SHORT COMMUNICATION Open Access

ISSN 2278-1404

International Journal of Fundamental & Applied Sciences

OMICS OF asnA GENE FROM *Enterobactor aerogenes* KCTC2190

Rajeswara Reddy Erva[1*], Satish Babu.R[2], CVK Reddy[1], Diwakar Reddy K[1], Sugunakar YJ[1]

[1]Department of Biotechnology,Jawaharlal Nehru Technological University Anantapur College of Engineering Pulivendula, Pulvendula 516390, Andhra Pradesh, INDIA, [2]Department of Biotechnology,National Institute of Technology, Warangal 506002,Andhra Pradesh,INDIA.

Abstract

Background & Aim: The present study was carried out to evaluate the *omics of asnA gene from Enterobactor aerogenes* strain KCTC2190. **Methods:** This was achieved by using various online computational tools for both genomic and proteomic analysis. **Results:** Genomic studies states that the pyramidine rich gene codes for 388 aminoacid peptide with approximately 37KD (ss DNA) containing both left & right primers, a hybridization probe and several restriction sites for restriction enzymes. The corresponding peptide primary structure reveals that it is an acidic, stable protein. Secondary structure defines that it mainly contains the random coils & alpha helix. The protein has a total of ten phosphorylation sites along with net N & O-Glycosylation sites. **Conclusion:** The present study states that the pyramidine rich gene has both left & right primers, Hybridization probe with sufficient GC content. The encoded acidic, stable peptide ahs the multiple sites for phosphorylation, glycosylation. Above results shows that the asnA from *Enterobactor aerogenes* strain KCTC2190 is much suitable for the experiments to enhance the L-Asparginase enzyme.

Keywords: asnA gene, *Enterobactor aerogenes,*Anticancer, Asparginase, Primary & Secondary structure, Glycosylation.

1. Introduction

Asparaginase (EC 3.5.1.1) is an enzyme that catalyzes the hydrolysis of asparagine to aspartic acid. Asparaginases are naturally occurring enzymes expressed and produced by microorganisms. Different types of asparaginases can be used for different industrial and pharmaceutical purposes. The most common use of asparaginases is as a processing aid in the manufacture of food. Marketed under the brand names Acrylaway and Prevent-ASe, asparaginases are used as a food processing aid to reduce the formation of acrylamide, a suspected carcinogen, in starchy food products such as snacks and biscuits[1]. A different asparaginase is marketed as a drug under the brand name "Elspar" for the treatment of acute lymphoblastic leukemia (ALL) and is also used in some mast cell tumor protocols[2]. Unlike most of other chemotherapy agents, it can be given as an intramuscular, subcutaneous, or intravenous injection without fear of tissue irritation. As a food processing aid, asparaginases can effectively reduce the level of acrylamide up to 90% in a range of starchy foods without changing the taste and appearance of the end product[3]. Complete acrylamide removal is probably not possible due to other, minor asparagine-independent formation pathways[1]. The rationale behind asparaginase is that it takes advantage of the fact that ALL leukemic cells and some other suspected tumor cells are unable to synthesize the non-essential amino acid asparagine, whereas normal cells are able to make their own asparagine; thus leukemic cells require high amount of asparagine. These leukemic cells depend on circulating asparagine. Asparaginase, however, catalyzes the conversion of L-asparagine to aspartic acid and ammonia. This deprives the leukemic cell of circulating asparagine, which leads to cell death[4]. This protein may use the morpheein model of allosteric regulation[5]. The discovery and development of asparaginase as an anti-cancer drug began in 1953, when scientists first observed that lymphomas in rat and mice regressed after treatment with guinea pig serum[6]. Later it was found out that it is not the serum itself which provoke the tumour regression, but rather the enzyme asparaginase[7].

Corresponding author

Full Address :

Department of Biotechnology,
JNTUA College of Engineering Pulivendula, Pulvendula 516390, Andhra Pradesh,INDIA.

Phone no. +919490905151

E-mail: *rajeshreddy.bio@gmail.com*

2. Methadology

2.1 Sequence Retrieval from NCBI

National Center for Biotechnology Information maintained by National Institutes of Health (NIH) and National Library of Medicine (NLM). Established in 1988 as a national resource for molecular biology information. The information is retrived through Entrez server.The nucleotide sequence of asnA gene from Enterobactor aerogenes KCTC2190 has been retrieved from the NCBI database and its gene ID is

gil336246508 .

2.2 Genomic Studies

2.2.1 Bioedit

BioEdit is a free program given by Brown lab (James W. Brown) which is intended to provide basic functions for protein and nucleic acid sequence editing, alignment, manipulation and analysis.

2.2.2 ORF finder

The ORF Finder (Open Reading Frame Finder) is a graphical analysis tool which finds all open reading frames in the given query sequence or in a sequence already in the database. This tool identifies all open reading frames using the standard or alternative genetic codes. The output is given in 6 frames i.e., three from forward strand and three from its complement. The output shows a list of all the possible ORF'S present in the query in the decreasing order of the length of the predicted ORF'S. The longest one is considered to be the right one for the process of transcription to produce a proper, complete and a functional protein, when translated.

2.2.3 Primer 3.0

Primer3 tool picks the possible primers present in the submitted sequence. It predicts the presence of the left and the right primers in the query. Thus enabling the user to design the complementary primers for the query.

2.2.4 Genscan

GENSCAN was developed by Chris Burge in the research group of Samuel Karlin, Department of Mathematics, Stanford University. Generally used to predict the complete gene structures in human DNA and genomic DNA. Predicts the locations of the exon-intron structures of genes in genomic sequences from a variety of organisms like vertebrates, arabidopsis and maize. It predicts the presence of the initial exon(INIT),internal exon(INTR), terminal exon (TERM),promoter (PROM).single exon(SNGL),poly A tail (PLYA).

2.3 Proteomic Studies

2.3.1 Primary Structure Prediction

2.3.1.1Bioedit

BioEdit is a free program given by Brown lab (James W. Brown) which is intended to provide basic functions for protein and nucleic sequence editing, alignment, manipulation and analysis.

2.3.1.2 Protparam

Protparam is the Protein Identification and Analysis Tools on the ExPASy Server. It computes various physico-chemical properties that can be deduced from a protein sequence. The query protein can either be specified as a Swiss-Prot/TrEMBL accession number or ID, or in form of a raw sequence.

2.3.2 Secondary Structure Prediction

2.3.2.1 GOR 4

The **GOR method** (**G**arnier-**O**sguthorpe-**R**obson) is an information theory-based method for the prediction of secondary structures in proteins[8]. The GOR method analyzes sequences to predict alpha helix, beta sheet, turn, or random coil secondary structure at each position.

2.3.2.2 SOPMA

Self-optimized prediction method (SOPM) has been described to improve the success rate in the prediction of the secondary structure of proteins. SOPM method (SOPMA) correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil)[9,10].

2.4 Post Translational Modifications

2.4.1 NetOGlyc

The NetOglyc server produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins. The best overall network used as input amino acid composition, averaged surface accessibility predictions together with substitution matrix profile encoding of the sequence.

2.4.2 NetNGlyc

The NetNglyc server predicts N-Glycosylation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr (where Xaa is not Pro)sequons. This consensus sequence is a prerequisite for the modification. Here generally the asparigine residues are glycosylated. Artificial neural networks can discriminate between the acceptor and non-acceptor sequences.

2.4.3 NetPhos 2.0

The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins. Version 2.0 was trained on a larger data set of known phosphorylation sites.

3. Results and Discussion

3.1 Genomics

3.1 Genomics

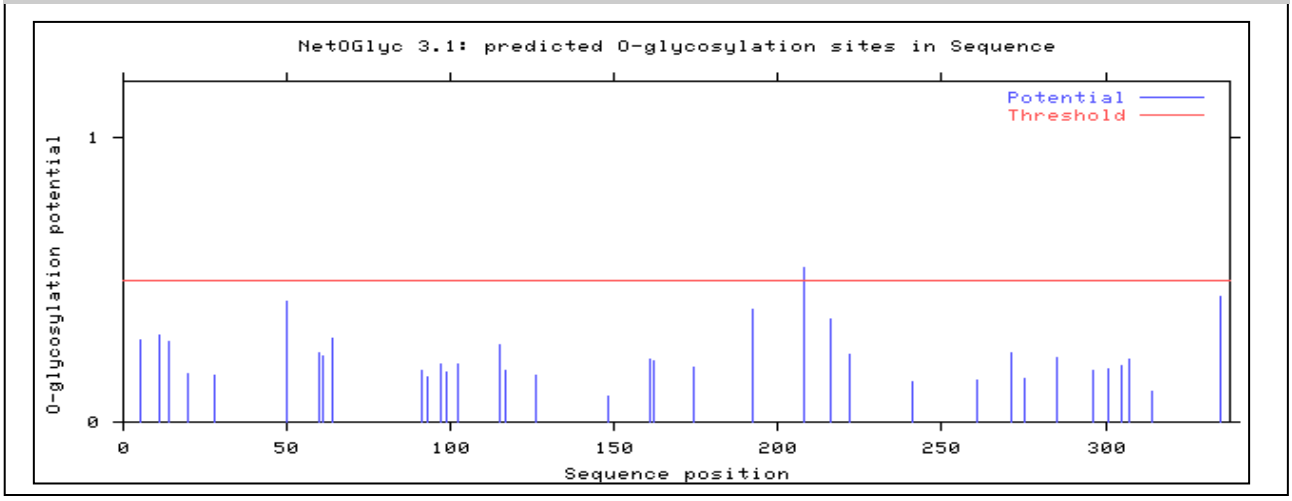
Nucleotide composition of asnA has been predicted by Bioedit & the results reveals that the sequence is pyramidine rich with themolecular weight of more than 3101KDa (ssDNA) containing 1017 base pairs. The CG & AT content was found to be 57.62% and 42.38 % respectively. ORF Finder predicted the presence of the possible protein coding region of sequence. It was been identified that the asnA codes for 9 encoded proteins present in both the + and – strands. The largest protein-coding region (exon) was identified in the 1st frame of the direct strand from the position 1 to 1016 of length 1017 bases. The gene contains the left and the right primers of length 20 residues in the oligonucleotide query. It gives the presence of the left primer starting from 65th position with GC content of 50% and the right primer at starting from 216th position with GC content of 50%. It also predicts the presence of the hybridization probe starting from 132nd position with GC content of 55%. GenScan predicted the presence of the exon in the direct strand from the position 52 to 1068. The probability of the predicted output for the exon is **0.991.**The predicted exon is categorized as a strong exon as the exon score is greater than 100 i.e., 159.98.

3.2 Proteomics

3.2.1 Primary Structure Prediction

Amino acid composition of asnA protein has been predicted by Bioedit & the results reveals that the protein is rich in Ala, Gly and Leu. The protein has a molecular weight protein of 36.951KDa containing 338 amino acids. The protein was acidic in nature, as it has more

Figure I: Net O Glycosylation sites of protein



negatively charged residues than the positively charged ones. The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo) and >10 hours (Escherichia coli, in vivo). The protein is classified as stable, as the instability index is computed to be 37.35. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 91.80. The protein is highly hydrophilic as the Grand Average of Hydropathicity value is –0.113, which is very much lesser than 0.05.

3.1.2 Secondary Structure Prediction

GOR4 (Garnier Method) methods was used to predict the secondary structure of asnA. It was found that 34.02% of amino acids fall on Alpha helix region, 18.05% of amino acid was found to be lays in beta sheet & beta turn and remaining 47.93% tends to form random coil. As per SOPMA (Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments) method it was found that 35.21% of amino acids fall on Alpha helix region, 18.34% of amino acid was found to be lays in beta sheet & beta turn and remaining 38.46% tends to form random coil.

3.1.3 Post Translational Modifications

The given protein contains only one O-Glycosylation site for Threonine & two Asparagine residues for N-Glycosylation at the positions 159 and 269. Netphos 2 results revealed that it contains the phosphorylation sites of Serine, Threonine & Tyrosine of 1, 5 & 4 respectively.

CONCLUSION

L-asparginase a medically important enzyme hydrolyzes L-asparagine (essential amino acid) to aspartic acid and ammoni,a have the anti cancer properties. The asn A gene of *Enterobactor aerogenes* strain KCTC2190 (gi|336246508:4643029-4644045) codes for the L-Asparginase enzyme. Genomics states that the pyramidine rich gene codes for 388 aminoacid peptide with approximately 37KD (ss DNA). The gene contains both left & right primers, a hybridization probe and several restriction sites for restriction sites. In proteomics primary structure says that it is an acidic, stable protein. Secondary structure reveals that it mainly contains the random coils & alpha helix. The protein has a total of ten phosphorylation sites (Serine-1, Threonine-5 & Tyrosine -4) along with net N (Asparagine at positions 159 & 269) & O-Glycosylation (Threonine at position 208) sites.

References

1. Kornbrust BA, Stringer MA, Lange NK, & Hendriksen HV, Asparaginase – an enzyme for acrylamide reduction in food products. In: Enzymes in Food Technology, *Wiley-Blackwell*, **2** (2010) 59-87.

2. Appel IM, Van Kessel-Bakvis C, Stigter R, & Pieters R, Influence of two different regimens of concomitant treatment with asparaginase and dexamethason on hemostasis in childhood acute lymphoblastic leukemia. *Leukemia*, **21** (11) (2007) 2377–80.

3. Hendriksen, HV, Kornbrust BA, Oestergaard PR, & Stringer MA, Evaluating the Potential for Enzymatic Acrylamide Mitigation in a Range of Food Products Using an Asparaginase from Aspergillus oryzae. *J Agri Food Chem*, **57** (2009). 4168–4176.

4. Broome JD, L-Asparaginase: Discovery and development as a tumor-inhibitory agent. *Cancer treatment reports*, **65** (1981)111–114.

5. Selwood T, & Jaffe EK, Dynamic dissociating homo-oligomers and the control of protein function. *Arch Biochem Biophys*, **519** (2011) 131–43.

6. Kidd JG, Regression of transplanted lymphomas induced in vivo by means of normal guinea pig serum. I. Course of transplanted cancers of various kinds in mice and rats given guinea pig serum, horse serum, or rabbit serum. *The J experimental med*, **98** (1953) 565–582.

7. Broome JD, Evidence that the L-asparaginase of guinea pig serum is responsible for its antilymphoma effects. I. Properties of the L-asparaginase of guinea pig serum in relation to those of the antilymphoma substance. *The J experimental med*, **118** (1963) 99–120.

8. Garnier J, Gibrat JF, & Robson B, GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol*, **266** (1996) 540-53

9. Geourjon C, & Deléage G, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, PMID:8808585,.

10. Geourjon C, & Deleage G, SOPMA : Significant improvement in protein secondary structure prediction by c prediction from alignments and joint prediction. *CABIOS*, **11** (1995) 681-684.