Author's personal copy

Genes Genom
DOI 10.1007/s13258-014-0258-x

RESEARCH ARTICLE

In silico approach to identify the role of a putative protein MAP1138c in the virulence of Johne's disease

Syed A. Hassan

Received: 18 October 2014 / Accepted: 30 November 2014
© The Genetics Society of Korea and Springer-Science and Media 2014

Abstract The complete sequencing of *Mycobacterium avium* subspecies *paratuberculosis* (MAP) strain K-10 genome has shown the existence of over 4,000 genes. This opens up many opportunities to study the interaction of MAP with its hosts and the environment. Understanding the immune response directed at MAP antigens at different stages of infection, would be enhanced by the characterization of putative antigens. In this context, our comprehensive in silico analysis of MAP1138c, a putative protein demonstrates its sequential, physicochemical, structural and functional homology with Rv1411c (LprG) lipoprotein. The InterPro Scan studies have also shown that MAP1138c protein is a member of LppX/LprAFG family of lipoprotein involved in cell wall biogenesis and pathogenesis of Mycobacterium species. The structure assessment tools reveal that the theoretical structure of MAP1138c protein generated by SWISS-MODEL server shows homology with the crystal structure of Rv1411c (LprG) lipoprotein with respect to the global, local and stereochemical properties. Additionally, the structure-based ligand interaction studies using AutoDock Vina 1.1.2 shows that the triacylated glycoprotein (Ac_1PIM_2) also interacts with the hydrophobic pockets in the 3D theoretical structure of MAP1138c protein. Similar interactions of Rv1411c (LprG)-Ac_1PIM_2 leads to Toll-like Receptor 2 (TLR-2) mediated evasion of immune responses within host macrophages in tuberculosis infection. Hence, these results support our hypothesis that the MAP1138c protein is probably involved in immune evasion within host macrophages leading to virulence and infection in ruminants and human, respectively.

Keywords MAP1138c · Rv1411c (LprG) · Homology modeling · Triacylated glycolipids · Structure-based ligand interaction · Immune evasion · Virulence

Electronic supplementary material The online version of this article (doi:10.1007/s13258-014-0258-x) contains supplementary material, which is available to authorized users.

S. A. Hassan (✉)
Faculty of Computing and Information Technology Rabigh, King Abdulaziz University, P.O Box 344, Rabigh 21911, Saudi Arabia
e-mail: asif_srmcbt@yahoo.com

Introduction

Johne's disease is of foremost concern to the dairy industry worldwide (Nielsen and Toft 2009). *Mycobacterium avium* subspecies *paratuberculosis* (MAP) is the causal agent of this alarming disease of the ruminants and is associated with the *Mycobacterium avium* complex (MAC). MAC includes mild environmental bacteria as well as unscrupulous pathogen of humans (Turenne et al. 2007). Like any other *Mycobacterial* species, MAP is also specialized in the blockage of phagosome-lysosome fusion, thereby sustaining its survival (Kuehnel et al. 2001). Furthermore, they provide anti-apoptotic signals lengthening the life span of their host cell. The interaction of MAP with macrophages are mediated through variety of receptors, as well as Toll-like Receptors (TLRs) (Ferwerda et al. 2007; Koets et al. 2010). MAP can amplify interleukin (IL)-10, an immunomodulatory cytokine that represses killing of MAP by macrophages. It also triggers Th1-type immune responses which are mainly preferred to combat intracellular infections by inhibition of IL-12 (Weiss et al. 2002, 2005). Macrophages are one of the important cells of the innate immune system, but also provide signals to induce adaptive immune responses. MAP can interfere with antigen presentation, hence consequent adaptive responses (Weiss and Souza

Published online: 09 December 2014

Author's personal copy

2008). Correspondingly, the host immune response to *Mycobacterial sp.,* namely *Mycobacterium tuberculosis (M tb)* is generally mediated by a TLR-2 in macrophages (Yoshida et al. 2009; Jo et al. 2007; Stenger and Modlin 2002). Mtb non-acylated lipoprotein LprG (Rv1411c) has a TLR-2 agonist action that is dependent on its association with triacylated glycolipids binding specifically with the hydrophobic pocket of Mtb LprG lipoprotein protein. The detection of a glycolipid carrier function has important implications on the role of LprG in Mycobacterial physiology (Kuehnel et al. 2001; Gehring et al. 2004). Continued exposure (>16 h) of human macrophages to LprG (acylated agonist of TLR2) results in noticeable inhibition of MHC-II Ag processing. Moreover, this inhibition depends on TLR-2. MHC-II Ag processing inhibition by mycobacterial lipoproteins may permit *M. tuberculosis* within infected macrophages to evade the innate immune response of the host that results in virulence and latency (Weiss et al. 2002; Drage et al. 2010). On the contrary, we know little about the evasion and survival mechanism of MAP within the host macrophages, a process common to all mycobacterial infections. Recently, it was proposed by Hassan et al. (2014) that MAP1138c a putative protein shares a structural and functional homology with Rv1411c lipoprotein. Therefore, a proper functional and structural characterization of the putative antigenic proteins will be useful in understanding the evasion and survival mechanism of MAP within host macrophages.

In this regard, the present comprehensive comparative proteomic study of MAP1138c and Rv1411c lipoproteins shows that the putative protein MAP1138c shares both structural and functional homology with Rv1411c or LprG lipoprotein. The structure-based interaction studies of the theoretical 3D structure of MAP1138c protein and triacylated glycolipid shows that the Ac_1PIM_2 (triacylated glycolipids) binds specifically to the hydrophobic pockets of MAP1138c protein in a similar way as it binds to the hydrophobic pockets of the crystal structure of Rv1411c or LprG lipoprotein. Thus, we propose that MAP1138c-Ac_1-PIM_2 binary complex acts as an agonist for TLR-2 receptor and eventually leads to TLR-2 dependent inhibition of antigen processing within host macrophages and eventually reduced recognition by $CD4^+$ T cells. This mechanism of action may lead to immune elusion and latency of MAP within host macrophages.

Materials and Methods

Sequence retrieval

The accession numbers for MAP1138c and Rv1411c proteins were acquired from UniProtKB (http://www.uniprot.org/). FASTA sequences of both MAP1138c and Rv1411c (LprG) proteins studied were obtained from protein database (www.ncbi.nih.gov/protein). The protein sequences in FASTA format were used to compare the physicochemical, structural and functional properties of MAP1138c and Rv1411c (LprG) proteins.

Sequence analysis

Gapped-Basic Local Search Tool (BLAST) proposed by Altschul and Madden (1997) was used by the SWISS-MODEL server to identify homologs of the query protein (MAP1138c) against the ExPDB template library of the Swiss-Model server.

Phosphorylation sites evaluation

Comparative evaluation of the serine and threonine phosphorylation locations in the orthologous proteins (MAP1138c and Rv1411c (LprG)) was executed using NetPhosBac 1.0 server (Miller et al. 2009) (http://www.cbs.dtu.dk/services/NetPhosBac 1.0).

Secretory nature profiling

The existence of signal sequence and signal peptide cleavage site in the orthologous proteins (MAP1138c and Rv1411c (LprG)) was evaluated and compared using SignalP 4.1 server (Petersen et al. 2011) (http://www.cbs.dtu.dk/services/SignalP).

Globular/order and disorder regions prediction

The globular structures/order and disorder regions between MAP1138c and Rv1411c (LprG) protein was compared using GlobPlot (Linding et al. 2003) (http://globplot.embl.de/cgiDict.py).

Hydropathy analysis

The Hydropathy plot of DNASTAR was used to predict and compare the flexible region (Karplus and Schulz 1985), surface probability (Emini et al. 1985) and hydrophilicity (Kyte and Doolittle 1982) and the antigenic index (Jameson and Wolf 1988) of MAP1138c and Rv1411c (LprG) proteins directly from its primary amino acid sequence.

Sequence feature annotation

Domain assignment

The member databases of InterPro (Zdobnov and Apweiler 2001) helps in the identification of protein domains as well as the determination of protein function. InterPro Scan sequence analysis and classification tool were used to

Author's personal copy

detect and compare protein domains and functional sites of the orthologous proteins (MAP1138c and Rv1411c (LprG)).

Secondary, disorder and transmembrane prediction

PSIPRED (Jones 1999) a secondary structure prediction server based on two feed-forward neural networks will be used to compare the secondary structure, disorder region and transmembrane segment of the orthologous proteins (MAP1138c and Rv1411c (LprG)).

Homology modeling and model quality estimation

Model building

Raw model structure of MAP1138c protein was built by means of manual protein modeling server "SWISS-MODEL Workspace" (Biasini et al. 2014; Arnold et al. 2006; Bordoli et al. 2009; Kiefer et al. 2009; Kopp and Schwede 2006; Guex et al. 2009). The N-terminal secretion signal sequence of MAP1138c protein was removed before submitting to the server for model building. The hypothetical model of MAP1138c protein was made based on the results of target template sequence alignment using Promod-II (Guex and Peitsch 1997) and Modeller (Sali and Blundell 1993).

Model quality estimation

Model goodness and reliability of the model MAP1138c protein structure was evaluated by means of QMEAN score (Benkert et al. 2008; Benkert et al. 2009a, b, 2011). The estimated model reliability ranges between 0–1. The absolute model quality was measured by QMEAN Z-score. The QMEAN score obtained by the theoretical model was evaluated against the scores of high-resolution crystal structures of the same size, and a Z-score of the contributing components of QMEAN was determined. The standard Z-score of the high-resolution structure is zero. Finally, obtained model was validated at the SWISS-MODEL server using PROCHECK validation tool. PROCHECK aims to evaluate the similarity of the geometry of the amino acid residues in the predicted 3D structure of MAP1138c protein, when compared to stereochemical parameters obtained from well-refined, high-resolution structures (Laskowski et al. 1993). The analysis of the super-secondary structural motifs in model MAP1138c protein structure was performed using PROMOTIF (Hutchinson and Thornton 1996).

Ligand-binding domain (LBD) analysis

The prediction of ligand-binding sites/domains of MAP1138c protein structure was performed using Fpocket 1.0 (Guilloux et al. 2009) at http://fpocket.sourceforge.net/run_online.html.

Ligand preparation

Structure of triacylated glycolipid (Ac_1PIM_2) was drawn using Marvin Sketch and energy minimization of the ligand was accomplished using MMFF94 force field. Energy minimization was performed to help the docking program for identifying the bioactive conformer from the local minima.

Molecular docking

The PDB of the 3D theoretical structure of MAP1138c protein generated by SWISS-MODEL server was used for docking using AutoDock Vina 1.1.2 (Trott and Olson 2010). Fpocket web server was employed to detect the binding pockets in MAP1138c protein structure. The docking procedure for the interaction studies of MAP1138c (protein) and Ac_1PIM_2 (ligand) was fixed to rigid conditions, and the dock grid was set encircling the protein–LBD of MAP1138c protein. Exhaustiveness was set to 20 with all other parameters set to default standards. The docking and simulation studies was performed using the system configuration namely; Windows® seven operating system (OS) with 8 GB of RAM and Intel® Core™ i7 processor. The binary complex (Ac_1PIM_2-MAP1138c) categorized by their binding energy values, was examined for geometry and docking. After docking, the preeminent binary complex model was selected on the basis of lowest binding energy. Whereas the most appropriate complex (Ac_1PIM_2-MAP1138c) conformation was chosen on the basis of hydrogen bond interactions between the protein and ligand near the ligand-binding site. The lowermost energy poses specify the maximum binding affinity since high energy produces the unstable conformations. For visualization purpose, AutoDock Vina was used to generate docking poses, which was loaded directly to PyMol (Seelier and Bert 2010). PyMol was used further to produce the images of the protein ligands complex models.

Results

Gapped-BLAST analysis

A gapped-BLAST was performed for MAP1138c protein sequence to obtain its homologs that provided significant functional and structural motifs for the query protein. The gapped-BLAST analysis of the query sequence leads to the identification of a possible template with known X-Ray structure (PDB ID: 3MH8A) having 70 % identity and

Author's personal copy

83 % positive match (Supplemental Fig. 1). The template was found to be the crystal structure of LprG from H37Rv strain of *Mycobacterium tuberculosis*.

Analysis of phosphorylation potential of orthologous proteins

The phosphorylation of serine and threonine residues plays a key role in the regulation of host-pathogen interactions and cell signaling. The potency of the phosphorylation reaction is based upon the phosphorylation potential. The amino acids (serine and threonine) phosphorylation potential was calculated using NetPhosBac 1.0 server. The amino acids are having a phosphorylation potential greater than a threshold value are considered effective site of phosphorylation (Miller et al. 2009). Gain of phosphorylation site was observed for MAP1138c protein, where new eight potent phosphorylation sites (one threonine and seven serine residues) were found when compared to Rv1411c (LprG) protein. Whereas, Rv1411c (LprG) protein shows only four serine residues having phosphorylation potential greater than the threshold value (0.5). The predicted phosphorylation site and their position are tabulated in Table 1.

Analysis of signal sequences of orthologous proteins

Estimation of the signal sequence in a protein is important to categorize the protein as secretory or non-secretory protein (Petersen et al. 2011). In this study, MAP1138c and Rv1411c (LprG) proteins were used as query proteins to detect appropriate signal sequence. The results of SignalP 4.1 server for each of the orthologous proteins (Rv1411c (LprG) and MAP1138c) showed that these are preproteins and are secretory by nature due to the presence of secretory signal sequences in the N-terminal region of preprotein (Supplemental Fig. 2a, b). The length of the signal peptide for Rv1411c (LprG) and MAP1138c proteins were found to be 30 amino acids each. The presence of twin-arginine at N-terminal region of the signal peptide for both Rv1411c (LprG) and MAP1138c proteins relates to their translocation across the membrane via Tat translocases. It can also be stated that after translocation there is a possibility of cleavage of the preprotein by SPase I (Paetzel et al. 2002).The mature peptide for MAP1138c and Rv1411c (LprG) proteins were found to be 31–238 and 31–235, respectively. Thus, we conclude that MAP1138c and Rv1411c (LprG) proteins are preproteins and are secretory by nature.

Table 1 Putative phosphorylation site in the query proteins [MAP1138c and H37Rv1411c (LprG)]

Sl.No.	Proteins	Number of potent phosphorylation site above threshold phosphorylation potential (0.500)	
		Threonine (T)	Serine (S)
1	MAP1138c	1	11
2	Rv1411c	–	4

Analysis of the globular and disorder domains in orthologous proteins

Globular and disorder domains in a protein plays significant roles in protein function (Linding et al. 2003). Thereby, a change in the profile of order and disorder region in a protein may lead to a change in the functionality of the protein. MAP1138c and Rv1411c (LprG) proteins of MAP and H37Rv exhibit similar profile of the globular domain. Figure 1a, b graphically represents the comparative GlobPlot profiles of both MAP1138c and Rv1411c (LprG) proteins. The GlobPlot profile of both MAP1138c and Rv1411c (LprG) show absence of low complexity,

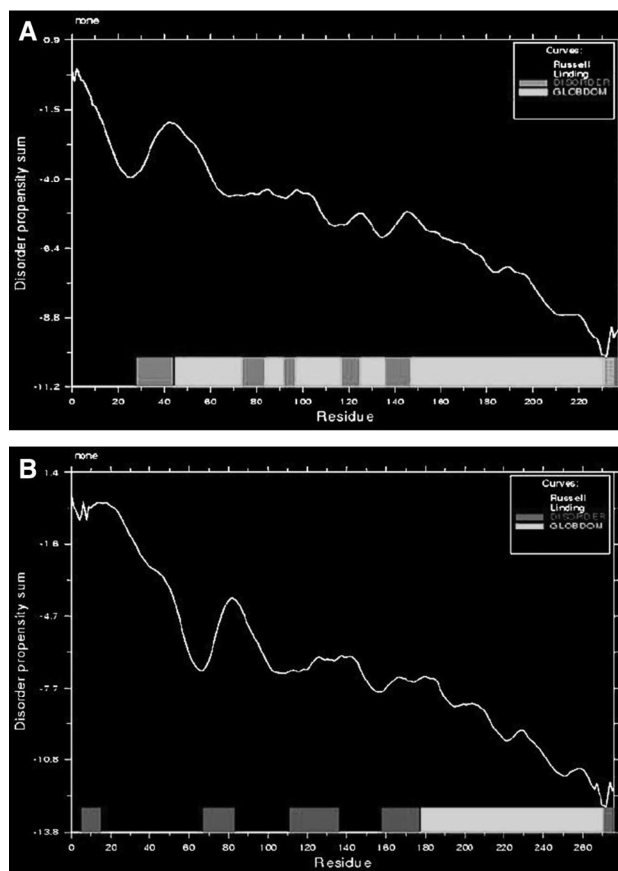

Fig. 1 a–b Disorder propensity of **a** MAP1138c and **b** H37Rv1411c (LprG) predicted by GlobPlot to detect the low complexity region (*yellow*), disorder region (*blue*), globular domain (displayed in *green*) and transmembrane region (striped column-*white*)

Author's personal copy

transmembrane and coil & coil region, thereby highlighting its role as a secretory globular protein.

Analysis of the hydropathy plot of orthologous proteins

Figure 2a, b shows the result of the hydropathy analysis of both MAP1138c and Rv1411c (LprG) proteins using DNASTAR software (Madison, WI, USA) which reveals regions of high antigenic index (potential antigenic determinants). In silico analysis of MAP1138c protein revealed the presence of high antigenic index regions (above zero-reticle): (31–60, 78–85, 90–95, 100–106, 120–134, 140–146, 155–180, 200–207, 210–220 and 230–238 aa). Similarly, for Rv1411c (LprG) protein (31–60, 78–92, 98–105, 118–134, 138–144, 152–175, 198–206, 210–218 and 222–236 aa) amino acids stretches with high antigenic index were observed. The comparative antigenic profiling of MAP1138c and Rv1411c (LprG) proteins were promising as the antigenic contours for both the proteins all along the sequence were nearly similar. Moreover, the flexible regions, surface probability and hydrophilicity plot of both the protein were also identical.

ProtScan domain and functional analysis of orthologous proteins

ProtScan Domain analysis showed that both MAP1138c and Rv1411c (LprG) proteins share a common domain DUF1396 ranging from 40–231 amino acids. This domain comprises of various lipoproteins and as a group; they are known as the LppX/LprAFG family from Mycobacterium species. The members of this family are involved in virulence and localization of complex lipids to the outer membrane of MTB (Bigi et al. 1997). According to MEMSAT3 (Jones et al. 1994), tool both proteins share a similar profile of transmembrane segment at their N-Terminal region. Figure 3a, b shows the presence of a transmembrane segment at the N-terminal region of both MAP1138c and Rv1411c (LprG) proteins. The transmembrane segment of MAP138c protein consists of an outside loop (1–9 a.a), inside helix cap (10–13 a.a), central transmembrane helix segment (14–22 a.a) and outside helix cap (23–26 a.a). Similarly, the transmembrane segment of Rv1411c protein constitute of an outside loop (1–10 a.a), inside Helix cap (11–14 a.a), central transmembrane helix

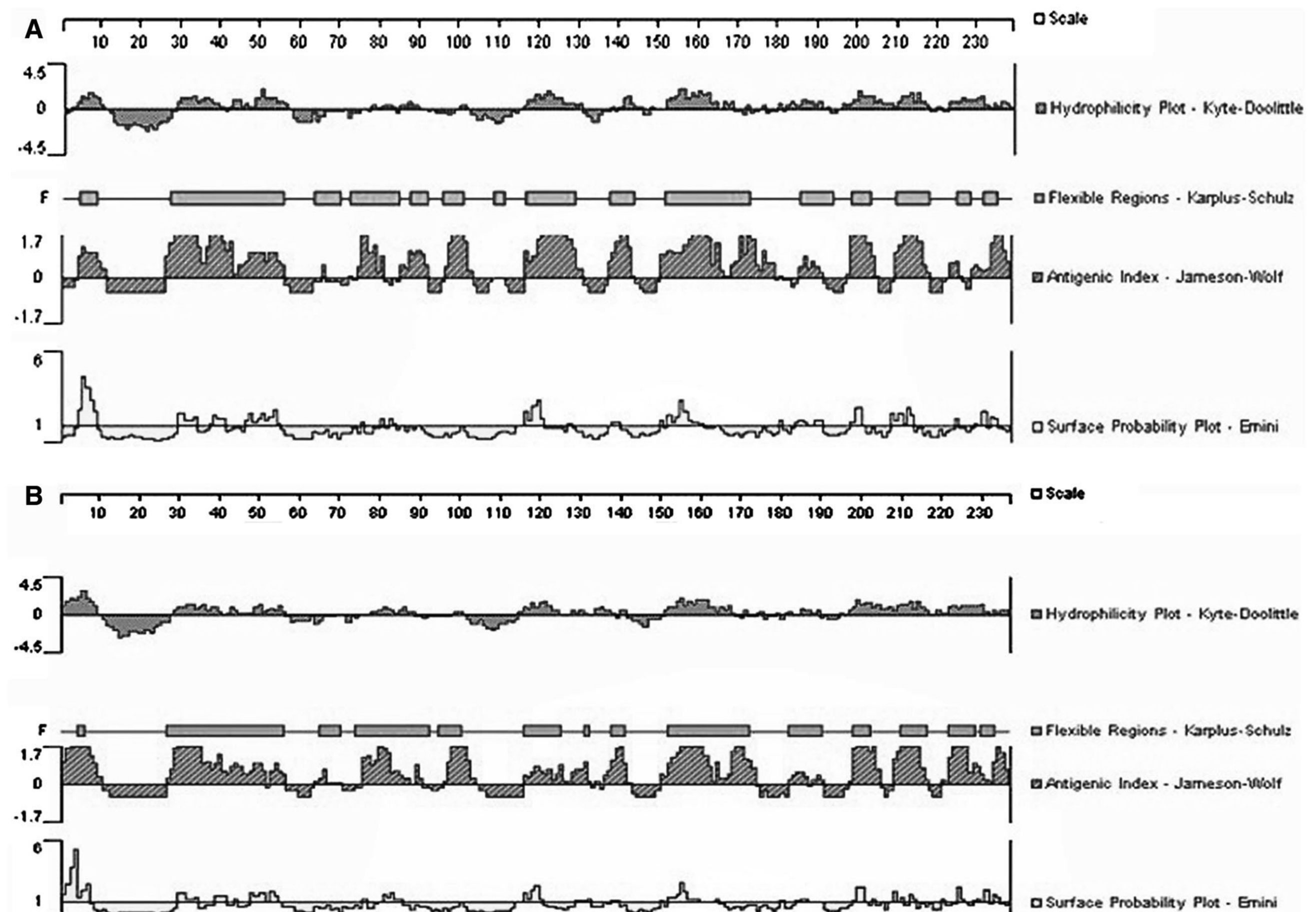

Fig. 2 **a–b** A comparative graphical representation of Hydropathy plot of **a** MAP1138c and **b** H37Rv1411c (LprG) proteins

Author's personal copy

Fig. 3 a–b Graphical representation of the ProProt Scan analysis of **a** MAP1138c and **b** Rv1411c (LprG) proteins

segment (15–23 a.a) and outside helix cap (24–27 a.a) for Rv1411c (LprG) protein. The N-terminal region of MAP1138c protein corresponds to the signal sequence as predicted by SignalP 4.1. Therefore, the presence of central transmembrane helix at the C-terminal region of the signal sequence of the protein is important for anchoring the protein to the outer membrane of the bacteria during secretion across the membrane.

Template search and homology modeling

A theoretical model of MAP1138c protein was generated using the automated protein-modelling SWISS-MODEL server. The coordinates of crystal structure of Rv1411c (LprG) protein from MTB (PDB ID: 3MH8A) was used as a template. Figure 4 shows the resultant 3D model of MAP1138c protein showing the secondary structure