

Computational biology and protein modeling of cyanobacteria using bioinformatics tools and techniques

Padhi S.B.*, Behera S., Swain P., Behura S., Behera G., Panigrahi M., Baidya S., Mishra A., Beza S., Panigrahi H., Dash N., Pradhan S. and Ratha N.

Algal Research Laboratory, Department of Botany, Berhampur University, Bhanja, 760007, Bihar, 09937095353, 09861115351, sailabalapadhi@gmail.com, ghana.rgnfs@rediffmail.com

Abstract- Computational biology is a term coined from analogy to the role of physical sciences, is now coming into its own as a major element of contemporary biological and biomedical research. In the sharp in this pattern, over past few years, experiments in life sciences in the academic institutions have begun to recognize the value of bioinformatics and computational biology in the field of algology. Cyanobacteria (also known as blue-green algae) are a group of extraordinarily diverse Gram-negative prokaryotes that originated 3.5 billion years ago. After the advent of bioinformatics in the field of algology, complete genome sequences of Cyanobacteria have been reported in more than 30 species and strains including unicellular. The filamentous cyanobacterium *Anabaena* sp. PCC 7120 (further referred to as *Anabaena* sp.) is a model system to study nitrogen fixation, cell differentiation, cell pattern formation and evolution of plastids. It is a multicellular photosynthetic microorganism consisting of two cell types, vegetative cells and nitrogen fixing heterocysts. The nucleotide sequence of the entire genome of a filamentous Cyanobacterium, *Anabaena* sp. Strain PCC 7120, was determined. This study focuses on the function and dynamics of the proteome of the Gram-negative outer membrane in *Anabaena* sp.

Key Words: - *Anabaena* sp, Bioinformatics, Computational biology, Cyanobacteria, Outer membrane.

Introduction

In looking at the origin and growth of computational biology, we see a parallel to molecular biology, which had its origins 40-50 years ago as the tools of multiple disciplines—especially physics, chemistry biophysics, microbiology and genetics were applied in new ways, a vigorous and often rigorous reductionism attack on central barriers to understanding the mechanisms by which simple bacteria and viruses function. The success of molecular biology is such that today it is no longer discipline but instead is deeply integrated into all bioscience research. Today, instead the revolution lies in the application of a new set of interdisciplinary tools, computational approaches will provide the underpinning for the integration of broad disciplines in developing a quantitative system approach, an integrative or synthetic approach to understanding how individual macromolecules interact with each other, how cells interact, how signaling, genetics and metabolic networks interact. Bioinformatics provides the “glue” for system biology [7]. It aims at establishing the language for diverse disciplines. With the advent of Bioinformatics, the genomics of different organisms was unzipped by means of different genome projects. In Cyanobacteria, complete genome sequences of *Anabaena* sp. PCC 7120 have been reported. The genome of *Anabaena* consisted of a single chromosome (6,413,771 bp) and six plasmids pCC7120 α (408,101 bp), pCC7120 β (186,614 bp), pCC7120 γ (101,965 bp), pCC7120 δ (55,414 bp), pCC7120 ϵ (40,340 bp), and pCC7120 ζ (5,584 bp); the chromosome bears 5368 potential protein-encoding genes, four sets of rRNA genes, 48 tRNA genes representing 42 tRNA species,

and 4 genes for small structural RNAs[9]. The outer membrane is a border of a Gram-negative cell, such as *Anabaena* sp. toward the surrounding medium. For this reason the structure and the function of the outer membrane are essential for establishing the contact and compound exchange with the surroundings, sensing the outside conditions and reacting to them [5]. Since the outer membrane proteins are crucial in all these processes, revealing the protein composition of the *Anabaena* sp. Outer membrane supports successful utilization of this Cyanobacterium as a model system in cell differentiation, nitrogen fixation and plastid evolution studies. It is also the step toward elucidating how the multicellularity and differentiation in *Anabaena* sp. reflect on the proteome composition and functions of the outer membrane. Gram-negative bacteria, including *Anabaena* sp., are surrounded by the outer and the inner membrane, separated by peptidoglycan containing periplasm (Fig 1). The inner membrane of classical Gram-negatives is a symmetrical bilayer composed of the three major phospholipids (phosphatidylethanolamine, phosphatidylglycerol and cardiolipin [10] and proteins. Proteins are either integral transmembrane proteins or lipoproteins anchored to the outer leaflet [20]. The transmembrane proteins are generally α -helical, hydrophobic and involved in transport of nutrients, protein translocation and lipid biosynthesis or provide energy for the cell processes by oxidative phosphorylation [15]. The outer membrane is an asymmetrical bilayer with an inner leaflet composed of the same phospholipids as the inner membrane and an outer leaflet composed

of lipopolysaccharides [18,8] chelated by divalent cations [13]. In the outer membrane two types of proteins are found: to the inner leaflet anchored lipoproteins and integral transmembrane β -barrel proteins [15]. Besides, β -barrels some proteins with transmembrane α -helical regions are also targeted to the outer membrane of these organelles. The β -barrel proteins contain from 8 to 22 β -strands, usually with tight turns on the periplasmic side and large quite flexible loops on the extracellular side of the outer membrane. The smallest 8-stranded β -barrels (e.g. structural outer membrane protein OmpA) have tightly packed residues closing the barrel lumen [1]. On the other hand, 16, 18 and 22-stranded β -barrels serve as the outer membrane transporters with large water filled pores. In the case of 22-stranded transporters the pores are occluded with the "plug" domain. The outer membrane β -barrel transporters are divided into four groups [3, 19]. The first group are general porins (e.g. OmpF, OmpC and PhoE), which are 16 stranded β -barrels structured as homotrimers with three barrels and three pores [21, 4]. The second group comprises specific porins with 18-stranded homotrimeric structures, importing sugars like maltose or sucrose. Their pores are partially constricted by three extracellular loops folding into each β -barrel [16]. In both cases substrates of limited size up to approximately 600 Da diffuse passively through the pores into the periplasm. In the periplasm they are bound by the substrate-specific binding proteins and shuttled further to the inner membrane ATP-binding cassette (ABC) transporter. ABC transporters utilize energy generated by ATP hydrolysis to perform an active transport across the inner membrane into the cytoplasm [3]. However, many compounds crossing the outer membrane are much bigger than 600 Da and their transport requires energy at both steps: for the active transfer over the inner membrane and for the active transfer over the outer membrane. These big substrates are transported across the outer membrane with a help of the third and the fourth group of the outer membrane transporters. The third group of the transporters mediates an active import of larger substrates into the periplasm and includes TonB-dependent iron or vitamin B12 transporters [6]. The fourth group of the transporters, presented by TolC protein, mediates the transport in reverse direction: an active secretion of compounds out of the cell. However, because of the porous structure of the outer membrane, membrane potential cannot be established across this membrane and energy rich compounds such as Adenosine triphosphate (ATP), Guanosine triphosphate (GTP), Nicotinamide adenine dinucleotide phosphate (NADPH) are not present in the periplasm. To solve this problem, transporters of TonB-dependent group and TolC-

like proteins [11], establish physical contact with the inner membrane components, which provide them with energy from the cytoplasmic sources to drive the active transport across the outer membrane. The significance of the outer membrane for the function of the Gram-negative bacteria cell is huge. The outer membrane provides additional protection against osmotic stress, antibiotics, detergents and other factors that may pose a danger for a bacterial cell. However, besides being a barrier, the outer membrane offers a number of mechanisms to ensure an adequate compound exchange with the surrounding medium and to sense and react to the outside conditions. The protein machineries, pores and channels involved in these processes are numerous and their investigation is essential for understanding of the Gram-negative cell functionality and endosymbiotic relations. In Cyanobacteria proteins are the expressed marker, whose regulation is influenced by experimental conditions [2]. Whole cell protein patterns can be useful in grouping large number of closely related organism including Cyanobacteria [12]. Characterize filamentous Cyanobacteria strains from different genera using SDS-PAGE whole cell proteins and PCR-RFLP of 16S rRNA gene [12]. Using the total protein pattern mutual relationships at the generic level was revealed. On the basis of these technique Anabaena clustered with Aphanizomenon whereas Nodularia clustered with Nostoc, The taxonomic value of cell protein profiles at the generic and even species level [14]. The comparison of the protein banding pattern of the isolated protein complexes shown that different protein patterns are found in different genera and even in different species of the same genera [17]. *Cyanobacterium Anabaena sp. PCC 7120* as a model system offers possibility to explore not only the features of a classical Gram-negative microorganism, but also photosynthesis, nitrogen fixation and one of the simplest known cell differentiation process all of it co-existing and functioning in the same cell. The outer membrane continuum and the periplasmic continuum along the *Anabaena sp. filament* are important. Attempts are made for analyzing the outer membrane proteomes and by closer insight in the outer membrane proteins.

Method and Tools

The complete genome sequence of *Anabaena sp. PCC 7120* was analyzed from the GenBank database with accession no. BA000019. It was determined that there are 6132 protein coding genes from one circular chromosome and six plasmids. It was found that out of 5368 protein coding genes of the chromosome only 80 protein coding genes play a vital role in formation of Cell envelop of *Anabaena sp. PCC 7120*. The plasmids play no role in cell wall or envelop

formation. Out of 80 only 22 of them, are assigned to the outer membrane. The amino acid sequences along with accession numbers of all the 22 protein coding genes namely alr 4550 represented as alr1, alr 2269 represented as alr1, all 4499 represented as all1, all 1861 represented as all2, alr 3608 represented as alr2, alr 0092 represented as alr3, alr 1819 represented alr4, all 4388 represented as all4, alr 2887 represented as alr5, all 4294 represented as all5, all 1140 represented as all6, all 2430 represented all7, alr 2588 represented alr7, alr 0397 represented as alr8, alr 4740 represented as alr9, all 3310 represented as all12, alr 0834 represented as alr10, alr 4067 represented as alr11, alr 1278 represented as alr12, all 1101 represented all13, all 4026 represented as all14 and all 7611 represented all15 were retrieved from the GenBank database. The sequences were converted to FASTA format for multiple sequence and phylogenetic analysis among the 22 protein coding genes. The multiple sequence analysis was done by using ClustalW and ClustalX.2 for getting the conserved regions. The output of ClustalW was saved in .txt format and gone for phylogenetic analysis using Phylip 3.69 version. As the sequences were amino acid sequences protpars tool was used for phylogenetic analysis. The output of ClustalW was again analyzed for phylogenetics using MEGA 4 tool. The bootstrap values along with distances of all the 22 sequences were found. The phylogenetic trees in different formats were obtained. For analyzing and viewing the structure of outer membrane protein PDB (Protein Data Bank) site was browsed. The structure of first group general porins (OmpF) and fourth group was downloaded and viewed in RasMol.

Results and discussion

There was no significant result obtained from multiple sequence alignment of 22 protein coding genes using ClustaX and ClustalW tools. No conserved regions were obtained by multiple sequence analysis. This depicts that the proteins are distantly related to each other. The phylogenetic tree (Fig 2) that was obtained from Phylip was an unrooted tree with different branch lengths. Two types of phylogenetic analysis were obtained from MEGA4 programme. Rectangular phylogenetic tree (Fig 3) with branch lengths and phylogenetic tree with Boot strapping (Fig 4). The trees that were obtained are unrooted trees. The structure of first group are general porins (OmpF) with accession number 2ZFG (Fig 5) was downloaded from PDB and viewed in RasMol. It was found that the structure consists of 340 residues Porin: 339 residues 3% helical (3 helices; 12 residues) 60% beta sheet (25 strands; 204 residues).

TolC-dependent protein secretion

The fourth group of the outer membrane transporters (Fig6) was described first. Among four known secretion systems of Gram-negative bacteria (type I-IV), type I is the simplest. This system requires only three proteins: (1) the substrate-specific inner membrane component (ABC transporter or proton antiporter), (2) the TolC channel-tunnel as an outer membrane component and (3) the so-called membrane fusion (MFP) or adaptor protein. Type I secretion system exports many large proteins, including some bacterial toxins such as 110 kDa *E. coli* hemolysin, 170 kDa *Bordetella pertussis* cyclolysin and different enzymes like proteases, lipases, nucleases, phosphatases, and glucanases. The 3D (3 dimensional) structure of TolC- dependent protein shown in Fig 7 represents a huge number of helices and strands facing towards opposite directions. TolC is a rather promiscuous protein, coupling with different inner membrane complexes in order to transport a variety of substrates. The substrate specificity is determined by the periplasmic and the inner membrane component. The contact between TolC and the inner membrane complex is established in the periplasm with a help of the MFP (Membrane Fusion protein) or adaptor protein

TonB-dependent transporters and iron uptake

The third group of the outer membrane transporters comprises TonB-dependent transporters (also known as TonB-dependent receptors). These proteins play a crucial role in iron uptake and iron regulation in the Gram-negative bacterial cell. In biological systems iron is one of the most important elements acting as a major redox mediator. As a representative of cyanobacteria, *Anabaena* sp. is a Gram-negative microorganism capable of oxygenic photosynthesis. In that purpose it contains intracellular thylakoid membrane system where photosynthetic apparatus is placed. Iron is important for the photosynthetic electron transfer as an essential metal of photosystem I, cytochrome b6f complex and photosystem II. Because of that cyanobacteria generally have exceptional requirements for iron, but also for copper and manganese in comparison to other nonphotosynthetic bacteria.

Conclusion

The advent of Bioinformatics has given us many genomic databases which give a clear picture to understand the genes and genomes that are involved in different function of an organism. The genes that are involved in structuralism and functionality of outer membrane of *Anabaena* sp was analyzed by using different bioinformatics tools. Found that the major functions of every bacterial cell envelope is to allow sufficient

transport of nutrients and metabolites into and out of the cell, a function especially delicate in cyanobacteria as their thick, multilayered envelopes form a considerable mechanical and permeability barrier for larger molecules. Consequently, cyanobacteria have developed different transport systems, and as in other bacteria these transport systems can be discriminated according to the energy source used for transport, the complexity of the transport machinery, or the chemical nature of the translocated substrates. It is too far for conclusion, because, although we are witnessing a remarkable change in the scale of molecular microbiological research and are entering an integrative “genomic science”, the access to genome information is still very limited. Perhaps, the information on the sequences, gene organization and phylogeny reviewed in the present perspective would provide us the possibilities of enumerating the other potentials that clue about the origin of the outer membrane genes in them. TonB-dependent group and TolC-like proteins establish physical contact with the inner membrane components, which provide them with energy from the cytoplasmic sources to drive the active transport across the outer membrane.

Acknowledgement

The authors express their deep sense of gratitude to Head, P.G. Department of Botany, and Berhampur University for laboratory facilities.

References

- [1] Arora A., Abildgaard F., Bushweller J.H. and Tamm L.K. (2001) *Nat. Struct. Biol.*, 8(4), 334-338.
- [2] Bhagwat A.A. and Apte S.K. (1989) *J.bacteriol.*,171, 5187-5189.
- [3] Buchanan S.K. (2001) *Trends Biochem. Sci.*, 26(1), 3-6.
- [4] Cowan S.W., Schirmer T., Rummel G., Steiert M., Ghosh R., Pauptit R.A., Jansonius J.N. and Rosenbusch J.P. (1992) *Nature*, 358(6389), 727-733.
- [5] Egbert H. and Alfred H. (2000) *Journal of Bacteriology*, 1191–1199.
- [6] Faraldo-Gomez J.D. and Sansom M.S. (2003) *Nat. Rev. Mol. Cell Biol.*, 4, 105-116.
- [7] John C.Wooley, (1999) *Journal of Computational Biology*, 6, 459-474.
- [8] Kamio Y. and Nikaido H. (1976) *Biochemistry*, 15(12), 2561-2570.
- [9] Kaneko T., Nakamura Y. and Wolk C.P. (2001) *DNA Res.*, 8,205-213.
- [10] Kanemasa Y., Akamatsu Y. and Nojima S. (1967) *Biochem. Biophys. Acta*, 144(2), 382-390.
- [11] Koronakis V., Eswaran J. and Hughes C. (2004) *Annu. Rev. Biochem.*, 73, 467-489.
- [12] Lyra G., Hantula J., Vainio E., Rapala J., Rouliainen L. and Sivonen, K. (1997) *Arch. Microbiol.*,168,176-184.
- [13] Nikaido H. (2003) *Microbiol. Mol. Biol. Rev.*, 67(4), 593-656.
- [14] Palinska K.A., Liesack, W., Rhid, E. and Kaunbein, W.E. (1996) *Arch. Microbiol.*, 166, 224-233.
- [15] Ruiz N., Kahne D. and Silhavy T.J. (2006) *Nat. Rev. Microbio.*, 4(1), 57-66.
- [16] Schirmer T., Keller T.A., Wang Y.F. and Rosenbusch J.P. (1995) *Science*, 267(5197), 512-514.
- [17] Schenk H. E.A. and Kuhfitting G. (1983) *Biochem. Sys.*, 1, 179-184.
- [18] Smit J., Kamio Y. and Nikaido H. (1975) *J. Bacteriol.*,124(2), 942-958.
- [19] Tamm L.K., Hong H. and Liang B. (2004) *Biochim. Biophys. Acta.* 1666 (1-2), 250-263.
- [20] Tokuda H. and Matsuyama S. (2004) *Biochim. Biophys. Acta.* 1694(1-3), IN1-9.
- [21] Weiss M.S., Abele U., Weckesser J., Welte W., Schiltz E. and Schulz G.E. (1991) *Science*, 254(5038), 1627-30.

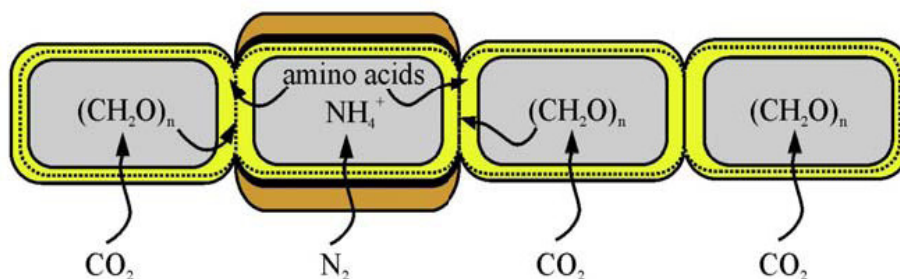


Fig. 1- Continuum of the outer membrane and the periplasm in *Anabaena* sp. PCC 7120 filament. The yellow area represents the periplasmic continuum surrounded by the continuous outer membrane and the inner membrane which is separate for every cell. The dotted line represents the peptidoglycan layer; the grey area is the cytoplasm. The arrows indicate metabolite exchange between vegetative cells and heterocysts through the continuous periplasm. Vegetative cells supply heterocysts with sugars (sucrose) as products of photosynthesis, while heterocysts supply vegetative cells with products of nitrogen fixation (amino acids).

One most parsimonious tree found:

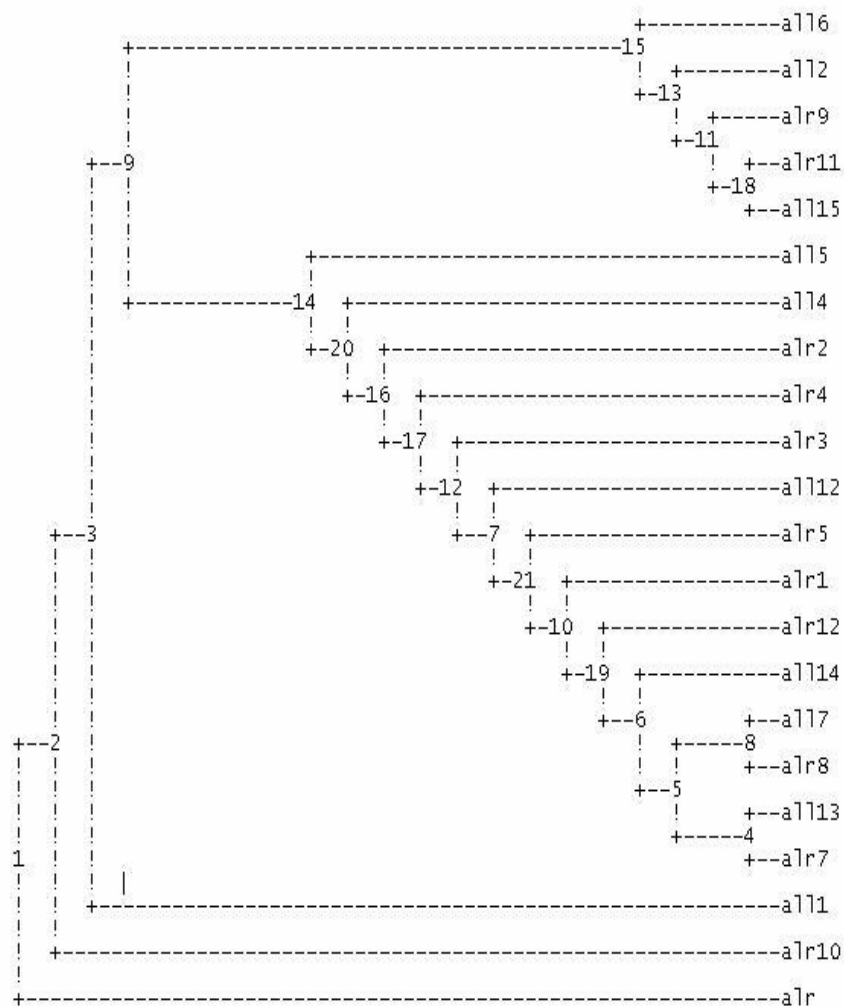


Fig. 2- The un-rooted phylogenetic tree obtained from Phylip programme with different branch lengths.

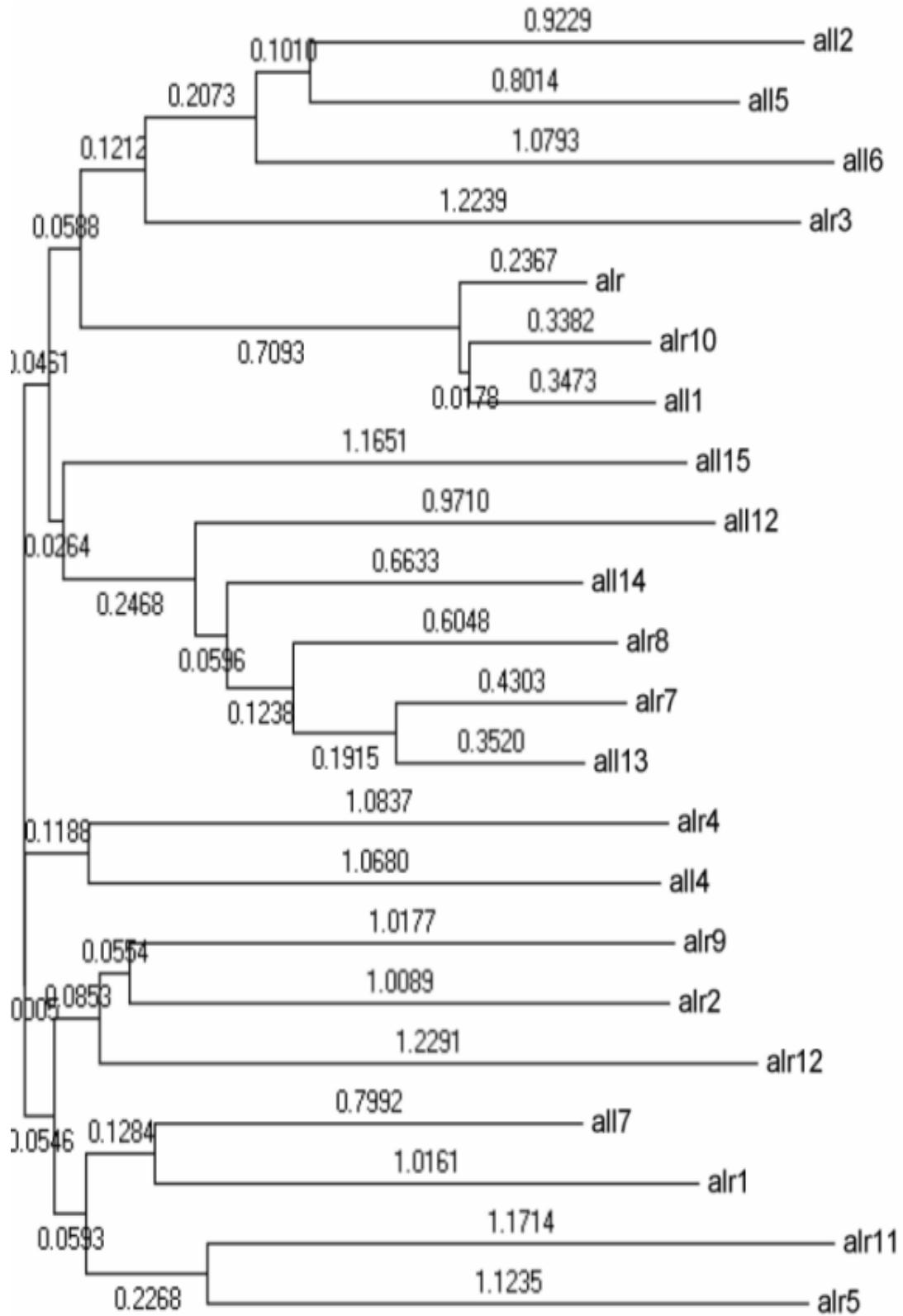


Fig. 3- The un-rooted rectangular phylogenetic tree with branch lengths obtained from MEGA4programme.

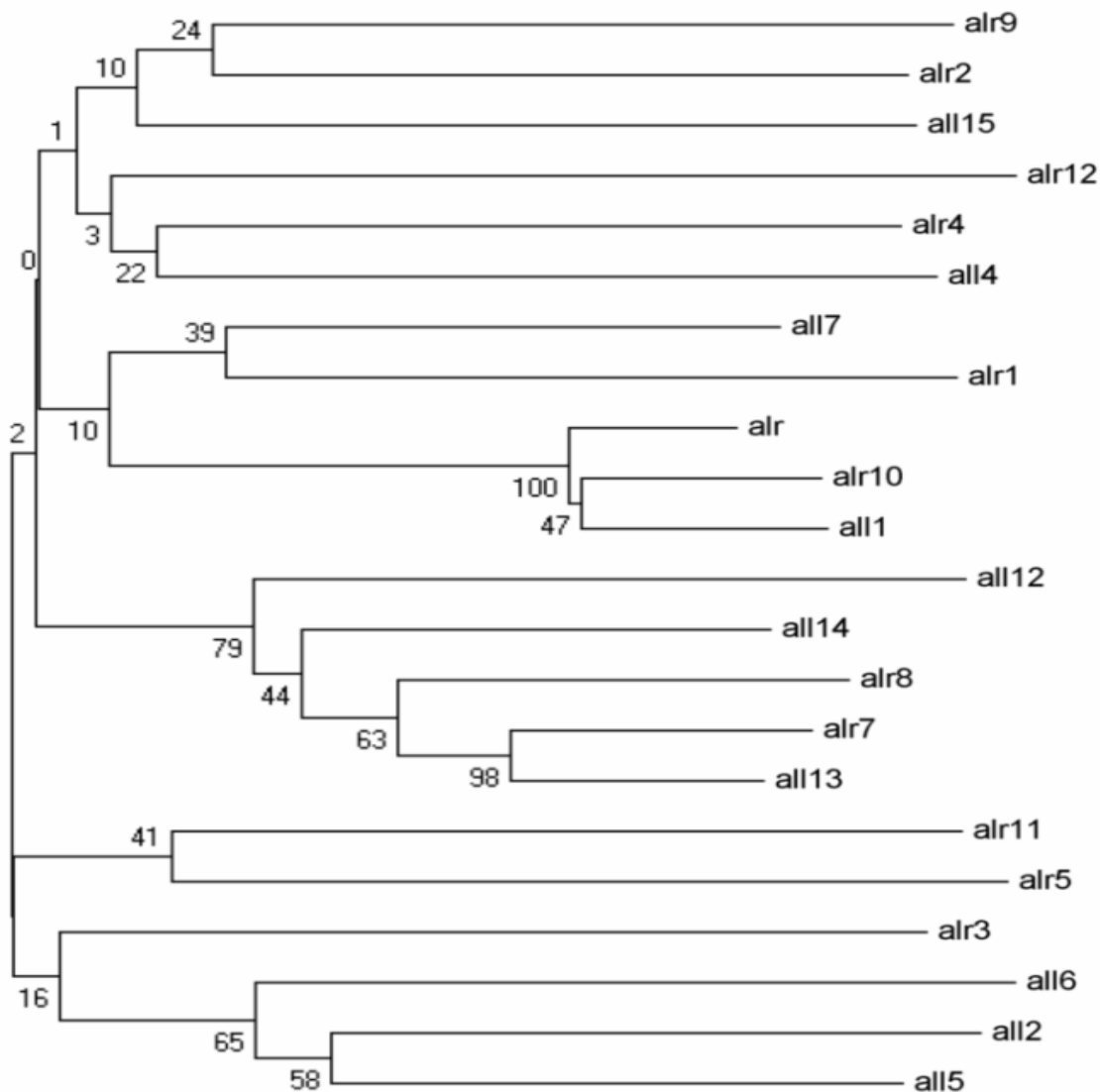


Fig. 4- The un-rooted phylogenetic tree with Boot strapping values obtained from MEGA4programme.

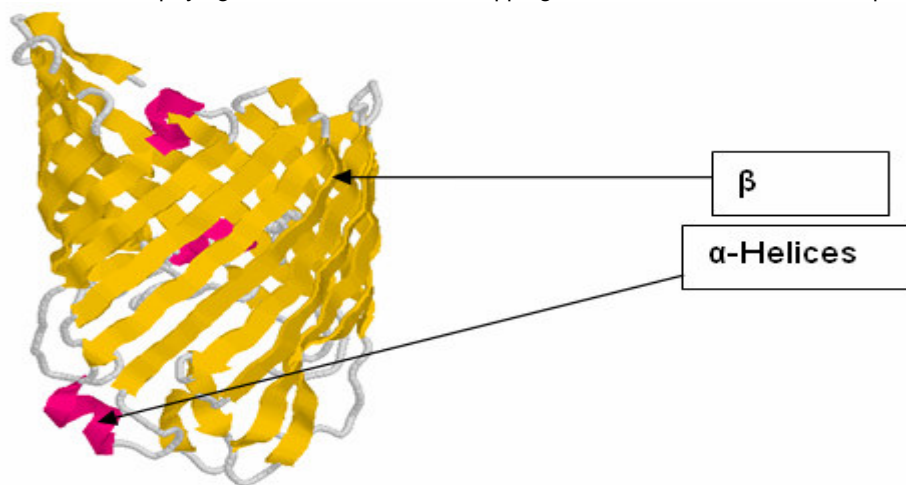


Fig. 5- 3D view of first group general porins (Outer membrane protein F (OmpF))

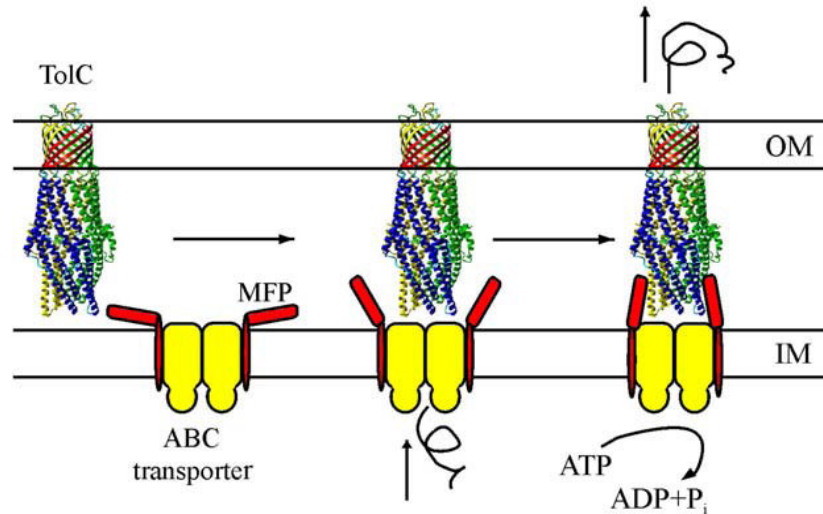


Fig. 6- Protein secretion through the TolC channel-tunnel. The contact between TolC and the inner membrane ABC transporter (yellow subunits) is established transiently with a help of the adaptor or membrane fusion protein (MFP, red subunits) when the protein substrate (curly line) binds to the substrate specific ABC transporter (according to Koronakis et al., 2004).

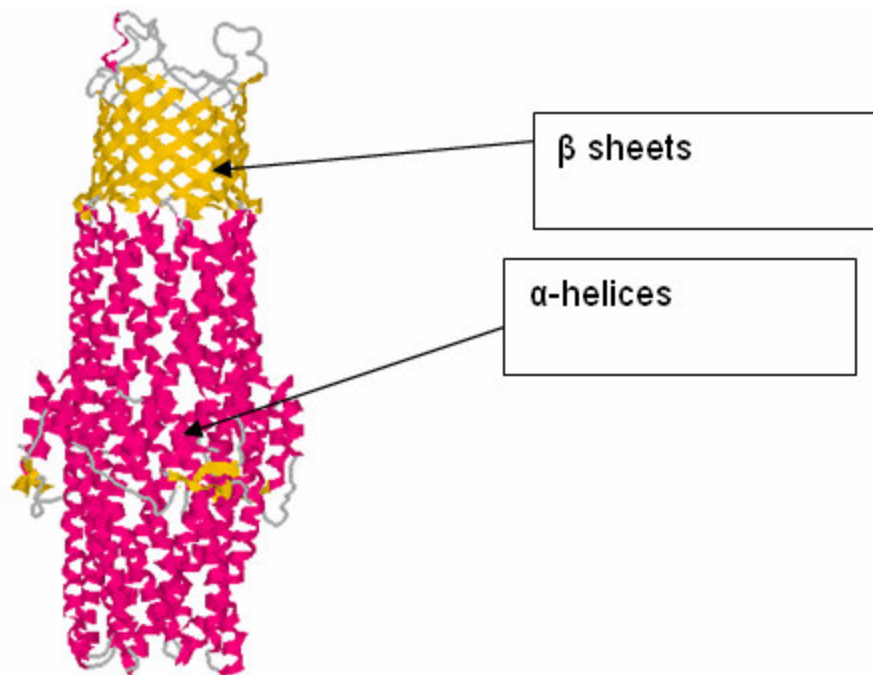


Fig. 7- 3D View of outer membrane protein TolC viewed by using RasMol.