Available online at www.ilns.pl

International Letters of Natural Sciences

5 (2013) 52-58 ISSN 2300-9675

Structural analyses of *Shigella* invasion proteins reveals non-conserved; intrinsically unstructured regions

Santanu Chakrabarti[1,*], Sayak Ganguli[2,]**

[1]Department of Zoology, Acharya Prafulla Chandra Roy Government College, Silguri 734 010, WB, India

[2]DBT Centre for Bioinformatics, Presidency University, Kolkata 700073, WB, India

*,**E-mial address: scwbes@gmail.com , sayakbif@yahoo.com

ABSTRACT

Shigella is one of the most common bacterial pathogens that are isolated from patients with diarrhea. Various attempts are being made worldwide with encouraging observations; still the emergence of multidrug-resistant *Shigella* strains and a continuous high disease incidence imply that shigellosis is an unsolved global health problem which can probably be solved only by developing a proper vaccine and a vaccine regime for the disease. The need of the hour is to foster the development of an effective vaccine which should not only serve to improve hygiene but also should be able to curb infections by the pathogen. This goal can only be achieved by gaining proper detailed knowledge underlying *Shigella* pathogenesis. The analyses of the *Shigella* invasion proteins which have been long been targeted to be potential candidate vaccines remains an open ended problem and forms the core of this present computational study which identifies the fact that long regions in the structure of the proteins are disordered having no distinct structural conformation; multiple alignments however, did not show any conserved stretches in the disordered regions. The results probably explain the ability of these proteins to interact with multiple cellular proteins and perform a diverse array of functions leading to successful pathogenesis.

***Keywords*:** Invasion proteins; intrinsically unstructured proteins; moonlighting; Kyle – Doolittle scale; Disordered region; functional promiscuity

1. INTRODUCTION

Among the first group of proteins secreted by the *Shigella* spp. type III secretion system, are the dominant immunogenic invasion plasmid antigens IpaA to IpaD (Hromockyj et.al. 1989; Yang et.al. 2007) of which IpaB, IpaC and Ipa D have been identified to be the key virulence factors. Invasion of epithelial cells by *Shigella* is an essential pathogenic feature of bacillary dysentery. Several workers have indicated that the prerequisite for *Shigella* internalization is the successful delivery of a viral load of a set of effector proteins through the type III secretion system of the bacteria into the epithelial cells of the host. This package is consisted of the *Shigella* invasion proteins such as IpaA, IpaB, IpaC, IpaD, IpgD and

virulence proteins such as VirA (Allaoui et al., 1993a; Ménard et al., 1993, 1994; Uchiya et al., 1995; Tran Van Nhieu et al., 1997; Tran Van Nhieu and Sansonetti, 1999). Though we have much to learn and envisage regarding the precise mechanisms of invasion recent studies have indicated that IpaA and IpaC once inside the host are capable of modulating dynamics of actin filaments as well as signal transduction pathways which are needed for the survival of the pathogen. (Tran Van Nhieu et al., 1997, 1999; Bourdet-Sicard et al., 1999; Tran Van Nhieu and Sansonetti, 1999).

Thus *Shigella* invasion proteins may serve as prime targets for vaccine development since they are an essential part of the triggering cascade. Their key functions are provided in Table 1 (Yoshida, 2006).

Table 1. Invasion plasmid antigens of *Shigella* and their key functions. (Yoshida, 2006). This table provides a brief idea about the multitasking abilities of the Shigella invasion proteins during their role in the pathogenesis.

IpaA	Increasing invasion, actin cytoskeleton rearrangements, disassembly of cell-matrix adherence
IpaB	Control of type III secretion, formation of Translocon, mediating phagosome escape, macrophage apoptosis
IpaC	Formation of Translocon and filopodium, mediating phagosome escape, disrupting the EC tight junctions
IpaD	Type III secretion control, membrane insertion of translocon

The present computational study tried to identify the fact that long regions in the structure of the Ipa proteins of *Shigella* are disordered i.e. having no distinct structural conformation.

2. MATERIALS AND METHOD

Uversky et al. (2000) has defined the mean net charge, <R>, as the absolute value of the difference between the numbers of positively and negatively charged residues at pH 7.0, divided by the total residue number, and the mean hydrophobicity, <H>, as the sum of all residue hydrophobicities, divided by the total number of residues, using the Kyte/Doolittle scale (Kyte and Doolittle, 1982), rescaled to a range of 0-1. Using the same algorithm, the threshold used for the analyses was kept using mean hydrophobicity at -1.16 and mean net charge at 2.785.

The FoldIndex equation was used as standard and all positive values indicated residues with less potential to get folded while negative values indicated propensity for unstructuredness. The results were then verified using established programs such as Fold Index (Prilusky 2005) and PONDR – Fit (Xue 2010) programs. A multiple sequence alignment followed by phylogenetic tree was generated using MUSCLE with standard input parameters.

3. RESULTS AND DISCUSSION

All four major *Shigella* Invasion proteins displayed large regions of unstructuredness according to our calculations (Fig 1). The verification of the results by the established prediction programs such as PONDR- FIT (Fig. 2) and Fold Index (Fig. 3) also showed similar degrees of unstructuredness in the *Shigella* invasion proteins.

Fig. 1. Predicted unstructuredness in the *Shigella* invasion proteins. (Amino acids in capital letters indicate disordered residues as predicted).

Protein Name	Disorder	Residues
IpaA	1-59, 63-90, 101-120, 124-174, 193-207, 214-239, 257-356, 374-499, 511-576, 584-633	MHNVNNTQAP TFLYKATSPS STEYSELKSK ISDIHSSQTS LKTPASVSEK ENFATSFNQk clDFLFSSSG KEDVLRSIYS NSMNAYAKSE ilefsnvlys LVHQNDLNFE NEKGLQKIVA qysELIIKDK LSQDSAFGPW SAKNKKLHQL RQNIEHRLAL LAQQHTSGEA LSLGqkllnt evssfiknni laELKLSNET VSSLKLDdlv daqAKLAFDS LRNQRKNTID SKGFGIGKLs rdlntvavfp ellrkvLNDI LEDIKDSHPI QDGLPTPPED MPDGGPTPGA NEKTSQPVIH YHINNDNRTY DNRVFDNRVY DNSYHENPEN DAQSPTSQTN DLLSRNGNSL LNPQRAlvqk vtsvlphsis davQTFANNS ALEKVFNHTP DNSDGIGSDL LTTSSQERST NNSLSRGHRP LNIQNSSTTP PLHPEGVTSS NDNSSDTTKS SASLSHRVAS QINKFNSNTD SKVLQTDFLS RNGDTYLTRE TIFEASKKVt nslsnlisli GTKSGTQERE LQEKSKDITK STTEHRINNK LKVTDANTIN YVTETNADTI DKNHAIYEKA KEVSSAIskv lskIDDTSAE LLTDDISDLK NNNDITAENN NIYKAAKDVT TSLSKVLKNI NKD
IpaB	1-10, 22-33, 46-74, 97-243, 259-305, 387-393, 448-460, 521-546, 559-571	MHNVSTTTTG fplakilast eLGDNTIQAA NDAanklfsl tiadlTANQN INTTNAHSTS NILIPELKAP KSLNassqlt llignliqil geksltALTN KITAWKSQQQ ARQQKNLEFS DKINTLLSET EGLTRDYEKQ INKLKNADSK IKDLENKINQ IQTRLSELDP ESPEKKKLSR EEIQLTIKKD AAVKDRTLIE QKTLSIHSKL TDKSMQLEKE IDSFSAFSNT ASAEQLSTQQ KSLtglasvt qlmatfiqLV GKNNEESLKN DLALFQSLQE SRKTEMERKS DEYAAEVRKA EELNRvmgcv gkilgallti vsvvaaafsg gaslalaavg lalmvtdaiv qaatgnsfme qalnpimkav ieplikllsd aftkmlEGLG VDSkkakmig silgaiagal vlvaavvlva tvgkqaaakl aenigkiigk tltdlipKFL KNFSSQLDDL itnavarlnk flgaagdevi skqiisthln qavllgesvn satqaggsva savfqnsast NLADLTLSKY QVEQLSKYIS EAIEKFgqlq eviadllaSM SNSQANRTDV Akailqqtta
IpaC	Disorder 1-40, 47-98, 141-145, 171-284, 291-335, 343-361	MEIQNTKPTQ ILYTDISTKQ TQSSSETQKS QNYQQIAAHI plnvgkNPVL TTTLNDDQLL KLSEQVQHDS EIIARLTDKK MKDLSEMSHT LTPENTLDis slssnavsli isvavllsal rtaetklgsq lsliafdatk SAAENivrqg laalsssitg avtqvgitgi GAKKTHSGIS DQKGALRKNL ATAQSLEKEL AGSKLGLNKQ IDTNITSPQT NSSTKFLGKN KLAPDNISLS TEHKTSLSSP DISLQDKIDT QRRTYELNTL SAQQKQNIGR ATMEtsavag NISTSGGRYA SALEEEEQLI SQASSKQAEE ASQVSKEASQ ATNQLiqkll niIDSINQSK NSTASQIAGN Ira
IpaD	1-164, 179-307	MNITTLTNSI STSSFSPNNT NGSSTETVNS DIKTTTSSHP VSSLTMLNDT LHNIRTTNQA LKKDLSQKTL TKTSLEEIAL HSSQISMDVN KSAQLLDILS KKEYPINKDA RELLHSAPKE AELDGYEMIS HRELWDKIAK SINNINEQYL KVYEHAVSSY TQMY qdfsav lsslagwi SP GGNDGNSVKL QVKSLKDELT KLKEKYKDKP LYPANNTVSK EQANKWLTEL GGTIGKVSEK NGGYVVNINM TPIDNTLKSL DNLGGNGEVV LDNAKYQAWN AGFSAEDETM KNNLQTLVQK YSNANSI fdn lvkvlsstis sctdtdklfl hf

The figure provides insights on the exact residues that are predicted to be in the intrinsically disordered region of the various *Shigella* invasion proteins.

Ganguli et.al. (2011) have reported the significance of structural analyses in developing an immuno-prophylactic measure such as vaccine identification and design in their works. Over the years numerous evidences has accumulated that many important proteins, in whole or in part, is unstructured in their native state (Amitai et.al. 2007).

Such proteins have been referred to as intrinsically unstructured proteins and rather than folding into a single, stable, 3D structure, these protein members exist as a conglomeration of rapidly changing conformations which are interchangeable and often appear like the denatured states of ordered proteins (Ma 1999; Tsai et.al. 1999; 2001, 2009).

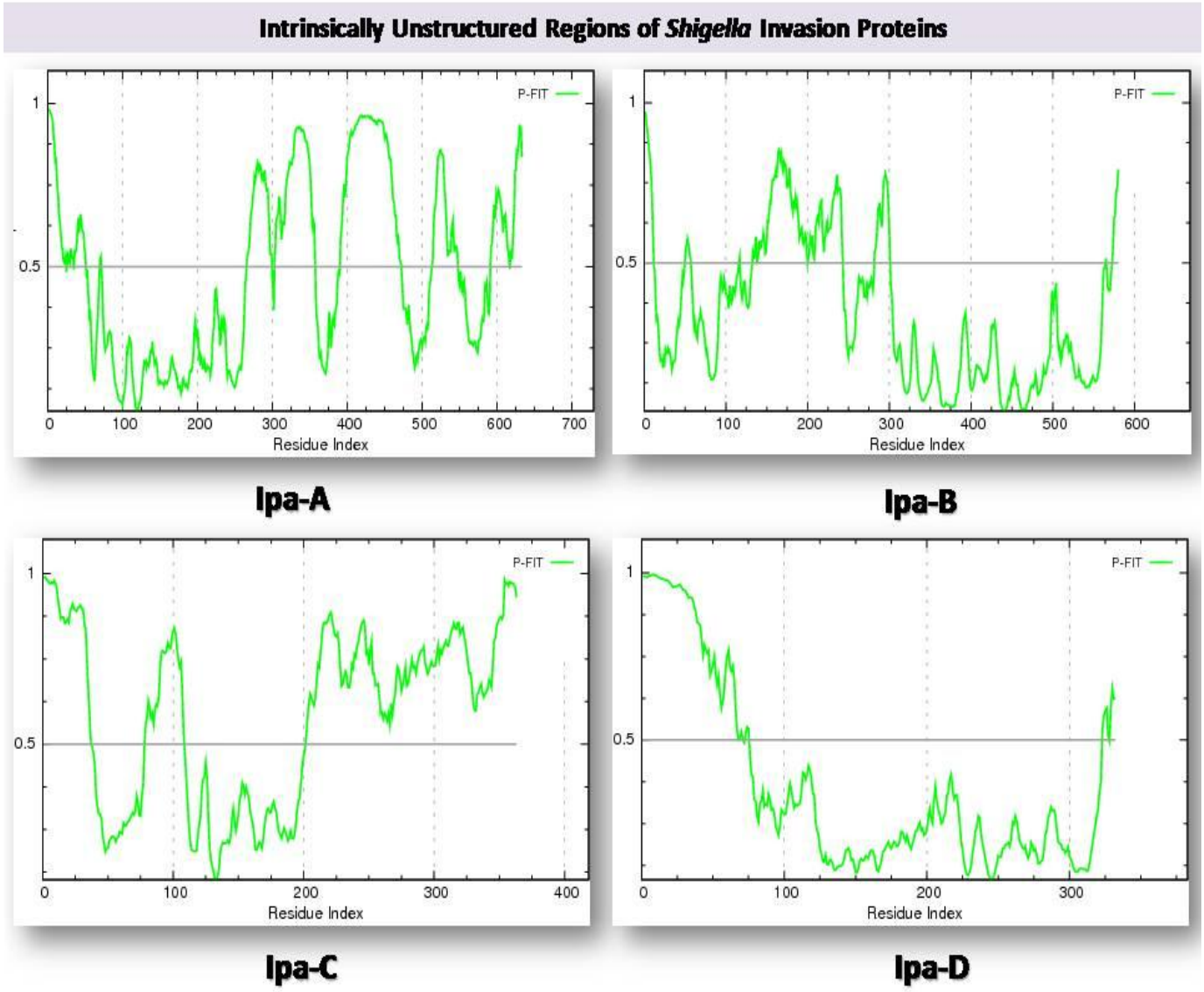


Fig. 2. Prediction of disordered region by PONDR – FIT (0.5 threshold) The above figure provides a graphical overview regarding the number of intrinsically disordered residues present in the four *Shigella* invasion proteins used in the study. The threshold used for analyses is 0.5 which is considered to be standard for bacterial systems.

Though most of these proteins have been reported to lack stable secondary or tertiary structural conformations, yet many of them have been implicated in crucial role plays in regulatory cellular events such as transcription regulation, mRNA processing, DNA condensation along with differentiation and apoptosis.

Some workers believe that there lies an intricate network of proteins which possess no definite structural conformation in part or in whole and functional promiscuity. Such proteins have been referred to as moonlighting proteins (Tompa 2005; Hernandez et.al. 2012) and it may be envisaged that the functional multiplicity may be as a result of the disordered regions.

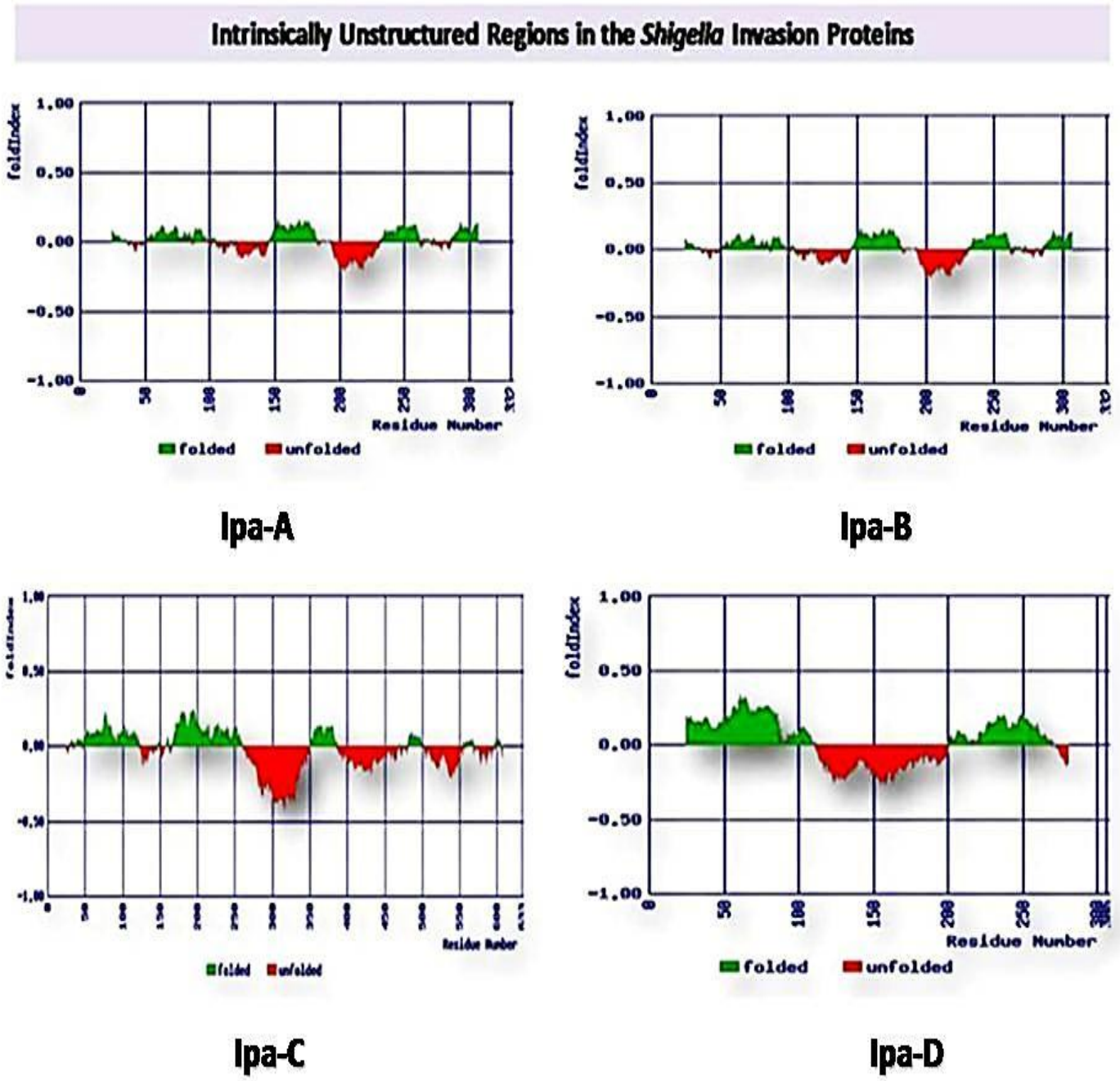


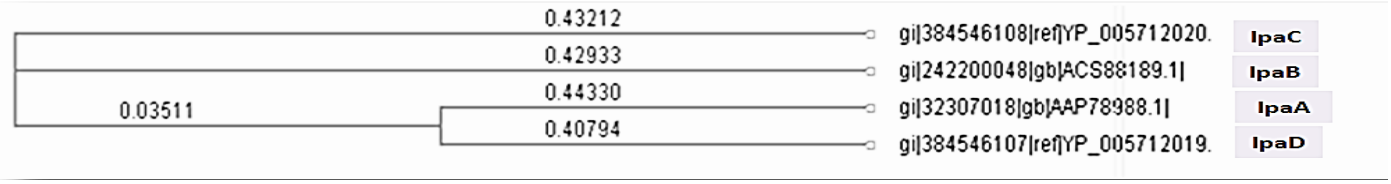
Fig. 3. Prediction of disordered region (indicated in red) by FOLD INDEX.

The above figure provides a graphical overview regarding the number of intrinsically disordered residues present in the four *Shigella* invasion proteins used in the study. The threshold used for analyses is 0.5 which is considered to be standard for bacterial systems.

The needle part of the Supramolecular structure of the *Shigella* type III secretion machinery has showed that it is changeable in length and essential for delivery of other effectors in *Shigella* invasion related pathomechanisms (Koichi Tamano et al, 2000).

The identification of multiple residues exhibiting lack of proper secondary structure or disorderedness in the four major *Shigella* invasion proteins possibly sheds light on the fact as to how all of these proteins perform more than one function in the pathogenesis cascade of *Shigella.* Whether they can be referred to as moonlighting proteins is a case for experimental establishment in control environments; however, it can be safely concluded that the intrinsic disordered regions of the four *Shigella* proteins under study play major role in the pathogenesis leading to functional diversity (Carayol et.al. 2013). The phylogenetic tree (fig 4) showed that IpaA and IpaD form a single sister group while IpaC and IpaB tend to evolve in a parallel path.

Fig. 4. Phylogenetic relationships among the invasion proteins of *Shigella*.



The phylogenetic tree provides an insight on the interrelationship amongst the four invasion proteins used for this study along with their evolutionary distances.

4. CONCLUSION

The identification of disordered regions in the *Shigella* invasion proteins provide insights to important structural dynamics of invasion proteins secreted by the type two secretion system of the bacterial pathogens. Most pathogens have been reported.

Acknowledgement
The authors acknowledge the DBT – BTBI scheme for provision of funds used to maintain the facility.

References

[1] Amitai G., Gupta R. D., Tawfik D. S., *HFSPJ* 1 (2007) 67-78.

[2] Buchrieser C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti, C. Parsot, *Mol. Microbiol.* 38 (2000) 760-771.

[3] Carayol N., Tran Van Nhieu G., *Current Opinion in Microbiology* 16(1) (2013) 32-37.

[4] Choudhari S. P., et al., *Protein Sci.* 22(5) (2013) 666-670.

[5] Ganguli S., Gupta D., Datta A., *Int. Jour. of Comp. Biol.* (2)1 (2011) 38-40.

[6] Hernández S., et al., *J. Proteomics Bioinform.* 5 (2012) 262-264.

[7] Hromockyj A. E., A. T. Maurelli, *Infect. Immun.* 57 (1989) 2963-2970

[8] Kyte J., Doolittle R. F., *J. Mol. Biol.* 157(1) (1982) 105-132.

[9] Ma B., Kumar S., Tsai C. J., Nussinov R., *Protein Eng.* 12 (1999) 713-720.

[10] Nandi T., Gupta S., Ganguli S., Datta A., *International Journal of Biology, Pharmacy and Allied Sciences* 1(9) (2012) 1 -2.

[11] Prilusky J., et al., *Bioinformatics* 21(16) (2005) 3435-8.

[12] Tompa P, Szász C, Buday L., *Trends Biochem Sci.* 30 (2005) 484-489.

[13] Tsai C. J., Ma B., Nussinov R., *Proc Natl Acad Sci* 96 (1999) 9970-9972.

[14] Tsai C. J., Ma B., Nussinov R., *Trends Biochem Sci.* 34 (2009) 594-600.

[15] Tsai C. J., Ma B., Sham Y. Y., Kumar S., Nussinov R., *Proteins* 44 (2001) 418-427.

[16] Uversky V. N., et al., *Proteins* 41 (2000) 415-427.

[17] Xue B., R. L. DunBrack, R.W. Williams, A. K. Dunker, V. N. Uversky, *Biochem. Biophys. Acta* 1804(4) (2010) 996-1010.

[18] Yang J., et al., *J. Mol. Evol*. 64 (2007) 71-79.

[19] Yoshida S., et al., *Science* 314 (2006) 985-989.

[20] Allaoui A., Ménard R., Sansonetti P. J., Parsot C. (1993a), *Infect. Immun.* 61 (1993) 1707-1714.

[21] Ménard R., Sansonetti P. J., Parsot C., *J. Bacteriol.* 175 (1993) 5899-5906.

[22] Ménard R., Sansonetti P. J., Parsot C., *EMBO J.* 13 (1994) 5293-5302.

[23] Uchiya K.-I., et al., *Mol. Microbiol.* 17 (1995) 241-250.

[24] Tran Van Nhieu G., Sansonetti P. J., *Curr. Opin. Microbiol.* 2 (1999) 51-55.

[25] Tran Van Nhieu G., Ben-Ze'ev A., Sansonetti P. J., *EMBO J.* 16 (1997) 2717-2729.

[26] Bourdet-Sicard R., et al., *EMBO J.* 18 (1999) 5853-5862.

[27] Koichi Tamano, et al., *EMBO J.* 19(15) (2000) 3876-3887.

(Received 11 November 2013; accepted 15 November 2013)