Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vaccine 23 (2005) 1265–1272
Vaccine
www.elsevier.com/locate/vaccine

Defining putative T cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential

M.G. Chaitra[a], Sridhar Hariharaputran[b], Nagasuma R. Chandra[b], M.S. Shaila[a], R. Nayak[a,*]

[a] *Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India*
[b] *Bioinformatics Centre, Indian Institute of Science, Bangalore 560012, India*

Received 11 June 2004; accepted 26 August 2004
Available online 12 October 2004

Abstract

The identification of T cell epitopes from immune relevant antigens of *Mycobacterium tuberculosis* is a critical step in the development of a vaccine covering diverse populations. Two multigene families, PE-PGRS and PPE make up about 10% of the *M. tuberculosis* genome. However, the functions of the proteins coded by these large numbers of genes are unknown. All possible nonameric peptide sequences from PE and PPE proteins were analysed in silico for their ability to bind to 33 alleles of class I HLA. These results reveal that of all PE and PPE proteins, a significant number of these peptides are predicted to be high-affinity HLA binders, irrespective of the length of the protein. The pathogen peptides that could behave as self or partially self-peptides in the host were eliminated using a comparative study with human proteome, thus reducing the number of peptides for analysis. The structural basis for recognition of the nonamers by the respective HLA molecules thus predicted was analyzed by molecular modeling. The structural analysis showed good correlation with the binding prediction. The analysis also led to an understanding of the binding profile of the peptides with respect to different alleles of class I HLA. The predicted epitopes can be tested experimentally for their inclusion in a potential vaccine against tuberculosis that is HLA haplotype-specific.

Keywords: PE and PPE proteins; T cell epitope; HLA class I binding prediction; Molecular modeling

1. Introduction

BCG is the only currently available vaccine for prevention of tuberculosis (TB), which has exhibited considerable variations in efficacy in clinical trials in geographically distinct populations [1–6]. Although BCG prevents disseminated tuberculosis in newborns, it fails to protect against most common form of the disease, pulmonary tuberculosis in adults [7]. So is the case of natural infection with *Mycobacterium tuberculosis*, which fails to protect against reinfection. [8]. Given the global incidence of tuberculosis with ∼8 million new cases and ∼3 million deaths each year [9], it is crucial to explore newer strategies to improve BCG or to develop a more effective vaccine than *M. bovis* BCG.

M. tuberculosis H37Rv contains two large glycine-rich gene families, which together account for about 10% of the coding capacity of the genome [10]. These glycine-rich gene families code for PE and PPE proteins. The names PE and PPE are derived from the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) found in most cases near N-terminus of these glycine and alanine-rich proteins. PE family has two subfamilies, PE and PE_PGRS. All the 99 members of PE family have a highly conserved N-terminal domain of 110 amino acid residues, whereas the C-terminal domain show marked heterogeneity, showing variation in size, sequence and repeat copy numbers. The members of PE_PGRS subfamily have a polyglycine-rich sequence at the C-terminus, along

Abbreviations: PE, proline-glutamic acid motif; PPE, proline-proline-glutamic acid motif

* Corresponding author. Tel.: +91 80 2293 2703; fax: +91 80 2360 2697.
E-mail address: nayak@mcbl.iisc.ernet.in (R. Nayak).

0264-410X/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.vaccine.2004.08.046

with the conserved amino terminus. The C-terminal extension is characterized by the presence of multiple tandem repetitions of Gly–Gly–Ala or Gly–Gly–Asn encoded by PGRS motif. The PPE family consists of 68 members and has a conserved N-terminal domain of 180 amino acid residues with varying carboxy terminal domains. The polymorphism of these two gene families is the major source of variation in *M. tuberculosis* complex in an otherwise genetically homogeneous bacterium [11]. Though the sub cellular localization of these proteins is still a mystery, a few of PE_PGRS proteins have been considered as possible virulence factors in *M. marinum* [12], and some are cell surface constituents, involved in interaction of mycobacteria and macrophage [13].

These two multigene families are of potential interest from immune response point of view, since they could function as a source of antigenic variation for *M. tuberculosis* in order to evade the host immune response [10,14] and as cell surface antigens [15]. Some of PE and PPE proteins have been shown to be potent B and T cell antigens. Two proteins from PE_PGRS subfamily, Rv1759c and Rv3367 are expressed during infection and show antibody response in humans and rabbits, respectively [16–18]. Rv1196 and Rv0915c from PPE family have been shown to be good T cell antigens [19,20]. Another study has shown that the PE domain of PE_PGRS protein Rv1818c upon immunization into mice induces good cell mediated immune response, whereas the PGRS domain is responsible for good humoral response [21].

There is abundant evidence in support of an important role for CD4 T cells in controlling *Mycobacterium tuberculosis* infection. However, several lines of evidence also suggest a role for CD8 T cells in controlling *M. tuberculosis* infection in the host [22]. At the same time, MHC class I restricted CD8 T lymphocytes specific for mycobacterial antigens have been observed in mouse models of TB [23,24] as well as in humans [25–28], thus emphasizing the role of T cell mediated responses. The identification of T cell epitopes from immunologically relevant antigens therefore remains a critical step in the development of vaccines.

Relatively few epitopes in mycobacterial antigens have so far been identified for human CD8 T cells [29]. In this regard, release of genome sequences of *M. tuberculosis* has provided an opportunity to identify proteins with vaccine potential that could give immune protection in individuals with different HLA backgrounds. In an effort to identify potential T cell antigens from PE and PPE family of proteins, we have carried out a systematic in silico analysis of the 167 different PE and PPE proteins. Employing immuno informatics approach [30], a set of HLA class I binding peptides have been identified from these proteins. Further, their binding abilities have been ascertained using independent methods such as molecular modeling and structural analysis methods. This study has led to the identification of potential T cell epitopes, which can be tested experimentally for inclusion in specific vaccine for global coverage.

2. Materials and methods

2.1. Prediction of MHC class I binding epitopes

The complete amino acid sequences of proteins that have been annotated as PE (99) and PPE (68) proteins, respectively from *M. tuberculosis* were obtained from the Tuberculist database [http://genolist.pasteur.fr/TubercuList/]. All possible overlapping nonamers from these proteins were screened for their potential to bind to thirty three different alleles of HLA class I molecules using a prediction algorithm HLA-BIND [http://bimas.dcrt.nih.gov/molbio/hla_bind], which identifies and ranks nonamer peptides that contain allele specific binding motifs for class I HLA alleles measured in terms of half time of dissociation ($T_{1/2}$) of β2 micro globulin [30].

The algorithm estimates the binding against 33 HLA Class I alleles, which include nine HLA-A alleles, 20 HLA-B alleles and four HLA-C alleles. As the optimum length of the peptide binding to MHC class I is nine amino acids, all possible overlapping nonamers were first generated from PE (8655 peptides from PE and 41569 peptides from PE_PGRS subfamilies) and PPE (42691) families of proteins. The binding was estimated in terms of half time of β2 micro globulin dissociation rates [cutoff, $T_{1/2}$ of $\geq$100 min]. Only those peptides, which were predicted to bind to any of the 33 alleles studied, were picked up for further analysis.

2.2. Identification of "self" peptides

Peptides predicted to bind to HLA were checked for similarity with each of the 47523 human proteins annotated so far, by using BLAST. Each of the binding nonamer was checked with each of the 47523 ORFs. Since both PE and PPE protein sequences show significant redundancy, only unique peptides were selected for BLAST search. The BLAST variables were tailored appropriately for this analysis, in view of the short size of the peptides. The BLAST results were then parsed with a perl script [indigenously developed] to identify those peptides that exhibited [a] 100% [9 aa] identity of the epitopes from *M. tuberculosis* with peptide nonamers from human proteome [b] 90% [8 aa] and [c] 80% or 7 aa similarities.

2.3. Feasibility analysis by molecular modeling

Three-dimensional crystal structure of peptide-MHC complex for alleles A_0201 [1DUZ], B_2705 [1HSA], B_3501 [1A9E], B_5101 [1E27], and Cw_0401 [1IM9] were obtained from the Protein Data Bank [35]. Based on the prediction analysis determined earlier, putative epitopes from Rv3018c, Rv3812 and Rv1818c were chosen for molecular modeling.

The peptides with highest as well as lowest $T_{1/2}$ for each allele were modeled on to their respective structural templates [1DUZ, 1HSA, 1E27, 1A9E, 1IM9], so as to replace the original peptides present in the crystal structures, and the complexes were subjected to energy minimization.

Model building and energy minimizations were carried out using INSIGHT-II and DISCOVER modules [Accelrys Inc.]. All energy minimizations were carried out with a 13 Å nonbonded cutoff and a distance-dependent dielectric constant of 4.0, initially using the steepest algorithm followed by conjugate gradients till the root mean square [rms] derivative was less than 0.4 kcal mol^{-1} $Å^{-1}$. An identical minimization with the original peptide was also carried out as a control in every case. The binding of the peptides was then estimated by analyzing the intra-molecular hydrogen bonds, electrostatic, van der Waals and hydrophobic interactions with the protein residues in the vicinity.

3. Results

Employing the in silico method, 8655 overlapping 9mer peptides were generated from 38 proteins belonging to PE subfamily proteins and about 17% (1519) of the peptides were identified by HLA-binding predictions. Similarly, of the peptides derived from 61 PE_PGRS proteins, only 6.4% bound to one or more of HLA alleles, which indicated that the PGRS part in these proteins may be responsible for low binding. It is also likely that PE_PGRS proteins may elicit more B-cell response. This property has been reported for Rv1818c, one of the proteins of this group [21]. Proteins of PPE family exhibited a predicted average binding of 14%. Within this range of values, the binding is independent of the length of the protein.

3.1. Specificity and promiscuity of peptide-HLA binding

Thirty-three alleles belonging to HLA A, B and C loci have been tested. The prediction analysis showed that the majority of peptides bind to a single allele, and a given nonamer can bind to a maximum of four alleles, out of the thirty-three alleles tested. This finding was not restricted to one locus but include all three loci: A, B and C. Thus, in the PE subfamily, out of 1519 total binding peptides, 1133 nonamers are found to be monoallelic binders, 162 nonamers bind to two alleles, 213 bind to three alleles, and only 11 bind to four alleles. Similar is the case with PE_PGRS and PPE families of proteins. The results have been summarized in Table 1.

Table 1
Promiscuity of the peptides from PE, PE_PGRS and PPE family of proteins in binding to any of the 33 HLA class I alleles studied here

No. of peptides binding to	PE subfamily	PE_PGRS subfamily	PPE family
One allele	1133	1022	1703
Two alleles	162	248	942
Three alleles	213	485	751
Four alleles	11	1	10

The number of peptides from PE, PE_PGRS and PPE proteins predicted to bind to one allele or more than one allele of HLA. Maximum number of alleles a given peptide can bind is four. Majority of the peptides are monoallelic binders.

3.2. Identification of proteins with binding peptides for a large number of alleles

Identifying proteins with many peptide sequences binding to HLA class I molecule is important, considering the polymorphic nature of HLA and its diversity in population of different geographical regions. Therefore, a good T cell antigen should have peptides recognized by many HLA alleles. The analysis revealed that in the PE family, Rv0151c binds to the highest number of alleles [19 out of 33]; the average number of alleles for all proteins is 12–13. Similarly Rv3343 from PPE family shows good binding to 22 alleles.

3.3. A few selected alleles predominate in peptide binding

The predicted binding peptides were also analyzed to find out the probability of a given allele recognizing a large number of peptides. For all the three classes of proteins, the allele B_5102 binds largest number of nonamers out of the total binding peptides followed by B_5101, B_5103 and B_2705. Few of the 33 alleles are not predicted to bind to any of the generated peptides at an arbitrary cut off value for $T_{1/2}$ fixed at 100 min. They are A_1101, A_3101, A_3302, B_3902, Cw_0602, and Cw_0702 Fig. 1. However, these alleles could bind peptides at lower $T_{1/2}$ values (data not shown).

3.4. B locus alleles show higher affinity of binding

The $T_{1/2}$ with which a peptide binds to HLA ranges from 100 to 15,000 min. About 204 from PE_PGRS, 145 peptides from PE subfamily and 691 peptides from PPE family of proteins are high-affinity binders, which bind with a $T_{1/2}$ of $\geq$500.

In general, the binding affinity of peptides to B locus alleles [especially B_2705 and B_5101] are higher as compared to alleles of A or C locus.

3.5. Pathogen peptides that could behave as self or partially self-peptide of host

The binding peptides from all 167 proteins of PE and PPE families were analyzed for similarities with each of 47523 ORFs from human proteome. Peptides were categorized into three classes, [a] those peptides having complete similarity with human peptides [self-peptides], [b] those with seven or [c] eight amino acid similarities with human proteins [partially self-peptides] and [d] those without any significant similarity [non-self peptides]. The results of this analysis are shown in Table 2.

A large number of predicted peptides from both PE and PPE proteins were observed, for which no self-peptides exist. This finding is consistent with the uniqueness of these classes of proteins in mycobacteria. It is important

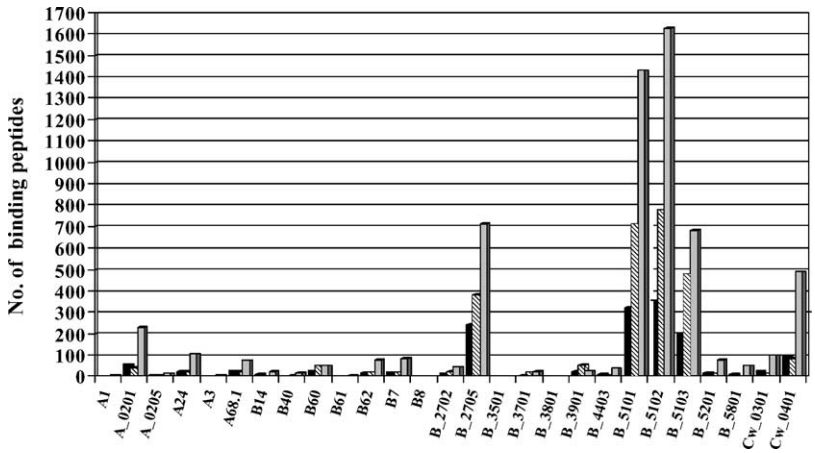

Fig. 1. Binding potential of 33 different HLA class I alleles to any of the peptides from PE and PPE proteins. Alleles from B locus are shown to bind to large number of peptides with high affinity compared with A and C loci. HLA B_5102 is binding to largest number of peptides. Some of the alleles like A_1101, A_3101, A_3302, B_3902, Cw_0602 and Cw_0702 do not show binding to any of the peptides (black bars: PE proteins, crossed bars: PE_PGRS proteins, grey bars: PPE proteins).

to identify partially self-peptides, since they could mount an autoimmune response in the host upon immunization. A nonameric peptide LRSLGATLK from Rv3477 [PE] shows complete homology with the human potassium voltage gated channel protein. Nonamers GQTGANGGR, GAGGAGGGV from Rv0278c and Rv0578c [PE_PGRS], respectively show 9 aa similarity with a peptide from lymphocyte activation-associated protein and hypothetical protein FLJ10210. AAAAAAAAV, a peptide from Rv0287 of PPE protein, is shared between 15 of human proteins. Out of the binding peptides from PE subfamily, 76 nonameric peptides are partially self to the human proteome, whereas only one of the peptides has complete homology with the human peptide. In PE_PGRS subfamily, about 247 binding peptides are partially self and two are self-peptides whereas from the PPE family, 229 and two peptides are partially self and self-peptides, respectively. These proteins containing self and partially self-peptides were removed from further analysis.

Table 2
Uniqueness/selfness of the nonameric peptides from PE and PPE family to the human proteome

No. of binding peptides	Self	8 aa similarity	7 aa similarity	Non-self
PE subfamily	1	21	55	623
PE_PGRS subfamily	2	*61*	180	739
PPE family	2	60	169	2105

The similarities of the binding peptides with each of 47,523 ORFs from human proteome. Peptides were categorized into three classes: (i) those peptides having complete similarity with human peptides (self-peptides), (ii) those with seven or (iii) eight amino acid similarities with human proteins (partially self-peptides) and (iv) those without any significant similarity (non-self peptides).

3.6. Structural basis for the recognition of nonamers by the respective HLA molecules by molecular modeling studies

The feasibility of binding of peptides to their respective class I HLA alleles was simultaneously investigated by molecular modeling and structural analysis. Superposition of the crystal structure of five alleles [HLA A_0201, B_2705, B_5101, B_3501 and Cw_0401] revealed that they are very similar with RMS [root mean square] deviations less than about 1 Å, as also in the modes with which they bind different peptides. The striking structural similarity of the overall structure of the whole α1 and α2 domains and β2 micro globulin can be easily exploited to unambiguously dock the nonamer on to the HLA molecule. At the same time, the regions at the binding sites are the most variable both in HLA molecules and in the various TCRs that recognize each one of these, which are responsible for the generation of diversity or polymorphism in a population. Samples from both high as well as low-binding peptides [12 for HLA A_0201, 13 for B_2705, 16 for B_5101, 9 for B_3501 and 16 for Cw_0401] were modeled on to their respective structural templates so as to replace the original peptides seen in the crystal structure. The analysis of several different HLA class I structures shows that there is an overall similarity in the mode of peptide binding. The modeled structure of HLA with high-affinity peptides shows more interactions in terms of hydrogen bonds when compared to peptides with low affinity (Fig. 2). This finding correlates well with the higher binding affinities of the predicted peptides. The complexes of chosen peptides indicate the feasibility of binding and to a large extent the affinities of the binding peptides for a given allele. This is illustrated by the intermolecular energies between the peptides and the alleles, and a comparison of intermolecular energies and $T_{1/2}$ values is given in

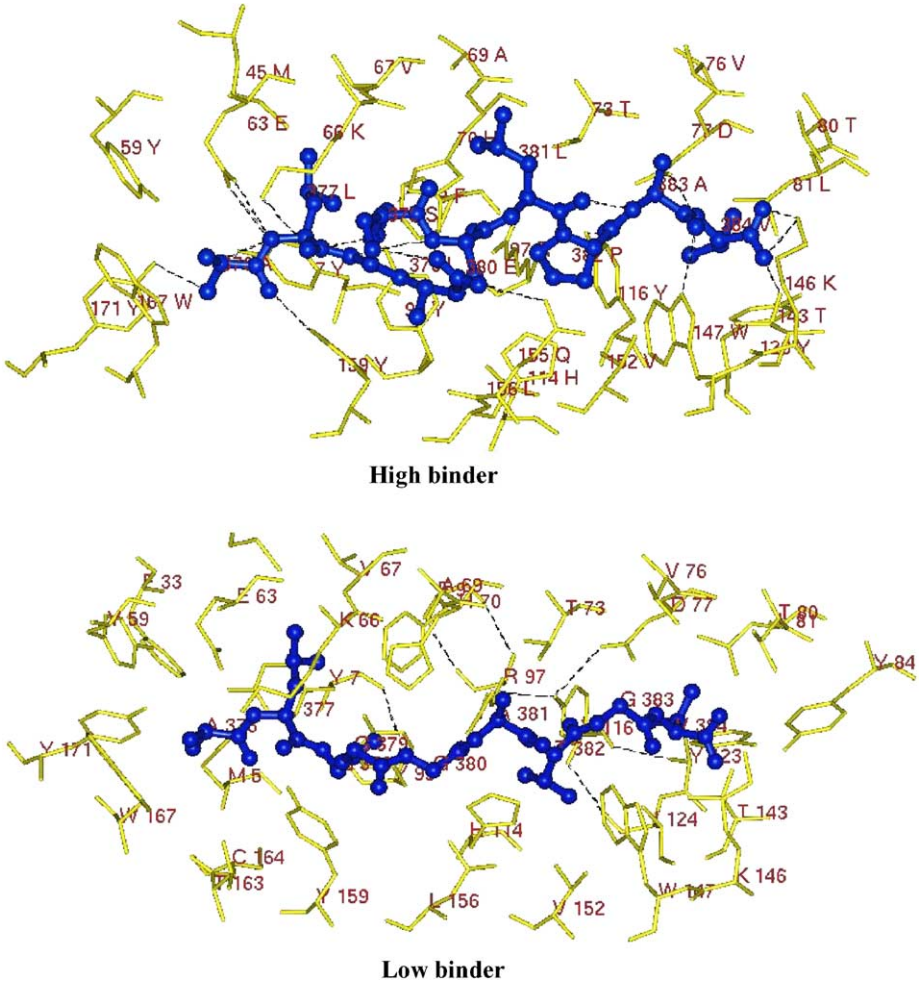

Fig. 2. Interaction between a peptide and the peptide-binding groove of HLA. Both high- and low-affinity peptides were modeled on to five different HLA templates. The figure shows an example of a high-affinity peptide-ALLSELPAV from Rv3018c displaying more hydrogen bonds as compared to a low-affinity peptide AQLLTEFAI also taken from Rv3018c.

Table 3. It can be seen that the peptides with high affinity as judged by the $T_{1/2}$ values also show high (more negative) interaction energies, for most of the peptides, thus falling in to a qualitative pattern. Such correlations were not seen in all the cases however, especially where the $T_{1/2}$ was less than 50. It must be mentioned, however, that seeking quantitative correlations between the matrices would be inappropriate, given that they both measure different parameters employing different methods, although ultimately both reflect ability to bind. Therefore, we feel that the correlation, to a large extent but not total, is not unnatural for an acceptable prediction. An examination of the data for class I binding peptides and the alleles they bind revealed that there are three promiscuous epitopes (a peptide that binds to more than one allele) for each of the proteins Rv3812 and Rv3018c (highlighted in Table 3). Thus, these proteins could now be tested experimentally for their ability to act as T cell antigens and the class I binding abilities of the peptides could also be verified.

4. Discussion

The design and development of new generation of vaccines have focused on individual proteins or DNA coding for a single or limited number of proteins. Peptide binding to HLA is essential for generation and maintenance of an immune response. Therefore, it becomes imperative that the proteins, which are identified as possible vaccine candidates, must generate peptides that are recognized by class I MHC for cytotoxic T cell response or by class II MHC to provide helper responses to both T and B cells. This becomes important, since humans can carry only a limited number of co-dominant HLA alleles in their genome, out of hundreds of polymorphic alleles that are present in the population. Therefore, a candidate vaccine must generate peptides that can bind to a wide range of HLA molecules to provide good population coverage. Without this, the protein even if it generates good primary response and memory response, will work only in a limited number of individuals. To address this problem

Table 3
Feasibility of the predicted peptides to bind to their respective class I HLA alleles

Allele	Gene no. (Rv no.)	aa Start position	Peptide	$T_{1/2}$	Intermolecular energy (Kcal/mol)
A_0201	1818	6	TIPEALAAV	90	−392
		385	ALGGGATGV	70	−267
		10	ALAAVATDL	21	−210
	3812	260	NLLVTGFDT	160	−213
		235	*LQLAFQQLL*	52	−216
		399	ALIDAPAHA	20	−355
	3018	219	ALLSELPAV	592	−361
		285	WVIGNLFGV	373	−320
		254	*AQLLTEFAI*	23	−266
B_2705	1818	114	GRPLIGNGA	200	−456
		70	ALFHEQFVR	75	−229
	3812	56	GRYGQEFQT	1000	−394
		235	*LQLAFQQLL*	200	−185
		479	*TPFMGMAPL*	50	−209
		158	WQQIAAALA	20	−247
	3018	145	RMWVQAATV	225	−335
		254	*AQLLTEFAI*	60	−159
		288	GNLFGVVPL	20	−154
B_5101	1818	110	LALLGRPLI	286	−231
		96	APLEGVLDV	242	−257
		275	GAGGNGGLL	55	−147
		336	AGGSGGSAL	22	−124
	3812	369	APFASLNAI	968	−226
		479	*TPFMGMAPL*	220	−137
		26	LGKAMTNLL	26	−260
		116	*APGGAYGQL*	121	−152
	3018	361	EPAPASTSV	440	−174
		9	*SPPEVHSAL*	121	−181
		314	AGLAGVAGL	22	−120
B_3501	3812	313	MPPSILRDM	40	−244
		479	TPFMGMAPL	20	−178
		116	*APGGAYGQL*	20	−152
	3018	*9*	*SPPEVHSAL*	40	−196
		197	*FPWHEIVQF*	30	−263
		414	LPGSWGPDL	20	−216
Cw_0401	1818	54	LFSGHAQAY	28	−191
		96	APLEGVLDV	24	−388
	3812	246	DYNAAVANL	528	−425
		490	NYIPQQLAL	240	−371
		479	*TPFMGMAPL*	115	−373
		116	*APGGAYGQL*	88	−345
		369	APFASLNAI	20	−167
	3018	212	AYDQYLSAL	600	−521
		71	AYVPYVAWL	576	−348
		9	*SPPEVHSAL*	96	−495
		197	*FPWHEIVQF*	80	−337
		146	MWVQAATVM	20	−209

This table depicts correlation of the binding energies (Kcal/mol) for the peptides chosen with the corresponding $T_{1/2}$ values as predicted by BIMAS algorithm. High as well as low-affinity peptides for each of five alleles chosen were modeled on to their respective templates and intermolecular energies have been calculated. Highlighted peptides (in italics) are the promiscuous epitopes binding to more than one allele.

of selection of appropriate proteins as candidate vaccines, we have carried out a systematic analysis of mycobacterial peptides derived from PE and PPE proteins, determination of their HLA binding specificity to narrow down the number of proteins containing sufficient peptides for total population coverage.

Several computational methods are now available for analysis of binding of peptides to MHC [29]. We have used the BIMAS binding tool available in public domain to study the binding of peptides to thirty-three HLA alleles. This method has been developed using experimental data of half time of dissociation of β2 micro globulin from peptide–HLA complex since β2 micro globulin dissociates from the complex when the peptide is released [30]. A $T_{1/2}$ of 100 min was chosen as a cut off point in order to select the relatively higher affinity binding peptides. It is possible that some good T cell

epitopes may be missed in the analysis, which bind with HLA with relatively less affinity and may still bind TCR with high affinity. It is the TCR–HLA–peptide complex, which is crucial in determining the T cell response.

The nonameric sequences from PE and PPE families of proteins were predicted to contain high percentage of binding peptides to human class I HLA, whereas PE_PGRS proteins show relatively low level of binding. This difference is seen in spite of PE and PE_PGRS being subfamilies of the same family, PE. Previous study has shown that PE domain of PE_PGRS but not the full length of the protein induces good cell mediated immune response, whereas the PGRS domain is responsible for good antibody response [21]. Most of the peptides were predicted to bind to alleles from B locus, and the affinity with which the peptides are binding to this locus is far higher than the alleles of A and C loci. This implies that the distribution of B loci alleles in the population may play a role in the susceptibility of the population for certain infection.

Many of the peptides are monoallelic binders i.e., they bind to a single allele. The T cell epitopes that are recognized in context of more than one HLA and more than one T cell clone are called promiscuous epitopes. Promiscuous peptides are of prime interest for vaccine design because of their relevance to higher proportions of human population. This in silico approach would help to predict some of the HLA-binding motifs, which could act as promiscuous epitopes.

There are a large number of binders from both PE and PPE proteins, for which either self or partially self-host peptides exist. This finding implicates the possible failure of certain epitopes as vaccine candidates. Partially self-peptides can mount an autoimmune response in the host upon immunization, whereas the inclusion of self-peptides has no obvious advantage. Employing the known crystal structures of HLA–peptide complexes, the ability of the peptides identified as high binders to replace the resident peptide in the crystal structure was assessed to provide additional evidence for the choice of epitopes.

A large number of workers have been addressing the identification of T cell epitopes of individual mycobacterial proteins [31]. For some proteins, the alleles and the corresponding peptides have been identified [32]. Computational methods have been used to identify peptides from proteins, which would bind to HLA [29,32]. The present work differs from these earlier investigations in that a genome-based approach has been followed and all overlapping peptides of a major class of *M. tuberculosis* protein have been analyzed to predict possible T cell reactive peptides. The epitope prediction should be followed by experimental verification of the binding of synthetic peptides with HLA in order to reduce the number of potential candidate peptides for the final analysis of recognition by peptide specific T cells. Binding of synthetic peptides to MHC molecules is usually studied by either reconstitution of MHC–peptide complexes on the cell surface [33] or stabilization of MHC–peptide complexes on TAP-deficient cells [34]. The predicted epitopes are experimentally verified by identification of T cells, which specifically recognize the naturally processed epitope in a HLA-restricted fashion. T cells are either obtained by in vivo or in vitro priming of human PBMCs or mouse splenocytes. T cell responses after stimulation can be measured by lytic activity in vitro proliferation and cytokine production [29].

In summary, we have generated a large dataset on possible HLA class I-binding peptides, which have been predicted to bind to different HLA alleles. Seventy-one high- as well as low-affinity peptides from both PE and PPE proteins have been analyzed for structural compatibility with crystal structures of HLA in terms of intermolecular energies and were found to correlate well with the corresponding affinities predicted by the BIMAS algorithm. Most of the peptides binding to HLA are specific with very few promiscuous binders. It is thus obvious that a large cocktail of proteins are required to achieve reasonable population coverage. Besides, this work suggests the feasibility of designing haplotype specific subunit vaccine, which can be given to individuals with known HLA haplotype. The haplotype specific vaccines can be combined to target a population where the distribution of HLA alleles is known. This work also indicates that use of single or limited number of genes in a DNA vaccine may not be suitable to cover a given population.

Acknowledgements

We thank the Indian Council of Medical Research for financial assistance.

References

[1] Roche PW, Triccas JA, Winter N. BCG vaccination against tuberculosis: past disappointments and future hopes. Trends Microbiol 1995;3:397–401.
[2] Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. Lancet 1995;346:1339–45.
[3] Ten Dam HG. Research on BCG vaccination. Adv Tuberc Res 1984;21:79–106.
[4] Wilson ME, Fineberg HV, Colditz GA. Geographic latitude and the efficacy of Bacillus Calmette-Guerin vaccine. Clin Infect Dis 1995;20:982–91.
[5] Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis-meta-analysis of the published literature. JAMA 1994;271:698–702.
[6] Kaufmann SHE. Is the development of new tuberculosis vaccine possible. Nat Med 2000;6:955–60.
[7] Sepulveda RL, Parcha C, Sorensen RU. Case-control study of the efficacy of BCG immunization against pulmonary tuberculosis in young adults in Santiago. Chile Tuberc Lung Dis 1992;73:327–72.
[8] Van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, et al. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med 1999;341:1174–9.
[9] Bloom BR, Murray CJL. Tuberculosis: commentary on a re-emergent killer. Science 1992;257:1055–64.

[10] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–44.
[11] Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci USA 1997;94:9869–74.
[12] Ramakrishnan L, Federspiel NA, Falkow S. Granuloma-specific expression of Mycobacterium virulence proteins from the glycine-rich PE-PGRS family. Science 2000;288:1436–9.
[13] Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, et al. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immunol 2001;69:7326–33.
[14] Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tuberc Lung Dis 1999;79:329–42.
[15] Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol Microbiol 2002;44:9–19.
[16] Abou-Zeid C, Garbe T, Lathigra R, Wiker HG, Harboe M, Rook GA, et al. Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. Infect Immunol 1991;59:2712–8.
[17] Espitia C, Laclette JP, Mondragon-Palomino M, Amador A, Campuzano J, Martens A, et al. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins. Microbiology 1999;145:3487–95.
[18] Singh KK, Zhang X, Patibandla AS, Chien PJr, Laal S. Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect Immunol 2001;69:4185–91.
[19] Dillon DC, Alderson MR, Day CH, Lewinsohn DM, Coler R, Bement T, et al. Molecular characterization and human T cell responses to a member of a novel *Mycobacterium tuberculosis* mtb39 gene family. Infect Immunol 1999;67:2941–50.
[20] Skeiky YA, Ovendale PJ, Jen S, Alderson MR, Dillon DC, Smith S, et al. T cell expression cloning of a *Mycobacterium tuberculosis* gene encoding a protective antigen associated with the early control of infection. J Immunol 2000;165:7140–9.
[21] Delogu G, Brennan MJ. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. Infect Immunol 2001;69:5606–11.
[22] Kaufmann SHE, Ladel ChH. Role of T cell subsets in immunity against intracellular bacteria: experimental infections of knockout mice with Listeria monocytogenes and Mycobacterium bovis BCG. Immunobiol 1994;191:509–19.
[23] Flynn JL, Goldstein MM, Triebold KJ, Koller B, Bloom BR. Major histocompatibility complex class I-restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection. Proc Natl Acad Sci USA 1992;89:12013–7.
[24] Ladel CH, Daugelat S, Kaufmann SH. Immune response to Mycobacterium bovis bacille Calmette Guerin infection in major histocompatibility complex class I- and II-deficient knock-out mice: contribution of CD4 and CD8 T cells to acquired resistance. Eur J Immunol 1995;25:377–84.
[25] Pathan AA, Wilkinson KA, Wilkinson RJ, Latif M, McShane H, Pasvol G, et al. High frequencies of circulating IFN-gamma-secreting CD8 cytotoxic T cells specific for a novel MHC class I-restricted *Mycobacterium tuberculosis* epitope in *M. tuberculosis*-infected subjects without disease. Eur J Immunol 2000;30:2713–21.
[26] Lalvani A, Brookes R, Wilkinson RJ, Malin AS, Pathan AA, Andersen P, et al. Human cytolytic and interferon gamma-secreting $CD8^+$ T lymphocytes specific for *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 1998;95:270–5.
[27] Lewinsohn DM, Alderson MR, Briden AL, Riddell SR, Reed SG, Grabstein KH. Characterization of human $CD8^+$ T cells reactive with *Mycobacterium tuberculosis*-infected antigen-presenting cells. J Exp Med 1998;187:1633–40.
[28] Canaday DH, Ziebold C, Noss EH, Chervenak KA, Harding CV, Boom WH. Activation of human $CD8^+$ alpha beta TCR^+ cells by *Mycobacterium tuberculosis* via an alternate class I MHC antigen-processing pathway. J Immunol 1999;162:372–9.
[29] Martin W, Sbai H, De Groot AS. Bioinformatics tools for identifying class I-restricted epitopes. Methods 2003;29:289–98.
[30] Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2-binding peptides based on independent binding of individual peptide side-chains. J Immunol 1994;152:163–75.
[31] Anderson P. TB vaccines: progress and problems. Trends Immunol 2001;22:160–8.
[32] Flower DR. Towards in silico prediction of immunogenic epitopes. Trends Immunol 2003;24:667–74.
[33] Sette A, Sidney J, del Guercio MF, Southwood S, Ruppert J, Dahlberg C, et al. Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. Mol Immunol 1994;31:813–22.
[34] Stuber G, Modrow S, Hoglund P, Franksson L, Elvin J, Wolf H, et al. Assessment of major histocompatibility complex class I interaction with Epstein-Barr virus and human immunodeficiency virus peptides by elevation of membrane H-2 and HLA in peptide loading-deficient cells. Eur J Immunol 1992;22:2697–703.
[35] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nuc Acids Res 2000;28:235–42.