BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 11(12)

Insights from the predicted epitope similarity between *Mycobacterium tuberculosis* virulent factors and its human homologs

Venkata Ravi Gutlapalli[1,2], Aparna Sykam[1, 2], Anuraj Nayarisseri[3], Sujai Suneetha[1], Lavanya M Suneetha[1]*

[1]CODEWEL Nireekshana-ACET, Hyderabad, Telangana-500029, India; [2]Centre for Biotechnology, AcharyaNagarjuna University, Nagarjuna Nagar, Guntur, Andhra Pradesh 522510, India; [3]Bioinformatics Research Laboratory, Eminent Biosciences, Vijaynagar, Indore - 452010, Madhya Pradesh, India; Lavanya M Suneetha – Email: drlavanyasuneetha@gmail.com; Phone: 9752295342; *Corresponding author

Received November 21, 2015; Revised December 13 2015; Accepted December 13, 2015; Published December 31, 2015

Abstract:
Mycobacterium tuberculosis is known to be associated with several autoimmune diseases such as systemic lupus erythematous, rheumatoid arthritis and multiple sclerosis. This is attributed to sequence similarity between virulent factors and human proteins. Therefore, it is of interest to identify such regions in the virulent factors to assess potential autoimmune related information. *M. tb* specific virulent factors were downloaded from the VFDB database and its human homologs were identified using the sequence comparison search tool BLASTP. Both virulent proteins and their corresponding human homologs were further scanned for epitopes (B cell and HLA class I and II allele specific) using prediction programs (BCPRED and NETMHC). Data shows the presence of matching 22 B-cell, 79 HLA class II and 16 HLA class I specific predicted epitopes in these virulent factors having human homologs. A known peptide (HAFYLQYKNVKVDFA) associated with autoimmune atopic dermatitis is shown in the superoxide dismutase homolog structures of the bacterium (PDB ID: 1IDS) and human (PDB ID: 2QKC). This data provides insight into the understanding of infection-associated auto-immunity

Background:
Pathogenic intracellular organisms have strategies of evading or suppressing the host's immune response. Strategies against acquired immunity include antigenic variation, immune suppression and molecular mimicry. Molecular mimicry is well documented in viruses such as HIV, monkey pox and cow pox and its primary function is camouflage **[1, 2].** Molecular mimicry can be defined as sequence or structural similarity between host and pathogen peptides resulting in immune evasion or cross reactivity leading to autoimmune response. Pathogens may also mimic host molecules to manipulate factors in signal transduction pathways via their receptors **[3–6].** Previous studies have shown that cross reactive antibodies are produced in response to bacterial infections causing tissue damage **[7–9].** Tuberculosis (TB) has been associated with several autoimmune diseases such as systemic lupus erythematous, rheumatoid arthritis and multiple sclerosis **[10–18]**. *M. tb* induced T cell reactivity with foreign and self-antigens lead to autoimmune responses **[11, 14, 15, 17].** Thus, detecting epitopes involved in cross reactivity could help in comprehending TB immuno-pathogenesis. The present study identified epitopes with sequence and structural similarities between *M. tuberculosis* virulent factors and host homologs for B-cells and T-cells (class I and II HLA alleles) specificity **[19, 20]**

Methodology:
Virulent Factors Database (VFDB)
M.tb specific virulent factors (number) were downloaded in FASTA format from VFDB (a database of virulent factors) **[21].**
Basic Local Alignment Search Tool (BLAST) - 2.2.28
The Basic Local Alignment Search Tool (BLAST) is used to find regions of local similarity between *M.tb* virulent factor and human proteome **[22].**

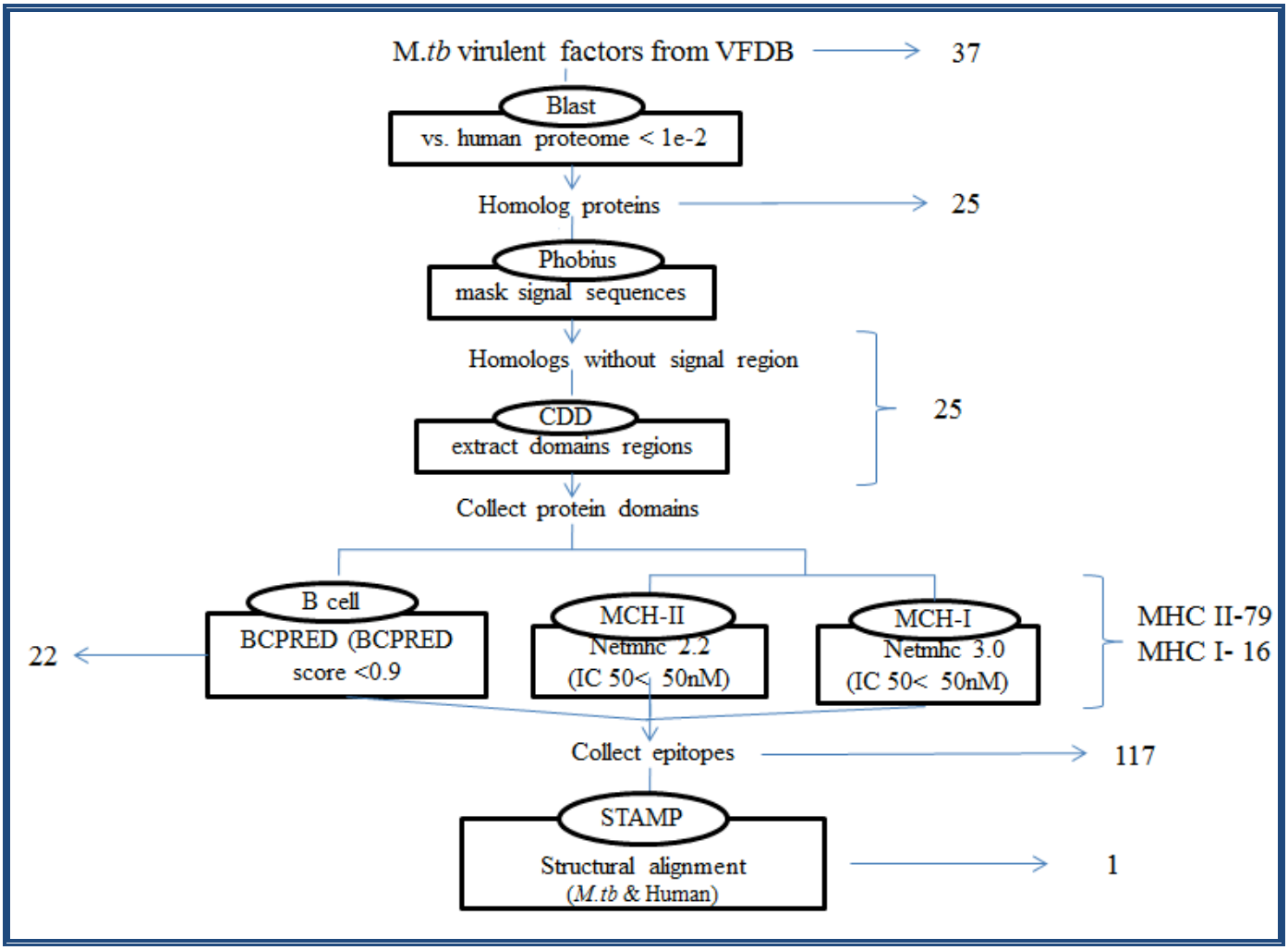

Figure 1: A workflow showing steps involved in the identification of epitopes in the virulent factors of *M. tb* having human homologs. Epitopes were predicted in both virulent factors and their corresponding homologs. Both virulent proteins and their corresponding human homologs were further scanned for epitopes (B cell (BCPRED with score > 0.9) and HLA class I and II alleles (NETMHC with binding score < 50nM) prediction.

Phobius 1.0.1
Phobius was used to identify and exclude the signal region of the homologous proteins. **[23].**

Conserved Domain Database (CDD)
This database was used to identify conserved domains in homologous proteins of *M.tb* virulent factors and human.

B-cell epitope prediction server (BCPREDS)
Prediction of B cell epitopes **(Table 1)** for *M.tb* specific virulent factors and its corresponding human homologs using B cell epitope prediction server (BCPREDS) **[24].**

T-cell epitope prediction
Prediction of HLA class-**I (Table 2)** and class-II **(Table 1)** T cell epitopes was completed for virulent factors and its corresponding human homologs using NetMHC 2.2 and 3.0 **[25].**

Visual Molecular Dynamics 1.9.1(VMD)
This program was used to visualize the 3D structures of *M. tb* and human superoxide dismutase **[26]**

Workflow
M. tb virulent factors are obtained from VFDB and human proteins from Ensembl. Virulent factors are BLAST searched against human proteome using BLAST (version 2.2.28). Then homologs are extracted at E-value ≤ 0.01. The remaining *M. tb* proteins were run through Phobius to remove predicted N-terminal signal peptides from the protein sequence. Then sequences are run through CDD for getting domain co-ordinates. Further the collected sequences are run through BCPRED server for B cell epitopes of 20 amino acids length and the classifier specificity was 75% and overlap filter was used for analysis. Based on prior BLAST results, regions of amino acids (small peptides) that were similar between the human and *M. tuberculosis* proteins were selected for further analysis. BCPRED score of greater than 0.9 is considered for blast matched peptides in both pathogen and host homologs.

NetMHC (version 2.2) was used for HLA class II and NetMHC (version 3.0) for HLA class I binding peptide prediction. Peptides were selected based on IC_{50} values <50 nM as high affinity, <500 nM as intermediate affinity and <5000 nM as low affinity **[27].** The matched peptides in both pathogen and host with a binding score less than IC50 ≤ 50 are considered as

strong binders. 3D structures of protein sequences matched to host are viewed and aligned structurally to find out whether these peptides are on the surface of the protein. The similarity between the predicted epitopes of virulent factors was found by multiple structural alignments using the STAMP algorithm in VMD. The detection of epitopes is shown in **Figure 1.** All calculations were performed using the local Linux server.

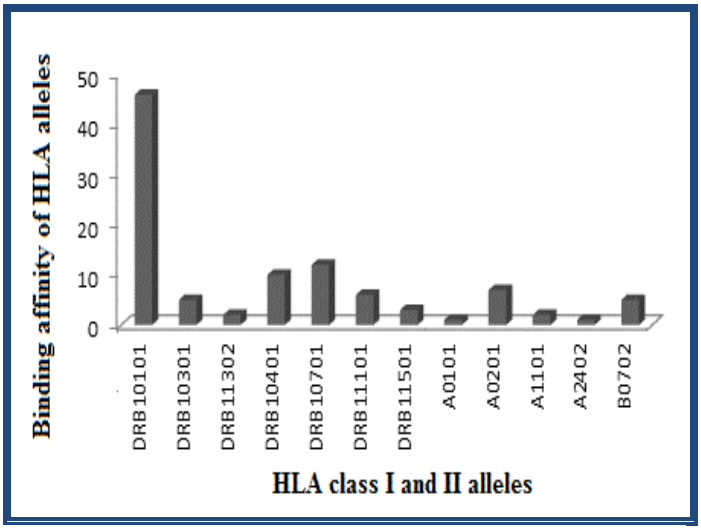

Figure 2: Frequency of epitopes with strong binding affinity for HLA class I and II T cell alleles that shared similarity with M.tb virulent factors. Legend:HLA class II host-pathogen epitopes binding to HLA-DRB10101 (57%) show highest affinity over HLA class Iepitopes indicating their significant role in pathogenesis of TB.

Results & Discussion:

The analysis of data obtained from the search between *M. tuberculosis* virulence factors and the human proteome revealed considerable similarities in sequences. A total of 25 best-hit homologous proteins with E-value cut off 1-E02 and similar regions of 9 or more amino acids were identified. The classification of the homologous virulent factor proteins are 21 metabolic proteins 3 membrane associated protein and a protein kinase **(Table 3).** Binding affinities of *M.tb* virulent factors vs. B cell epitopes and HLA class II and I alleles were measured by BCPRED and NetMHC. A peptide was considered having significant affinity to virulent factors if it had a BCPRED score ≥ 0.9 for B cell epitope **(Table 4)** and IC50 value ≤ 50 for HLA class II and I epitopes. The analysis of binding affinities of HLA class II peptides is 83% as compared to HLA class I (17%). Of 79 HLA class II host-pathogen epitopes highest affinity was to HLA-DRB10101 (57% followed byHLA-DRB10701 (14%), HLA-DRB10401 (11%), HLA-DRB10301 (6%), HLA-DRB11101 (5%), HLA-DRB10302 (2.5%) and HLA-DRB11501 (2.5%). The analysis of HLA class I peptides indicated a maximum affinity of peptides binding to allele A*0201 (44%) followed by B*0702 (31%), 2 for A*1101 (13%) and 6% each for A*0101 & A*2402 **(Figure 2, Table 1 & 4).** HLA class II has significant number of high affinity binding peptides which could be involved in dys-regulation of T cell function and or autoimmunity **[28].** The virulent factors binding to host tissue antigens could influence signaling and immune evasion **[29].**

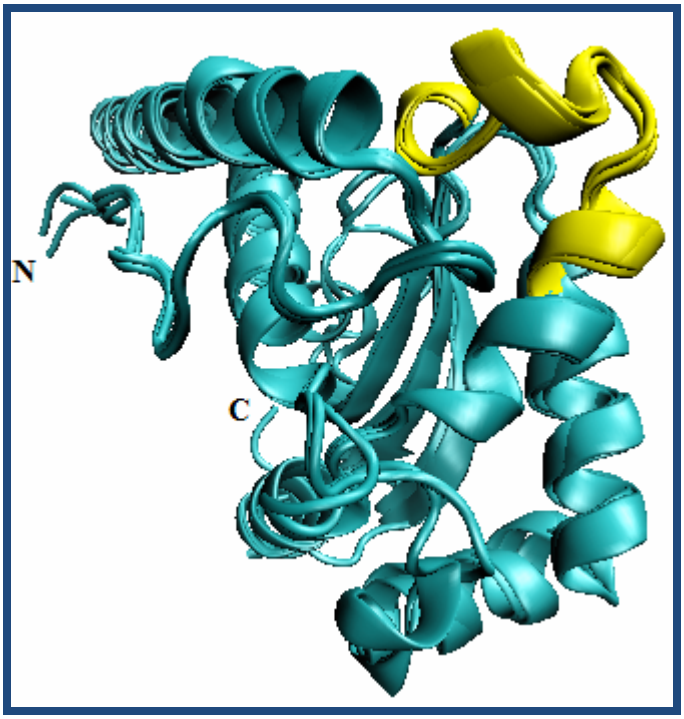

Figure 3: Multiple structural alignments of M.tbsuperoxide dismutase (1IDS), Aspergillus fumigatus Mn superoxide dismutase (1KKC) and human Mn superoxide dismutase (2QKC) by STAMP in VMD-multiseq window. Legend:visualization panel of VMD shows structurally conserved epitope 'HAFYLQYKNVKVDFA' in yellow among M.tbsuperoxide dismutase (1IDS), Aspergillus fumigatus Mn superoxide dismutase (1KKC) and human Mn superoxide dismutase (2QKC)

The myco-bacterial virulent proteins of this study were classified into categories such as structural, metabolic, catalytic, kinases and transport proteins **(Table 4, 1 & 2**). Majority of the virulent factor epitopes having binding affinity for B and HLA class I and II alleles were involved in (i) lipid, protein and nucleotide metabolism/degradation pathways (ii) free radical mediated damage pathway (iii) ion transport (iv) degradation of proteins glycosylation/phosphorylation pathways. These similarities could impact metabolic rate of reactions, interfere in homeostasis of cell and could trigger cell damage by free radical mediated reactions **[29].** Peptides, which have binding capacity to more than one allele of HLA class-I and-II, are called promiscuous peptides. Promiscuous peptides for HLA class II were 24% (19/79) and none for HLA class I molecules. Interestingly, the presence of promiscuous peptides for HLA class II suggest that these peptides could have role in presentation of antigens for immune recognition and amplification of response against *M.tb* **(Table 3 and 4) [30].** Genetic studies on HLA class I and II alleles are associated with susceptibility to disease and the present study indicates their similarities/binding to *M.tb* virulent factors.

This study identified a host peptide human manganese superoxide dismutase (*MnSOD)* sharing structural similarity with *M. tb* Superoxide dismutase (*M. tb SOD*) virulent factor. This epitope was previously implicated in diseases such as atopic dermatitis (AD), autoimmune hepatitis (AIH), Epstein-Barr virus (EBV) infection and fumigatus-allergy **[31–34].** A peptide from *M.tb* Superoxide dismutase (*M.tb SOD*)

HAFYLQYKNVKVDFA, bound to allele HLA-DRB1*15:01 allele with high affinity is identified **(Table 3).** HLA-DRB1*15:01 is known to be responsible for susceptibility to tuberculosis. Further there was a high structural similarity of *M.tb SOD* and human *MnSOD* at both primary and tertiary structure level **Figure 3 [35]**. Clinical studies identified *MnSOD* cross-reactive autoimmune antibodies in patients with atopic dermatitis (AD) and has been implicated in disease pathogenesis **[31].** This epitope is conserved and well investigated in *Aspergillus fumigatus Mn SOD* (1KKC) in relation to various autoimmune conditions **[31–34].** Identifying the key homologous peptides of host pathogen similarity could help us design highly selective peptide blockers, which would be a valuable addition to complement the understanding of autoimmune diseases.

PDB crystal structures of superoxide dismutase *M.tb*, human and *Aspergillus fumigatus* were available. Superposition revealed a high measure of structural conservation and similarity with low RMSD having Qres value of 0.9 and showing high measure of the similarity of the 'C-C alpha' distances between residues of aligned proteins **(Figure 3) [36]**. These structurally similar regions of these three epitopes (which is known to cause atopic dermatitis) could be significant in tuberculosis in causing immune inflammatory processes characteristic of TB (**Figure 3).** It can be noted that many other mycobacterial antigens have been associated with autoimmune diseases **[37–39]**. There is no clear evidence that *M.tb* virulent factors are involved and further clinical investigations on epitope specificities involved in autoimmunity are warranted.

Although, computational tools have been used in the past to examine molecular mimics in other diseases **[40];** the understanding of these epitopes need to be further probed. Utilizing these methods, we have identified potential auto-reactive B cell, HLA class II and class I epitopes that may elicit autoimmune response during *M. tuberculosis* infection. The findings of this study are as follows: (i) there were 95 auto reactive B cell, HLA class II and class I epitopes that are similar to peptides of myco-bacterial virulent factors; (ii) 22% of similarities were promiscuous that are binding to HLA class II cell epitopes (iii) high Qres score of 0.9 suggesting structural similarity between *M.tb SOD* and human *Mn SOD* and the epitope has an established evidence of autoimmunity. The similarities were observed across the spectrum of metabolic activities of host cell suggesting *M.tb* could use multiple split approach in causing tuberculosis.

Conclusions:

We report regions in the *M.tb* virulent factors having human homologs sharing predicted B-cell and T-cell epitopes. Data shows the presence of 22 B-cell, 79 HLA class II and 16 HLA class I specific predicted peptides in these virulent factors having human homologs. A known peptide (HAFYLQYKNVK VDFA) associated with autoimmune atopic dermatitis is shown in the superoxide dismutase homolog structures of the bacterium (PDB ID: 1IDS) and human (PDB ID: 2QKC). This data provides insights in understanding infection-associated auto-immunity.

Acknowledgement:

The authors gratefully thank CODEWEL Nireekshana-ACET for support.

References:

[1] Lambris JD *et al. Nat Rev Microbiol* 2008 **6:** 132 [PMID: 18197169]
[2] Srinivasappa J *et al. J Virol* 1986b **57**: 397 [PMID: 3753614]
[3] Hide G *et al. Mol Biochem Parasitol* 1989 **36**: 51 [PMID: 2682237]
[4] Ghansah TJ *et al. J Eukaryot Microbiol* 2002 **49**: 383 [PMID: 12425525]
[5] Spiliotis M *et al. Gene* 2003 **323:** 57 [PMID: 14659879]
[6] Vicogne J *et al. J Biol Chem* 2004 **279**: 37407 [PMID: 15231836]
[7] Sfriso P *et al. J Leukoc Biol* 2010 **87**: 385 [PMID: 20015961]
[8] Shahrizaila N & Yuki N, *J Biomed Biotechnol*. 2011 **2011**: 829129 [PMID: 21197269]
[9] Vardhini D *et al. Infect Genet Evol.* 2004 **4**: 21 [PMID: 15019586]
[10] Birnbaum G & Kotilinek L, *Ann N Y Acad Sci*. 1997 **835**: 157 [PMID: 9616771]
[11] Birnbaum G *et al. Ann Neurol* 1993 **34:** 18 [PMID: 8517675]
[12] Ghosh K *et al. Rheumatol Int* 2009 **29:** 1047 [PMID: 19360412]
[13] Mor F & Cohen IR, *J Clin Invest* 1992 **90**: 2447 [PMID: 1281835]
[14] Res PC *et al. Lancet* 1988 **2**: 478 [PMID: 2457140]
[15] Salvetti M *et al. J Autoimmun*. 1992 **5**: 691 [PMID: 1489484]
[16] Salvetti M *et al. J Neuroimmunol* 1996 **65:** 143 [PMID: 8964896]
[17] van Eden W *et al. Proc Natl Acad Sci U S A* 1985 **82**: 5117 [PMID: 3927297]
[18] van Eden W *et al. Nature* 1988 **331:** 171 [PMID: 2448638]
[19] Agrewala JN & Wilkinson RJ, *Eur J Immunol* 1999 **29**: 1753 [PMID: 10382737]
[20] Gowthaman U & , Agrewala JN, *J Proteome Res.* 2008 **7:** 154 [PMID: 18034454]
[21] Chen L *et al. Nucleic Acids Res* 2012 **40:** D641 [PMID: 22067448].
[22] Boratynet GM *et al. Nucleic Acids Res*. 2013 **41**: W29 [PMID: 23609542]
[23] Käll L *et al. J Mol Biol*. 2004 **338:** 1027 [PMID: 15111065].
[24] El-Manzalawyet Y *et al.* Comput Syst Bioinformatics Conf. 2008 **7:** 121 [PMID: 19642274].
[25] Nielsen M *et al. Bioinformatics* 2004 **20:** 1388 [PMID: 14962912]
[26] Humphrey W *et al. J Mol Graph*. 1996 **14**: 33 [PMID: 8744570].
[27] Lundegaard C *et al. Bioinformatics* 2008 **24**: 1397 [PMID: 18413329]
[28] Sundberg EJ *et al. Semin. mmunol.* 2007 **19:** 262 [PMID: 17560120]
[29] Schmid-Hempel P, *Philos Trans R Soc Lond B Biol Sci.* 2009 **364**: 85 [PMID: 18930879]
[30] Mustafaet AS, *PLoS One.* 2014 **9**: e103679 [PMID: 25136958]
[31] Schmid-Grendelmeier P *et al. J Allergy Clin Immunol*. 2005 **15:** 1068 [PMID: 15867868]
[32] Crameri R *et al. J Exp Med* 1996 **184:** 265 [PMID: 8691141]
[33] Miyata M *et al. Clin Rheumatol* 1995 **14:** 673 [PMID: 8608687]
[34] Dalpke AH *et al. J Med Virol*. 2003 **71**: 408 [PMID: 12966547]
[35] Flückiger S *et al. J Immunol*. 2002 **168**: 1267 [PMID: 11801664]
[36] Eargle J *et al. Bioinformatics* 2006 **22**: 504 [PMID: 16339280]
[37] Wucherpfennig KW, *J Clin Invest* 2001 **108**: 1097 [PMID:

11602615]
[38] Esaguy N *et al. Infect Immun*. 1991 **59**: 1117 [PMID: 1900057]
[39] Van Eden W *et al. Proc Natl Acad Sci U S A*. 1985 **82**: 5117 [PMID: 3927297]
[40] Kovvali G *et al. FEBS Lett*. 2005 **579**: 2261 [PMID: 15848155]

Edited by P Kangueane
Citation: **Gutlapalli *et al.*** Bioinformation 11(12): 517-524 (2015)

BIOINFORMATION

open access

Supplementary material:

Table 1: Predicted MHC class II restricted peptides in virulent factors having human homologs using NETMHC 2.2 (affinity score: IC50 < 50).

pos	matched peptide	affinity(nM)	Allele
7	NYLAQTYSVLVTSAA	30.3	HLA-DRB10701
81	LLAFTNPTVNSYKRL	28.2	HLA-DRB10101
11	YLQYTSGSTRTPAGV	38.8	HLA-DRB10401
85	FCLFSSGAALLGSPG	7.5	HLA-DRB10101
68	RNTVQFAAAVQAAME	6.8	HLA-DRB10101
102	RNTVQFAAAVQAAME	35.6	HLA-DRB10701
17	YLQYTSGSTRTPAGV	38.8	HLA-DRB10401
5	LGHFAAVSAATGLVV	3.7	HLA-DRB10101
8	QLTYRELDALADRLA	20.8	HLA-DRB10401
194	GNTVAMRLRPQSAMS	15.2	HLA-DRB10101
8	YARYLAEHGARRIVL	4.8	HLA-DRB10701
11	RHRLLSDVTRALADE	17.9	HLA-DRB10301
148	ILAELGMDTTTLVAA	34.4	HLA-DRB10101
12	LAELGMDTTTLVAAL	17.9	HLA-DRB10301
6	ERHTAINSLVTATHG	29.1	HLA-DRB10401
20	LDIFAALRSGGAIVV	4.9	HLA-DRB10101
157	LAYVLFTSGSTGEPK	13.4	HLA-DRB10101
109	TLLLADVLAAAPAEF	10.5	HLA-DRB10101
195	VQEALRSAAGHSAYP	15.7	HLA-DRB10101
219	GGHFYLNDHLDAVAR	17.5	HLA-DRB10101
43	KSLVANDVDTFVVQY	39	HLA-DRB10301
303	AAQRAVRAALNGRTA	28.1	HLA-DRB10101
45	YHDMGLILGICAPLV	12.6	HLA-DRB10101
64	RFDYEQLTAGQARPC	18	HLA-DRB10101
120	LPLYHDMGLILGICA	34.2	HLA-DRB10101
40	AVYRAALAAAGVQPE	5.4	HLA-DRB10101
115	KLMTRIAGAGAMGSV	8.8	HLA-DRB10101
18	VVTGLNNSVASGRIA	16.5	HLA-DRB11302
43	LPFYHDMGLVIGICA	14	HLA-DRB10101
50	VDYRLIPKHSLGMAL	17.8	HLA-DRB10101
6	DSAGGYLALALAQRL	15.7	HLA-DRB10701
2	HAFYLQYKNVKVDFA	5.9	HLA-DRB10101
19	WDYGALEPHISGQIN	10	HLA-DRB11501
31	YGALEPHISGQINEL	15.8	HLA-DRB10101
50	PDLDWDYGALEPHIS	38.4	HLA-DRB10101
41	SLRLLQVGGSKLEPE	7.6	HLA-DRB10101
154	EGELLVRGPYTLNGY	40.6	HLA-DRB10101
169	PGDRVLLQLPNGCQF	47.7	HLA-DRB10101
4	QWSRLLAQRAKALDS	3.8	HLA-DRB10101
111	LLATAAARMVTAWRR	8.8	HLA-DRB10101
163	DSIVALSVVQAARRR	12.2	HLA-DRB10101
165	QAAYVIFTSGTTGTP	12.5	HLA-DRB10101
392	AYVIFTSGTTGTPKG	34.4	HLA-DRB10101
438	GELVTAVAEQTLGAL	42.3	HLA-DRB10101
129	DLDRIRARVAAALPE	10.4	HLA-DRB10101
54	GVDLDRIRARVAAAL	22.9	HLA-DRB10301
390	YLIYTSGTTGLPKGV	41.8	HLA-DRB10101
59	ICVSYRITGDIDLAR	24.4	HLA-DRB10301
111	VDALSANIVSAAVAD	13	HLA-DRB10101
34	RGERFVDALSANIVS	9.8	HLA-DRB10701
41	GERFVDALSANIVSA	30.3	HLA-DRB11302
115	AANRLLDVMAAQLRA	13.4	HLA-DRB10101
263	GHSLGEVAAAYLAGS	40.6	HLA-DRB10101
90	VGHSLGEVAAAYLAG	39.8	HLA-DRB10701
102	LTVDTSCSSALAAFH	39.6	HLA-DRB10101
164	VQFVGPLSVVDSALA	42.6	HLA-DRB10101
103	SFAILHPKKYEEIVR	24.3	HLA-DRB10101
53	GKTTIARVVANILAG	12	HLA-DRB10101
133	TLLLGPPGTGKTSVA	39	HLA-DRB10101

Table 2: Predicted MHC class I restricted peptides in virulent factors having human homologs using NETMHC 2.2 (affinity score: IC 50 < 50).

pos	matched peptide	affinity(nM)	Allele
18	ALLSGLLRA	26	A*0201
394	FIDEAYALV	5	A*0201
438	AIINTLLLY	26	A*1101
4	FLITVALAL	9	A*0201
281	LLAEAQAEL	17	A*0201
499	LPRTSSGKL	7	B*0702
332	APAGRGPLL	14	B*0702
272	TPATPATPV	20	B*0702
236	YPAVLTSPV	9	B*0702
685	PSDPTALAY	14	A*0101
862	SLVGYVTPA	13	A*0201
218	FYHDMGLVI	42	A2402
166	YLQYKNVKV	9	A*0201
1580	ALAAILADV	13	A*0201
697	TSGSTGEPK	41	A*1101
16	LPRRLAIAA	33	B*0702

Table 3: Human homologs of M. tuberculosis virulent factors identified using BLASTP search (cut-off E value < 0.01)

M. tuberculosis		Human		Sequence identity (%)
VFG	GenBank Description	Gene ID	GenBank Description	
1384	Pyrroline-5-carboxylase (proC)	NP_075566	Pyrroline-5-carboxylate reductase 3	32
1387	SAICAR synthetase (purC)	NP_001072993	Multi-functional protein ADE2 isoform 1	21
1393	Fatty acyl-AMP ligase (fadD33)	XP_005257758	Acyl-CoA synthetase family member 2	25
1399	Glutamine synthetase (GlnA1)	NP_057655	Lengsin isoform A	24
1403	Phenyloxazoline synthase MbtB	XP_005265782	Acyl-CoA synthetase isoform X5	27
1408	Myco-cerosic acid synthase (Mas)	NP_004095	Fatty acid synthase	30
1415	Esterase/lipase (LipF)	NP_997248	Aryl acetamide deacetylase-like 2 precursor	34
1811	fbpC2	NP_001975	S-formyl glutathione hydrolase	24
1812	Secreted antigen 85-B FbpB (85B)	XP_005266335	S-formylglutathione hydrolase isoform X1	31
1817	Salicyl-S-ArCP synthetase	NP_001230208	Acyl-CoA synthetase family member 3	27
1820	Peptide synthetase (MbtF)	NP_115890	Acetyl-coenzyme A synthetase 2-like	23
1821	Peptide synthetase (MbtE)	XP_005265781	Acyl-CoA synthetase isoform X4	26
1822	Poly ketide synthetase (MbtD)	NP_004095	Fatty acid synthase	26
1823	Polyketidesynthetase (MbtC)	NP_004095	Fatty acid synthase	33
1826	GTP pyrophosphokinase (RelA)	NP_940929	3'-pyrophos-phohydrolase MESH1	39
2380	ESX-1 type VII secretion protein	XP_005264430	ATPase family AAA domain 2B isoform X2	31
1404	Alkyl hydroperoxidereductase C	NP_054817	Thio-redoxin-dependent peroxide reductase	37
1407	Fatty-acid-AMP ligase (FadD26)	NP_055789	Disco-interacting protein 2 homolog C	26
1409	Fatty-acid-AMP ligase (FadD28)	XP_005252487	Disco-interacting protein 2 homolog C isoform X5	25
1421	Superoxide dismutase [FE] (SodA)	NP_001019636	Superoxide dismutase [Mn]	52
1825	Sensor histidine kinase (DevS)	NP_001137311	cGMP-dependent 3',5'-cyclic phosphodiesterase	27
2391	Membrane-anchored mycosin (MycP1)	NP_777596	Proprotein convert a ses-ubtilisin	27
2401	ESX conserved EccA5	NP_054828	ATPase family AAA domain-containing protein 2	30
1380	Cu, Zn Superoxide dismutase (sodC)	NP_005116	Copper chaperone for superoxide dismutase	28
1397	Heat shock protein (HspX)	NP_077721	Outer dense fiber protein 1	34

Table 4: Predicted B cell epitopes in virulent factors and its corresponding human homologs using BCPRED (cut-off score < 0.9).

VFG	Epitope *(M. tb)*	Human AC #	Matched peptide (human)	Domain name
1393	STCAVTVPVPGIGLLADRVI	XP_005257758	NSPVTFAHFPEDTVEQKAES	Malonyl-CoA synthetase
1399	DAISGWWNTGAATEADGSPN	NP_057655	QELVDGLYHTGANVESFSSS	Glutamine synthetase
1403	FAGLGGATETAVHATIFEVQ	XP_005265782	IFNVYGITEVSSWATIYRIP	Adenylate forming
1403	ADSGDDCPDWVAGELWVSGR	XP_005265782	RDTNGFTIQEGSGQVFLGGR	Adenylate forming
1404	IVDPNNEIQFVSATAGSVGR	NP_054817	IIDPNGVIKHLSVNDLPVGR	Peroxiredoxin (PRX)
1407	PKQTAQVFDAKLVDPAPAAP	NP_055789	GDESLQSDHFNSRLSFGDTQ	Adenylate forming
1408	AIVEEAPAEASAPESSPGDA	NP_004095	IILRPNTQPPPAPAPHATLP	polyketide synthases
1409	VHGDNVANGYWQKPDESERT	XP_005252487	VCAVATGTSYYGLSGMTKNT	Adenylate forming
1421	AEYTLPDLDWDYGALEPHIS	NP_001019636	KHSLPDLPYDYGALEPHINA	Fe/Mn superoxide dismutases
1811	MPAWLQANKGVSPTGNAAVG	NP_001975	LPQLINANFPVDPQRMSIFG	S-formylglutathione hydrolase

1817	LLRAGAIPVMCLPGHRAAEL	NP_001230208	SWMSGGVAVPLYRKHPAAQ	Malonyl-CoA synthetase
1820	RVAEILRQTSAPVVIDEGVF	NP_115890	RVVELKKIVDEAVKHCPTVQ	Malonyl-CoA synthetase
1820	DENALAAINVTEGPATPPQT	NP_115890	MRRLLRKIITSEAQELGDTT	Malonyl-CoA synthetase
1821	TSGTTGLPKGVAVPHRPVAE	XP_005265781	TSGTTGIPKIVRVPHKCIVP	Adenylate forming domain
1821	PKINTTMHLLDDSLQPVPTG	XP_005265781	PLLGTVVEVRDTNGFTIQEG	Adenylate forming
1823	LAGHDVGCYVGASALEYGPA	NP_004095	LRGTHTGVWVGVSGSETSEA	Polyketide synthases
1823	GMVEGHGTATRLGDRTELRS	NP_004095	EYIEAHGTGTKVGDPQELNG	Polyketide synthases
2380	DEAYALVQERDGRTDPFGQE	XP_005264430	DEIDGLAPVRSSRQDQIHSS	ATPase family AAA
2389	SPPPPDVPTLVVPSPGTPGT	NP_055853	TPEEPASPAAAVPTPEEPTS	EspG family
2389	PIPGAPVTPITPTPGTPVTP	NP_055853	PEEPTSPAAAVPTPEEPTSP	EspG family
2391	DPRGWNNVQTVVTPAWYAPL	NP_777596	APEVITVGATNAQDQPVTLG	Peptidase S8 & S53
2401	HTLHEKGYSQGDPYGNAIIN	XP_005264430	DGLAPVRSSRQDQIHSSIVS	ATPase family AAA