Assessing the precision of high-throughput computational approaches for the genome-wide subcellular localization of putative proteins from Vibrio cholerae

P Somvanshi, V Singh, P Seth

Citation

Abstract

Vibrio cholerae have evolved mechanism to become pathogenic to humans with a potential to cause the severe life-threatening diarrhea disease, cholera. Cholera can emerge as explosive outbreaks in the human population. V. cholerae illness is produced primarily through the expression of a potent toxin within the human intestine. Several proteins are involved in the pathogenesis and regulation of bacterial cell activity. The insilico prediction of protein subcellular localization was used to distinguish the actual location in the cells. Total 52 V. cholerae proteins were analyzed with the help of PSLPred. The subcellular localization of these proteins was five in the cells like cytoplasm, periplasm, inner membrane, outer membrane and extracellular space. They have widespread applications in function of proteins in the host cell and in designing the drugs.

INTRODUCTION

Worldwide 1.3 billion cases of acute diarrhea occur in children below 5 years annually of which more than 3 million die and 80 per cent of these deaths are in children below 2 years of age (Sur and Bhattacharya 2003). Acute diarrhea was caused by various numbers of bacterial, viral or parasitic agents. The most important bacterial agents causing outbreaks of acute diarrhoea are V. cholerae O1 and O139. Diarrheal disease outbreaks were causes of major public health emergencies in India. West Bengal located in the Gangetic delta has been hailed as the "homeland of cholera", with frequent localized outbreaks being reported (Fule et al 1990). Vibrio cholerae O1 and V. cholerae O139 are etiological agents of epidemic cholera. However, V. cholerae O1 strains that do not produce cholera toxin, i.e., that are nontoxigenic (NT), and non-O1/non-O139 strains have also been associated with cholera, gastroenteritis, septicemia or other intestinal infections (Morris et al 1984; Mukhopadhyay et al 1995). Outbreaks of cholera were reported in Brazil during the third (1853 to 1854), fourth (1866 to 1868), and fifth (1893 to 1895) pandemics (Barua, 1992).

The computational prediction of the subcellular localization of bacterial proteins is an important step in genome annotation and in the search for novel vaccine or drug targets. V. cholerae have five major subcellular localization sites like cytoplasm, periplasm, inner membrane, outer membrane and extracellular space. The subcellular location of a protein can provide valuable information about its function. With the rapid increase of sequenced genomic data, the need for an automated and accurate tool to predict subcellular localization becomes increasingly important. Because of its simplicity, this approach can be easily extended to other organisms and should be a useful tool for the high-throughput and large-scale analysis of proteomic and genomic data. Generally, existing methods of subcellular localization developed for eukaryotic proteins like TSSub, LOCSVMPSI, ESLpred, Euk-Ploc (Guo and Lin 2006; Xie et al 2005; Bhasin and Raghava 2004; Shen et al 2007). Since, a range of bioinformatics tools were available for prediction of subcellular localization of prokaryotic proteins viz. PSORTb, PSLpred (Gardy et al 2005; Bhasin et al 2005).

Computational SCL investigation of the growing number of complete bacterial genomes or individual proteins allow researchers to screen for vaccine/drug candidates, automatically annotate gene products or select proteins for further study. The pathogenicity of V. cholerae O1 and O139 strains depends on a combination of properties, including

enterotoxin (cholera toxin [CT], ctxA) and the ability to adhere to, and colonize, the small intestine (colonization factor, tcpA) (Herrington et al 1988). The major virulence-associated factors are present in clusters with at least three regions in the V. cholerae chromosome. Several factors have been associated with entero pathogenicity which, includes a El Tor-like hemolysin (hlyA) (Yamamoto et al 1984), heat-stable enterotoxin (stn/sto) (Arita et al 1986), hemagglutinins (Datta-Roy et al 1986), outer membrane protein (ompU) (Sperandio et al 1996), Shiga-like toxin (stx) (Kaper et al 1994), a ToxR regulatory protein (Miller et al 1987), and a zonula occludens toxin (zot) (Fasano et al 1991). The application of this study was to predict the sub cellular localization of putative proteins of V. cholerae O1 strain as specific subcellular localization of bacterial proteins was useful in designing the therapeutics against bacterial proteins that can not affect the host cellular activity. Subcellular localization of proteins is crucial for the pathogenesis and survival of bacteria as they might be useful for targeting drugs.

MATERIALS AND METHODS

COLLECTION OF SEQUENCES

The complete nucleotide and protein sequences were extracted from biological database National Centre for Biotechnology Information (NCBI) cited at http://www.ncbi.nlm.nih.gov

ANALYSIS OF PHYSICO CHEMICAL PROPERTIES

The physico-chemical properties of proteins were analyzed viz. total number of amino acids, molecular weight and isoeletric point with Generunner, DNAstar and ExPaSy tools.

PREDICTION OF SUB CELLULAR LOCALIZATION OF PROTEINS

Total 1302 bacterial proteins have been used to develop the PSLpred tool. The five localization and numbers of proteins (248 cytoplasmic, 268 inner membrane, 244 periplasmic, 352 outer membrane and 190 extracellular) have been included. Machine-learning technique, SVM, has been used for the prediction of subcellular localizations of prokaryotic proteins. The prediction of subcellular localizations is a multi-class classification problem. The performance of the SVM modules developed in the present study was evaluated through 5-fold cross-validation technique. In this technique, the relevant dataset is partitioned randomly into five equal sized sets. The training and testing was carried out five times, using one distinct set for each testing and the remaining (four sets) for the training. In order to assess the predictive performance, accuracy and Matthew's correlation coefficient (MCC) (Matthews, 1975) have been calculated (Bhasin et al 2005).

RESULTS AND DISCUSSION

In this study we had selected fifty two putative protein of V. cholerae O1 and their physico chemical nature was analyzed theoretically. The molecular weight and isoelectric point of all these proteins was deduced (Table 1).

Figure 1

Table 1: Physico-chemical properties and subcellular localization of putative proteomes of O1 strain.

Sl. No	Protein Designation	Accession No.	Sequence (Amino acid)	Molecular weight (Da)	pI	Subcellular localization of proteins
1	Inner membrane protein	AAF93182.1	592	60627.1	6.46	Inner membrane
2	ABC transporter, ATP-binding protein	AAF93186.1	331	27306.8	8.14	Inner membrane
3	ABC transproter, permease protein	AAF93187.1	308	24956.8	9.25	Inner membrane
4	ABC transporter, periplasmic amino acid-binding portion	AAF93188.1	341	27306.2	5.12	Periplasmic
5	Zinc binding Alcohol dehydrogenase	AAF93204.1	348	34508.1	6.36	Cytoplasmic
6	Hemolysin, putative	AAF93218.1	215	23666.3	9.18	Inner membrane
7	Multidrug resistance protein	AAF93247.1	214	43098.7	9.69	Inner membrane
8	Universal stress protein A	AAF93254.1	427	15708.5	4.69	Cytoplasmic
9	alkaline serine protease	AAF93333.1	607	56867.7	6.24	Extracelular
10	Multidrug resistance protein, putative	AAF93340.1	1087	111711.6	5.28	Inner membrane
11	Peptide ABC transporter, ATP-binding protein	AAF93346.1	656	63528.6	7.09	Inner membrane
12	Peptide ABC transporter, periplasmic peptide-binding protein"	AAF93347.1	642	60621.9	5.82	Periplasmic
13	Peptide ABC transporter, permease protein	AAF93349.1|	1038	34318.5	6.40	Periplasmic
14	Transposase, putative	AAF93361.1	472	47647.4	9.29	Cytoplasmic
15	Hemolysin secretion ATP-binding protein, putative	AAF93375.1	611	68787.1	6.96	Inner membrane
16	Lipopolysaccharide biosynthesis protein RfbV	AAF93434.1	461	46843.2	8.91	Cytoplasmic
17	Iron-containing alcohol dehydrogenase family protein RfbM"	AAF93426.	886	41234.0	5.65	Cytoplasmic
18	Bacterioferritin-associated ferredoxin	AAF93537.1	113	6839.8	8.89	Cytoplasmic
19	Sulfite reductase (NADPH) flavoprotein	AAF93557.1	665	67710.0	5.22	Cytoplasmic
20	Malate dehydrogenase	AAF93605.1	422	36962.8	7.19	Cytoplasmic
21	Membrane-bound lytic murein transglycosylase C	AAF93623.1	482	45576.1	9.48	Periplasmic
22	Iron-regulated outer membrane virulence protein, TonB receptor family	AAF93648.1	703	71677.2	4.88	Outer membrane
23	Protease, insulinase family/protease, insulinase family"	AAF93722.1	1003	106367.7	5.79	Periplasmic
24	Hemolysin, putative	AAF93726.1	492	47412.9	6.57	Cytoplasmic
25	protease DO"	AAF93734.1	515	48372.2	5.74	Periplasmic
26	Hemolysin, putative	AAF93746.1	255	20580.2	9.18	Periplasmic
27	Lipoprotein	AAF93749.1	704	73489.8	5.23	Inner membrane
28	Carbonic anhydrase, putative	AAF93753.1	295	25227.2	6.08	Cytoplasmic
29	Sugar fermentation stimulation protein	AAF93764.1	337	28288.7	8.31	Cytoplasmic
30	Multidrug resistance protein, putative	AAF93795.1	1154	117904.6	4.77	Inner membrane protein
31	Outer membrane protein OmpU	AAF93799.1	419	37663.6	4.34	Outer membrane
32	Ferredoxin	AAF93881.1	143	9280.5	4.41	Cytoplasmic
33	Protease, putative	AAF93882.1	531	52483.3	5.83	Cytoplasmic
34	Protein-export membrane protein SecF	AAF93909.1	366	34333.2	4.99	Inner membrane
35	Toxin co-regulated pilus biosynthesis protein B	AAF93992.1	513	47124.2	5.25	Cytoplasmic
36	Toxin co-regulated pilus biosynthesis protein Q	AAF93993.1	234	17244.2	4.43	Cytoplasmic
37	Outer membrane protein OmpV	AAF94476.1	327	28114.3	5.23	Outer membrane
38	Multidrug transporter, putative	AAF94549.1	543	50141.7	9.79	Inner membrane
39	Multidrug resistance protein, putative	AAF94566.1	565	52137.5	4.89	Outer membrane
40	Multidrug resistance protein VceA	AAF94567.1	484	43602.6	6.22	Cytoplasmic
41	Multidrug resistance protein VceB	AAF94568.1	588	55895.6	8.23	Inner membrane
42	Transposase OrfAB, subunit A	AAF94632.1	186	13414.0	9.87	Cytoplasmic
43	Transposase OrfAB, subunit B	AAF94633.1	358	33558.1	10.07	Inner membrane
44	Multidrug resistance protein NorM,putative	AAF94694.1	545	49859.8	9.18	Inner membrane
45	Outer membrane protein TolC, putative	AAF94719.1	496	46741.5	5.09	Outer membrane
46	Multidrug transporter, putative	AAF94751.1	552	50809.6	9.89	Inner membrane
47	Outer membrane protein, putative	AAF94775.1	286	23468.9	4.71	Outer membrane
48	Multidrug resistance protein	AAF94785.1	475	42920.1	9.61	Inner membrane
49	Multidrug resistance protein, putative	AAF94825.1	424	37743.4	6.20	Cytoplasmic
50	Peptidase, putative	AAF95131.1	682	70022.4	5.52	Cytoplasmic
51	Outer membrane lipoprotein LolB	AAF95326.1	283	24381.5	6.19	Cytoplasmic
52	Cholera toxin secretion protein EpsM	AAF95864.1	217	18784.9	9.55	Cytoplasmic

It indicates protein stability in a particular isoelectric point (pI). An online PSLpred server was used to predict protein subcellular localization within bacteria or targeting the host. We investigate putative proteome of V. cholerae and their specific subcellular location (Table1). V. cholerae elaborates zonula occludens toxin (Zot), a protein that increases the permeability of small intestinal mucosa by opening intercellular tight junctions. The zot gene is located, together with the genes encoding CT and Ace enterotoxins, within the genome of V. cholerae filamentous phage CTX small ef, Cyrillic. Zot localizes in the V. cholerae cell envelope with M(r); 45 kDa which is consistent with the predicted primary translation product from the first methionine of zot (44.8 kDa). A second molecule, corresponding to the 33 kDa N-terminal region of Zot, was also detected. Both molecules are exposed at the bacterial cell surface (Uzzau et al 1999). The cellular location of the haemolysin of Vibrio cholerae El Tor strain 017 has been analyzed. This protein is found both in the periplasmic space and the extracellular medium in V. cholerae (Mercurio et al 1985).

V. cholerae O1 biotype El Tor produces and secretes a 65-kDa cytolysin/hemolysin into the culture medium. We cloned the structural gene (hlyA) for the cytolysin from the total DNA of a V. cholerae O1 El Tor strain, N86. Sequence analysis of hlyA protein of 741 amino acids with a molecular weight of 81,961 was done. Consistent with this, a 79-kDa protein was identified as the product of hlyA by maxicell analysis in Escherichia coli. N-terminal amino acids of this 79-kDa HlyA protein and those of a 65-kDa El Tor cytolysin purified from V. cholerae were Asn-26 and Asn-158, respectively. The 82- and 79-kDa precursors of the 65-kDa mature cytolysin were found in V. cholerae by pulse-chase labeling and Western blot (immunoblot) analysis of hlyA products. V. cholerae, the 82-kDa preprotoxin synthesized in the cytoplasm is secreted through the membranes into the culture medium (Yamamoto et al 1990).

In previous studies thirty-nine putative proteins of Mycobacterium tuberculosis H37Rv strain were predicted for four locations viz cytoplasmic, integral membrane, secretory and protein attached to membrane by Lipid anchor in the subcellular localization (Somvanshi et al 2008).Also prediction of motif in Dengue virus serotypes polyprotein was done and several significant motifs which involves in regulation, activity and stability of virus were observed for targeting or designing the antiviral inhibitors against these motifs of Dengue virus (Somvanshi and Seth 2008).

In conclusion, we include the specified prediction of subcellular localization results in the most putative proteins of strain of V. cholerae O1. This initiative may help in annotating newly sequenced or hypothetical proteins. Thus, the search for a potential vaccine/ drug target for an important bacterial pathogen by in vitro researchers will greatly be appended by this prediction.

References

1. Allan, E. and Wren, B.W. (2003) Genes to genetic immunization: identification of bacterial vaccine candidates. Methods, 31, 193-198.
2. Arita, M. T., T. Takeda, T. Honda, and T. Miwatani. 1986. Purification and characterization of Vibrio cholerae non-O1 heat-stable enterotoxin. Infect. Immun. 52:45-49.
3. Barua, D. 1992. History of cholera, p. 1-35. In D. Barua and W. B. Greenough III (ed.), Cholera. Plenum Medical Book Company, New York, N.Y.
4. Bhasin M, A. Garg and G. P. S. Raghava. 2005. PSLpred: prediction of subcellular localization of bacterial proteins. 21(10): 2522-2524.
5. Datta-Roy, K., K. Banerjee, S. P. De, and A. C. Ghose. 1986. Comparative study of expression of hemagglutinins, hemolysins, and enterotoxins by clinical and environmental isolates of non-O1 of Vibrio cholerae in relation to their enteropathogenicity. Appl. Environ. Microbiol. 52:875-879.
6. Fasano, A., B. Baudry, D. W. Pumplin, S. S. Wasserman, B. D. Tall, J. M. Ketley and J. B. Kaper. 1991. Vibrio cholerae produces a second enterotoxin which affects intestinal tight junctions. Proc. Nat. Acad. Sci. USA 88:5242-5246.
7. Fule RP, Power RM, Menon S, Basutkar SH, Saoji AM. Cholera epidemic in Solapur during July to August 1988. Indian J Med Res 1990; 91: 24-6.
8. Herrington, D. A., R. H. Hall, G. Losonsky, J. J. Mekalanos, R. K. Taylor, and M. M. Levine. 1988. Toxin, toxin-coregulated pili and the toxR regulon are essential for Vibrio cholerae pathogenesis in humans. J. Exp. Med. 168: 1487-1492.
9. Kaper, J. B., A. Fasano, and M. Trucksis. 1994. Toxins of Vibrio cholerae, p. 145-176. In I. K. Wachsmuth, P. A. Blake, and O. Olsvik (ed.), Vibrio cholerae and cholera: molecular to global perspectives. ASM Press, Washington, D.C.
10. Mekalanos, J. J. 1985. Cholera toxin: genetic analysis, regulation, and role in pathogenesis. Curr. Top. Microbiol. Immunol. 118:97-118.
11. Mercurio A, Manning PA 1985. Cellular localization and export of the soluble haemolysin of Vibrio cholerae El Tor. Mol Gen Genet. 200(3):472-5. Uzzau S, Cappuccinelli P, Fasano A. 1999. Expression of Vibrio cholerae zonula occludens toxin and analysis of its subcellular localization. Microb Pathog. 27(6):377-85.
12. Miller, V. L., R. K. Taylor, and J. J. Mekalanos. 1987. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. Cell 48:271-279.
13. Mora, M., Veggi, D., Santini, L., Pizza, M. and Rappuoli, R. (2003) Reverse vaccinology. Drug Discov. Today, 8, 459-464.
14. Morris, J. G., Jr., J. L. Picardi, S. Lieb, J. V. Lee, A. Roberts, M. Hood, R. A. Guun, and P. Blake. 1984. Isolation of nontoxigenic Vibrio cholerae O group 1 from a patient with severe gastrointestinal disease. J. Clin Microbiol. 19:196-197.

15. Mukhopadhyay, A. K., P. K. Saha, S. Garg, S. K. Bhattacharya, T. Shimada, T. Takeda, Y. Takeda, and G. B. Nair. 1995. Distribution and virulence of Vibrio cholerae belonging to serogroups other than O1 and O139: a nation survey. Epidemiol. Infect. 114:65-70.

16. Paine,K. and Flower,D.R. (2002) Bacterial bioinformatics: pathogenesis and the genome. J. Mol. Microbiol. Biotechnol., 4, 357-365.

17. Somvanshi P, Singh V and Seth P. K 2008. In silico analysis of subcellular localization of putative proteins of Mycobacterium tuberculosis H37Rv strain. Int J Health (In Press).

18. Somvanshi P and Seth P. K 2008. Comparative proteome analysis of distinct variants of Dengue virus using insilico methods. Indian J of med Res (In Press).

19. Sperandio, V., C. Bailey, J. A. Giron, V. J. DiRita, W. D. Silveira, A. L. Vettore, and J. B. Kaper. 1996. Cloning and characterization of the gene encoding the OmpU outer membrane protein of Vibrio cholerae. Infect. Immun. 64:5406-5409.

20. Sur D, Bhattacharya SK. Research and development of vaccines against diarrhoeal diseases. In: Ghosh TK, Kalra A, editors. Infectious diseases in children and newer vaccines. Part II. 2003 p. 158-65.

21. Yamamoto K, Ichinose Y, Shinagawa H, Makino K, Nakata A, Iwanaga M, Honda T, Miwatani T. 1990. Two-step processing for activation of the cytolysin/hemolysin of Vibrio cholerae O1 biotype El Tor: nucleotide sequence of the structural gene (hlyA) and characterization of the processed products. Infect Immun. 58(12):4106-16.

22. Yamamoto, K., M. Al-Omani, T. Honda, Y. Takeda, and T. Miwatani. 1984. Non-O1 Vibrio cholerae hemolysin: purification, partial characterization, and immunological relatedness to El Tor hemolysin. Infect. Immun. 45:192-196.

Author Information

Pallavi Somvanshi
Biotech Park

Vijai Singh
Aquatic Microbes Section, National Bureau of Fish Genetic Resources

P.K. Seth
CEO, Biotech Park