

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

Potential T cell epitopes of *Mycobacterium tuberculosis* that can instigate molecular mimicry against host: implications in autoimmune pathogenesis

BMC Immunology 2012, **13**:13 doi:10.1186/1471-2172-13-13

Sathi Babu Chodiseti (sathibabu@imtech.res.in)
Pradeep K Rai (pradeeprai@imtech.res.in)
Uthaman Gowthaman (gowthaman.uthaman@yale.edu)
Susanta Pahari (susanta86@imtech.res.in)
Javed N Agrewala (javed@imtech.res.in)

ISSN 1471-2172

Article type Research article

Submission date 1 October 2011

Acceptance date 21 March 2012

Publication date 21 March 2012

Article URL <http://www.biomedcentral.com/1471-2172/13/13>

Like all articles in BMC journals, this peer-reviewed article was published immediately upon acceptance. It can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to

<http://www.biomedcentral.com/info/authors/>

**Potential T cell epitopes of *Mycobacterium tuberculosis* that can instigate
molecular mimicry against host: implications in autoimmune
pathogenesis**

**Sathi Babu Chodiseti^{1,†}, Pradeep K Rai^{1,†}, Uthaman Gowthaman^{1,†}, Susanta Pahari¹,
Javed N Agrewala^{1*}**

¹Immunology Laboratory, CSIR-Institute of Microbial Technology, Chandigarh-160036,
India.

Email:

Sathi Babu Chodiseti - sathibabu@imtech.res.in

Pradeep K Rai – pradeeprai@imtech.res.in

Uthaman Gowthaman – gowthaman.uthaman@yale.edu

Susanta Pahari - susanta86@imtech.res.in

Javed N Agrewala* – javed@imtech.res.in

[†] **Equal contributions**

***Address correspondence:** Dr. Javed N. Agrewala, Institute of Microbial Technology,
Sector 39A, Chandigarh-160036, India. E-mail: javed@imtech.res.in

ABSTRACT

Background: Molecular mimicry between microbial antigens and host-proteins is one of the etiological enigmas for the occurrence of autoimmune diseases. T cells that recognize cross-reactive epitopes may trigger autoimmune reactions. Intriguingly, autoimmune diseases have been reported to be prevalent in tuberculosis endemic populations. Further, association of *Mycobacterium tuberculosis* (*M. tuberculosis*) has been implicated in different autoimmune diseases, including rheumatoid arthritis and multiple sclerosis. Although, *in silico* analyses have identified a number of *M. tuberculosis* specific vaccine candidates, the analysis on prospective cross-reactive epitopes, that may elicit autoimmune response, has not been yet attempted. Here, we have employed bioinformatics tools to determine T cell epitopes of homologous antigenic regions between *M. tuberculosis* and human proteomes.

Results: Employing bioinformatics tools, we have identified potentially cross-reactive T cell epitopes restricted to predominant class I and II alleles of human leukocyte antigens (HLA). These are similar to peptides of mycobacterial proteins and considerable numbers of them are promiscuous. Some of the identified antigens corroborated with established autoimmune diseases linked with mycobacterial infection.

Conclusions: The present study reveals many target proteins and their putative T cell epitopes that might have significant application in understanding the molecular basis of possible T cell autoimmune reactions during *M. tuberculosis* infections.

BACKGROUND

Although, the immune system efficiently discriminates between self and non-self, the occurrence of autoimmune diseases is a testimony to the fact that such discrimination may be imprecise [1]. Understanding the etiology of autoimmune diseases has been a great challenge to immunologists. The existence of central tolerance mechanism ensures the clonal deletion of autoreactive T cells and B cells. Nonetheless, there are ample evidences signifying that a considerable number of such cells can escape these “failsafe” mechanisms [1, 2]. Immunological insults like exposure to pathogenic bacteria, viruses, aberrant expression of self proteins and exposure to cryptic antigens, etc., have been implicated to trigger and amplify the immune reactions that culminate into autoimmune diseases [3-5]. Antigenic determinants/epitopes present in pathogens, which resemble the host proteins, can potentially be a threat in activating the cells of immune system, resulting in autoimmunity [3, 4]. This resemblance is popularly termed as molecular mimicry.

Many different autoimmune diseases have been hypothesized to be a result of this mistaken identity. As a result of molecular mimicry, the immune cells attack the host tissues [3, 5]. The sharing of similar epitopes between the host and the pathogens may instigate autoaggression by stirring autoreactive T cells and B cells. Usually, autoreactive T cells are quiescent in the periphery, since they may recognize cryptic or low affinity epitopes. Pathogenic organisms express pathogen associated molecular patterns (PAMPs) that are perceived by the immune system as “danger signals” through Toll Like Receptors (TLRs) [6]. Hence, the “TLR licensed” antigen presenting cells (APCs) can potentially activate the self-reactive T cells, since they present antigens along with inflammatory signals.

Antigenic presentation in such a context may result in high avidity interactions between autoreactive T cells and the APCs that eventually break tolerance [6]. Antigens like the pulD protein from *Klebsiella sp.*, nuclear antigen-1 from Epstein-Barr virus and OSP-A from *Borrelia sp.* have been associated with diseases like ankylosing spondylitis, systemic lupus erythematosus (SLE) and Lyme arthritis, respectively [7-9]. Importantly, T cells play a pivotal role in autoimmune reactions, since they may directly attack the host tissues or help B cells to produce autoantibodies [10]. Molecular mimicry has been demonstrated in T cell specific autoimmune diseases such as multiple sclerosis (MS), myocarditis, diabetes, etc. One of the early, classic studies by Strominger's group showed that the T cells reacting to immunodominant peptide of myelin basic protein (MBP) could cross-react with viral antigens [3].

M. tuberculosis infects about two million people annually. In TB-endemic areas, it is estimated that almost one-third of the population is infected with *M. tuberculosis* [11]. Interestingly, an abundant presence of autoimmune diseases has been reported in these populations [12, 13]. TB has been associated with many different autoimmune diseases like SLE, rheumatoid arthritis (RA), MS, etc [4, 13-21]. There are ample evidences to suggest that TB reactive T cells can potentially recognize self antigens [14, 17-19, 21]. This has been demonstrated in animal models and in TB affected individuals. For example, T cells responding to the 65 kDa antigen of *M. tuberculosis* have been shown to be present in the synovia of arthritis patients [14, 17]. Hence, during a chronic state of disease, T cells that cross-react with mycobacterial and self antigens are activated, leading to detrimental autoimmune responses. Identification of such cross-reactive epitopes may be of immense

scope in understanding the pathogenesis of autoimmunity. In this era of informatics, *in silico* analyses have identified a number of *M. tuberculosis* specific T cell epitopes that could be potentially used as vaccines [22-25]. However, it warrants the information on prospective cross-reactive epitopes that may elicit autoimmune responses. Cytotoxic CD8 T cells and helper CD4 T cells recognize peptides in the context of HLA class I and class II molecules, respectively. Both the subsets have been implicated in mediating autoimmune responses. Here, we have used bioinformatics tools to identify *M. tuberculosis* and human cross-reactive T cell epitopes, restricted to predominant HLA class I and class II alleles [26, 27]. Interestingly, we could identify several epitopes exhibiting similarity between human and *M. tuberculosis* proteins that may be molecular triggers of autoimmunity.

METHODS

Alleles used in the study

Predominantly occurring MHC (major histocompatibility complex) alleles in human population for HLA class I (A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, B*07:02, B*08:01) and HLA class II (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:02, DRB1*11:01, DRB1*13:02, DRB1*15:01) were chosen for the study [26-30].

Programs and databases

Netmhc2.2: NetMHC 2.2 server predicts binding of peptides to various human HLA class II alleles using artificial neural networks (ANNs) [30].

Netmhc3.0: NetMHC 3.0 server predicts binding of peptides to a number of different HLA class I alleles using artificial neural networks and weight matrices [30].

HAMAP: HAMAP (High-quality Automated and Manual Annotation of microbial Proteomes) automatically annotates a significant percentage of proteins originating from microbial genome sequencing projects [31]. HAMAP uses annotation templates for protein families to propagate annotations to all members of manually defined protein families.

Expasy: Expasy (Expert Protein Analysis System), is a proteomics server and allows browsing through a number of data bases as well as other cross-referenced ones [32]. It also allows access to many analytical tools for the identification of proteins, analysis of their sequence and the prediction of tertiary structure.

UniProt: UniProt is the world's most comprehensive catalogue of information on proteins [33]. It is a central repository of protein sequences and functions created by joining the information contained in UniProt/Swiss-Prot/TrEMBL. The UniProt knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation.

Identification and analysis of homologous *M. tuberculosis* peptides in humans

Sequences of curated proteins of *M. tuberculosis* were obtained from UNIPROT database employing HAMAP search. A total of 444 well characterized proteins of *M. tuberculosis* H37Rv were selected from the database. The four classifications of the proteins of *M. tuberculosis*, namely structural, secretory, antigenic and metabolic have been directly adopted from UNIPROT database. The proteins chosen were compared for similarity with the human proteome using BLAST program [34] from ExPASy server. Based on the BLAST results, regions of nine or more amino acids (small peptides) that were similar between the human and *M. tuberculosis* proteins were selected for further analysis. The

selected human peptides were assessed for binding to different predominant HLA class I and class II alleles by using the NetMHC server. IC₅₀ values were selected based on the binding scores of peptide core regions (9 amino acids length) to each allele. The peptides were classified based on predicted IC₅₀ values as strong binders (IC₅₀ ≤ 500), weak binders (500 ≤ IC₅₀ ≤ 5000) and non-binders (IC₅₀ ≥ 5000) [35-37]. The results were then analyzed by considering the binding affinity of peptides to HLA alleles, nature of antigens, allelic associations of autoimmune diseases and tuberculosis.

RESULTS

Mycobacterial proteome contains T cell epitopes that cross-react with human proteins. The analysis of the data obtained from the search between *M. tuberculosis* and human proteomes revealed considerable similarities in sequences. Many regions, that were part of metabolic proteins, were homologous (Additional file 1). A total of 460 similar regions of 9 or more amino acids were identified. We then utilized *in silico* tools to analyze the ability of these peptides to bind to different HLA class I and class II alleles. Any such common peptide that had a binding score of IC₅₀ < 5000 was considered as a HLA-binder and thus was likely to be a T cell epitope. Further, peptides having an IC₅₀ of ≤ 500 were considered to be strong binders [35-37]. We found considerable number of peptides (epitopes henceforth) binding to HLA class I and class II alleles (Figure 1). Our analysis revealed that the greatest number of CD4 T cell epitopes were restricted to DRB1*01:01 (291) followed by DRB1*07:01 (229) and DRB1*15:01 (191) (Figure 1). Intriguingly, DRB1*15:01 is also a susceptibility allele for tuberculosis [38, 39]. The analysis of CD8

epitopes indicated maximum (84) number of peptides binding to allele A*02:01 followed by 81 for A*11:01 and 62 for A*03:01 (Figure 1).

The mycobacterial proteins for this study were classified into four different categories; antigenic, secretory, structural and metabolic proteins. The HLA binding analysis showed that the largest fraction of CD4 T cell epitopes was from metabolic proteins (Figure 2 and 3). For almost all the alleles analyzed in the current study, metabolic peptides represented >80% of the total CD4 T cell epitopes obtained for that particular HLA allele. However, the absolute number of epitopes restricted to each allele varied greatly (Figure 2). The highest number of epitopes homologous to antigenic, structural and metabolic proteins were restricted to DRB1*01:01 (Figure 2). Interestingly, although most of the alleles had equivalent fractions of CD4 T cell epitopes from metabolic proteins, the relative representation of other categories of epitopes varied considerably (Figure 2 and 3). Highest fractions of antigenic, structural and secretory epitopes were noted for DRB1*03:01, DRB1*01:01, and DRB1*13:02, respectively (Figure 3). Next, we analyzed the fraction of strong binders among the total epitopes belonging to the different categories of proteins. We considered peptides with predicted IC_{50} value ≤ 500 as strong binders. Although, a majority of the homologous peptides were weak binders, a fraction of about 20%-40% were of higher affinities (Figure 4). We next analyzed for the presence of promiscuous epitopes among the HLA binding peptides. A peptide that could bind three or more HLA alleles was considered promiscuous. We could identify a total of 242 promiscuous peptides restricted to HLA class II molecules (Additional file 1). Interestingly, most of the promiscuous peptides (200) were from the metabolic proteins.

One of the most striking observations was that the total number of peptides binding to HLA class I were considerably lesser than HLA class II alleles (Figure 1). Nevertheless, the homologous peptides screened from the different classes of proteins (metabolic, antigenic, structural and secretory), were abundantly restricted to certain HLA alleles (Figure 5). For instance, allele A*02:01 bound to the highest number of antigenic (eight) and metabolic epitopes (70). Further, A*11:01 and B*07:02 indicated binding to highest numbers (21 and 19) of peptides from structural class of antigens (Figure 5). None of the HLA alleles bound to the homologous epitopes were from secretory antigens. Among the various HLA class I alleles examined, least number of CD8 epitopes exhibited binding to A*01:01 (Figure 1). Interestingly, similar to CD4 T cell epitopes, most of the peptides that were restricted to HLA class I molecules were from the metabolic proteins of *M. tuberculosis* (Figure 5 and 6). However, there was profound difference in the relative representative fractions of peptides from different classes of proteins. Peptides from all the three different classes of proteins bound to majority of the tested HLA class I or II alleles. Nevertheless, just two classes of peptides displayed binding to majority of HLA class I alleles (Figure 6). Further, we analyzed the fraction of strong binders among the total peptides for each class of proteins (Figure 7). Similar to HLA class II molecules, majority of the HLA class I alleles showed weak binding, with the exception of antigenic peptides for A*11:01 and structural peptides for A*03:01. Next we analyzed the occurrence of promiscuous CD8 T cell peptides among the autoantigenic epitopes similar to *M. tuberculosis* proteins (Additional file 2). Our analysis for promiscuous epitopes revealed that although there were fewer

promiscuous CD8 T cell epitopes (44), compared to CD4 T cell epitopes (242), most of them were from the metabolic proteins (26) (Additional file 1, 2).

Having recognized the possible autoreactive CD4 and CD8 T cell epitopes, we next addressed whether the host antigens containing cross-reactive epitopes had any known associations in autoimmune reactions. We identified that some of the host proteins (sharing similarity with *M. tuberculosis* proteins) were known to be implicated in diseases like MS, RA, Stiff-man syndrome, etc [4, 19, 40]. MS is an autoimmune demyelinating disorder that severely compromises patients. The precise etiology for the onset of the disease remains obscure. One of the hypotheses that explain the initiation of the disease is the occurrence of antecedent infections that ultimately lead to the pathogenesis of MS [4, 41]. Intriguingly, individuals positive for HLA-DRB1*15:01 are known to be susceptible to both MS and tuberculosis [39, 41]. One of the earlier observations suggested the presence of T cells in MS patients that cross-react to mycobacterial heat shock proteins (HSP) [18, 19, 21, 41]. Our analysis showed several peptides from mycobacterial HSP60 (also known as HSP65) homologous to human HSP60 and its relatives, to bind to many different alleles (Table 1). One particular peptide, KPLVIIAEDVDGEALSTLVLN, promiscuously bound to many alleles including HLA-DRB1*15:01 with high affinity. This is suggestive of the fact that such cross-reactive epitopes may initiate the pathogenesis of MS. Another autoimmune disease that has been frequently associated with the occurrence of tuberculosis is RA [14, 15, 17]. Intriguingly, HSP60 of *M. tuberculosis* is also implicated in the pathogenesis of RA [14, 15, 17]. Our analysis also concurs with this hypothesis (Table 1). There are reports that suggest the presence of HSP60 reactive T cell clones which can also be reactive to the

cartilage in the synovia of RA patients [14]. Hence, HSP60 cross reactive epitopes may play a significant role in the etiology of RA and MS. Further, we could also identify many T cell epitopes from different antigens like gephyrin, triosephosphate isomerase, etc., which are associated with Stiff-man syndrome, arthritis, etc. The putative autoimmune epitopes with the restricting alleles for known autoantigens that shared sequence similarity with *M. tuberculosis* proteins are listed in Table 1. The results discussed in the current section (Table 1) were solely derived from proteins known to be linked with autoimmune diseases [4, 14-18, 20]. However, other host antigens, whose putative T cell autoreactive epitopes are identified here, may be involved in the etiology of many other autoimmune diseases that need to be clinically verified.

DISSCUSSION

Molecular mimicry between antigenic determinants present in pathogenic organisms and host proteins could potentially trigger autoimmune reactions [42]. Interestingly, many mycobacterial antigens have been associated with autoimmune diseases [14, 16, 17]. This prompted us to investigate the occurrence of peptides in mycobacteria that share sequence similarity with human antigens; and identify T cell epitopes that probably would be responsible for autoimmunity.

Although, *in silico* tools have been used in the past to examine molecular mimics in other diseases [5]; the knowledge of such epitopes from mycobacteria still needs to be explored. Here, utilizing *in silico* methods, we have identified potential autoreactive CD4 and CD8 T cell epitopes that may act as molecular mimics and result in autoimmune response during

M. tuberculosis infection. The following major findings have emerged from the present study: (i) there is an extensive number of potentially autoreactive CD4 and CD8 T cell epitopes that are similar to peptides of mycobacterial antigens; (ii) the majority of such epitopes are similar to the antigens from the metabolic proteins of mycobacteria; (iii) a considerable number of promiscuous CD4 T cell epitopes could be detected; (iv) some of the identified antigens were corroborated with established autoimmune diseases linked with mycobacterial infection, thus validating the approach. We believe that this study would be a suggestive starting point for future investigation that whether mycobacterial infections and molecular mimics may elicit T cell autoimmune reactions.

Autoimmune reactions occur as a consequence of the breakdown of self-tolerance. Even though the immune system has central and peripheral tolerance mechanisms to deter the presence of autoreactive T cells, the very occurrence of autoimmune diseases signifies that this may not totally eliminate the presence or detrimental activity of host reactive T cells [6]. Autoimmune diseases develop as a result of multifactorial influences like genetic, hormonal, and environmental factors [43, 44]. One of these key elements that substantially influence the development of autoimmunity is the occurrence of antecedent infections [45]. Almost every autoimmune disease investigated is assumed to be linked to one or more such infections. One of the classical evidences arguing for this hypothesis is the autoimmune acute rheumatic fever, which is associated with the infection with *Streptococcus pyogenes*. Molecular resemblance between the bacterial M-protein and human glycoproteins results in a breakdown of self-tolerance [46]. The molecular mimicry hypothesis proposes that shared epitopes between the host and pathogen can break tolerance and elicit autoreactivity. The

degeneracy of antigen recognition by the T cell receptor may also help in such cross-reactivity [47]. Similarly, pathogen specific antibodies also can cross-react with host proteins [47]. Hence, molecular mimicry and consequent epitope spreading is now a generally accepted phenomenon influencing autoimmune reactions [3, 42, 45, 47]. The idea of molecular mimicry was strongly put forward by Fujinami and Oldstone, where they argued that molecular mimicry could contribute pathogenesis of MS [48]. The criteria for this mechanism includes that the pathogen must be associated with the onset of the autoimmune reactions, the antigens from the pathogen must provoke an immune response that cross-reacts with host proteins. Further, the cross-reactive epitopes should induce disease, if tested in an animal model [45]. There are many reports that act as evidences to satisfy each of these criteria [2, 7, 44, 45].

The presentation of certain antigenic epitopes (that are mimics of host antigens) to T cells by the pathogen encountered “TLR licensed” APCs may initiate the autoreactive responses [6]. Thus, homologous antigens from pathogens can potentially “revive” the otherwise non-responding autoreactive T cells. The inflammatory cytokines present during such priming may also imprint tissue migratory properties. When such autoreactive T cells come across the “cognate” host antigens, they will destruct tissues by their effector mechanisms. Hence, infection not only activates the autoreactive T cells but also may empower them to migrate to distant tissues thus initiating a process that ultimately escalates in to a full-bloomed disease. Tuberculosis has been associated with autoimmune reactions [13, 16]. Classical studies have demonstrated the occurrence of mycobacterium reactive T cells that cross-react to antigens associated with MS, RA, etc [18-21]. Interestingly, in many cases the

antigen was found to be HSP60 [14, 18, 20]. The present study also corroborates with these findings. In addition, our extensive comparative analysis of the proteomes of *M. tuberculosis* and humans followed by T cell epitope identification has revealed many more such possible target proteins and their putative epitopes. In the present scenario, where identifying the etiology of autoimmune diseases remains a great challenge, we believe that the outcome of the present study would open up extensive future investigations into molecular basis of possible T cell autoimmune reactions during mycobacterial infections.

CONCLUSIONS

In essence, this study indicates the existence of considerable number of potential cross-reactive T cell epitopes between *M. tuberculosis* and the human proteome, which may elicit molecular mimicry and result in autoimmune responses during *M. tuberculosis* infection. Some of the epitopes were promiscuously binding to predominantly occurring HLA alleles and corroborated well with established autoimmune diseases. The identified target proteins and their putative T cell epitopes may have significant implications that will open up extensive investigations in understanding the molecular basis of autoimmune reactions during *M. tuberculosis* infection.

COMPETING INTERESTS: The authors have no financial conflict of interest.

AUTHOR'S CONTRIBUTIONS

JA and UG designed the study; SBC, PKR and SP performed the experiments; UG and SBC analyzed the data; JA, UG and SBC wrote the manuscript. All the authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank Council of Scientific and Industrial Research (CSIR) and Department of Biotechnology (DBT), India, for financial support. SBC and PKR are recipients of fellowship of CSIR, and UG of DBT.

REFERENCES

1. Mueller DL: **Mechanisms maintaining peripheral tolerance.** *Nat Immunol* 2010, **11**(1):21-27.
2. Bach JF: **Infections and autoimmune diseases.** *J Autoimmun* 2005, **25 Suppl**:74-80.
3. Wucherpfennig KW, Strominger JL: **Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein.** *Cell* 1995, **80**(5):695-705.
4. Birnbaum G, Kotilinek L: **Heat shock or stress proteins and their role as autoantigens in multiple sclerosis.** *Ann N Y Acad Sci* 1997, **835**:157-167.
5. Kovvali G, Das KM: **Molecular mimicry may contribute to pathogenesis of ulcerative colitis.** *FEBS Lett* 2005, **579**(11):2261-2266.
6. Goverman J: **Autoimmune T cell responses in the central nervous system.** *Nat Rev Immunol* 2009, **9**(6):393-407.
7. Benoist C, Mathis D: **Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry?** *Nat Immunol* 2001, **2**(9):797-801.
8. Fielder M, Pirt SJ, Tarpey I, Wilson C, Cunningham P, Ettelaie C, Binder A, Bansal S, Ebringer A: **Molecular mimicry and ankylosing spondylitis: possible role of a novel sequence in pullulanase of *Klebsiella pneumoniae*.** *FEBS Lett* 1995, **369**(2-3):243-248.
9. McClain MT, Heinlen LD, Dennis GJ, Roebuck J, Harley JB, James JA: **Early events in lupus humoral autoimmunity suggest initiation through molecular mimicry.** *Nat Med* 2005, **11**(1):85-89.
10. Gowthaman U, Chodisetti SB, Agrewala JN: **T cell help to B cells in germinal centers: putting the jigsaw together.** *Int Rev Immunol* 2010, **29**(4):403-420.
11. Singh V, Gowthaman U, Jain S, Parihar P, Banskar S, Gupta P, Gupta UD, Agrewala JN: **Coadministration of interleukins 7 and 15 with bacille Calmette-Guerin mounts enduring T cell memory response against *Mycobacterium tuberculosis*.** *J Infect Dis* 2010, **202**(3):480-489.
12. Acevedo-Vasquez E, Ponce de Leon D, Gamboa-Cardenas R: **Latent infection and tuberculosis disease in rheumatoid arthritis patients.** *Rheum Dis Clin North Am* 2009, **35**(1):163-181.
13. Ghosh K, Patwardhan M, Pradhan V: ***Mycobacterium tuberculosis* infection precipitates SLE in patients from endemic areas.** *Rheumatol Int* 2009, **29**(9):1047-1050.
14. Res PC, Schaar CG, Breedveld FC, van Eden W, van Embden JD, Cohen IR, de Vries RR: **Synovial fluid T cell reactivity against 65 kD heat shock protein of mycobacteria in early chronic arthritis.** *Lancet* 1988, **2**(8609):478-480.
15. Van Eden W, Thole JE, van der Zee R, Noordzij A, van Embden JD, Hensen EJ, Cohen IR: **Cloning of the mycobacterial epitope recognized by T lymphocytes in adjuvant arthritis.** *Nature* 1988, **331**(6152):171-173.
16. Esaguy N, Aguas AP, van Embden JD, Silva MT: ***Mycobacteria* and human autoimmune disease: direct evidence of cross-reactivity between human lactoferrin and the 65-kilodalton protein of tubercle and leprosy bacilli.** *Infect Immun* 1991, **59**(3):1117-1125.

17. Van Eden W, Holoshitz J, Nevo Z, Frenkel A, Klajman A, Cohen IR: **Arthritis induced by a T-lymphocyte clone that responds to Mycobacterium tuberculosis and to cartilage proteoglycans.** *Proc Natl Acad Sci U S A* 1985, **82**(15):5117-5120.
18. Salvetti M, Buttinelli C, Ristori G, Carbonari M, Cherchi M, Fiorelli M, Grasso MG, Toma L, Pozzilli C: **T-lymphocyte reactivity to the recombinant mycobacterial 65- and 70-kDa heat shock proteins in multiple sclerosis.** *J Autoimmun* 1992, **5**(6):691-702.
19. Birnbaum G, Kotilinek L, Albrecht L: **Spinal fluid lymphocytes from a subgroup of multiple sclerosis patients respond to mycobacterial antigens.** *Ann Neurol* 1993, **34**(1):18-24.
20. Salvetti M, Ristori G, Buttinelli C, Fiori P, Falcone M, Britton W, Adams E, Paone G, Grasso MG, Pozzilli C: **The immune response to mycobacterial 70-kDa heat shock proteins frequently involves autoreactive T cells and is quantitatively dysregulated in multiple sclerosis.** *J Neuroimmunol* 1996, **65**(2):143-153.
21. Mor F, Cohen IR: **T cells in the lesion of experimental autoimmune encephalomyelitis. Enrichment for reactivities to myelin basic protein and to heat shock proteins.** *J Clin Invest* 1992, **90**(6):2447-2455.
22. Mustafa AS, Al-Attayah R, Hanif SN, Shaban FA: **Efficient testing of large pools of Mycobacterium tuberculosis RD1 peptides and identification of major antigens and immunodominant peptides recognized by human Th1 cells.** *Clin Vaccine Immunol* 2008, **15**(6):916-924.
23. Mustafa AS: **Th1 cell reactivity and HLA-DR binding prediction for promiscuous recognition of MPT63 (Rv1926c), a major secreted protein of Mycobacterium tuberculosis.** *Scand J Immunol* 2009, **69**(3):213-222.
24. Mustafa AS: **In silico binding predictions for identification of HLA-DR-promiscuous regions and epitopes of Mycobacterium tuberculosis protein MPT64 (Rv1980c) and their recognition by human Th1 cells.** *Med Princ Pract* 2010, **19**(5):367-372.
25. Gowthaman U, Agrewala JN: **In silico methods for predicting T-cell epitopes: Dr Jekyll or Mr Hyde?** *Expert Rev Proteomics* 2009, **6**(5):527-537.
26. Agrewala JN, Wilkinson RJ: **Influence of HLA-DR on the phenotype of CD4+ T lymphocytes specific for an epitope of the 16-kDa alpha-crystallin antigen of Mycobacterium tuberculosis.** *Eur J Immunol* 1999, **29**(6):1753-1761.
27. Gowthaman U, Agrewala JN: **In silico tools for predicting peptides binding to HLA-class II molecules: more confusion than conclusion.** *J Proteome Res* 2008, **7**(1):154-163.
28. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusic V: **Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research.** *BMC Immunology* 2008, **9**:8.
29. Lin HH, Zhang GL, Tongchusak S, Reinherz EL, Brusic V: **Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research.** *BMC Bioinformatics* 2008, **9 Suppl 12**:S22.
30. Nielsen M, Lundegaard C, Worning P, Hvid CS, Lamberth K, Buus S, Brunak S, Lund O: **Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach.** *Bioinformatics* 2004, **20**(9):1388-1397.

31. Lima T, Auchincloss AH, Coudert E, Keller G, Michoud K, Rivoire C, Bulliard V, de Castro E, Lachaize C, Baratin D *et al*: **HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot.** *Nucleic Acids Res* 2009, **37** (Database issue):D471-478.
32. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: The proteomics server for in-depth protein knowledge and analysis.** *Nucleic Acids Res* 2003, **31**(13):3784-3788.
33. **The universal protein resource (UniProt).** *Nucleic Acids Res* 2008, **36** (Database issue):D190-195.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
35. Gowthaman U, Chodisetti SB, Parihar P, Agrewala JN: **Evaluation of different generic in silico methods for predicting HLA class I binding peptide vaccine candidates using a reverse approach.** *Amino Acids* 2010, **39**(5):1333-1342.
36. Tong JC, Zhang GL, Tan TW, August JT, Brusica V, Ranganathan S: **Prediction of HLA-DQ3.2beta ligands: evidence of multiple registers in class II binding peptides.** *Bioinformatics* 2006, **22**(10):1232-1238.
37. Harrison LC, Honeyman MC, Trembleau S, Gregori S, Gallazzi F, Augstein P, Brusica V, Hammer J, Adorini L: **A peptide-binding motif for I-A(g7), the class II major histocompatibility complex (MHC) molecule of NOD and Biozzi AB/H mice.** *J Exp Med* 1997, **185**(6):1013-1021.
38. Casanova JL, Abel L: **Genetic dissection of immunity to mycobacteria: the human model.** *Annu Rev Immunol* 2002, **20**:581-620.
39. Bellamy R, Hill AV: **Genetic susceptibility to mycobacteria and other infectious pathogens in humans.** *Curr Opin Immunol* 1998, **10**(4):483-487.
40. Butler MH, Hayashi A, Ohkoshi N, Villmann C, Becker CM, Feng G, De Camilli P, Solimena M: **Autoimmunity to gephyrin in Stiff-Man syndrome.** *Neuron* 2000, **26**(2):307-312.
41. Lutton JD, Winston R, Rodman TC: **Multiple sclerosis: etiological mechanisms and future directions.** *Exp Biol Med (Maywood)* 2004, **229**(1):12-20.
42. Wucherpfennig KW: **Mechanisms for the induction of autoimmunity by infectious agents.** *J Clin Invest* 2001, **108**(8):1097-1104.
43. Shoenfeld Y, Gilburd B, Abu-Shakra M, Amital H, Barzilai O, Berkun Y, Blank M, Zandman-Goddard G, Katz U, Krause I *et al*: **The mosaic of autoimmunity: genetic factors involved in autoimmune diseases--2008.** *Isr Med Assoc J* 2008, **10**(1):3-7.
44. Shoenfeld Y, Zandman-Goddard G, Stojanovich L, Cutolo M, Amital H, Levy Y, Abu-Shakra M, Barzilai O, Berkun Y, Blank M *et al*: **The mosaic of autoimmunity: hormonal and environmental factors involved in autoimmune diseases--2008.** *Isr Med Assoc J* 2008, **10**(1):8-12.
45. Kivity S, Agmon-Levin N, Blank M, Shoenfeld Y: **Infections and autoimmunity--friends or foes?** *Trends Immunol* 2009, **30**(8):409-414.
46. Fae KC, da Silva DD, Oshiro SE, Tanaka AC, Pomerantzeff PM, Douay C, Charron D, Toubert A, Cunningham MW, Kalil J *et al*: **Mimicry in recognition of cardiac**

- myosin peptides by heart-intralesional T cell clones from rheumatic heart disease. *J Immunol* 2006, **176**(9):5662-5670.**
47. Oldstone MB: **Molecular mimicry and immune-mediated diseases.** *Faseb J* 1998, **12**(13):1255-1265.
48. Fujinami RS, Oldstone MB: **Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity.** *Science* 1985, **230**(4729):1043-1045.
49. Sabater L, Gomez-Choco M, Saiz A, Graus F: **BR serine/threonine kinase 2: a new autoantigen in paraneoplastic limbic encephalitis.** *J Neuroimmunol* 2005, **170**(1-2):186-190.
50. Sato S, Kuwana M, Hirakata M: **Clinical characteristics of Japanese patients with anti-OJ (anti-iso-leucyl-tRNA synthetase) autoantibodies.** *Rheumatology (Oxford)* 2007, **46**(5):842-845.
51. Ohosone Y, Ishida M, Takahashi Y, Matsumura M, Hirakata M, Kawahara Y, Nishikawa T, Mimori T: **Spectrum and clinical significance of autoantibodies against transfer RNA.** *Arthritis Rheum* 1998, **41**(9):1625-1631.
52. Matsushita T, Hasegawa M, Fujimoto M, Hamaguchi Y, Komura K, Hirano T, Horikawa M, Kondo M, Orito H, Kaji K *et al*: **Clinical evaluation of anti-aminoacyl tRNA synthetase antibodies in Japanese patients with dermatomyositis.** *J Rheumatol* 2007, **34**(5):1012-1018.

FIGURE LEGENDS

Figure 1: *Identification of HLA class I and II restricted T cell epitopes from host proteins that shared similarity with M. tuberculosis antigens.* Protein sequences of *M. tuberculosis* were subjected to BLAST search with the human proteome for identifying similarity regions. The peptides were analyzed for HLA binding for predominantly occurring HLA class I and II alleles using NetMHC server as described in methods. Peptides binding to HLA class I and class II molecules were considered to be CD8 T cell and CD4 T cell epitopes respectively. The peptides were classified based on predicted IC₅₀ value as strong binders (IC₅₀ ≤ 500), weak binders (500 ≤ IC₅₀ ≤ 5000) and non-binders (IC₅₀ ≥ 5000). The total number of binders (strong and weak) for each allele is represented as bar diagrams for (A) HLA class II and (B) HLA class I molecules.

Figure 2: *CD4 T cell epitopes of host proteins share resemblance with different classes of M. tuberculosis antigens.* The absolute number of host-peptides restricted to different HLA class II alleles that showed similarity with the different classes of mycobacterial proteins is represented as pie charts for (A) antigenic; (B) structural; (C) metabolic proteins of *M. tuberculosis*.

Figure 3: *The relative fractions of autoreactive CD4 T cell epitopes exhibiting similarity with different categories of antigens of M. tuberculosis.* Bar diagram indicates the relative percentages of HLA class II binders from host-antigens sharing sequence corresponding to different categories of antigenic proteins of *M. tuberculosis*.

Figure 4: *A considerable fraction of peptides from autoantigens demonstrate strong HLA class II binding.* Bar diagram depicts the percentage of strong binders ($IC_{50} \leq 500$) among the total binders from host peptides that shared similarity to *M. tuberculosis* antigens.

Figure 5: *CD8 T cell epitopes of host antigens display similarity with different classes of M. tuberculosis antigens.* The absolute number of host-peptides restricted to different HLA class I alleles that showed similarity with the different classes of *M. tuberculosis* proteins is represented as pie charts for (A) antigenic; (B) structural; (C) metabolic proteins.

Figure 6: *The relative fractions of autoreactive CD8 T cell epitopes exhibiting cross-reactivity to different categories of antigens of M. tuberculosis.* Bar diagram indicates the relative percentages of HLA class I binders of host-antigens sharing sequence similarity with the different categories of antigenic proteins of *M. tuberculosis*.

Figure 7: *The peptides identified from host proteins reveal strong HLA class I binding.* Bar diagram depicts the percentage of strong binders ($IC_{50} \leq 500$) among the total identified binders from host-peptides that shared cross-reactivity with *M. tuberculosis* antigens.

TABLE 1: Putative CD4 T cell epitopes of established antigens involved in autoimmune diseases that share similarity with *M. tuberculosis* proteins.

Protein from <i>M. tuberculosis</i> .	Disease	Homologous Human Proteins (known to be associated with autoimmunity)	Putative T cell epitope	HLA Class II Alleles	
HSP60	Multiple Sclerosis [4, 18, 19, 21]	HSP60	IGAKLVQDVA	HLA DRB1* 01:01	
			EGMKFDRGYIS	HLA DRB1* 01:01, 03:01,04:01,07:01,11:01,15:01	
			VAVKAPGFGD	HLA DRB1*15:01	
			KPLVIIAEDVDGEALSTLVN	HLADRB1*01:01,03:01,04:01,07:01,08:02,11:01,13:01,15:01	
		cDNA FLJ54912	EGMKFNRGYIS	HLA DRB1*01: 01, 04: 01, 07: 01, 01: 01, 13: 01, 15: 01	
			KPLVIIAEDVDGEALSTLV	HLA DRB1*01: 01, 03: 01, 04: 01, 07: 01, 08: 02, 11: 01,15: 01	
		60 kDa chaperonin (Fragment) [HSPD1]	EGIKFDRGYIS	HLA DRB1*01: 01, 03: 01, 04: 01, 07: 01, 08: 02, 11: 01, 13: 01, 15: 01	
			LKFDRGYVS	HLA DRB1*01: 01, 03: 01, 04: 01, 07: 01, 11: 01, 13: 01	
		Rheumatoid Arthritis [14, 15, 17]	Putative uncharacterized protein HSPD1	GEALSTLVN	HLA DRB1*01: 01
				KPLVIIAEDVDGEALSTLVN	HLA DRB1*01: 01, 03: 01, 04: 01, 07: 01, 08: 02, 11: 01, 13: 01, 15: 01
		T-complex protein 1 subunit beta	LALVTGGEI	HLA DRB1*01: 01, 07: 01, 13: 01, 15: 01	
		T-complex protein 1 subunit epsilon	LDKISDSVL	HLA DRB1*01: 01, 07: 01	
Molybdopterin biosynthesis Mog protein	Stiff Man's Syndrome [40]	Gephyrin	LNLILTGGTG	HLA DRB1*01: 01, 04: 01, 07: 01, 11: 01, 13: 01, 15: 01	
		Highly similar to Gephyrin	GKTLIINLPGS	HLA DRB1*01: 01, 04: 01, 07: 01, 11: 01, 15: 01	
Serine/threonine-protein kinase pknD	Paraneoplastic Limbic Encephalitis [49]	BR serine/threonine-protein kinase-2	HRDLKPENLLL	HLA DRB1*01: 01, 07: 01, 13: 01	
Isoleucyl-tRNA synthetase	Arthritis [50, 51] SLE [51] Interstitial lung disease [50, 52]	Isoleucyl-tRNA synthetase	GLPHYGHIL	HLA DRB1*01: 01, 15: 01	

ADDITIONAL FILES:*Additional file 1***Title: List of putative cross reactive epitopes from human proteome restricted to HLA class II alleles**

Description: Data sets represent a list of human peptide regions showing sequence similarity with *M. tuberculosis*. These peptides were assessed for binding to predominant HLA class I and HLA class II alleles. Based on IC_{50} values, they were classified as strong binders ($IC_{50} \leq 500$), weak binders ($500 \leq IC_{50} \leq 5000$) and non-binders ($IC_{50} \geq 5000$). Peptides binding to more than three different alleles were considered as promiscuous.

*Additional file 2***Title: List of putative cross reactive epitopes from human proteome restricted to HLA class I alleles**

Description: Data sets represent a list of human peptide regions showing sequence similarity with *M. tuberculosis*. These peptides were assessed for binding to predominant HLA class I and HLA class II alleles. Based on IC_{50} values, they were classified as strong binders ($IC_{50} \leq 500$), weak binders ($500 \leq IC_{50} \leq 5000$) and non-binders ($IC_{50} \geq 5000$). Peptides binding to more than three different alleles were considered as promiscuous.

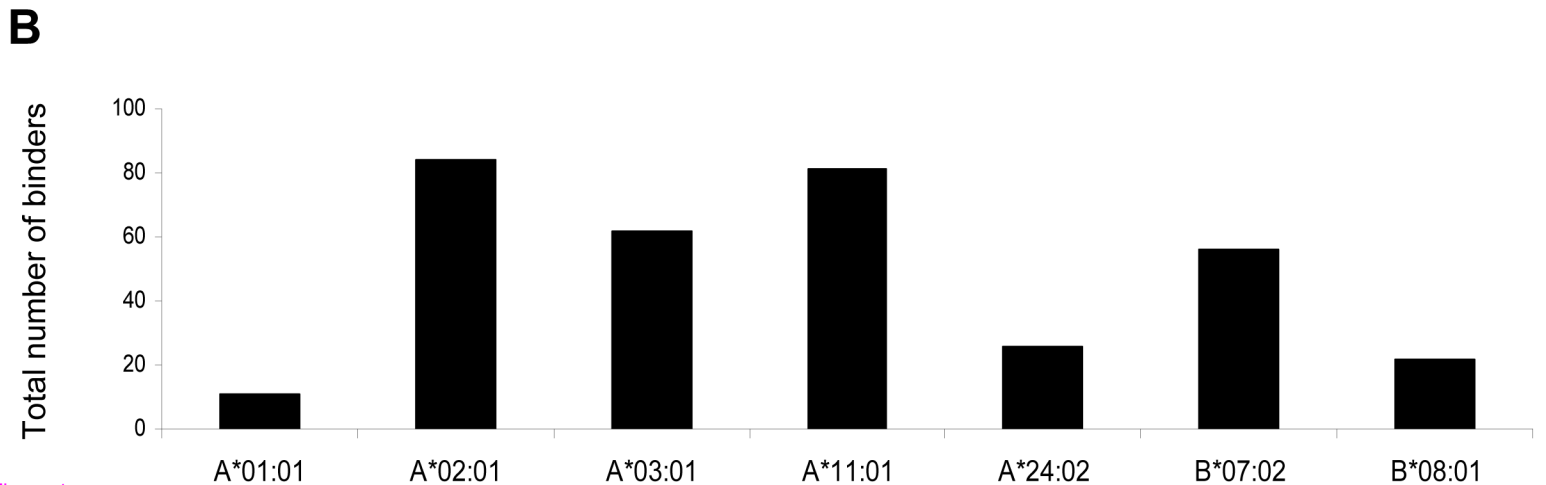
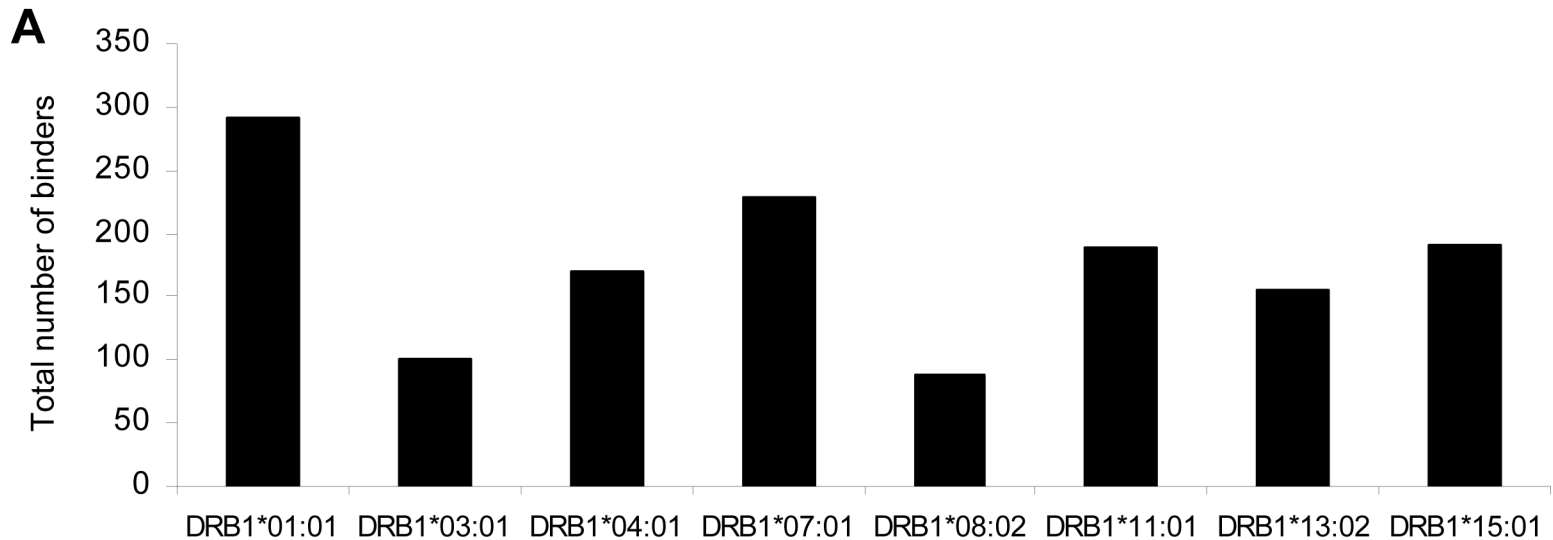
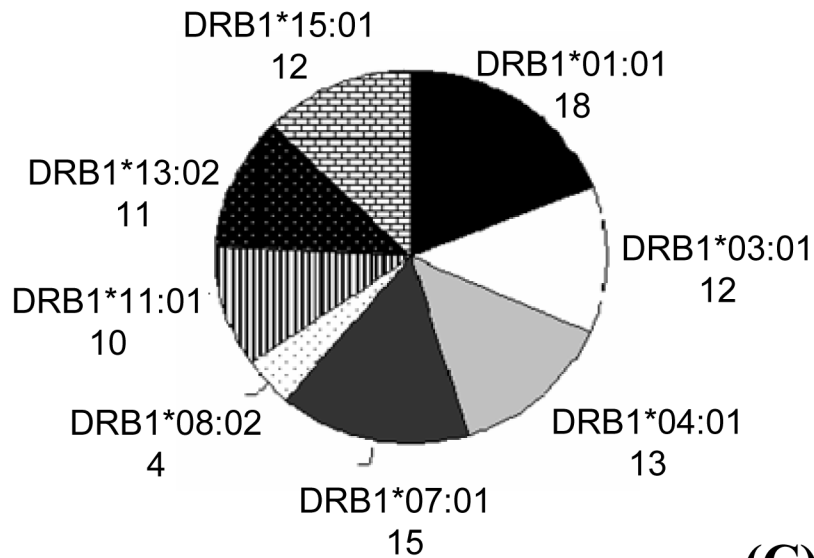
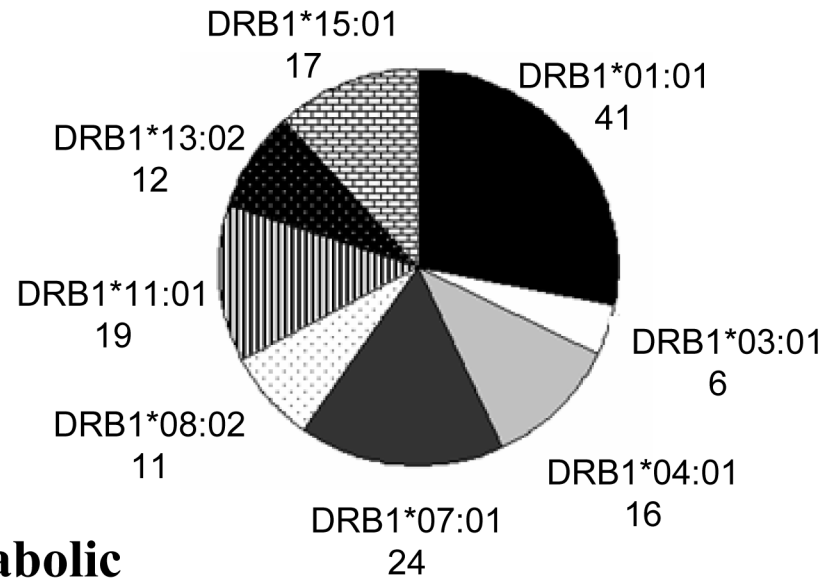


Figure 1

(A) Antigenic



(B) Structural



(C) Metabolic

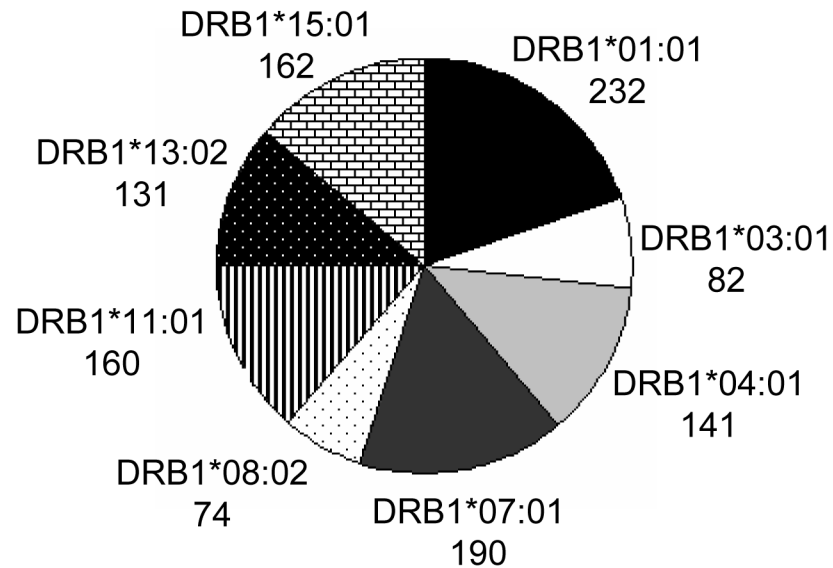


Figure 2

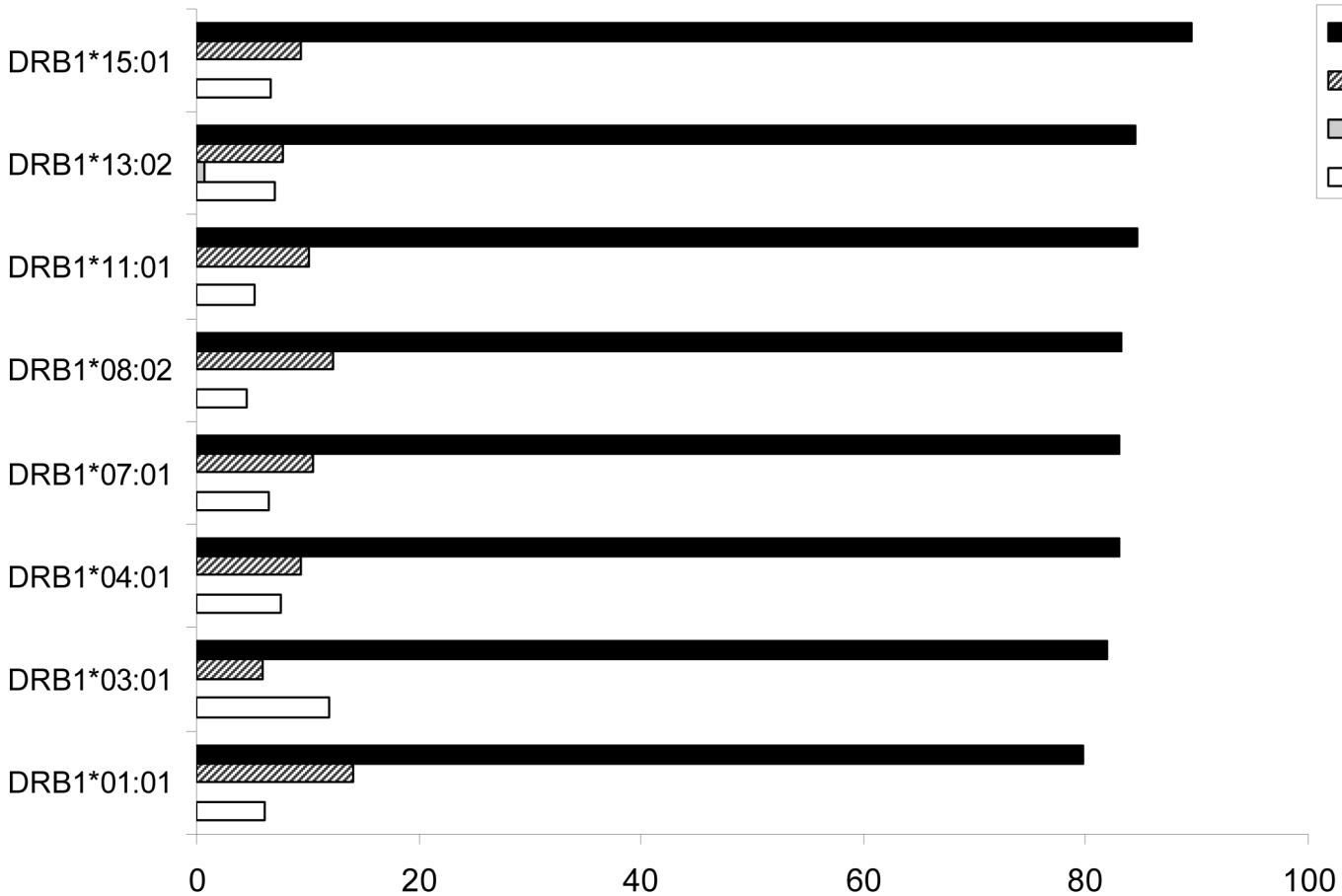
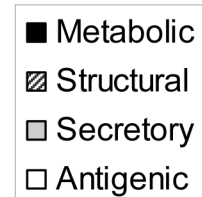


Figure 3

Percentage

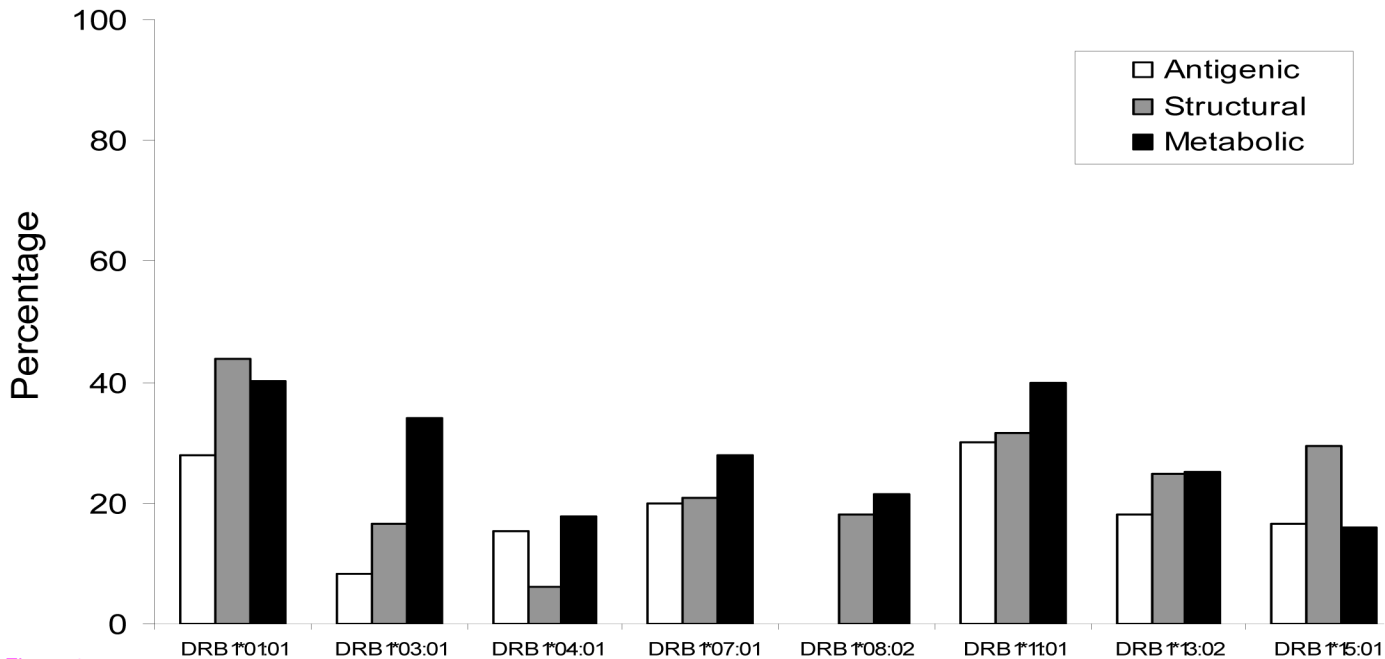
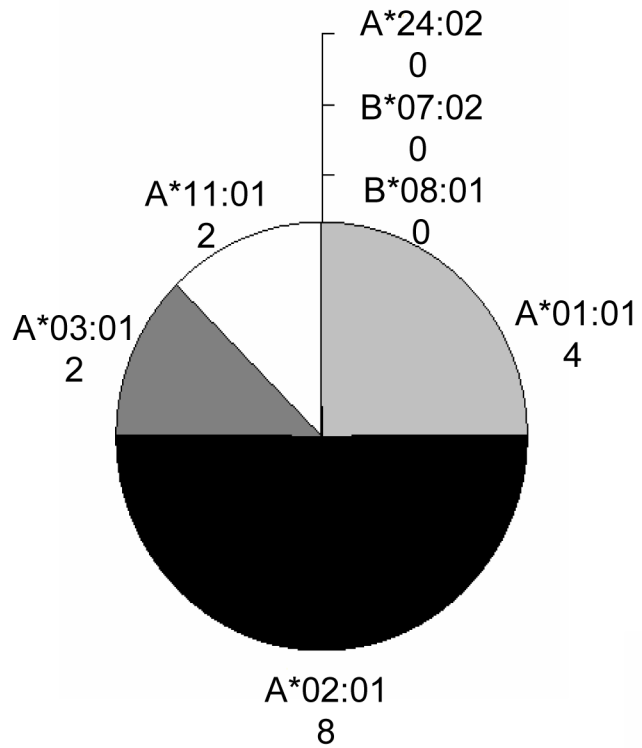
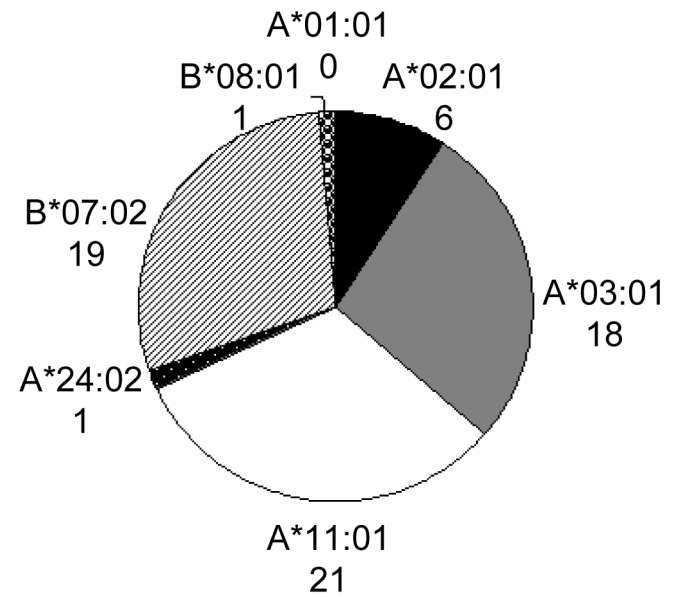


Figure 4

(A) Antigenic



(B) Structural



(C) Metabolic

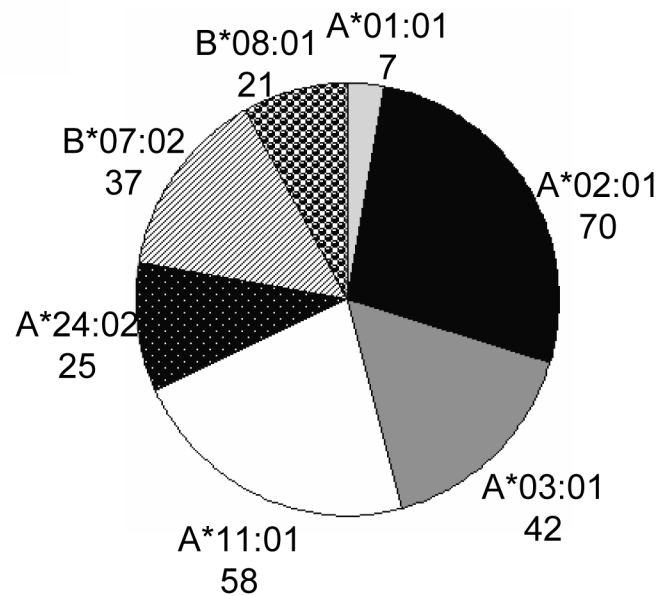


Figure 5

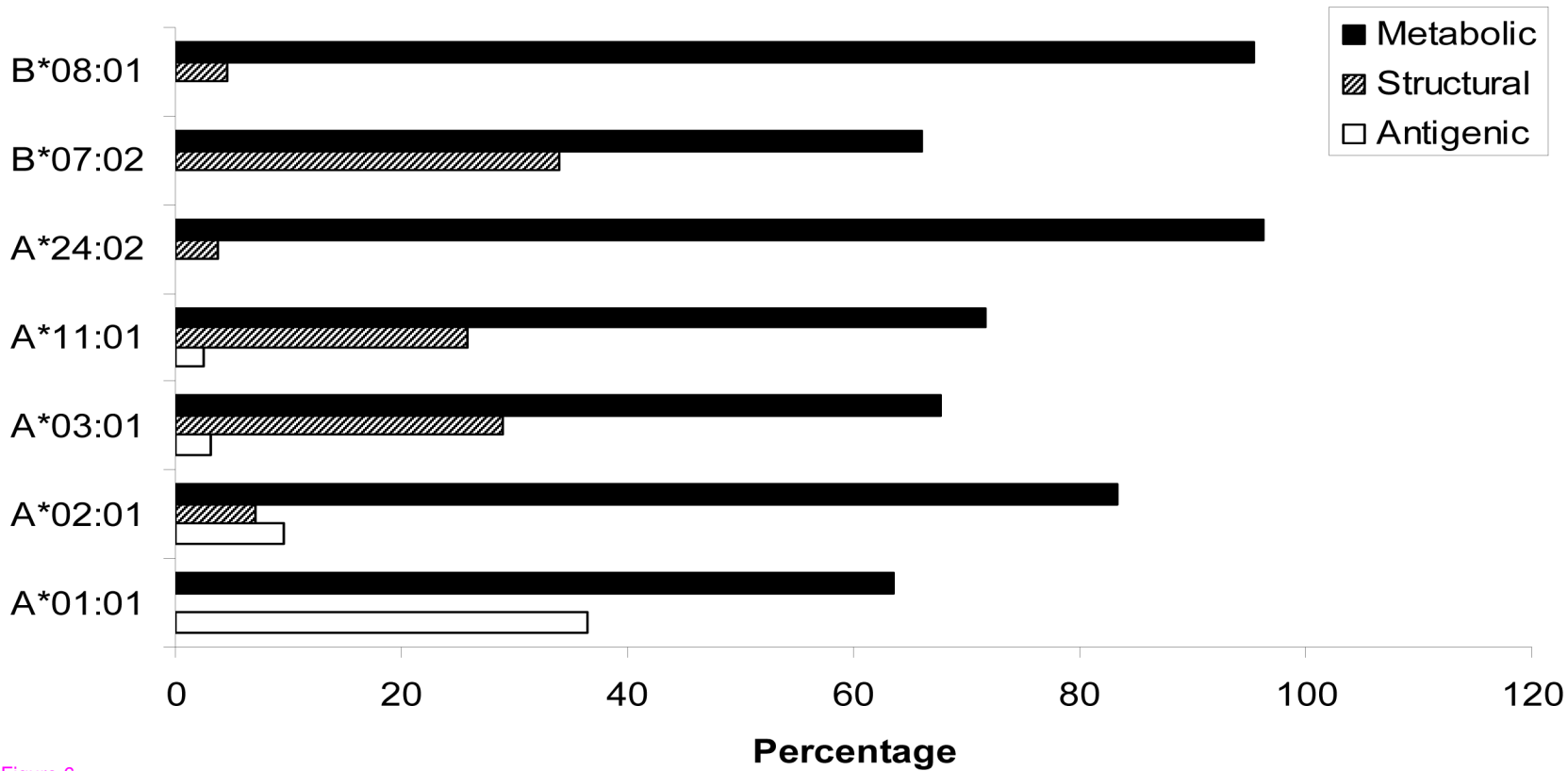


Figure 6

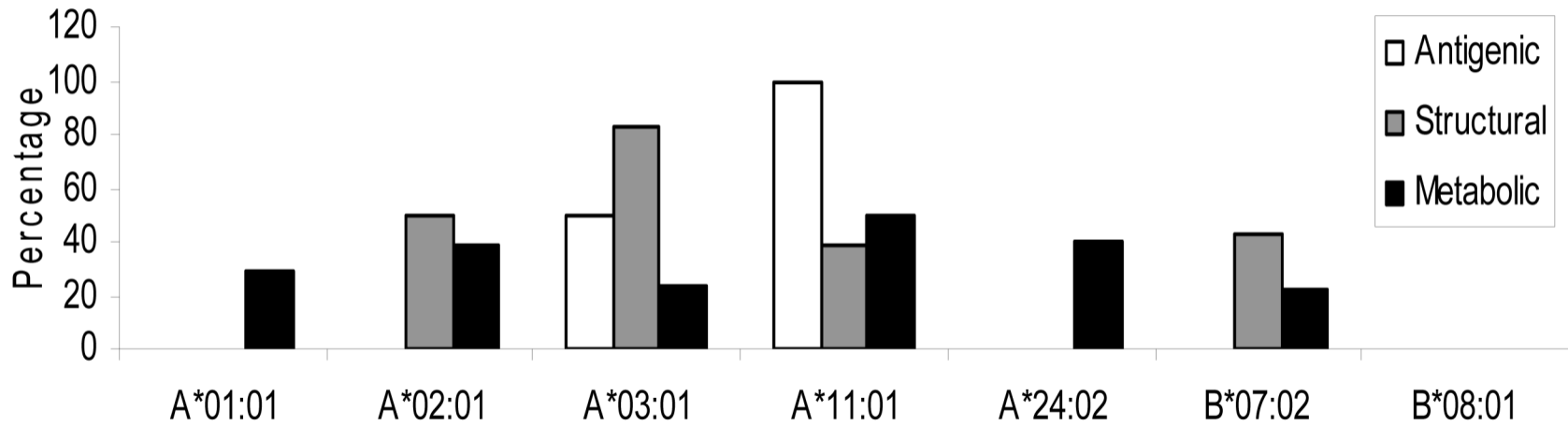


Figure 7

Additional files provided with this submission:

Additional file 1: Additional file 1.xlsx, 53K

<http://www.biomedcentral.com/imedia/1325736214678992/supp1.xlsx>

Additional file 2: Additional file 2.xlsx, 45K

<http://www.biomedcentral.com/imedia/1533878574675843/supp2.xlsx>