Send Orders for Reprints to reprints@benthamscience.ae

Current Proteomics, 2015, *12*, 000-000

Functional insight into Putative Conserved Proteins of *Rickettsia rickettsii* and their Virulence Characterization

Mohd. Shahbaaz[1], Krishna Bisetty[1], Faizan Ahmad[2] and Md. Imtaiyaz Hassan[2,*]

[1]*Department of Chemistry, Durban University of Technology, Durban – 4000, South Africa* [2]*Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi – 110025, India*

Md. Imtaiyaz Hassan

Abstract: *Rickettsia rickettsii* is an aerobic, Gram-negative and non-motile coccobacillus known to cause Rocky Mountain spotted fever. The sequenced genome of its 'Sheila Smith' strain contains 1,343 protein-coding genes, 3 rRNA genes and 33 transfer RNA genes. There are 680 hypothetical proteins (HPs) present in the genome of *R. rickettsii*. Since functions of these proteins are not validated experimentally, characterization of these HPs may play a significant role in understanding the pathogenic mechanisms of *R. rickettsii*. Hence, functions of these HPs were annotated by *in silico* methods based on sequence similarity, protein clustering and protein-protein interactions. We have successfully predicted functions of 214 proteins among 680 HPs present in *R. rickettsii*. These annotated proteins were further classified into 88 enzymes, 59 transport and membrane proteins, 35 binding proteins, 12 structural motifs and the rest of the protein families. Moreover, we identified HPs involved in virulence among 214 functionally annotated proteins. 15 HPs were classified as virulence factors and two proteins with the highest scores were selected for further analyses. Additionally, molecular dynamics simulations were performed on these selected virulent HPs in order to observe their conformational behaviors. These analyses can further be utilized in the identification of new drug targets for development of better therapeutic agents against the infections caused by *R. rickettsii*.

Keywords: *Rickettsia rickettsii*, Molecular pathogenesis, Hypothetical proteins, Function prediction, Virulence, Molecular dynamics simulation.

INTRODUCTION

Rickettsia rickettsii is an obligatory, Gram-negative and pleomorphic coccobacillus. It belongs to Rickettsiaceae family and is the primary causative means of Rocky Mountain spotted fever in human [1]. The Rickettsiaceae family is subdivided into two groups on the basis of serological characteristics, which are typhus and spotted fever [2]. The Rocky Mountain spotted fever shows high prevalence in children aged 5–9 years and for adults between 40–64 years [1]. In the case of severe infections the manifestations appeared in systemic, pulmonary, cutaneous, gastrointestinal, cardiac, renal, ocular and neurological pathways [1]. The sequenced genome of virulent Sheila Smith strain (Accession number - NC_009882.1) was selected in this study [3]. Its 1.26 Mb genome contains 32% GC content and 1379 genes which express into 1,343 proteins [3].

680 proteins were found to be functionally uncharacterized among the group of 1343 proteins referred to as hypothetical proteins (HPs) due to unavailability of experimental validations at the level of protein expression [4]. A sizeable fraction of various sequenced bacterial genomes was included under the respective category. The unavailability of their functions poses a challenge to the comprehension of current genomic information [5, 6]. The function annotations of these HPs are necessary for the detection of new metabolic pathways, structure-function relationships and other undiscovered facts [6-9]. The process of drug design and discovery will be improved significantly by utilizing these HPs as markers and potential targets [10-17].

The function predictions of HPs in the genome of *R. rickettsii* were performed by integrating varieties of sequence-based function prediction platforms. For this purpose, various heuristic algorithms such as BLAST [18] were used for the identification of functional related homologous proteins in different available biological databases [4]. The predictions performed by these tools form the basis for the identification of their characteristics roles in the biochemical processes [19]. Among the set of 680 uncharacterized proteins, we have successfully predicted functions of 214 HPs. Furthermore, virulence factors were classified in the annotated group of 214 HPs using a variety of virulence prediction servers. 15 HPs were categorized as putative virulence factors. Moreover, two HPs with the highest virulence scores were selected and subjected to Molecular Dynamics (MD) simulations in order to better understand the conformational behaviors of these proteins under explicit solvent conditions.

MATERIALS AND METHODS

The previously formulated protocols were used for the characterization of the HPs in the genome of *R. rickettsii* [4,

*Address correspondence to this author at the Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi – 110025, India; Tel: +91-9990-323217; E-mail: mihassan@jmi.ac.in

1570-1646/15 $58.00+.00

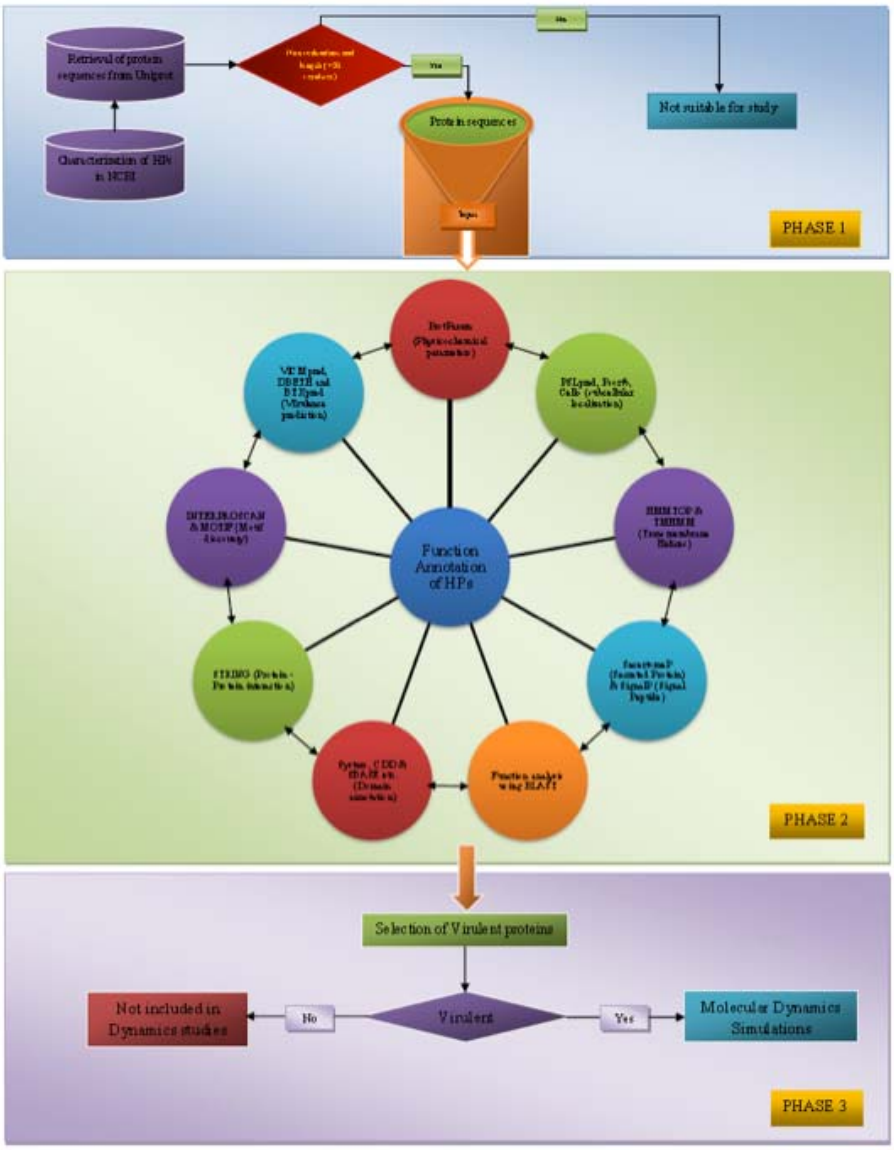

Fig. (1). The computational framework adopted for functional characterization of 680 HPs from *R. rickettsii.* First phase comprises annotation of HPs and sequence retrieval from various protein databases. Second phase incorporates calculations of various sequence based parameters by utilizing a variety of Bioinformatics tools. Third phase integrates the function predictions as well as selection of the virulent HPs. The latter were subjected to MD simulations in order to understand the conformational preferences.

20-22] (illustrated in Fig. **1**). Primarily, the HPs were characterized in the NCBI database (http://www.ncbi.nlm.nih.gov/genome/) and their sequences were retrieved from the Uniprot [23]. Then predictions based on physicochemical properties, subcellular localization, function and the characteristic domain annotation as well as interaction network analyses were performed on HPs. Finally, the virulence factors were classified from the set of 214 HPs and two HPs with highest virulence scores were subjected to MD simulations. The methodology used in each stage followed by the molecular dynamics simulations (MD) is explained below in detail.

HPs Characterization

The genome of *R. rickettsii* submitted in the NCBI database contains a list of 698 HPs (http://www.ncbi.nlm.nih.gov/genome/). The "Protein product ID" of each HPs present in NCBI database was used as keyword to search the Uniprot [23], in order to retrieve the Uniprot IDs and the protein sequences. Two or more proteins showing similar results for the Uniprot ID were considered as redundant. Similarly, HPs (<50 residues) were assumed to be insignificant as prior studies showed that the functions of these proteins were not predicted reliably [4, 24]. The HPs which were satisfying the above criterion were considered for further analyses.

Initial Sequence Analyses

ProtParam server (http://web.expasy.org/protparam/), one of the Expasy's utilities was used to calculate the molecular weight (M_w), isoelectric point (pI), instability index and extinction coefficient. The calculated physicochemical parameters were listed in Table **S1**. Furthermore, in order to classify the HPs as drug or vaccine targets, information about their

sub cellular localization is necessary [25]. The membrane proteins are generally utilized in the production of vaccine [26]. The integral membrane and cytoplasmic proteins were considered as the drug targets [25, 27]. A variety of servers were used for predicting subcellular localization due to unavailability of experimental data. The support vector machine (SVM) based methods such as PSLpred [28] and CELLO [29, 30] were used for predicting the localization of 680 HPs in the cellular framework of *R. rickettsii*. These annotations were further validated by using the outcomes of PSORTb [31]. The predictions performed by the CELLO server were found to be highly accurate as it is based on two-level SVM based model, which were produced using the dataset of about 1444 bacterial as well as 7589 eukaryotic protein sequences [29, 30]. Furthermore, the signal peptides in the HPs were identified using SignalP 4.1 [32] while their involvement in the secretory mechanisms was predicted by SecretomeP [33]. Similarly, membrane bound HPs that are involved in transport mechanisms were characterized by the identification of trans-membrane helices using HMMTOP [34] and TMHMM [35]. (Table **S2**)

Function Annotations

We assumed that the primary step in the function prediction of the proteins was identification of functional homologues in the variety of databases using a variety of heuristic methods like BLAST [18]. The proteins were considered as a functionally related homologue if they showed high sequence identities (>40%) with e-value (>0.005). Similarly, those which do not satisfy this criterion and showed the sequence identity (≤ 20%) and query coverage (<50%) were annotated as the distant homologues. These proteins were not considered as the reliable hit (Table **S3**). Similarly, varieties of bioinformatics tools and databases like Pfam [36], SYSTERS [37], SMART [38], CDD [39] and SBASE [40] were used to perform the domain identification and motif discovery (Tables **S3** and **S4**). The open source database present in the NCBI such Conserved Domain Database (CDD) utilized the information about annotated conserved domain footprints to perform the functional characterization of unknown protein sequences [39]. Another attribute of this database contains the hierarchical arrangement of the diverse manually curated models of domains [39]. Since CDD imports the information regarding domain family models from diverse external resources, it showed moderate redundancy in the observed outcomes of function predictions [39]. Hence, in order to reduce the complexity arising due to the process of function annotation of proteins, these respective models were arranged into numerous protein superfamilies [39]. Furthermore, SMART (Simple modular architecture research tool) was used that searches similar domain architecture in the databases such as SP-TrEMBL [41], Swiss-Prot [42] and stable Ensembl [43]. Moreover, InterProScan [44] and MOTIF (http://www.genome.jp/tools/motif/) were used to identify the conserved motifs in the sequences of HPs.

Since the extent of acuteness of particular infection depends upon the intensity of virulence factors (VFs) [45], they are considered as the putative targets in drug design and discovery [45]. The classification of these undiscovered VFs present in this Gram-negative bacterium was performed using VICMpred [46]. Furthermore, extent of the toxic nature of the predicted VFs was estimated using BTXpred [47] and DBETH [48] servers.

The function of a protein is often controlled by interaction pathways in a biological network. Thus, understanding the interaction of proteins is necessary in order to predict their functions precisely. The STRING database [49] was used to predict the interacting partners of HPs present in *R. rickettsii*. The prediction of this database is given by quantitative integration of the data regarding physical, experimental and functional associations for a sizeable scale of organisms, and the information connecting these respective organisms is rearranged wherever needed to predict the interacting proteins.

MD Simulations

HP A8GTM0 and HP A8GR55 showed highest virulence scores and were selected for further studies. The three dimensional (3-D) structures of these proteins were predicted using the *ab initio* based protocols present in the I-TASSER [50]. After the assessment of the predicted models on the basis of TM score [51] and Root Mean Square Deviation (RMSD) values, the best models were selected. The accuracy of the predicted models was assessed using the structure alignment protocols of DALI [52] and the PPM server [53] was used for the prediction of membrane bound residues. MD studies of the respective HPs were performed in explicit water conditions using GROMACS [54] (version – 4.6.5). The initial topological parameters were generated using available Optimized Potentials for Liquid Simulations - All Atoms (OPLS-AA) force field in GROMACS. Then, proteins were solvated using Single Point Charge (SPC) water model in a cubic system. The solvated structures of these HPs were energetically minimized by means of a convergence standard of 0.005 kcal mol^{-1}using 2000 steps of steepest descent algorithm. The minimized structures were equilibrated for 1 ns in the solvent environment using "isothermal-isochoric" (NVT) and "isothermal-isobaric" (NPT) ensemble conditions. The LINCS algorithm present in the GROMACS was used to perform the final MD simulations for 100 ns with a time step of 2 fs. 10000 conformations were generated after each MD simulations. These conformations were clustered using Javis Patrick algorithm present in g_cluster module of GROMACS with cut-off values was set at 0.1 nm so as to organize the nearest structures. The central conformation of most densely populated clustered group was selected as the reference structure for the calculation of Root Mean Square Fluctuation (RMSF) and RMSD. The supplementary analyses were performed using the additional GROMACS utilities.

RESULTS AND DISCUSSION

The consensus amongst the resulted predictions was formed because of the similarities in the results produced by variety of tools in the adopted pipeline (Fig. **1**). If three or more prediction tools showed similar functions for a given HP, then it was assumed as possible functions for the respective HPs. Therefore, functions of 214 HPs present in the genome of *R. rickettsii* were successfully predicted (Tables **1**-**3**). These proteins were further classified into 88 enzymes,

59 transporter and membrane proteins, 14 DNA binding proteins, 12 structural motifs, 6 tetratricopeptide repeats, 3 virus related proteins, *etc*. (Fig. **2**). Due to similarities with proteins of known functions, these HPs were considered to be essential for the survival of *R. rickettsii*. The major identified functional classes are described here in details.

HPs as Enzymes

88 HPs were grouped in different classes of enzymes (Table **1**). There were 9 HPs annotated as methyltransferase in the category of 19 predicted transferases. The methytransferases have been identified to be significant for the virulence of *Rickettsia prowazekii* [55]. This class of Rickettsiaceae is the major cause of widespread of typhus, a possible agent of bioterrorism [55]. The OmpB (Outer membrane protein B) is an adhesin which mainly acts as the defensive envelop and as the primary cause of immuno-dominance [55]. There is an observed association among the methylation of lysine residues present in the OmpB protein and virulence of bacteria [55]. This suggested the significance of the methyltransferase enzymes in understanding the virulence characteristic of *R. rickettsii* [55]. Similarly, HP A8GQG3, an oxidoreductase, is found to be a homolog of heme oxygenase in the set of eight proteins. Heme is the primary source of iron for the Gram-negative bacteria, and heme oxygenases catalyze the degradation of tetrapyrrole ring present in the heme. This process leads to the release of free iron for this pathogen [56]. Furthermore, HP A8GRT6 was found to be a disulfide bond formation protein DsbB, while HP A8GS27 showed the presence of DsbA-like thioredoxin domain. The DsbB deoxidizes DsbA which catalyzes the formation of disulfide bond [57]. The bacterial virulence factors are generally produced through secretion, which are mainly transported from the cytoplasmic and periplasmic regions. Majority of these proteins undergo post-translation modification by the formation of the disulfide bond [57]. Therefore, these predicted Dsb proteins may be involved in the pathogenesis.

The cell wall biosynthesis of bacteria is usually targeted by the β-lactam antibiotics, which were considered as primary sources in the treatment against the bacterial pathogenesis [58]. The pathogenic Gram-negative bacteria have developed the resistance against these beta-lactam antibiotics [58]. The development of these multidrug resistance strains poses a challenge to the health care facilities worldwide [58]. The enzymes are responsible for the development of resistances and belong to the class of β-lactamases through their action on the β-lactam ring present in these antibiotics [58]. HP A8GTL1 showed the presence of β-lactamase superfamily domain indicating that it may be involved in the development of the multidrug resistance in this pathogen. Therefore, it can be used as a drug target for the treatment against *R. rickettsii*, which is naturally resistant to β-lactam antibiotics [59, 60]. Moreover, HP A8GRS7 and HP A8GU55 showed the presence of N-acetylmuramoyl-L-alanine amidase-like activities. Hence these proteins may be involved in the process of cell separation and is important for the virulence [61]. Among six characterized lyase enzymes, HP A8GSK4 showed high similarity to lytic transglycosylase. The Gram-negative bacteria release soluble peptidoglycan molecules during their growth phase, which are responsible for its pathogenesis [62]. The lytic transglycosylase is an enzymes involved in the formation of such peptidoglycan-derived molecules that causes the cytotoxicity in the mammalian host [62].

In the metabolism of nitrogen, glutamine derivatives play a central role [63]. The latter molecules are synthesized from ammonia, L-glutamate, and ATP [63]. It also acts as nitrogen contributor for many bio-molecules [63]. The glutamate synthetase pathway is the key producer of the glutamine, and under growth phase with limited nitrogen this pathway results in the assimilation of the ammonia in a variety of pathogenic bacteria [63]. Since, glutamate synthetase plays a crucial role in the metabolism of nitrogen, therefore it undergoes a series of feedback inhibitions and complex regulatory control [63]. This enzyme is essential for the growth of bacteria like *Mycobacterium tuberculosis* and is necessary for its virulence [63]. For this reason, HP A8GSA3 was considered significant for the nitrogen assimilation, and it may act as a target protein for drug design. Likewise, HP A8GT86 was predicted to be equally important as it showed the presence of SurA N-terminal domain. The SurA is a protein involved in the process of protein folding as well as in the formation of adhesins [64].

HPs as Transporter and Membrane Proteins

Since metabolic processes such as absorption and excretion of essential nutrients are regulated by the transportation mechanisms, these HPs are necessary for the survival and pathogenesis of such Gram-negative bacteria [4]. Consequently, 59 HPs were clustered in this category (Table **2**), of which two proteins (HP A8GRS8 and HP A8GR55) were annotated as ATP-binding cassette (ABC) transporter. These proteins may play an important role in the virulence as this category of transporters is involved in the absorption of metal ions such as zinc, iron and manganese [65]. Additionally, these proteins are also involved in the attachment of the bacteria to the mucosal cell's surfaces, so they are considered as the putative drug targets for the development of better therapeutic agents [65]. The development of multidrug resistance (MDR) against available pharmaceutical agents acts as a barrier in the treatment of a variety of bacterial disease [66]. The action of protein families such as ABC transporters and the major facilitator superfamily (MFS) is one of the several causes of MDR in the pathogenic bacteria [66]. HP A8GRC8 was found to be an MFS superfamily transporter and may be responsible for the development of MDR in *R. rickettsii*.

Similarly, RND (Resistance Nodulation Division) is another category of transporters found in the Gram-negative bacteria responsible for the development of resistance against antimicrobial agents [67]. Furthermore, these proteins are also involved in the manufacturing of optimal quantity of virulence factors and colonization process [67]. HP A8GRP4 is predicted to be an RND efflux pump membrane fusion protein. Additionally, HP A8GS40 was annotated as N-ethylmaleimide-sensitive factor attachment protein receptors (SNARE) associated Golgi protein which may be crucial for fusion of sections in the cells [68]. On the basis of the presence of centrally conserved residues in the SNARE motif region, this group of receptors was classified into glu-

Table 1. List of predicted enzymes among the group of HPs from the genome of *R. rickettsii* (Functional classes are shown with Tan background).

S. No	Uniprot ID	Predicted Function
Transferase		
1.	A8GQG0	Methyltransferase domain
2.	A8GQK4	S -adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), classI
3.	A8GRP1	Glycosyltransferase family 28 C-terminal domain
4.	A8GRP2	Methyltransferase domain
5.	A8GS11	Methylthiotransferase radical SAM domain signature.
6.	A8GSB1	glutathione S-transferase N-terminal domain/ glutathione S-transferase C-terminal domain
7.	A8GSC9	Methyltransferase domain
8.	A8GSG5	Aminotransferase class-V/cystein desulfurase
9.	A8GSH3	Acetyltransferase (GNAT) family
10.	A8GSJ2	Glycine cleavage T-protein C-terminal barrel
11.	A8GSL5	Acetyltransferase (GNAT) family
12.	A8GSN7	DNA methylase, N-6 Adenine-specific,conserved site
13.	A8GTA1	MobA-like NTP transferase domain
14.	A8GTF1	Ribosomal RNA small subunit methyltransferase D, RsmJ
15.	A8GTK2	RsmI AdoMet-dependent methyltransferase protein family signature.
16.	A8GTR2	Methyltransferase domain
17.	A8GTW9	Acyltransferase 3
18.	A8GU20	Phosphatidate cytidylyltransferase
19.	A8GRB8	Diguanylate phosphodiesterase, EAL domain
Oxidoreductase		
20.	A8GQF8	Thioredoxin like fold
21.	A8GQG3	Heme oxygenase- like, multihelical
22.	A8GQX4	Flavin reductase like domain
23.	A8GRT6	Disulfide bond formation protein DsbB
24.	A8GRV9	Cytochrome C oxidase ,subunit -1 domain
25.	A8GS27	DSBA-like thioredoxin domain
26.	A8GSA4	Multi-copper polyphenol oxidoreductase, laccase
27.	A8GRX8	Intradiol ring- cleavage dioxygenase C- terminal
Hydrolase		
28.	A8GRP6	Alpha/beta hydrolase family-1
29.	A8GS52	Alpha/beta hydrolase family
30.	A8GSE7	Alpha/beta hydrolase family
31.	A8GTZ1	Alpha/beta hydrolase family
32.	A8GSP2	Glycoprotease family
33.	A8GSV2	HAD-superfamily hydrolase, subfamily IIA, hypothetical3

Table 1. Contd……..

S. No	Uniprot ID	Predicted Function
34.	A8GSV5	guanosine polyphosphate pyrophosphohydrolase/synthetase-like protein
35.	A8GTD4	Ubiquitin-like protease family profile
36.	A8GTE3	Gylocoside hydrolase superfamily
37.	A8GTF5	guanosine polyphosphate pyrophosphohydrolase/synthetase-like protein
38.	A8GTJ0	Phospolipase/carboxylesterase/thioesterase
39.	A8GTL1	Beta-lactamase superfamily domain
40.	A8GSL4	NTP pyrophosphohydrolase MazG ,putative catalytic core
41.	A8GQD8	Polysaccharide deacetylase
42.	A8GR80	Peptidase M16 inactive domain
43.	A8GTW0	Phospholipase D phosphodiesterase active site profile.
44.	A8GRZ1	Peptidase family M23
45.	A8GSI7	LD-carboxypeptidase
46.	A8GTF0	Peptidase M23
47.	A8GU30	Aspartyl protease
48.	A8GTA8	TatD deoxyribonuclease family signature 1
49.	A8GRG6	Pin domain
50.	A8GRU2	PIN domain
51.	A8GS31	PIN domain
52.	A8GSA6	PIN domain
53.	A8GU10	PIN domain
54.	A8GU38	Phosphoesterase, PA-phosphatase
55.	A8GRS7	N-acetylmuramoyl-L-alanine amidase
56.	A8GU55	N-acetylmuramoyl-L-alanine amidase
57.	A8GU02	Uracyl DNA glycosilase
Endonuclease		
58.	A8GR78	Plasmid encoded toxin Txe
59.	A8GRM1	PD-(D/E)XK nuclease family transposase
60.	A8GRY5	Restriction endonuclease type II-like / PD-(D/E)XK nuclease
61.	A8GTJ9	PD-(D/E)XK nuclease family transposase
62.	A8GS09	PD-(D/E)XK nuclease family transposase
63.	A8GTU2	PD-(D/E)XK nuclease family transposase
Dehydrogenase		
64.	A8GSX9	Flavinator of succinate dehydrogenase
65.	A8GTG1	NADH ubiquinone oxidoreductase subunit NDUFA12
66.	A8GTM0	Bacterial NAD-glutamate dehydrogenase
67.	A8GTS3	NADH dehydrogenase subunit G

Table 1. Contd……..

S. No	Uniprot ID	Predicted Function
Lyase		
68.	A8GSZ6	Radical SAM
69.	A8GTW2	3-octaprenyl-4-hydroxybenzoate carboxy-lyase
70.	A8GR12	6-pyruvoyl tetrahydropterin synthase/QueD family
71.	A8GSK4	Lytic transglycosylase like- SLT domain
72.	A8GSN9	Fructose/tagarose-bisphosphate aldolase class II.
73.	A8GSF0	Tetrapyrrole biosynthesis Uroporphyrinogen-III synthase
Ligase		
74.	A8GSA3	Glutamine synthetase, catalytic domain
75.	A8GSB2	Biotin—acetyl-CoA-carboxylase ligase
76.	A8GSJ6	5-formyltetrahydrofolate cyclo-ligase
ATPase		
77.	A8GQE0	AAA ATPase domain
78.	A8GQF7	Putative F0F1-ATPase subunit (ATPase_gene1)
79.	A8GQW8	ATPase MipZ/Nubp2/Cfd1
80.	A8GSL7	AAA ATPase
81.	A8GT94	AAA ATPase
Kinase		
82.	A8GRG0	(p)ppGpp synthetase
83.	A8GRL0	(p)ppGpp synthetase
84.	A8GQC2	Bifunctional kinase pyrophosphorylase
85.	A8GTW3	SurA N-terminal domain
Isomerase		
86.	A8GU03	Rlu family of pseudouridine synthase signature
87.	A8GRX7	peptidyl-prolyl cis-trans isomerase ,FKBP-type ,domain
88.	A8GT86	SurA N-terminal domain

Table 2. List of predicted transporters and membrane proteins among the group of HPs from the genome of *R. rickettsii*

S. No	Uniprot ID	Predicted Function
Transporters and membrane Proteins		
1.	A8GQE4	Competence protein
2.	A8GQN2	Porin domain
3.	A8GQN9	Outer membrane protein OmpW
4.	A8GQP0	EamA-like transporter family
5.	A8GQP5	Autotransporter beta-domain profile
6.	A8GQQ7	Toluene tolerance, Ttg2/ phospholipid binding protein mlac

Table 2. contd….

S. No	Uniprot ID	Predicted Function
7.	A8GQS5	TrbL/VirB6 plasmid conjugal transfer protein
8.	A8GQS6	TrbL/VirB6 plasmid conjugal transfer protein
9.	A8GQS7	TrbL/VirB6 plasmid conjugal transfer protein
10.	A8GQS8	TrbL/VirB6 plasmid conjugal transfer protein
11.	A8GQS9	Type IV secretion system protein VirB6
12.	A8GR26	Lipopolysaccharide-assembly LptE
13.	A8GR31	Conjugal transfer TrbC/type 4 secretion VIRB2
14.	A8GRB7	BolA-protein
15.	A8GRC8	Major Facilitator Superfamily
16.	A8GRD9	Outer membrane protein Omp1
17.	A8GRH4	Bacterial virulence protein VirB8
18.	A8GRM3	Delta endotoxin central region - like domain
19.	A8GRN2	Band 7 protein
20.	A8GRP4	RND efflux pump,membrane fusion protein
21.	A8GRQ7	AsmA-like C-terminal region
22.	A8GRR3	Haemolytic domain
23.	A8GRR9	O-antigen ligase like membrane protein
24.	A8GRS8	ABC transporter substrate binding protein
25.	A8GRT8	Membrane transport protein
26.	A8GRU0	Mgr1-like, i-AAA protease complex subunit
27.	A8GRU3	AbrB-like transcriptional regulator
28.	A8GRX2	BON domain
29.	A8GRX6	Cupredoxin-like domain
30.	A8GS04	AcrB/AcrD/AcrF family
31.	A8GS06	AcrB/AcrD/AcrF family
32.	A8GS19	Septum formation initiator
33.	A8GS34	Prokaryotic cytochrome b561
34.	A8GS36	Tim44-like domain
35.	A8GS40	SNARE associated Golgi protein
36.	A8GS41	RDD Family
37.	A8GS76	Outer membrane usher protein
38.	A8GS77	Spore Coat Protein U
39.	A8GSA2	ribonuclease BN
40.	A8GSB4	BioY family
41.	A8GSE8	HemY N-terminal
42.	A8GSF1	Lysine type exporter protein (LYSE/YGGA)

Table 2. contd….

S. No	Uniprot ID	Predicted Function
43.	A8GSG6	Transcriptional regulator Rrf2-like
44.	A8GSH9	Transcriptional regulatory protein, C-terminal - like domain
45.	A8GSK3	Na+ dependent nucleoside transporter, C-terminal - like domain
46.	A8GSP9	Transcriptional regulator Rrf2-like
47.	A8GT91	Sugar efflux transporter for intercellular exchange/ Mitochondrial 18 KDa protein (MTP18)
48.	A8GTD2	FtsX-like permease family
49.	A8GTN9	permease YjgP/YjgQ family
50.	A8GTP4	Lipoprotein SmpA / OmlA
51.	A8GTR1	Outer membrane protein / Outer membrane enzyme, beta-barrel
52.	A8GTU4	SH3- like domain, bacterial-type
53.	A8GTW4	Autotransporter beta-domain profile
54.	A8GTX5	Outer membrane protein/Outer membrane enzyme pagP beta-barrel
55.	A8GTX6	Outer membrane protein/Outer membrane enzyme pagP beta-barrel
56.	A8GU24	Outer membrane lipoprotein carrier protein LolA-like
57.	A8GU41	Predicted permease YjgP/YjgQ family
58.	A8GR55	ABC transporter
59.	A8GRX5	Iron permease FTR 1

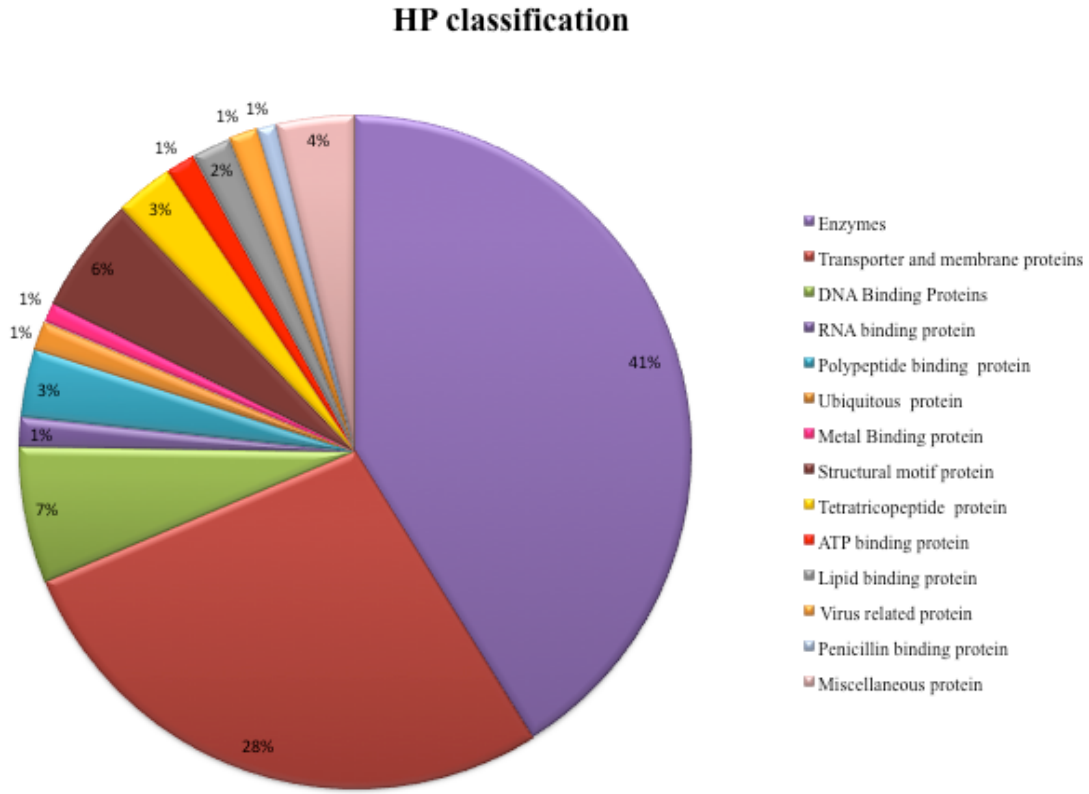

Fig. (2). The graphical representation of the observed functional categories in the set of 214 HPs from the genome of *R. rickettsii*. 41% of HPs are enzymes, 28% of the HPs are characterized as transporters and membrane associated proteins, and remaining HPs may involve in a variety of cellular processes like DNA binding, Transcription regulation and so on (the points represent different protein functional classes).

Table 3. List of predicted HPs involved in various cellular processes from the genome of *R. rickettsii*

S. No	Uniprot ID	Predicted Function
DNA binding		
1.	A8GQF9	CarD-like/TRCF domain
2.	A8GR77	type II toxin-antitoxin system, antitoxin Phd/YefM
3.	A8GRG7	Antitoxin Phd_YefM, type II toxin-antitoxin system
4.	A8GRM4	YceG-like family
5.	A8GSA1	Cro/C1-type Helix-Turn- Helix domain
6.	A8GSH7	Cro/C1-type Helix Turn Helix domain
7.	A8GST7	Metal binding domain of Ada
8.	A8GT76	type II toxin-antitoxin system, Antitoxin Phd_YefM
9.	A8GTD5	CBS domain profile
10.	A8GU11	Type II toxin-antitoxin system, antitoxin Phd/YefM
11.	A8GU26	Nucleoid-associated protein YbaB
12.	A8GU45	Type I restriction modification DNA specificity domain
13.	A8GU47	DNA methylase
14.	A8GTC7	HicB family
RNA binding		
15.	A8GS49	RNA-metabolising metallo-beta-lactamase
16.	A8GU07	tRNA threonylcarbamoyladenosine biosynthesis protein SUA5
17.	A8GSR2	Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily
Polypeptide binding		
18.	A8GR70	Ankyrin repeats
19.	A8GR93	Ankyrin repeat region circular profile
20.	A8GRT2	Ankyrin repeat-containing protein
21.	A8GSJ0	Ankyrin repeat
22.	A8GSW6	TraX protein
23.	A8GT01	Ankyrin repeats
24.	A8GTK3	Periplasmic binding protein –like I
Metal binding		
25.	A8GT81	Scaffold protein Nfu/NifU N terminal
26.	A8GTV7	RNA polymerase- binding transcription factor DksA
ATP binding		
27.	A8GT36	Bacterial dnaA protein
28.	A8GTW8	FtsK/SpoIIIE family
29.	A8GR45	Universal stress protein family
Lipid binding		
30.	A8GTY2	Phosphatidylethanolamine-binding protein PEBP

Table 3. Contd……

S. No	Uniprot ID	Predicted Function
31.	A8GT87	Organic solvent tolerance protein.C- terminal
32.	A8GS92	Organic solvent tolerance –like,N -terminal
33.	A8GS93	Lipopolysaccharide-assembly, LptC-related
Penicillin binding		
34.	A8GTT5	Penicillin binding protein transpeptidase domain
35.	A8GTT6	Penicillin binding protein transpeptidase domain
Ubiquitous protein		
36.	A8GTG1	NADH ubiquinone oxidoreductase subunit NDUFA12
37.	A8GR99	Pyrrolo-quinoline quinine beta –propeller repeat
38.	A8GT13	Ubiquinone biosynthesis protein CQQ9
Structural motif		
39.	A8GQM9	Sporulation -related domain
40.	A8GR32	zinc finger/thioredoxin putative
41.	A8GSK5	Bacterial chemotaxis sensory transducer
42.	A8GSN6	Pentapeptide repeats
43.	A8GSR9	Pentapeptide repeats
44.	A8GSY0	Rhodanese domain profile
45.	A8GTB8	DNA recombination RmuC
46.	A8GTE4	MerR type HTH domain
47.	A8GTK9	Cro/C1-type HTH DNA-binding domain
48.	A8GTL9	Helix-turn-helix domain
49.	A8GTP4	Lipoprotein SmpA / OmlA
50.	A8GTR4	Cro/C1-type HTH domain profile
Tetratricopeptide repeats		
51.	A8GQG9	Tetratricopeptide- like helical
52.	A8GQQ5	Tetratricopeptide TPR-1
53.	A8GR18	Tetratricopeptide like helical
54.	A8GR97	Tetratricopeptide _ like helical
55.	A8GRQ4	Tetratricopeptide TPR-1
56.	A8GU61	Tetratricopeptide repeat-containing domain
Virus related		
57.	A8GTY4	Arenavirus nucleocapsid protein
58.	A8GT28	Invasion protein locus B
59.	A8GSD8	Head to tainl joining protein , podovirus type
Miscellaneous protein		
60.	A8GRU1	Adenylate cyclase associated (CAP) N terminal
61.	A8GSA7	Class II Aldolase/ Adducin N-terminal domain

Table 3. Contd……

S. No	Uniprot ID	Predicted Function
62.	A8GST3	DNA alkylation repair enzyme
63.	A8GTG3	Stringent starvation protein B
64.	A8GTX2	RadC-like JAB domain
65.	A8GR46	NMT1/THI5 like
66.	A8GRW2	Fic/DOC family
67.	A8GSE5	Host attachment protein

tamine enclosing SNAREs (Q-SNAREs) and arginine enclosing SNAREs (R-SNAREs) [69]. These proteins are the targets for the pathogens responsible for host intracellular traffic destabilization, for they are regulating the fusion of the membrane regions [69]. This HP can be involved in the regulation of intracellular transfer by the replication of the SNAREs.

Cellular Processes' HPs

67 HPs were grouped in this category which includes 35 different binding proteins, three ubiquitous protein, 12 structural motifs, six tertratricopeptide repeats, 3 virus related proteins and the rest involved in other cellular activities (Table **3**). Many DNA-binding proteins regulate gene expression, *i.e.*, some of them will be important for virulence. Four HPs were predicted to be the components of bacterial toxin-antitoxin system. These systems are involved in promoting the bacterial survival in the conditions of environmental stress and stabilization of the bacterial plasmids [70]. Due to their role in the bacteria survival in the stress conditions these proteins were considered as crucial for the bacterial pathogenesis [70]. Similarly, sarZ proteins present in the pathogenic strains of bacteria contain a structurally conserved winged-helix-turn-helix motif which plays a role in the activation of *cvf* gene through the process of binding. Such motifs are known to play an important role in virulence as the expression of latter gene resulted in the formation of hemolysins [71]. The transcription regulatory protein Srv present in disease causing bacterial strains, shows that the presence of characteristic helix-turn-helix (HTH) structural motif that is mainly responsible for regulation of processes such as DNA binding, protein-DNA interaction and virulence [72]. Furthermore, HPs contribute to the RNA binding process and are significant for the survival of the pathogen in the mammalian host cells through the regulation of the virulence factors [73]. There were three HPs that may be involved in the RNA binding mechanisms. Moreover, six proteins are comprised of a tetratricopeptide repeat (TPR) in their framework, so these proteins may take part in the assembly of diverse protein complexes. Proteins that showed the presence of these structural motifs are mainly involved in virulence-related cellular processes [74]. Consequently, these proteins involved in a variety of cellular processes may be significant for the survival and virulence of *R. rickettsii*.

MD Simulations

The VICMpred is based on the training set of protein sequences obtained from the Gram-negative bacteria [46]. It was therefore used as the basis for the classification of the proteins relevant for virulence from the group of 214 functionally annotated HPs (Table **4**). As a result, 15 HPs were predicted to be virulence factors with HP A8GTM0 and HP A8GR55 showing highest scores. The sequence of HP A8GTM0 contains 1584 residues. ITASSER permits the modeling of protein sequences $\leq$ 1500 residues. Therefore, its sequence was trimmed using the modules of Discovery Studio (DS). The generated model of HP A8GTM0 showed 95.5 % of the residues in the allowed region and 4.5% in disallowed region of the Ramachandran plot. The structural comparisons using DALI showed high similarities to glutamate dehydrogenases with RMSD ranges from 1.2-2.0 Å, indicated the reliability of the generated 3-D models. Furthermore, the predicted structure of HP A8GTM0 showed characteristic glutamate dehydrogenase-like domain extending from Gly1-Gly204 (Fig. **3A**) in which five stranded mixed β-sheet is surrounded by two α-helices on one side and six helices on the other side. The other type of domain was NAD(P)-binding Rossmann-fold domain which contains four stranded mixed β-sheet surrounded by α-helices on both sides. This structure was subjected to 100 ns MD simulations under explicit water conditions, which showed unstable nature (Figs. **3B-D**). This is indicated by the variations in the plots of RMSD and radius of gyration. These predictions were complemented by the RMSF plot in which majority of residues showed higher fluctuation (Figs. **3B-D**). The average RMSD and Rg values for HP A8GTM0 were calculated to be 0.56 nm and 2.31 nm, respectively.

Similarly, HP A8GR55 showed 94.8% in the allowed region of Ramachandran plot, and structural alignments performed by DALI indicated high similarities to ABC transporters. The structure of HP A8GR55 illustrated characteristic trans-membrane helix α2 spanning form Thr23-Leu56 (Fig. **4A**). HP A8GR55 also showed an unstable behaviour when subjected to MD simulations in water solvent as projected by the GROMACS utilities (Figs. **4B-D**). The RMSD curve showed sharp fluctuations up to 30 ns while there was a decrease in the amplitude after this time interval with an average value of 1.97 nm with an average R*g* value of 2.03 nm.

Table 4. The characterized virulence factors among the group of 214 confidently predicted protein (most virulent are shown in green).

S.NO	UNIPROT ID	VICMpred	Toxin Protein (DBETH server)	BTXpred Server: Prediction of Bacterial Toxins
1.	A8GSC9	(+) 0.64	(+)	Exotoxin
2.	A8GRB8	(+) 1.06	(+)	Exotoxin
3.	A8GSV2	(+) 0.83	(+)	Exotoxin
4.	A8GTD4	(+) 1.26	(+)	Exotoxin
5.	A8GTA8	(+) 0.42	(-)	Exotoxin
6.	A8GTM0	(+) 13.48	(+)	Exotoxin
7.	A8GQE0	(+) 0.95	(+)	(-)
8.	A8GTW3	(+) 0.61	(-)	Exotoxin
9.	A8GQN9	(+) 0.74	(+)	Exotoxin
10.	A8GRP4	(+) 0.77	(-)	Exotoxin
11.	A8GRQ7	(+) 1.33	(+)	Exotoxin
12.	A8GRX2	(+) 1.001	(+)	Exotoxin
13.	A8GR55	(+) 3.50	(+)	Exotoxin
14.	A8GRM4	(+) 0.79	(+)	(-)
15.	A8GSR9	(+) 2.66	(+)	Exotoxin

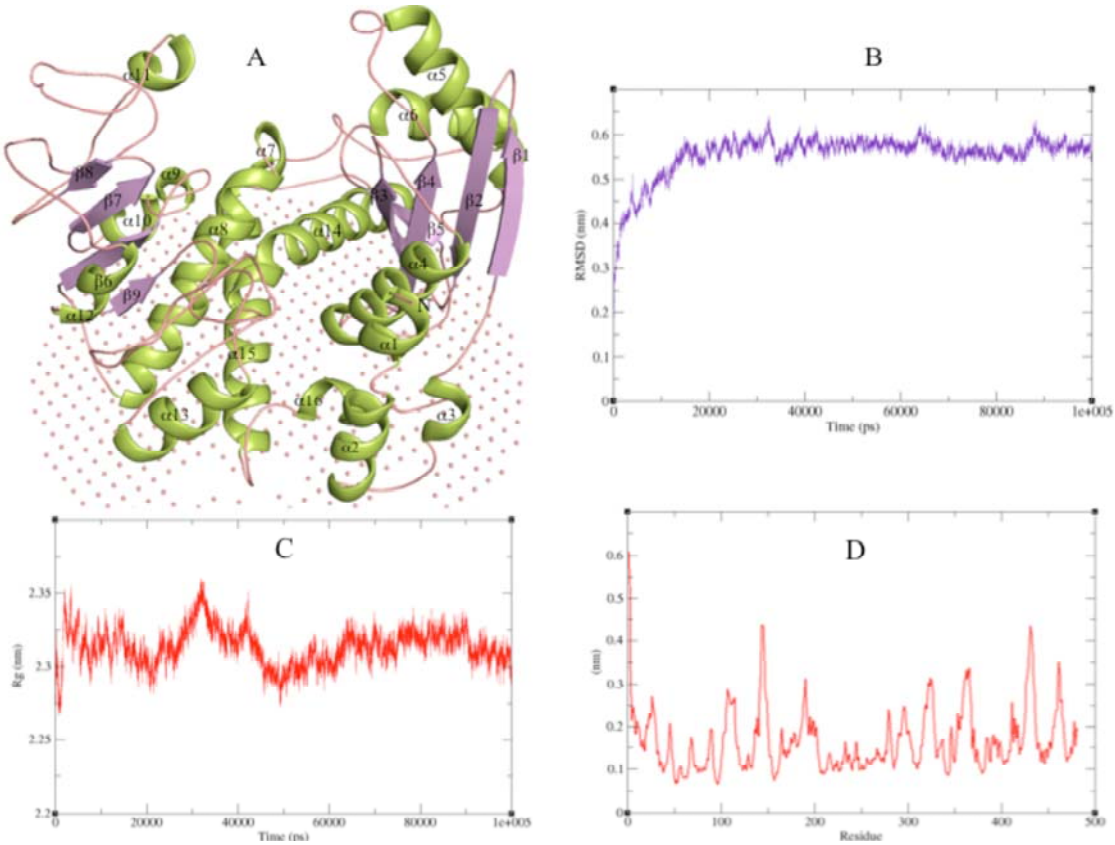

Fig. (3). (A) The description of glutamate dehydrogenase domain of HP A8GTM0. **(B)** RMSD curve illustrating variation during 100 ns MD simulation. **(C)** R_g curve plotted against the time intervals demonstrating the unstable nature of this protein. **(D)** RMSF plot of Cα atoms screening the high fluctuations of the constituent residues.

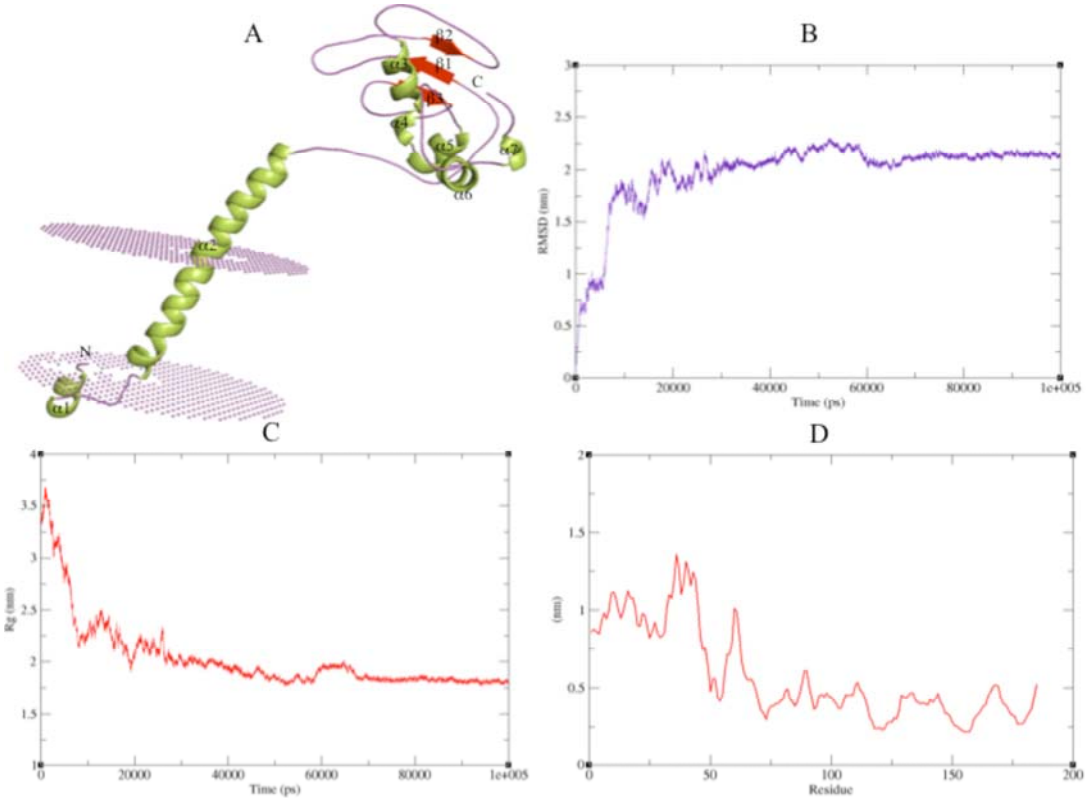

Fig. (4). (A) Predicted 3-D structure of HP A8GR55 showing a trans-membrane helix region. **(B)** The curve of RMSD scores with well-defined variations during 100 ns MD simulation in explicit water solvent. **(C)** Radius of gyration curve specifying HP A8GR55 is highly unstable. **(D)** RMSF plot of Cα atoms displays substantially higher fluctuations in the fundamental residues.

CONCLUSION

Functions of 214 gene of *R. rickettsii* genome were successfully predicted. The undiscovered functions and virulence nature of the HPs have been revealed in this study. These annotations were complimented by predicting their involvement in the non-classical secretory pathways. These characterizations can be utilized as the basis for further experimental validation, as the exact molecular functions cannot be identified using the *in silico* approaches. However, the conformational behaviors of classified virulence factors were further analyzed using the MD simulations which revealed that HP A8GTM0 and HP A8GR55 were unstable due to significant variations observed during the 100 ns of MD simulations.

CONFLICT OF INTEREST

Authors declare no conflict of interest regarding any financial and personal relationships with other people or organizations that could inappropriately influence (bias) this work.

ACKNOWLEDGEMENTS

Authors sincerely thank Indian Council of Medical Research for financial assistance (Project No. BIC/12(04)/ 2012). We express our gratitude towards Centre for high performance computing, South Africa for providing the computational infrastructure. We also thank high-speed server facility provided by Jamia Millia Islamia in the Central Instrumentation Facility.

REFERENCES

[1] Dantas-Torres, F.**;** Rocky mountain spotted fever; *Lancet Infect Dis.*, **2007**, *7*, 724-732.

[2] Ogata, H.; La Scola, B.; Audic, S.; Renesto, P.; Blanc, G.; Robert, C.; Fournier, P.E.; Claverie, J.M. and Raoult, D. Genome sequence of Rickettsia bellii illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genet.*, **2006**, *2*, e76.

[3] Ellison, D.W.; Clark, T.R.; Sturdevant, D.E.; Virtaneva, K.; Porcella, S.F. and Hackstadt, T.; Genomic comparison of virulent Rickettsia rickettsii Sheila Smith and avirulent Rickettsia rickettsii Iowa. *Infect. Immun.*, **2008**, *76*, 542-550.

[4] Shahbaaz, M.; Hassan, M.I.; Ahmad, F.; Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. *PLoS One*, **2013**, *8*, e84263.

[5] Loewenstein, Y.; Raimondo, D.; Redfern, O.C.; Watson, J.; Frishman, D.; Linial, M.; Orengo, C.; Thornton, J. and Tramontano, A. Protein function annotation by homology-based inference. *Genome Biol.*, **2009**, *10*, 207.

[6] Nimrod, G.; Schushan, M.; Steinberg, D.M. and Ben-Tal, N. Detection of functionally important regions in "hypothetical proteins" of known structure. *Structure*, **2008**, *16*, 1755-1763.

[7] Sinha, A.; Ahmad, F.; Hassan, I.; Structure based functional annotation of putative conserved proteins from treponema pallidum: search for a potential drug target. *Lett. Drug Design Discov.*, **2015**, *12*, 46-59.

[8] Kumar, K.; Prakash, A.; Anjum, F.; Islam, A.; Ahmad, F. and Hassan, M.I. Structure-based functional annotation of hypothetical proteins from Candida dubliniensis: a quest for potential drug targets. *3 Biotech.*, **2014**, 1-16.

[9] Naqvi, A.A.T.; Ahmad, F. and Hassan, M.I. Identification of functional candidates amongst hypothetical proteins of

Mycobacterium leprae BR4923, a Causative Agent of Leprosy. *Genome*, **2015**, *58*, 25-42.

[10] Minion F.C.; Lefkowitz E.J.; Madsen M.L.; Cleary B.J.; Swartzell S.M. and Mahairas G.G. The genome sequence of mycoplasma hyopneumoniae strain 232, the agent of swine mycoplasmosis. *J. Bacteriol.*, **2004**, *186*, 7123-7133.

[11] Lubec, G.; Afjehi-Sadat, L.; Yang, J.W. and John, J.P. Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog. Neurobiol.*, **2005**, *77*, 90-127.

[12] da Fonseca, M.M.; Zaha, A.; Caffarena, E.R. and Vasconcelos, A.T. Structure-based functional inference of hypothetical proteins from Mycoplasma hyopneumoniae. *J. Mol. Model*, **2012**, *18*, 1917-1925.

[13] Shahbaaz, M.; K. B.; Ahmad F.; Hassan M.I.**;** Towards New Drug Targets? Function prediction of putative proteins of Neisseria meningitidis MC58 and their virulence characterization; *OMICS: A Journal of Integrative Biology*, **2015**; PMID: 26076386.

[14] Amresh, P.; Kumar, K.; Islam, A.; Hassan, I. and Ahmad, F.; Receptor chemoprint derived pharmacophore model for development of CAIX inhibitors. *J. Carcinog. Mutagen S.*, **2013**, *8*, 17-20.

[15] Hassan, M.; Kumar, V.; Somvanshi, R.K.; Dey, S.; Singh, T.P. and Yadav, S. Structure-guided design of peptidic ligand for human prostate specific antigen. *J. Pept. Sci.*, **2007**, *13*, 849-855.

[16] Hassan, M.I.; Kumar, V.; Singh, T.P. and Yadav, S. Structural model of human PSA: a target for prostate cancer therapy. *Chem. Biol. Drug Design*, **2007**, *70*, 261-267.

[17] Thakur, P.K. and Hassan, M.I. Discovering a potent small molecule inhibitor for gankyrin using de novo drug design approach. *Int. J. Computat. Biol. Drug Design*, **2011**, *4*, 373-386.

[18] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W. and Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.*, **1990**, *215*, 403-410.

[19] Rost, B. and Valencia, A. Pitfalls of protein sequence analysis. *Curr. Opin. Biotechnol.*, **1996**, *7*, 457-461.

[20] Shahbaaz, M.; Ahmad, F. and Hassan M.I. Structure-based functional annotation of putative conserved proteins having lyase activity from Haemophilus influenzae. *3 Biotech.*, **2015**, *5*, 317-336.

[21] Shahbaaz, M.; Ahmad, F.; Hassan, M. Structure-based function analysis of putative conserved proteins with isomerase activity from Haemophilus influenzae. *3 Biotech.*, **2014**.

[22] Naqvi, A.A.; Shahbaaz, M.; Ahmad, F. and Hassan, M.I. Identification of functional candidates amongst hypothetical proteins of treponema pallidum ssp. pallidum. *PLoS One*, **2015**, *10*, e0124177.

[23] Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M.J.; Natale, D.A.; O'Donovan, C.; Redaschi, N. and Yeh, L.S. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.*, **2004**, *32*, D115-D119.

[24] Kumar, K.; Prakash, A.; Tasleem, M.; Islam, A.; Ahmad, F.; Hassan, M.I. Functional annotation of putative hypothetical proteins from Candida dubliniensis. *Gene*, **2014**, *543*, 93-100.

[25] Vetrivel, U.; Subramanian, G. and Dorairaj, S. A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. *Hugo J.*, *5*, 25-34.

[26] Varela, J.N.I.J.M.; de Hollanda, L.M. and Lancellotti, M. Membrane protein as novel targets for vaccine production in haemophilus influenzae and Neisseria meningitidis. *J. Vaccin Vaccination*, **2012**.

[27] Dunworth, W.P. and Caron, K.M. G protein-coupled receptors as potential drug targets for lymphangiogenesis and lymphatic vascular diseases. *Arterioscler Thromb. Vasc. Biol.*, **2009**, *29*, 650-656.

[28] Bhasin, M.; Garg, A. and Raghava, G.P. PSLpred: prediction of subcellular localization of bacterial proteins. *Bioinform.*, **2005**, *21*, 2522-2524.

[29] Yu, C.S.; Chen, Y.C.; Lu, C.H. and Hwang, J.K. Prediction of protein subcellular localization. *Proteins*, **2006**, *64*, 643-651.

[30] Yu, C.S.; Lin, C.J. and Hwang, J.K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.*, **2004**, *13*, 1402-1406.

[31] Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M. and Foster, L.J. Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, **2010**, *26*, 1608-1615.

[32] Emanuelsson, O.; Brunak, S.; von Heijne, G. and Nielsen H. Locating proteins in the cell using Target P, Signal P and related tools. *Nat. Protoc.*, **2007**, *2*, 953-971.

[33] Bendtsen, J.D.; Kiemer, L.; Fausboll, A.; Brunak, S. Non-classical protein secretion in bacteria. *BMC Microbiol.*, **2005**, *5*, 58.

[34] Tusnady, G.E. and Simon I.; The HMMTOP transmembrane topology prediction server; *Bioinformatics*, **2001**, *17*, 849-850.

[35] Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.; Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **2001**, *305*, 567-580.

[36] Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S.R.; Griffiths-Jones, S.; Howe, K.L.; Marshall, M. and Sonnhammer, E.L. The Pfam protein families database. *Nucleic Acids Res.*, **2002**, *30*, 276-280.

[37] Meinel, T.; Krause, A.; Luz, H.; Vingron, M. and Staub, E. The SYSTERS Protein family Database in 2005. *Nucleic Acids Res.*, **2005**, *33*, D226-D229.

[38] Letunic, I.; Doerks, T. and Bork, P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*, **2012**, *40*, D302-D305.

[39] Marchler-Bauer, A.; Lu, S.; Anderson, J.B.; Chitsaz, F.; Derbyshire, M.K.; DeWeese-Scott, C.; Fong, J.H.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Jackson, J.D.; Ke, Z.; Lanczycki, C.J.; Lu, F.; Marchler, G.H.; Mullokandov, M.; Omelchenko, M.V.; Robertson, C.L.; Song, J.S.; Thanki, N.; Yamashita, R.A.; Zhang, D.; Zhang, N.; Zheng, C. and Bryant, S.H. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.*, **2011**, *39*, D225-D229.

[40] Vlahovicek, K.; Kajan, L.; Murvai, J.; Hegedus, Z. and Pongor, S.; The SBASE domain sequence library, release 10: domain architecture prediction. *Nucleic Acids Res.*, **2003**, *31*, 403-405.

[41] Bairoch, A. and Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **2000**, *28*, 45-48.

[42] Gasteiger, E.; Jung, E.; Bairoch, A. SWISS-PROT: connecting biomolecular knowledge *via* a protein database. *Curr. Issues Mol. Biol.*, **2001**, *3*, 47-55.

[43] Hubbard, T.; Barker, D.; Birney, E.; Cameron, G.; Chen, Y.; Clark, L.; Cox, T.; Cuff, J.; Curwen, V.; Down, T.; Durbin, R.; Eyras, E.; Gilbert, J.; Hammond, M.; Huminiecki, L.; Kasprzyk, A.; Lehvaslaiho, H.; Lijnzaad, P.; Melsopp, C.; Mongin, E.; Pettett, R.; Pocock, M.; Potter, S.; Rust, A.; Schmidt, E.; Searle, S.; Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M. The Ensembl genome database project; *Nucleic Acids Res*, **2002**;*30*, 38-41.

[44] Quevillon E.; Silventoinen V.; Pillai S.; Harte N.; Mulder N.; Apweiler R., *et al.***;** InterProScan: protein domains identifier; *Nucleic Acids Res*, **2005**;*33*, W116-20.

[45] Baron C.; Coombes B.**;** Targeting bacterial secretion systems: benefits of disarmament in the microcosm; *Infect Disord Drug Targets*, **2007**;*7*, 19-27.

[46] Saha S.; Raghava G.P.**;** VICMpred: an SVM-based method for the prediction of functional proteins of Gram-negative bacteria using amino acid patterns and composition; *Genomics Proteomics Bioinformatics*, **2006**;*4*, 42-7.

[47] Saha S.; Raghava G.P.**;** BTXpred: prediction of bacterial toxins; *In Silico Biol*, **2007**;*7*, 405-12.

[48] Chakraborty A.; Ghosh S.; Chowdhary G.; Maulik U.; Chakrabarti S.**;** DBETH: a Database of Bacterial Exotoxins for Human; *Nucleic Acids Res*, **2012**;*40*, D615-20.

[49] Szklarczyk D.; Franceschini A.; Kuhn M.; Simonovic M.; Roth A.; Minguez P., *et al.***;** The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored; *Nucleic Acids Res*;*39*, D561-8.

[50] Roy A.; Kucukural A.; Zhang Y.**;** I-TASSER: a unified platform for automated protein structure and function prediction; *Nat Protoc*, **2010**;*5*, 725-38.

[51] Zhang Y.; Skolnick J.**;** TM-align: a protein structure alignment algorithm based on the TM-score; *Nucleic Acids Res*, **2005**;*33*, 2302-9.

[52] Holm L.; Rosenstrom P.**;** Dali server: conservation mapping in 3D; *Nucleic Acids Res*, **2010**;*38*, W545-9.

[53] Lomize M.A.; Pogozheva I.D.; Joo H.; Mosberg H.I.; Lomize A.L.**;** OPM database and PPM web server: resources for positioning of proteins in membranes; *Nucleic Acids Res,* **2012**;*40*, D370-6.

[54] Van Der Spoel D.; Lindahl E.; Hess B.; Groenhof G.; Mark A.E.; Berendsen H.J.**;** GROMACS: fast, flexible, and free; *J Comput Chem,* **2005**;*26*, 1701-18.

[55] Abeykoon A.H.; Chao C.C.; Wang G.; Gucek M.; Yang D.C.; Ching W.M.**;** Two protein lysine methyltransferases methylate outer membrane protein B from Rickettsia; *J Bacteriol,* **2012**;*194*, 6410-8.

[56] Murray G.L.; Srikram A.; Henry R.; Puapairoj A.; Sermswan R.W.; Adler B.**;** Leptospira interrogans requires heme oxygenase for disease pathogenesis; *Microbes Infect,* **2009**;*11*, 311-4.

[57] Lasica A.M.; Jagusztyn-Krynicka E.K.**;** The role of Dsb proteins of Gram-negative bacteria in the process of pathogenesis; *FEMS Microbiol Rev,* **2007**;*31*, 626-36.

[58] Thomson J.M.; Bonomo R.A.**;** The threat of antibiotic resistance in Gram-negative pathogenic bacteria: beta-lactams in peril!; *Curr Opin Microbiol,* **2005**;*8*, 518-24.

[59] Rolain J.M.; Maurin M.; Vestris G.; Raoult D.**;** *In vitro* susceptibilities of 27 rickettsiae to 13 antimicrobials; *Antimicrob Agents Chemother,* **1998**;*42*, 1537-41.

[60] Rolain J.M.; Stuhl L.; Maurin M.; Raoult D.**;** Evaluation of antibiotic susceptibilities of three rickettsial species including Rickettsia felis by a quantitative PCR DNA assay; *Antimicrob Agents Chemother,* **2002**;*46*, 2747-51.

[61] Kajimura J.; Fujiwara T.; Yamada S.; Suzawa Y.; Nishida T.; Oyamada Y., *et al.***;** Identification and molecular characterization of an N-acetylmuramyl-L-alanine amidase Sle1 involved in cell separation of Staphylococcus aureus; *Mol Microbiol,* **2005**;*58*, 1087-101.

[62] Cloud K.A.; Dillard J.P.**;** A lytic transglycosylase of Neisseria gonorrhoeae is involved in peptidoglycan-derived cytotoxin production; *Infect Immun,* **2002**;*70*, 2752-7.

[63] Tullius M.V.; Harth G.; Horwitz M.A.**;** Glutamine synthetase GlnA1 is essential for growth of Mycobacterium tuberculosis in human THP-1 macrophages and guinea pigs; *Infect Immun,* **2003**;*71*, 3927-36.

[64] Obi I.R.; Francis M.S.**;** Demarcating SurA activities required for outer membrane targeting of Yersinia pseudotuberculosis adhesins; *Infect Immun,* **2013**;*81*, 2296-308.

[65] Garmory H.S.; Titball R.W.**;** ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies; *Infect Immun,* **2004**;*72*, 6757-63.

[66] Kumar S.; Varela M.F.**;** Biochemistry of bacterial multidrug efflux pumps; *Int J Mol Sci,* **2012**;*13*, 4484-95.

[67] Bina X.R.; Provenzano D.; Nguyen N.; Bina J.E.**;** Vibrio cholerae RND family efflux systems are required for antimicrobial resistance, optimal virulence factor production, and colonization of the infant mouse small intestine; *Infect Immun,* **2008**;*76*, 3595-605.

[68] Jahn R.; Scheller R.H.**;** SNAREs--engines for membrane fusion; *Nat Rev Mol Cell Biol,* **2006**; *7*, 631-43.

[69] Fasshauer D.; Sutton R.B.; Brunger A.T.; Jahn R.**;** Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs; *Proc Natl Acad Sci U S A,* **1998**;*95*, 15781-6.

[70] De la Cruz M.A.; Zhao W.; Farenc C.; Gimenez G.; Raoult D.; Cambillau C., *et al.* A toxin-antitoxin module of Salmonella promotes virulence in mice; *PLoS Pathog,* **2013**;*9*, e1003827.

[71] Kaito C.; Morishita D.; Matsumoto Y.; Kurokawa K.; Sekimizu K.**;** Novel DNA binding protein SarZ contributes to virulence in Staphylococcus aureus; *Mol Microbiol,* **2006**;*62*, 1601-17.

[72] Doern C.D.; Holder R.C.; Reid S.D.**;** Point mutations within the streptococcal regulator of virulence (Srv) alter protein-DNA interactions and Srv function; *Microbiology,* **2008**;*154*, 1998-2007.

[73] Ariyachet C.; Solis N.V.; Liu Y.; Prasadarao N.V.; Filler S.G.; McBride A.E.**;** SR-like RNA-binding protein Slr1 affects Candida albicans filamentation and virulence; *Infect Immun*;*81*, 1267-76.

[74] Kondo Y.; Ohara N.; Sato K.; Yoshimura M.; Yukitake H.; Naito M., *et al.***;** Tetratricopeptide repeat protein-associated proteins contribute to the virulence of Porphyromonas gingivalis; *Infect Immun*;*78*, 2846-56.