Brief Original Article

In silico identification of cytotoxic T lymphocyte epitopes encoded by RD5 region of *Mycobacterium tuberculosis*

Jun-Wei Zhao[1], Min Yan[2], Guang Shi[1], Shu-Lin Zhang[3], Liang Ming[1]

[1] *Clinical Laboratory, The First Affiliated Hospital of Zhengzhou University, Key Clinical Laboratory of Henan Province, Department of Laboratory Medicine of Zhengzhou University, Zhengzhou, Henan, China*
[2] *Key Laboratory of Medical Immunology, Henan University of Traditional Chinese Medicine, Zhengzhou, Henan, China*
[3] *Department of Immunology and Microbiology, Shanghai Jiao Tong University School of Medicine, Shanghai, China*

Abstract
Introduction: While CD8+ T cells (cytotoxic T lymphocytes, CTLs) play important roles in immunity against *Mycobacterium tuberculosis*, only a small number of human leukocyte differentiation antigen (HLA) class I-restricted CTL epitopes for TB have been identified. The current study evaluates CTL epitopes of Rv3117 and Rv3120 proteins, two newly found *M. tuberculosis* region-diffference-5 (RD5)-encoded antigens, and their population coverage.
Methodology: The amino acid sequences of the two proteins were subjected to epitope analysis under HLA-A2, A3 and B7 supertype restriction using NetCTL, SYFPEITHI, BIMAS, NetCTLPan, IEDB, NetMHC and NetMHCPan prediction online servers.
Results: Eight RD5-encoded CTL epitopes were identified in the two proteins and the average population coverage of these epitopes was 87.2% among populations worldwide.
Conclusion: These CTL epitopes that were identified *in silico* and may have potential use for CD8+ T cell-mediated TB vaccine design.

Key words: tuberculosis; region of difference 5; cytotoxic T lymphocyte epitope; population coverage.

Introduction

Tuberculosis (TB), caused by intracellular pathogen *Mycobacterium tuberculosis* (*M. tuberculosis*) [1,2], remains one of the world's biggest threats [3].The Bacillus Calmette–Guerin (BCG) vaccine is an attenuated strain of *Mycobacterium bovis* and represents the world's most widely used live vaccine; however, this vaccine lacks efficacy in preventing pulmonary TB in adults [4]. To combat this ongoing scourge, the development of a more effective vaccine for TB is a global priority.

Effective cell-mediated immunity is essential to control *M. tuberculosis* infection [2], and T cell-mediated adaptive immunity plays important role in controlling initial *M. tuberculosis* infection and preventing reactivation of latent *M. tuberculosis* bacilli residing in granulomas [1]. Although it is generally accepted that CD4+ T cells are essential component of protective immunity against TB [2], recent studies have shown that major histocompatibility complex (MHC) class I-restricted CD8+ T cells (cytotoxic T lymphocytes, CTLs) are also important to immunity against *M. tuberculosis* [5-7]. In humans, MHC is also well known as human leukocyte differentiation antigen (HLA). HLA Class I molecules have variable epitope-binding specificities, and the frequency of HLA Class I polymorphism is unique across ethnicities [8]. However, only a small number of HLA class I-restricted CTL epitopes for TB have been identified; therefore, it is urgent to identify more promiscuous CTL epitopes and estimate the population coverage of these epitopes for design of an effective TB vaccine.

Comparative genomic studies have identified several *M. tuberculosis*-specific genomic regions of differences (RDs) that are absent in avirulent *M. bovis* BCG strains [9] and have paved a way for the identification of new candidate antigens for protective TB vaccines. Rv3117 and Rv3120 were found to be two *M. tuberculosis* RD5-encoded B-cell antigens in our previous study [10], while Rv3117 was able to evoke specific humoral and cellular immune responses in C57BL/6 mice [11]. Based on these findings, we

hypothesized that Rv3117 and Rv3120 proteins may have several CTL epitopes capable of inducing a robust CD8+ T cell immune response in humans. However, the CTL epitopes and their population coverage of Rv3117 and Rv3120 have not been investigated.

In the current study, based on the distribution of characteristics of HLA Class I alleles across all populations worldwide, we developed a combined immunoinformatics approach to identify potential CTL epitopes of Rv3117 and Rv3120 and population coverage of these epitopes.

Methodology

Protein sequence retrieval and analysis

The amino acid sequences of Rv3117 and Rv3120 of *M. tuberculosis* H37Rv (NC_000962) were downloaded from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) and compared with those of *M. tuberculosis* and human proteins using BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The molecular weights and isoelectric points of the Rv3117 and Rv3120 proteins were calculated using Accelrys Gene 2.5 software (Accelrys Inc., San Diego, USA).

Epitope prediction

Nine-mer CTL epitopes of *M. tuberculosis* Rv3117 and Rv3120 proteins were predicted based on HLA Class I supertypes using six prediction online servers (Table 1). First, the NetCTL 1.2 Server was used, and the threshold value for epitope identification was set at 0.75 [12]. In this study, HLA-A2, A3, and B7 supertypes were included in this analysis, as they were the most common HLA Class I supertypes across ethnicities, with an estimated frequency of > 80% [8]. Second, the predictive CTL epitopes by NetCTL 1.2 server were analyzed by the rest five prediction online tools: SYFPEITHI [13], BIMAS [14], NetCTLPan [15], Immune Epitope Database (IEDB) Analysis Resource [16], NetMHC [17], and NetMHCPan [18]. The default prediction value was selected. If there was no HLA Class I supertype list in these servers, the representative HLA Class I alleles were selected based on a pervious study [19] (Supplementary Table 1). Any 9-mer peptide, which was predicted by all six prediction online servers was identified as a potential CTL epitope in Rv3117 and Rv3120 proteins. Finally, potential CTL epitopes were evaluated for inclusion in the CD4+ T cell epitopes of Rv3117 and Rv3120 proteins by NetMHCIIpan 2.2 server (available at http://www.cbs.dtu.dk/services/NetMHCII/), one of the most accurate prediction servers based on an artificial neural network algorithm [20]. Ten HLA-DR alleles (HLA-DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*0801, DRB1*1101, DRB1*1301, DRB1*1302, DRB1*1501, and DRB5*0101) were selected for the prediction due to their inclusions in an HLA-DR supertype [21].

Population coverage calculation

For the short-listed potential CTL epitopes, population coverage was calculated using tools from IEDB source available at http://tools.immuneepitope.org/tools/population/iedb_input [22].

Results

Rv3117 and Rv3120 protein analysis

Rv3117 and Rv3120 protein sequences encoded by RD5 of *M. tuberculosis* were conserved in the *M. tuberculosis* complex and exhibited no more than 35% identity to human proteins. The Rv3117 and Rv3120 proteins had molecular weights of 31.0 KDa (isoelectric point 4.97) and 21.8 KDa (isoelectric point 5.10) respectively.

**Table 1.** Epitope prediction software online servers.

Name	Publish year	Algorithm	Website
NetCTL 1.2 server	2007	Artificial neural networks, Matrix	http://www.cbs.dtu.dk/services/NetCTL/
NetCTLpan 1.1 server	2010	Artificial neural networks, Matrix	http://www.cbs.dtu.dk/services/NetCTLpan/
NetMHC 3.2 server	2008	Artificial neural networks	http://www.cbs.dtu.dk/services/NetMHC/
NetMHCpan 2.8 server	2007	Artificial neural networks	http://www.cbs.dtu.dk/services/NetMHCpan/
SYFPEITHI	1999	Matrix	http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm
BIMAS	1994	Matrix	http://www-bimas.cit.nih.gov/molbio/hla_bind/
IEDB Analysis Resource	---	IEDB recommended	http://tools.immuneepitope.org/mhci/

Epitope prediction and population coverage

A total of eight potential CTL epitopes in *M. tuberculosis* RD5-encoded Rv3117 and Rv3120 proteins were successfully predicted against three HLA class I supertypes (Table 2). The number of predictive CTL epitopes in Rv3117 protein was more than that in Rv3120 protein. Five CTL epitopes were identified in Rv3117 protein, while only three CTL epitopes were found in Rv3120 protein. The scores from all six CTL epitope prediction tools showed that P99-107 (KLYGHEWVK) held the greatest potential in Rv3117 protein and was restricted to HLA-A3 supertype. On the other hand, p165-173 (KPGQRESEL) held the greatest potential in Rv3120 protein and was restricted to HLA-B7 supertype. There was no predictive CTL epitope restricted to HLA-A3 in Rv3120 protein. In addition, for Rv3117 protein, 9-mer CTL epitopes were overlapped with CD4+ T cell epitope (Supplementary Table 2). For example, p267-275 (SLVGAPIEL) was included in the CD4+ T cell epitope P261-277 (SWTEYGSLVGAPIELGS) restricted to HLA-DRB1*1501 and HLA-DRB1*0701 (Supplement Table 2). However, there was no predictive CTL epitope included in CD4+ T cell epitopes restricted to any HLA Class II alleles for Rv3120 protein.

Population coverage rates of eight CTL epitopes were shown in Figure 1. The combination of eight CTL epitope sequences from *M. tuberculosis* RD5-encoded Rv3117 and Rv3120 proteins provide a population coverage rate exceeding 75.0% for people in most areas, and 2.9% and 69.1% for people in Central America and Oceania, respectively. The cumulative worldwide population coverage of these epitopes was 87.2% (Figure 1).

Figure 1. The population coverage rates (%) were shown for the eight CTL epitopes as predicted by the six prediction online servers. CTL, cytotoxic T lymphocyte.
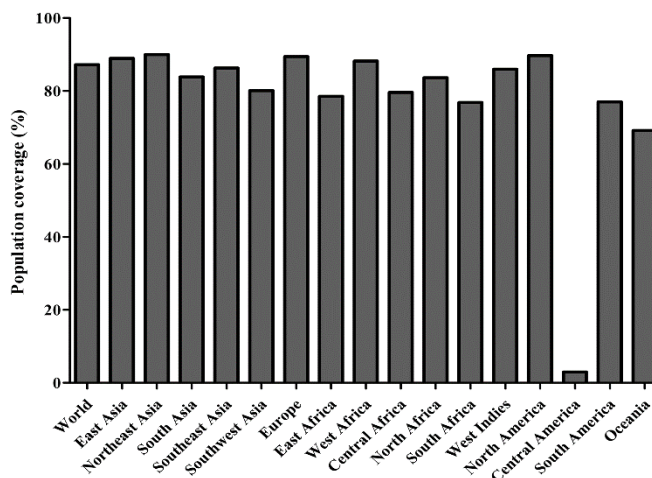
Discussion

A more effective vaccine for TB should generate both strong CD8+ T cell and CD4+ T cell responses, and a BCG vaccine with additional CTL epitopes may induce more balanced T cell response to enhance efficacy. In this study, we developed a combined immunoinformatics approach to predict eight RD5-encoded HLA Class I-restricted CTL epitopes for TB in Rv3117 and Rv3120 proteins. Three of the eight epitopes were even presented CD4+ T cell epitopes. In addition, eight CTL epitopes in Rv3117 and Rv3120 proteins were found to have 87.2% worldwide population coverage, indicating that they may be sufficient candidates for a novel TB vaccine design. To the best of our knowledge, this is the first report on population coverage of CTL epitopes.

The number of predictive epitopes between these proteins varied, which might be because of the variation

Table 2. CTL epitopes of Rv3117 and Rv3120 proteins were predicted using NetCTL, SYFPEITHI, BIMAS, NetCTLPan, IEDB, NetMHC, and NetMHCPan online servers.

Protein	Surpertype	Peptide	Sequence	Score			Percentile rank		Bind level[a]	
				NetCTL	SYFPEITHI	BIMAS	NetCTLPan	IEDB	NetMHC	NetMHCPan
Rv3117	A2	p267-275	SLVGAPIEL	1.1327	26	21.362	0.8	1.4	SB	SB
	A3	p99-107	KLYGHEWVK	1.5624	33	225.000	0.2	0.2	SB	SB
		p91-99	AAYAYWYFK	1.4742	19	9.000	0.1	0.3	SB	SB
	B7	p186-194	RPGHIPGAI	1.5366	19	8.000	0.8	0.4	SB	SB
		p19-27	APKVVFVEV	1.0586	21	12.000	0.8	1.5	WB	WB
Rv3120	A2	p93-101	ELADRLTLV	0.9865	24	49.191	1.0	1.8	WB	WB
	B7	p165-173	KPGQPESEL	1.3317	25	120.000	0.8	0.5	SB	SB
		p46-p54	VPEGPVLEL	0.9737	28	36.000	0.8	0.8	SB	WB

[a] SB, strong bind; WB, weak bind.

in their sizes. The relatively large protein Rv3117 with a molecular mass of 31.0 KDa was found to have more CTL epitopes compared with the Rv3120 protein. In addition, some of the CTL epitopes identified *in silico* were promiscuous. For an example, the CD4+ T cell epitope P85-105 (GNNNWFAAYAYWYFKLYGHEK) which includes CTL epitope p91-99 (AAYAYWYFK), was restricted to not only HLA-A3, but also HLA-DRB1*1501, HLA-DRB1*0701, and HLA-DRB1*1101 alleles (Supplement Table 2).

There are about 30 tools available for the prediction of CTL epitopes [23]; however, no tool has an overall 100.0% prediction sensitivity or specificity. The six prediction online servers in this study were applied independently or in integration to predict peptide HLA Class I-binding, proteasomal C-terminal cleavage, TAP transport efficiency, and the peptide half-time dissociation to HLA Class I molecules. Thus, theoretically, integrated application of six servers will improve the accuracy of *in silico* identification of CTL epitopes in Rv3117 and Rv3120 proteins. In fact, these online servers have been successfully used for the identification of T cell epitopes against different pathogens [24,25] and cancers [26]. Since NetCTL provides epitope prediction with 54-89% sensitivity and 94-99% specificity, we first employed NetCTL to predict CTL epitopes in Rv3117 and Rv3120 proteins at the default prediction value 0.75, which had 80% sensitivity and 97% specificity in a previous study [12]. Predictive CTL epitopes with high scores were further evaluated by SYFPEITHI, BIMAS, NetCTLPan, IEDB, NetMHC, and NetMHCPan online servers. Using this methodology, we aimed to predict CTL epitopes in Rv3117 and Rv3120 proteins that may bind to multiple HLA alleles promiscuously and serve as potential CTL epitopes for novel TB vaccine development. In fact, P118 (LIASNVAGV) in RD11-encoded protein Rv3425, a CTL epitope, was predicted to be a possible epitope for lysing target cells using three of the above six online servers [27].

In this study, eight identified *in silico* CTL epitopes could cover most individuals worldwide, making them favorable candidates for novel TB vaccine design. However, further experiments such as peptide-sensitized peripheral blood mononuclear cells and isolated CD8+ CTL responses *in vitro* and *in vivo*, are necessary to define these CTL epitopes.

Conclusions

Taken together, these eight *in silico* identified CTL epitopes would be used as potential candidates for the development of novel TB vaccine, based on CD8+ T cell immune responses. Further studies are warranted to investigate the immunogenicity and efficacy of a TB vaccine containing these epitopes *in vitro* and *in vivo*.

References
1. Philips JA, Ernst JD (2012) Tuberculosis pathogenesis and immunity. Annu Rev Pathol 7: 353-384.
2. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP (2013) The immune response in tuberculosis. Annu Rev Immunol 31: 475-527.
3. Xiao Y, Sha W, Tian Z, Chen Y, Ji P, Sun Q, Wang H, Wang S, Fang Y, Wen HL, Zhao HM, Lu J, Xiao H, Fan XY, Shen H, Wang Y (2016) Adenylate kinase: a novel antigen for immunodiagnosis and subunit vaccine against tuberculosis. J Mol Med 94:823-834.
4. von Reyn CF (2012) New studies of BCG: implications for tuberculosis vaccines. Lancet Infect Dis 12: 259-260.
5. Chaitra MG, Shaila MS, Chandra NR, Nayak R (2008) HLA-A*0201-restricted cytotoxic T-cell epitopes in three PE/PPE family proteins of *Mycobacterium tuberculosis*. Scand J Immunol 67: 411-417.
6. Jeong YH, Jeon BY, Gu SH, Cho SN, Shin SJ, Chang J, Ha SJ (2014) Differentiation of antigen-specific T cells with limited functional capacity during *Mycobacterium tuberculosis* Infection. Infect Immun 82: 132-139.
7. Behar SM (2013) Antigen-specific CD8(+) T cells and protective immunity to tuberculosis. Adv Exp Med Biol 783:141-163.
8. Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50: 201-212.
9. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284: 1520-1523.
10. Zhang MM, Zhao JW, Sun ZQ, Liu J, Guo XK, Liu WD, Zhang SL (2012) Identification of RD5-encoded *Mycobacterium tuberculosis* proteins As B-cell antigens used for serodiagnosis of tuberculosis. Clin Dev Immunol 2012: 738043.
11. Zhao JW, Sun ZQ, Pei H, Ye J, Chen CC, Samten B, Zhang SL, Guo XK (2013) Immunological evaluation of a novel *Mycobacterium tuberculosis* antigen, Rv3117, absent in *Mycobacterium bovis* BCG. Mol Med Rep 8: 1587-1593.
12. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8: 424.

13. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics 50: 213-219.
14. Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152: 163-175.
15. Stranzl T, Larsen MV, Lundegaard C, Nielsen M (2010) NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics 62: 357-368.
16. Moutaftsi M, Peters B, Pasquetto V, Tscharke DC, Sidney J, Bui HH, Grey H, Sette A (2006) A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. Nat Biotechnol 24: 817-819.
17. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 36: W509-W512.
18. Zhang H, Lundegaard C, Nielsen M (2009) Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. Bioinformatics 25: 83-89.
19. Sidney J, Peters B, Frahm N, Brander C, Sette A (2008) HLA class I supertypes: a revised and updated classification. BMC Immunol 9: 1.
20. Nielsen M, Justesen S, Lund O, Lundegaard C, Buus S (2010) NetMHCIIpan-2.0 - Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. Immunome Res 6: 9.
21. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chesnut RW, Grey HM, Sette A (1998) Several common HLA-DR types share largely overlapping peptide binding repertoires. J Immunol 160: 3363-3373.
22. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 7: 153.
23. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V (2008) Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. BMC Immunol 9: 8.
24. Yao Y, Huang W, Yang X, Sun W, Liu X, Cun W, Ma Y (2013) HPV-16 E6 and E7 protein T cell epitopes prediction analysis based on distributions of HLA-A loci across populations: an *in silico* approach. Vaccine 31: 2289-2294.
25. Dervillez X, Qureshi H, Chentoufi AA, Khan AA, Kritzer E, Yu DC, Diaz OR, Gottimukkala C, Kalantari M, Villacres MC, Scarfone VM, McKinney DM, Sidney J, Sette A, Nesburn AB, Wechsler SL, Benmohamed L (2013) Asymptomatic HLA-A*02:01-restricted epitopes from *Herpes Simplex* virus glycoprotein B preferentially recall polyfunctional CD8+ T cells from seropositive asymptomatic individuals and protect HLA transgenic mice against ocular herpes. J Immunol 191: 5124-5138.
26. Yoshimura S, Tsunoda T, Osawa R, Harada M, Watanabe T, Hikichi T, Katsuda M, Miyazawa M, Tani M, Iwahashi M, Takeda K, Katagiri T, Nakamura Y, Yamaue H (2014) Identification of an HLA-A2-restricted epitope peptide derived from hypoxia-inducible protein 2 (HIG2). PLoS One 9: e85267.
27. Chen F, Zhai MX, Zhu YH, Qi YM, Zhai WJ, Gao YF (2012) *In vitro* and *in vivo* identification of a novel cytotoxic T lymphocyte epitope from Rv3425 of *Mycobacterium tuberculosis*. Microbiol Immunol 56: 548-553.

Corresponding author
Ming Liang.
Clinical Laboratory, The First Affiliated Hospital of Zhengzhou University, Key Clinical Laboratory of Henan Province, Department of Laboratory Medicine of Zhengzhou University, 1 E. Jianshe Road, Zhengzhou 450052, Henan, China.
Phone: (086) 0371-61905678
Fax: (086) 0371-61905678
E-mail: mingliang3072@163.com

Conflict of interests: No conflict of interests is declared.

Annex – Supplementary Items

Supplement table 1. HLA Class I supertypes and their representative HLA Class I alleles.

HLA Class I supertype	Representative HLA Class I alleles
HLA-A2	HLA-A*02:01, A*02:02, A*02:03, A*02:04, A*02:05, A*02:06, A*02:07, A*02:14, A*02:17, A*68:02 and A*69:01
HLA-A3	HLA-A*03:01, A*11:01, A*31:01, A*33:01, A*33:03, A*66:01, A*68:01 and A*74:01
HLA-B7	HLA-B*07:02, B*07:03, B*07:05, B*15:08, B*35:01, B*35:03, B*42:01, B*51:01, B*51:02, B*51:03, B*53:01, B*54:01, B*55:01, B*55:02, B*56:01, B*67:01, and B*78:01

Supplement table 2. CD4+ T cell epitopes from Rv3117 and Rv3120 proteins were predicted using NetMHCPan 2.2 online server.

Protein	HLA-II Alleles	Peptide	Sequence [a]	Consensus sequence	Corresponding HLA-I supertype
Rv3117	DRB1*1501	P261-277	SWTEYG**SLVGAPIEL**GS	SLVGAPIEL	A2
		P85-105	GNNNWF**AAYAYWYFK**LYGHEK	AAYAYWYFK	A3
	DRB1*0701	P261-277	SWTEYG**SLVGAPIEL**GS	SLVGAPIEL	A2
		P85-105	GNNNWF**AAYAYWYFK**LYGHEK	AAYAYWYFK	A3
		p13-27	AESNLH**APKVVFVEV**	APKVVFVEV	B7
	DRB1*1101	P85-105	GNNNWF**AAYAYWYFK**LYGHEK	AAYAYWYFK	A3

[a] The underlined residues are the 9-mer CTL epitopes overlapped with the CD4+ T cell epitopes.