Journal of Advanced Bioinformatics Applications and Research ISSN 0976-2604
Vol 2, Issue 3, 2011, pp 206-210 http://www.bipublication.com

IN-SILICO PREDICTION OF STRUCTURAL AND FUNCTIONAL ASPECTS OF A HYPOTHETICAL PROTEIN OF *CAPNOCYTOPHAGA CANIMORSUS CC5*

Harshada H. Pilley

Bioinformatics Center of BTISNet, Department of Biotechnology, Sant Gadge Baba Amravati University, Amravati – 444602, Maharashtra, India

ABSTRACT:

Large-scale sequence analysis has increased the number of accessible genes tremendously in the last few years. An *in-silico* technique was initiated to characterize a hypothetical protein to deduce its structural and functional information. With the advent of functional genomics and the availability of comprehensive publicly accessible *Capnocytophaga canimorsus Cc5* genomic sequence database, the present study has been carried out *in-silico* to predict the structure and function of a hypothetical protein of *Capnocytophaga*. The hypothetical protein analyzed in the present study showed conserved domain characteristics of Osmotically inducible protein C (OsmC) superfamily which include characteristics of OsmC/Ohr family. OsmC is a stress–induced protein. This family also contains an organic hydroperoxide (Ohr) detoxification protein that has a novel pattern of oxidative stress regulation. The modelled protein exhibited a maximum number of random coils (47.06%) with alpha helix (36.03%) and extended strands (16.91%) as secondary structural elements. The existence of OsmC like domains in the hypothetical protein showed the importance as stress–induced protein of the organism.

Keywords*: Capnocytophaga canimorsus, hypothetical protein, OsmC, Modeller9v7*

1. INTRODUCTION

Capnocytophaga canimorsus is a gram-negative bacillus bacterium that causes a zoonotic disease, most commonly in asplenic patients [1, 2]. It causes fulminate sepsis with disseminated intravascular coagulation. *Capnocytophaga spp*. is normal inhabitants of the oropharyngeal flora and *C. canimorsus* exhibits robust growth when it is in direct contact with mammalian cells, including phagocytes [2]. They are also involved in periodontal diseases or animal bites, complicated by septicaemia with dissemination to a great variety of sites, both in immune-competent and immune-compromised hosts [3,4]. The genome sequence of *C. canimorsus* demonstrates the key role of surface-exposed protein deglycosylation systems in growth and survival in host tissues [1]. The large scale genome sequencing project has generated a plethora of information both in terms of genes and proteins. A broad spectrum of genetic, biochemical and computational approaches is being employed for annotating the physiological function of hypothetical proteins encoded by orphan genes [5]. There are however, a vast amount of proteins whose function and structure has not been unearthed yet. There is, therefore, an urgent need to characterize hypothetical proteins whose only primary information in the form of sequence is available. 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) [6], the present study was undertaken using various tools and software for the modelling of protein and the deduction of three dimensional structure of the hypothetical

protein for further research in various fields of biology [7, 19].

2. MATERIALS AND METHODS

2.1. Sequence retrieval

To analyze the hypothetical protein and assign its functional and structural role, various tools and software were used. The primary sequence of the hypothetical protein(Acc.No.gi|340622360|ref|YP_00474081 2.1|) was obtained from the GenBank [8] at National Centre for Biotechnology Information (NCBI) [9].

2.2. Sequence analysis

The sequence was compared for detecting homologous sequences found in databases using Basic Local Alignment Search Tool (BLAST) [10, 11]. Using the primary sequence, the physicochemical properties of the protein were calculated with the aid of the tool ProtParam [12] ***(Table 1)***. The motifs were identified using the tool Motif Search [13].

Amino acid	No. of residues	% of residues
Ala (A)	11	8.1%
Arg (R)	4	2.9%
Asn (N)	7	5.1%
Asp (D)	8	5.9%
Cys (C)	2	1.5%
Gln (Q)	6	4.4%
Glu (E)	7	5.1%
Gly (G)	9	6.6%
His (H)	5	3.7%
Ile (I)	13	9.6%
Leu (L)	7	5.1%
Lys (K)	11	8.1%
Met (M)	7	5.1%
Phe (F)	4	2.9%
Pro (P)	6	4.4%
Ser (S)	6	4.4%
Thr (T)	13	9.6%
Trp (W)	2	1.5%
Tyr (Y)	1	0.7%
Val (V)	7	5.1%

Table 1: Physicochemical properties of the protein

2.3. Secondary structure prediction

GOR IV server [14] was used for secondary structure prediction from hypothetical protein sequence ***(Table 2)***. BLAST [10] from NCBI was used to compare the query sequence with the database sequence to find its homologues. Conserved domains were detected from the BLAST analysis ***(Table 3)***. Protein fold was recognized using SCOP (Structural Classification of Proteins) tool [15].

Structural elements	Number of residues	Percentage of residues
Alpha helix (Hh)	49	36.03%
310 helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand (Ee)	23	16.91%
Beta turn (Tt)	0	0.00%
Bend region (Ss)	0	0.00%
Random coil (Cc)	64	47.06%
Ambigous states (?)	0	0.00%
Other states	0	0.00%

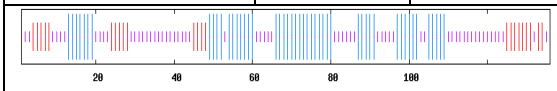

Table 2: Secondary structure elements of the protein

Accession ID	Similar hits	Score	E-value
YP_004740812.1	>gb|AEK23705.1|	285	9e-99
YP_004788814.1	>gb|AEM71392.1|	184	1e-58
YP_001194726.1	>gb|ABQ05407.1|	179	6e-57
ZP_02163039.1	>gb|EDP95425.1|	176	2e-55
YP_004735028.1	>emb|CAZ94637.1|	167	6e-52

Table 3: Putative conserved domains search using BLAST

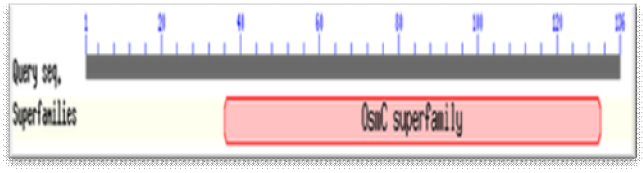

Fig 1: Conserved domain search (CDD) using BLAST

2.4. Homology modelling

Suitable template protein was searched through Geno3D server [16] and PSI-BLAST [11]. The sequence of hypothetical protein was found to

resemble with a template protein named crystal structure of a putative osmotic stress induced and detoxification response protein (psyc_0566) from *Psychrobacter arcticus* [pdb id: 2PN2] chain A. Amino acid sequence alignment of target and template proteins and Rough 3-D models (20 models) were constructed from the sequence alignment between hypothetical protein and the template protein using Modeller9v7 [17].

Fig 2: Three dimensional structure of the modeled hypothetical protein using Modeller9v7.

2.5. Model validation

All the generated structures of the protein model ***(Fig 2)*** were subjected to a series of tests for testing its internal consistency and reliability. Backbone conformations were evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis ***(Fig 3)*** [18].

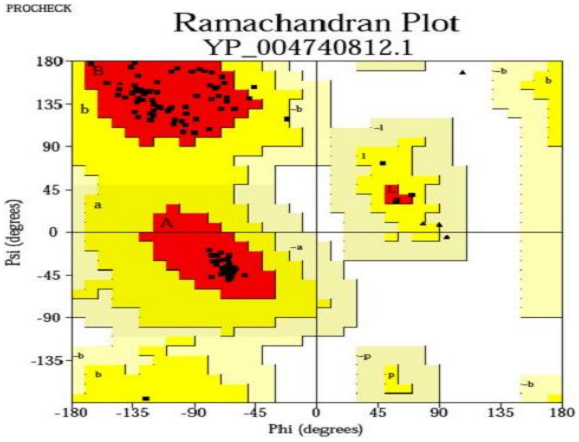

Fig 3: *Ramachandran Plot analysis using PROCHECK*

3. RESULTS

The primary sequence of the hypothetical protein of *Capnocytophaga canimorsus* having 136 amino acid retrieved from GenBank was analyzed for homology using BLASTP suite by choosing protein database. Sequence producing significant alignment with conserved domain characteristics of Osmotically inducible protein C (OsmC) family. The domain identified in the protein sequence was positioned at 36 to 132 residues. For classification of protein, Structural Classification of Protein (SCOP) algorithm was used. It was found that the fold is formed of swapped dimer of beta (3)-alpha-beta (2)-alpha (2)-beta subunits; mixed beta-sheet. The super families for the domain were Ohr/OsmC resistance proteins (***Table 1***). It was found that about 130 motifs were present in the test sequence.

The ProtParam results exhibited the physicochemical parameters of the hypothetical protein (***Table 1***).There are about 136 amino acids in the sequence. Its molecular weight was 15051.3 and the theoretical pI was 6.96. The maximum number of amino acids present in the sequence was found to be that of threonine (9.6%) and isoleucine (9.6%) followed by alanine (8.1%) and lysine (8.1%). Tyrosine exhibited a minimum frequency (0.7%) of 1 residue. The sequence contains about 15 negatively charged residues (Asp + Glu) and 15 positively charged residues (Arg + Lys). The atomic composition exhibited that the protein has 2114 atoms comprising carbon (659), hydrogen (1062), nitrogen (184), oxygen (200) and sulfur (9). Thus $C_{659}H_{1062}N_{184}O_{200}S_9$ has been arrived as the molecular formula for the hypothecated protein. The aliphatic index was calculated as 80.37. The instability index of the protein was computed to be 41.55. This classified the protein to be unstable. The grand average of hydropathicity (GRAVY) was calculated to be -0.304 (***Table 1***). The secondary structural analysis of the protein was done and

random coil was found to be most frequent (47.06%), followed by alpha helix (36.03%). Extended strand (Ee) was found to be least frequent (16.91%) (***Table 2***).The dominance of the coiled regions indicates the high level of conservation and stability of the protein structure [19]. The structure for the hypothetical protein was deduced by homology modeling. The structural information was obtained by the template from PDB. The identity of this template was 50%, which was good score to begin modelling. The modeled protein structure showed 1052 atoms, 93 H-bonds but no S-S bonds with 3 helices, 9 strands, and 11 turns. The homology modeling was done using Modeller9v7 comparing with templates retrieved from Protein Data bank and finally visualized using Rasmol (***Fig.2***).

4. DISCUSSION

Result analysis of the hypothetical protein showed the conserved domain characteristics of Osmotically inducible protein C (OsmC) superfamily. This family also contains an organic hydroperoxide (Ohr) detoxification protein that has a novel pattern of oxidative stress regulation but, experimental data showed that Ohr and OsmC define two functionally distinct subfamilies with distinct pattern of regulation [20]. OsmC is a stress–induced protein found in *Escherichia coli*. The transcription of the OsmC gene of *Escherichia coli* is regulated as a function of the phase of growth and is induced during the decelerating phase, before entry into stationary phase [21, 22]. Osmotically inducible protein C (OsmC) is involved in the cellular defense mechanism against oxidative stress caused by exposure to hyperoxides or elevated osmolarity [23, 24]. The osmotically inducible expression of the gene suggests that it could have some kind of role in the bacterial osmotic-stress response [20]. It has been experimentally demonstrated that OsmC preferentially metabolizes organic hydroperoxides over inorganic hydrogen peroxide and on the basis of structural and enzymatic similarities; the OsmC catalytic mechanism is analogous to that of the Ohr proteins [25]. Mutants in OsmC, a member of the rpoS regulon, exhibited reduced survival and increased sensitivity to oxidative stress when survival of *Escherichia coli* was investigated during long-term starvation in rich media [26]. The biochemical function of the envelope protein OsmC remains unknown [20, 24], but present data indicated that it participates, directly or indirectly, in the defense against oxidative compounds [25]. Studies revealed that, when Osm C gene from *Thermococcus kodakaraensis* (KOD1) was cloned and expressed in *E.coli*, TkOsmC has a significant peroxidase activity towards both organic and inorganic peroxides in high, but not in low temperature [23]. As, OsmC is widely distributed cluster in both Gram-positive and Gram-negative bacteria [20], further research involving development of appropriate strategies to study the protein biochemical function and regulation at various temperatures to reveal the adaptability and survival mysteries of bacteria to changing environments.

ACKNOWLEDGEMENT

The author is thankful to the Coordinator and Staff members at P. G. Department of Bioinformatics, Shri Shivaji Science College, Amravati and Bioinformatics Center of BTISNet, Department of Biotechnology, Sant Gadge Baba Amravati University, Amravati (M.S.), India for their encouragement, support and research facilities.

REFERENCES

1. Heimdahl A, Nord C (1988) Antimicrobial agents in the treatment of periodontal diseases: special aspects on tetracycline and doxycycline. *Scand J Infect Dis Suppl* 53: 35–45.
2. Mally M, Shin H, Paroz C, Landmann R, Cornelis G (2008) *Capnocytophaga canimorsus:* A Human pathogen feeding at the surface of epithelial cells and phagocytes. *PLoS Pathogens* Volume 4, Issue 9, e1000164.

3. Chraibi D, Girond S, Michel G. (1990) Evaluation of the activity of four antimicrobial agents using an in vitro rapid micromethod against oral streptococci and various bacterial strains implicated in periodontitis. *J Periodontal Res* 25 (4): 201–6.
4. Gougeon A, Sixou J, Shacoori Z, Mallet M (2007) Antimicrobial treatment of *Capnocytophaga* infections. *Journal of Antimicrobial Agents* Volume 29, Issue 4, Pages 367-373.
5. Lubec G, Afjehi- Sadat, Yang J, Pradeep John J (2005) Searching for hypothetical proteins: Theory and practice based upon original data and literature. *Progress in Neurobiology* 77 90–127.
6. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000) The Protein Data Bank. *Nucleic Acids Research Vol.28, No.1.*
7. Bhattacharjee A, Choudhury H, Maheswari U, Joshi S (2008) In-silico prediction of structural and functional aspects of a hypothetical protein of Arabidopsis thaliana (L) Heynh. *Advanced Biotech* pp 13-17.
8. Lipman D, Ostell J, Benson D, Boguski M (1994) Genbank. *Nucleic Acids Res* pages 22: 3441-3444.
9. Pruitt K, Tatusova T, Klimke W, Maglott D (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* Jan; 37(Database issue):D32-6.
10. Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
11. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
12. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel R, Bairoch A Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook.*
13. Thakallapally R, Kibbe W, Lang D, Korber B (2000) Motifscan: A Web-based Tool to Find HLA Anchor Residues in Proteins or Peptides. *Theoretical Biology and Biophysics.*
14. Garnier J, Gibrat J, Robson B (1996) GOR secondary structure prediction method version IV. *Methods in Enzymology* R.F. Doolittle Ed., vol 266, 540-553.
15. Murzin A, Brenner S, Hubbard T, Chothia C (1995) SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.* 247, 536–540.
16. Combet C, Jambon M, Deléage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics* 18:213-214.
17. Pilley H, Atre N (2011) Homology Modelling based Protein Structure Prediction of Iron regulated Peptidyl-Prolyl cis-trans Isomerase A from *Mycobacterium tuberculosis* strains H37Rv. *J Comp. Intell. Bioinfo. (JCIB*Vol 4 (1) pp.79 – 86.
18. Laskowski R, MacArthur M, Moss D, Thornton J (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst* 26, 283-291.
19. Neelamathi E, Vasumathi E, Bagyalakshmi S, Kannan R (2009) Insilico prediction of structure and functional aspects of a hypothetical protein of Neurospora crassa. *Journal of Cell and Tissue Research* Vol. 9(3), pp 1889-1894.
20. Atichartpongkul S, Loprasert S, Vattanaviboon P, Whangsuk W, Helmann J, Mongkolsuk S (2001) Bacterial Ohr and OsmC paralogues define two protein families with distinct functions and patterns of expression. *Microbiology* 147, 1775–1782.
21. Gordia S, Gutierrez C (2003) Growth-phase-dependent expression of the osmotically inducible gene *osmC* of *Escherichia coli* K-12. *Molecular Microbiology* Volume 19, Issue 4, pages 729–736.
22. Gutierrez C, Devedjian J (1991) Osmotic induction of gene *osmC* expression in *Escherichia coli* K12. *Journal of Molecular Biology* Volume 220, Issue 4, Pages 959-973.
23. Park S, Pham B, Van D, Jia B, Lee S, Yu R, Han S, Yang J, Hahm K, Cheong G (2008) Structural and functional characterization of osmotically inducible protein C (OsmC) from *Thermococcus kodakaraensis* KOD1. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* Volume 1784, Issue 5, Pages 783-788.
24. Villarejo M, Barron A, May G, Bremer E (1986) Regulation of Envelope Protein Composition during Adaptation to Osmotic Stress in Escherichia coli. *Journal of Bacteriology* pp. 433-438.
25. Lesniak J, Barton W, Nikolov D (2003) Structural and functional features of the Escherichia coli hydroperoxide resistance protein OsmC. *Protein Science* 12:2838–2843.
26. Conter A, Gangneux C, Suzanne M, Gutierrez C (2001) Survival of Escherichia coli during long-term starvation: effects of aeration, NaCl, and the rpoS and osmC gene products. *Res Microbiol.* Jan-Feb, 152(1):17-26.