Journal of Cell and Tissue Research Vol. 9(3) 1889-1894 (2009)
ISSN: 0974- 0910 *(Available online at www.tcrjournals.com)* **Original Article**

INSILICO PREDICTION OF STRUCTURE AND FUNCTIONAL ASPECTS OF A HYPOTHETICAL PROTEIN OF *NEUROSPORA CRASSA*

NEELAMATHI, E.,? VASUMATHI, E.,[1] BAGYALAKSHMI, S. AND KANNAN, R.[2]

P. G.. Department of Biotechnology, [1]P. G.. Department of Zoology, Nehru Memorial College, Puthanampatti 621 007, Tiruchirapalli, [2]Department of Botany, NGM College, Pollachi 642 001, Coimbatore, Tamilnadu. E. mail: neelamathimsc@yahoo.co.in

Received; September 12, 2009: Accepted: October 14, 2009

Abstract: Large-scale sequence analysis has increased the number of accessible genes tremendously in the last few years. In recent times, the proteomic studies in general and prediction and structural elucidation of protein, in particular adapt an in silico approach. With the advent of functional genomics and the availability of comprehensive publicly accessible Neurospora genomic sequence database, the present study has been carried out in silico to predict the structure and function of a hypothetical protein of Neurospora. The hypothetical protein analyzed in the present study showed conserved domain characteristics of FNR (Fumarate and Nitrate Reduction) like protein. FNR is a DNA binding protein that regulates a large family of genes involved in cellular respiration and carbon metabolism in anaerobic condition. The modelled protein exhibited a maximum number of random coils (47.48%) with alpha helix (25.79%) and extended strands (26.73%) as secondary structural elements. The existence of FNR like domains in the hypothetical protein showed the importance of oxidoreductase activity in the metabolism of the organism.

Key words: *Neurospora crassa,* hypothetical protein.

INTRODUCTION

Neurospora crassa is a type of red bread mold of the phylum Ascomyta. It is used as a model organism owing to its easy growth and haploid life cycle that eventually makes genetic analysis easy and construction of genetics maps. Analysis of genetic recombination is facilitated by the ordered arrangement of the products of meiosis in *Neurospora* ascospores. Its entire genome of seven chromosomes has been sequenced [1]. The genome measures about 43 megabases long and includes approximately 10,000 genes. There is a project underway to produce strains containing knockout mutants of every *N. crassa* gene [2]. *Neurospora* is known for its ability to germinate after forest fires and it helps to decompose burnt organisms. It is important in the elucidation of molecular events involved in circadian rhythms, epigenetics and gene silencing, cell polarity, cell development, as well as
many aspects of cell biology and biochemistry.

A great deal of genetic research in *Neurospora* has been done at the molecular level. For example, studies illustrate similarities between four plasmids in wild-type mtDNA of *N. crassa* and other fungal mitochondrial genomes [3]. Some researchers have used the genetic structure of *Neurospora* to better understand biological processes. The mutant species *N. crassa* was used to better understand cell fusion and cell-to-cell communication [4].

The MIPS *N. crassa* database currently holds more than 16 Mb of non-redundant data of the chromosomes II and V, analysed by the German *Neurospora* Genome Project. This represents more

than one-third of the genome. Open reading frames (ORFs) have been extracted from the sequence and the deduced proteins have been annotated extensively. 41% of analysed proteins are related to known proteins, 30% are hypothetical proteins with no match to a database entry. The entire genome is expected to comprise some 13,000 protein coding genes, more than twice as many as found in yeasts, and reflects the high potential of filamentous fungi to cope with various environmental conditions [5].

The *N. crassa* system offers many advantages over other systems for production of recombinant protein. In contrast to mammalian cell culture, *N. crassa* can be grown in a rapid and economic manner, generating large amounts of recombinant protein in simple, defined medium. Vaccines, therefore, can be produced more rapidly and at lower cost than conventional cell culture or egg-based systems. Influenza haemagglutinin and neuraminidase antigens have been produced and purified from *N. crassa*. This has important applications in tailoring the seasonal vaccine supply and responding to new pandemics [6].

In *Neurospora* transformation is efficient. 100 % homologous integration can be achieved, allowing targeted gene inactivation using a variety of selectable markers. Considering these characteristics and with the advent of functional genomics and the availability of large number of *Neurospora* mutants an attempt has been made in the present study *in silico* to predict the structure and function of a hypothetical protein in *Neurospora*. Since only the primary sequence of the protein is available in the NCBI, and no structural information like x-ray crystallographic data were available in the Protein Data Bank (PDB), the present study was undertaken using various tools and softwares for the modelling of protein and the deduction of three dimensional structure of the protein for further research in various fields of biology.

MATERIALS AND METHODS

The *in silico* prediction of structure and function of a hypothetical protein of *Neurospora* requires a variety of tools and softwares. Initially the sequence of a hypothetical protein of *Neurospora* (Acc.No. gi|11281244|pir||T48809) was retrieved from the GenBank at National Centre for Biotechnology Information (NCBI). The sequence was compared with the sequences in the databases using the Basic Local Alignment Search Tool (BLAST) to find out the homologous sequences. The conserved domains in the hypothetical protein were identified using CD Search [7,8]. The motifs were identified using the tool Motif Finder. Using the primary sequence, the physicochemical properties of the protein were calculated with the aid of the tool ProtParam [9]. The secondary structure was analyzed using the tools like GOR [10,11]. The structural classification of the hypothetical protein was done using the SCOP [12] (Structural Classification of Proteins) database. Homology modeling was done using SWISS MODEL [13] and the tertiary structure of the hypothetical protein was deduced by comparing with two suitable templates retrieved from Protein Data Bank. The protein structure was observed in RASMOL [14].

RESULTS AND DISCUSSION

The primary sequence of the hypothetical protein of *Neurospora* having 318 aminoacid retrieved from GenBank was analyzed for homology using BLAST P suite by choosing non-redundant protein sequences from the SwissProt database. Sequences producing significant alignments with its score (bits) and E-value were identified by the protein-protein BLAST. The sequence codes for the oxidoreductase NAD-binding

Table 1: BLAST result of the hypothetical protein

Accession Number	Similar hits	Score	E-value
gi|75039950|sp|Q58DM7.1|	Oxidoreductase NAD-binding domain containing protein	98.6	5e-20
gi|82188088|sp|Q7T0X7.1|		96.7	2e-19
gi|74731959|sp|Q96HP4.1|		92.0	4e-18
gi|158513831|sp|A4IHY0.1|		92.0	4e-18
gi|75040938|sp|Q5R4D2.1|		90.1	2e-17
gi|158513405|sp|A3KP77.1|		88.2	6e-17
gi|158563930|sp|Q8VE38.2|		87.4	1e-16

Table 2: Aminoacid composition of the hypothetical protein

Aminoacid	Number of residues	Percentage of residues
Ala (A)	25	7.9%
Arg (R)	25	7.9%
Asn (N)	11	3.5%
Asp (D)	22	6.9%
Cys (C)	5	1.6%
Gln (Q)	12	3.8%
Glu (E)	11	3.5%
Gly (G)	24	7.5%
His (H)	5	1.6%
Ile (I)	22	6.9%
Leu (L)	23	7.2%
Lys (K)	12	3.8%
Met (M)	12	3.8%
Phe (F)	13	4.1%
Pro (P)	18	5.7%
Ser (S)	20	6.3%
Thr (T)	13	4.1%
Trp (W)	6	1.9%
Tyr (Y)	4	1.3%
Val (V)	35	11.0%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Fig. 1. Putative conserved domains detected from *Neurospora* hypothetical protein sequence

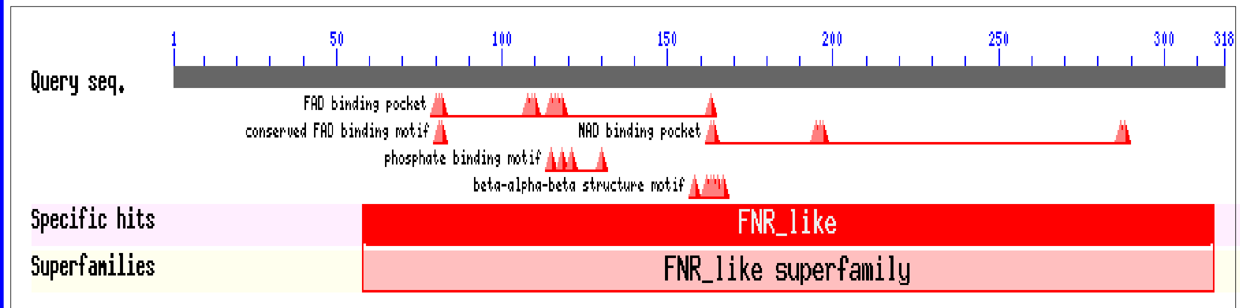

Fig. 2. Secondary elements of the protein exhibited by GOR.

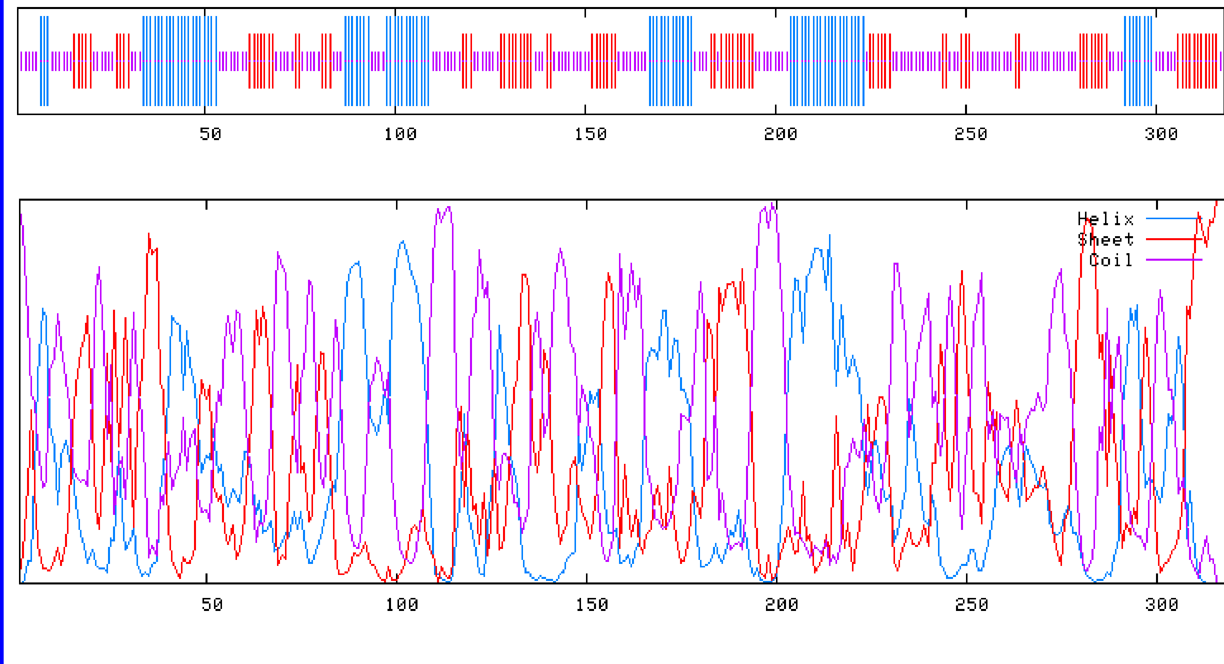

Table 3: Details of secondary structure for the hypothetical protein (using GOR IV)

Secondary elements	Number of strands	Percentage
Alpha helix (Hh)	82	25.79%
Beta helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand (Ee)	85	26.73%
Beta turn (Tt)	0	0.00%
Bend region (Ss)	0	0.00%
Random coil (Cc)	151	47.48%
Ambiguous states	0	0.00%

domain-containing protein precursor showed maximum alignments with a score of 98.6 and an e-value of 5e-20 (Table 1). The hits score with maximum identity was observed in seven sequences, which indicates that the particular protein with similar function exists in seven different organisms. FAD binding pocket, conserved FAD binding motif, NAD binding pocket, phosphate binding motif, beta-alpha-beta structure motifs were identified as putative conserved domains for the query sequence (Fig. 1).

The specific hits showed that the hypothetical protein forms homologous to FNR (fumarate and nitrate reductase) like protein. Similarly sequence analysis of *Rhizobium* DNA region revealed the presence of two complete open reading frames, *orf240* and *orf114*. The deduced amino acid sequence of *orf240* showed significant homology to *E. coli* FNR and *R. meliloti* FixK. A similar protein domain is also present in *E. coli* FNR and is essential for the oxygen regulated activity of this protein [15]. It was found that about 113 motifs were present in the test sequence.

The ProtParam results exhibited the physicochemical parameters of the hypothetical protein (Table 2). There are about 318 aminoacids in the sequence. Its molecular weight was 35176.7 and the theoretical pI was 8.83. Likewise, the deduced molecular weight (27947) of the FNR protein of *E.coli* was in agreement with that of the protein identified by the maxicell procedure, and the primary structure contained regions of homology with several transcriptional regulator proteins [16]. Of the 20 aminoacids, valine showed higher frequency (35) followed by alanine and arginine (25 each). Tyrosine exhibited a minimum frequency of 4 residues. The sequence contains about 33 negatively charged residues (Asp + Glu) and 37 positively charged residues (Arg + Lys). The atomic composition exhibited that the protein has 4969 atoms comprising carbon (1563), hydrogen (2500), nitrogen (444), oxygen (445) and sulfur (17). Thus $C_{1563}H_{2500}N_{444}O_{445}S_{17}$ has been arrived as the molecular formula for the hypothecated protein The instability index (II) (34.00) confirmed the protein as stable. The aliphatic index was calculated as 94.97 and the grand average of hydropathicity (GRAVY) was 0.015.

The secondary structure of the hypothetical protein deduced by the tool GOR (Fig 2) reveals the presence of alpha helices (25.79%), extended strands (26.73%) and random coils (47.48%). The dominance of coiled regions indicates the high level of conservation and stability of the protein structure. There was no beta helix, pi helix, beta bridges and beta turns in the secondary structure (Table 3). The SCOP analysis represents that the hypothetical protein belongs to the class of alpha proteins with globin-like fold. The homology modeling was done in SWISS MODEL by comparing with two templates retrieved from Protein Data bank and finally visualized using RASMOL (Fig. 3). In the modelled structure there were 293 H bonds but no S-S bonds.

CONCLUSION

To conclude the hypothetical protein considered in the study showed conserved domain characteristics of FNR (regulator for fumarate and nitrate reduction) which is a global regulator to regulate the transcription of overlapping modules of target genes in response to anaerobiosis and carbon source in *Escherichia coli*. FNR contains a redox/O2-sensitive element for detecting the anaerobic state. Further the synthesis of the anaerobic ribonucleotide reductase (RNR) enzyme within the organism is under direct control of a protein sensing oxygen concentration (termed FNR). Whenever oxygen is missing, this protein binds to DNA and activates the synthesis of the anaerobic enzyme. In addition, the FNR protein is able to respond to subtle variations in oxygen concentration by modulating the amount of synthesized RNR, therefore optimizing microbial resources [17]. The results of the study help to understand the genetic regulation of FNR (like essential enzymes) which can aid in designing specific antimicrobial drugs to restrain microbial proliferation through the course of an infection. Further research involving the development of appropriate strategies for studying these proteins could be of significance in relation to the genes encoding the domains and its functions.

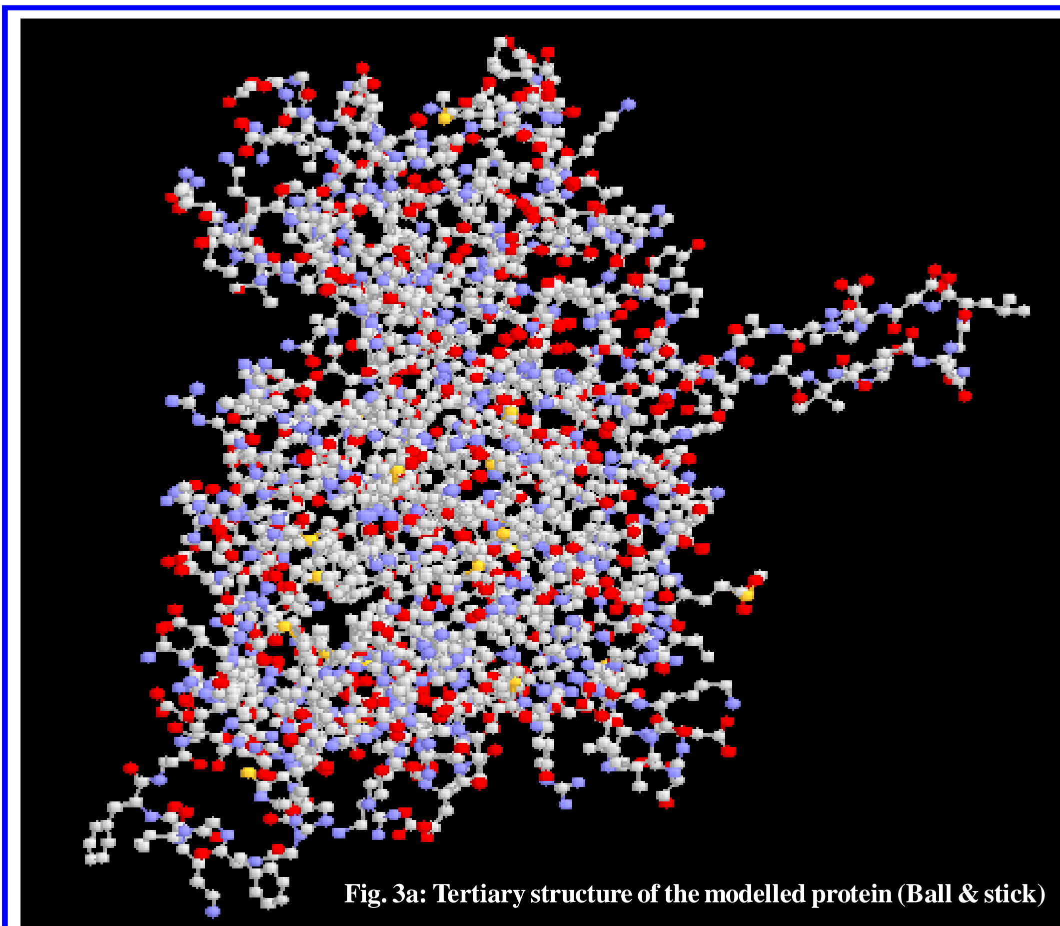

Fig. 3a: Tertiary structure of the modelled protein (Ball & stick)

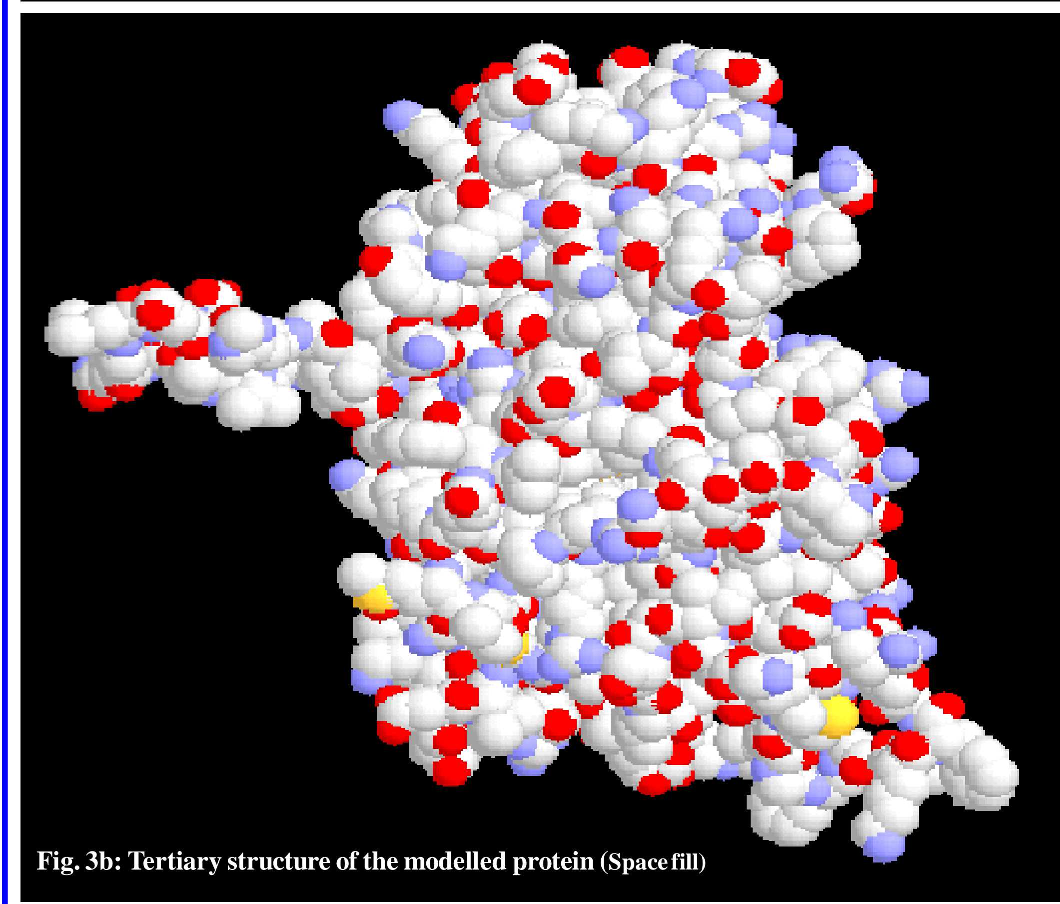

Fig. 3b: Tertiary structure of the modelled protein (Space fill)

ACKNOWLEDGEMENTS

The authors thank the authorities of the Nehru Memorial College, Puthanampatti and NGM College, Pollachi for their encouragement and support.

REFERENCES

[1] James E. Galagan, Sarah E. Calvo, Katherine A. Borkovich, Eric U. Selker, Nick D. Read, David Jaffe, William FitzHugh, Li-Jun Ma, Serge Smirnov, Seth Purcel, Bushra Rehman, Timothy Elkins, Reinhard Engels, Shunguang Wang, Cydney B. Nielsen, Jonathan Butler, Matthew Endrizzi, Dayong Qui, Peter Ianakiev, Deborah Bell-Pedersen, Mary Anne Nelson, Margaret Werner-Washburne, Claude P. Selitrennikoff, John A. Kinsey, Edward L. Braun, Alex Zelter, Ulrich Schulte, Gregory O. Kothe, Gregory Jedd, Werner Mewes, Chuck Staben, Edward Marcotte, David Greenberg, Alice Roy, Karen Foley, Jerome Naylor, Nicole Stange-Thomann, Robert Barrett, Sante Gnerre, Michael Kamal, Manolis Kamvysselis, Evan Mauceli, Cord Bielke, Stephen Rudd, Dmitrij Frishman, Svetlana Krystofova, Carolyn Rasmussen, Robert L. Metzenberg, David D. Perkins, Scott Kroken, Carlo Cogon, Giuseppe Macino, David Catcheside, Weixi Li, Robert J. Pratt, Stephen A. Osmani, Colin P. C. DeSouza, Louise Glass, Marc J. Orbach, J. Andrew Berglund, Rodger Voelker, Oded Yarden, Michael Plamann, Stephan Seiler, Jay Dunlap, Alan Radford, Rodolfo Aramayo, Donald O. Natvig, Lisa A. Alex, Gertrud Mannhaupt, Daniel J. Ebbole, Michael Freitag, Ian Paulsen, Matthew S. Sachs, Eric S. Lander, Chad Nusbaum and Bruce Birren: Nature 422: 859-868 (2003).

[2] Colot, H.V.,Park, G., Turner, G.E., Ringleberg, C., Crew, C.M., Litvinkova, L., Weiss, R.L., Borkovitch, K.A. and Dunlap, J.C.: Proceed. National Acad. Sci., USA, 103: 10352-10357 (2006).

[3] Cahan, P. and Kennell, J.C.: Mol. Genet. Genomics, 5: 13 (2005).

[4] Fleissner, A., Sarkar, S., Jacobson, D.J., Roca, M.G., Read, N.D. and Glass, N.L.: Eukaryot Cell. 4: 920-930 (2005).

[5] Schulte, I., Becker, I.H., Mewes, M. and Mannhaupt, G.: J. Biotechnol, 94: 3-13 (2002).

[6] Allgaier, S. Taylor, R.D., Brudnaya, Y., Jacobson, D.J., Cambareri, E. and Stuart, D.W.: Biologicals, 37: 128-132 (2009).

[7] Marchler-Bauer, A., Marchler-Bauer, A. John B. Anderson, Farideh Chitsaz, Myra K. Derbyshire, Carol DeWeese-Scott, Jessica H. Fong, Lewis Y. Geer, Renata C. Geer, Noreen R. Gonzales, Marc Gwadz, Siqian He, David I. Hurwitz, John D. Jackson, Zhaoxi Ke, Christopher J. Lanczycki, Cynthia A. Liebert, Chunlei Liu, Fu Lu, Shennan Lu, Gabriele H. Marchler, Mikhail Mullokandov, James S. Song, Asba Tasneem, Narmada Thanki, Roxanne A. Yamashita, Dachuan Zhang, Naigong Zhang and Stephen H. Bryant: Nucleic Acids Res.,.37: 205-210 (2009).

[8] Marchler-Bauer, A., Bryant, S. H.: Nucleic Acids Res. 32: 327-331 (2004).

[9] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A.: Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook* (Walker, J.M. ed), Humana Press Inc., Totowa, New Jersey, pp 571-607 (2005).

[10] Combet, C., Blanchet, C., Geourjon, C. and Deleage, G.: Trends in Biochem. Sci., 25: 147-150 (2000).

[11] Garnier, J., Gibrat, J.F. and Robson, B.: GOR secondary structure prediction method. In: *Methods in Enzymology* (Doolittle R.F. ed.), Version IV, pp 540-553 (1996).

[12] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C.: J. Mol. Biol. 247: 536-540 (1995).

[13] Arnold, K., Bordoli, L., Kopp, J. and Schwede, T.: Bioinformatics, 22: 195-201 (2006).

[14] Sayle, R. and James Milner-White, E.: Trends in Biochem. Sci., 20: 374 (1995).

[15] Colonna-Romano, S., Arnold, W., Schlüter, A., Boistard, P., Pühler, A. and Priefer, U.B.: Mol. General Genetics, 223: 138-147 (1990).

[16] Shaw, D.J. and Guest, J.R.: Nucleic Acids Res, 10: 6119-6130 (1982).

[17] Roca, I., Ballana, E., Panosa, A. and Torrents E.: Intern. Microbiol., 11: 49-56 (2008).