.

Figure 4: Verify 3D for Predicted Jack Bean Urease.

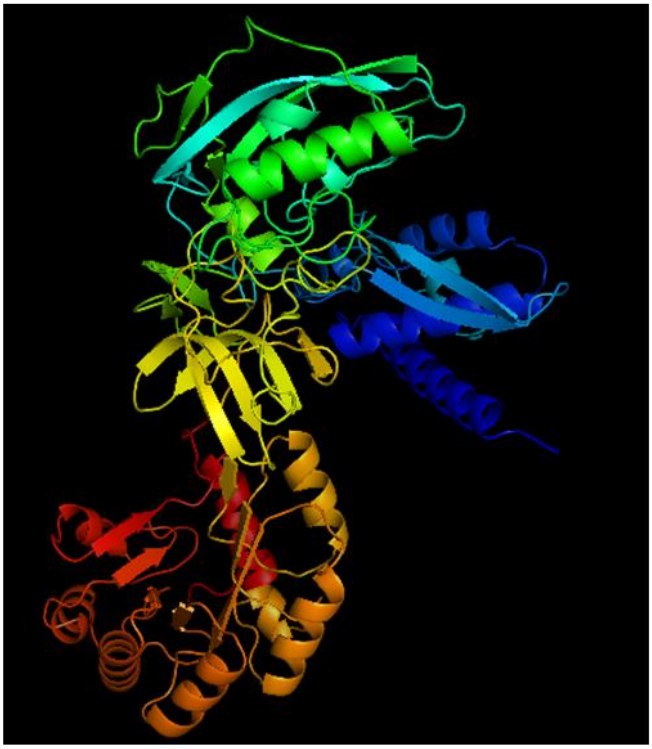

Figure 5: Predicted 3D structure by I-TASSER, visualized by pymol.

Figure 5: Predicted 3D structure by I-TASSER, visualized by pymol

REFERENCES

- Ciurli, S., Benini, S. and Rypniewski., W.R., et al. (1999): Structural properties of the nickel ions in urease: novel insights into the catalytic and inhibition mechanisms. Coord. Chem. Rev., 192:331-355.
- Collins, C.M. and D'Orazio, S.E. (1993): Bacterial ureases: structure, regulation of expression and role in pathogenesis. Mol. Microbiol., 9:907-913.
- Colovos, C. and Yeates, T.O. (1993): Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci., 2: 1511-1519.
- Eisenberg, D., Luthy, R. and Bowie, J.U. (1997): VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods in Enzymol, 277:396-404.
- Follmer, C. (2008): Insights into the role and structure of plant ureases. Phytochemistry, 69:18-28.
- Gasteiger, E., Gattiker, A. and Hoogland, C., et al. (2003): ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31:3784-3788.
- Grundy, W.N. (1998): Homology detection via family pairwise search. J. Comput. Biol. 5(3), 479-491.
- Gupta, R. and Brunak, S. (2002): Prediction of glycosylation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput., 310-322.
- Jabri, E., Lee, M.H. and Hausinger, R.P., et al. (1992): Preliminary crystallographic studies of urease from jack bean and from Klebsiella aerogenes. J. Mol. Biol., 227:934-937.
- Krogh, A., Larsson, B. and Von, H.G., et al. (2001): Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol., 305 (3): 567-580.
- Laskowski, R. A. (2001): PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res., 29:221-222.

- Laskowski, R.A., MacArthur, M.W. and Moss, D.S., et al. (1993): PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst., 26:283-291.
- Bauer, M.A., Anderson, J.B. and Derbyshire, M.K., et al. (2007): CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res., 35:D237-D240.
- Mobley, H.L. and Hausinger, R.P. (1989): Microbial ureases: significance, regulation, and molecular characterization. Microbiol. Rev., 53 (1): 85-108.
- Mobley, H.L., Island, M.D. and Hausinger R.P. (1995): Molecular biology of microbial ureases. Microbiol. Rev. 59 (3):451-480
- Nielsen, H., Engelbrecht, J. and Brunak, S., et al. (1997): Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng., 10 (1):1-6.
- Severo, O. D., Wassermann, G. E., and Carlini, C. R. (2006): Ureases display biological effects independent of enzymatic activity: Is there a connection to diseases caused by urease-producing bacteria?. Braz. J. Med. Biol. Res., 39:851-861
- Rice, P., Longden, I. and Bleasby, A. (2000): EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics, 16: 276-277.
- Sheridan, L., Wilmot, C. M. and Cromie, K. D., et al. (2002): Crystallization and preliminary X-ray structure determination of jack bean urease with a bound antibody fragment. Acta Crystallogr. D. Biol. Crystallogr., 58(Pt 2): 374-376.
- Steentoft, C., Vakhrushev, S.Y. and Joshi, H.J., et al. (2013): Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBOJ, 32: 1478-1488.
- Yu, C.S., Lin, C.J. and Hwang, J.K. (2004): Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci., 13: 1402-1406.
- Zhang, Y. (2008): I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9: 40.
- Zhao, F., Peng, J. and Debartolo, J., et al. (2010): A Probabilistic and Continuous Model of Protein Conformational Space for Template-Free Modeling. J. Comput. Biol., 17(6):783-798.