

Activation of the immune system by self and bacterial lipids

Peter Reinink

Activation of the immune system by self and bacterial lipids

Activatie van het immuunsysteem door
lichaamseigen en bacteriële lipiden

© 2019 Peter Reinink

All rights reserved. Published papers were reprinted in this publication with permission from the publisher or under CC licence. No part of this publication may be reproduced or transmitted in any form or by any means, electronical or mechanical, without prior permission of the author.

ISBN: 978-90-393-7129-9

The research described in this thesis was supported through travel grants of the European Molecular Biology Organization (EMBO) and Boehringer Ingelheim Fonds.

Research described in this thesis was performed at:

- the Department of Infectious Diseases and Immunology, Division of Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands.
- the Department of Rheumatology, Immunology & Allergy, Brigham and Women's Hospital, Boston, USA

Cover design: Aniek Lotterman-Reinink.

Artistic representation of a TLC plate loaded with whole lipid extracts of 12 *Salmonella* species.

Layout: Proefschrift-aio.nl

Printing: Proefschrift-aio.nl

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
donderdag 25 april 2019 des ochtends te 10.30 uur

door

Peter Reinink

geboren op 22 september 1985
te Nieuwegein

Promotor:

Prof. dr. W. van Eden

Copromotor:

Dr. I. van Rhijn

Table of content

Chapter 1

General introduction

7

Chapter 2

Mammalian CD1 and MR1 genes

21

Chapter 3

CD1b presents self and *Borrelia burgdorferi* diacylglycerols to human T cells

39

Chapter 4

TRBV4-1 (V β 7.1) bias among T cells recognizing CD1b regardless of the presented lipid antigen

67

Chapter 5

A T-cell receptor escape channel allows broad T-cell response to CD1b and membrane phospholipids

99

Chapter 6

Discovery of immunogenic trehalose phospholipids in *Salmonella* species reveals functional convergence with mycobacteria

127

Chapter 7

Discussion

163

Appendix

Nederlandse samenvatting

Dankwoord

Curriculum vitae

List of publications

175



Chapter 1

General introduction

This thesis describes the discovery of lipids that stimulate the human immune system. Three main areas of immunology are covered: stimulation of the innate immune system, stimulation of the adaptive immune system by pathogen-specific lipids, and stimulation of the adaptive immune system by “self” lipids. Specifically, a previously unknown lipid with macrophage-activating capacity was discovered in *Salmonella Typhi* and the mechanism of T cell recognition of common cell membrane lipids was elucidated.

Vaccines

Nowadays, numerous vaccines for infectious diseases are available and these vaccines contribute greatly to public health. Four main types of vaccines exist. Live vaccines, like the BCG vaccine, contain a live attenuated pathogen. Inactivated vaccines contain a killed version of the pathogen. Subunit vaccines or conjugate vaccines contain one or more pure antigens. In conjugate vaccines a weak immunogen, for instance a polysaccharide, is linked to a strong immunogen, the carrier, which is not necessarily from the same pathogen¹. Lastly, toxoid vaccines create immunity to the parts of the pathogen that cause a disease instead of the pathogen itself, like the vaccine against tetanus.

Besides the antigen, adjuvants are also an important part of vaccines. Adjuvants enhance the immune response, enabling a reduction of the dose per vaccination or vaccination frequency, and in some cases improve the stability of the vaccine formulation². The most frequently used adjuvant consists of aluminium salts: they are used in 36% of the vaccines approved by the US Food and Drug Administration³. Complete Freund's adjuvant contains mineral oil and killed *M. tuberculosis*. Among the many compounds in *M. tuberculosis* that stimulate the innate immune system, one of the strongest is trehalose-6,6'-dimycolate (TDM)⁴. Macrophage inducible C-type lectin (Mincle) is activated by TDM⁵, and by several other lipids from *M. tuberculosis*^{6,7}. Because of safety concerns Complete Freund's adjuvant is not used in vaccines for humans². As an alternative, the CAF01 adjuvant activates Mincle by means of a synthetic analog of *M. tuberculosis* TDM: trehalose-6,6-dibehenate (TDB) which is incorporated in liposomes composed of dimethyldioctadecylammonium (DDA)⁸.

Vaccination can trigger three different effector mechanisms. First, antibodies, which are produced by B cells, can clear extracellular pathogens in various ways. They can activate the complement cascade, block viral entry into the cell, and/or promote phagocytosis of extracellular bacteria. Second, cytotoxic T cells, which are mostly CD8⁺, can directly kill infected cells. Third, CD4⁺ T-helper cells play a role in the reduction, control and clearance of intracellular and extracellular pathogens, mainly by producing cytokines. There are many T-helper subsets known, among which the most well-known are the follicular T-helper cells

which help B cells in their antibody production, T-helper 1 cells which are mainly involved in the reaction to intracellular pathogens, and T-helper 2 cells which are responding to helminths^{9,10}. Major Histocompatibility Complex (MHC) class I or class II molecules present peptides to CD8⁺ or CD4⁺ T cells, respectively, which recognize the peptide MHC complex with their T cell receptor (TCR).

Unfortunately, not all vaccines provide perfect protection. Some vaccines are unable to elicit a strong enough immune response to provide full protection, some only provide short term protection, and some vaccines generate a suboptimal type of immune reaction, for instance a T-helper 1 response instead of T-helper 2 response or vice versa. Furthermore, immunization may not work in all individuals and immunizations can be effective only for a subset of strains of a pathogenic microbial species. Also, pathogens can escape the immune system by mutation.

MHC antigen presentation.

Already in 1936 MHC genes were described in mice¹¹⁻¹³. Later, it was named the H2 complex¹⁴. Since then, MHC molecules are studied in almost all vertebrate species and their way of antigen presentation is well understood. We discriminate two classes of MHC molecules: MHC class I molecules which consist of three α domains, that are non-covalently bound to a β 2-microglobulin, and the MHC class II molecules which are heterodimers of two chains both consisting of two domains (Figure 1). MHC class I molecules are present on the cell surface of each nucleated cell, and MHC class II molecules are only present on the cell membrane of so-called professional antigen-presenting cells. MHC class I molecules are recognized by the T cells that co-express the CD8 protein and MHC class II molecules are recognized by CD4 expressing T cells. Despite these differences between the two classes of MHC molecules, they have two major things in common. First, they present peptide antigens to T cells, and second, many genetic variants of these genes exist. In other words: MHC genes are polymorphic.

CD1 antigen presentation

Like MHC molecules, CD1 molecules also present antigens to T cells. CD1 molecules are structurally more related to MHC class I than to MHC class II molecules (Figure 1). The two most important differences between MHC class I molecules and CD1 molecules are related to the ligands they bind and their genetic variation. First, instead of peptides, CD1 presents lipids to T cells. The second major difference is the lack of polymorphism of CD1

molecules. The result is that all individuals of one species express an identical set of CD1 molecules. This is different from MHC, which is so polymorphic that it is extremely rare that unrelated individuals express identical sets of MHC molecules.

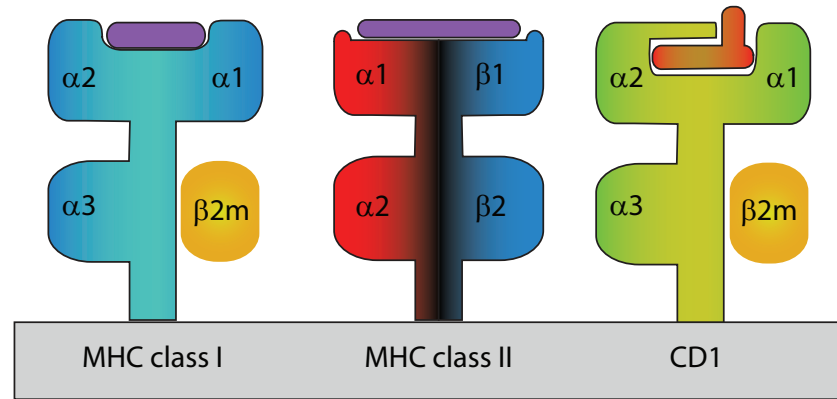


Figure 1: Schematic overview of the structures of MHC class I, MHC class II, and CD1 molecules.

MHC class I molecules and CD1 molecules both consist of three α domains and an $\beta 2m$ molecule. MHC class II molecules are heterodimers formed by an α chain and a β chain. Both MHC class I and MHC class II bind peptides (purple) whereas CD1 molecules present lipids (orange).

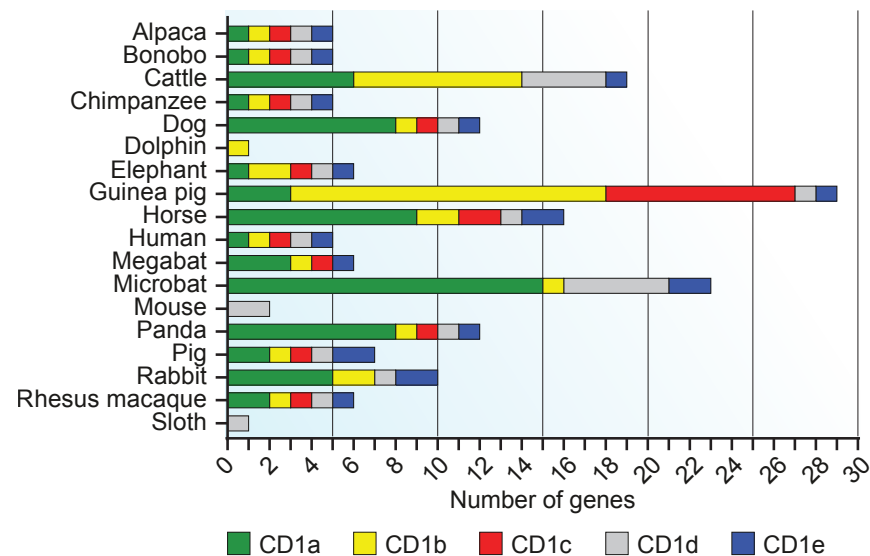


Figure 2: Overview of the predicted numbers of CD1 genes per mammal species.

The number of CD1 molecules per species per CD1 isoform. Functional as well as pseudogenes are included in this graph.

Adapted from Reinink & Van Rhijn, BioRxiv, 2016¹⁶

The family of CD1 molecules consists of five isoforms, CD1a, CD1b, CD1c, CD1d, and CD1e. CD1 genes are identified in all mammalian species studied so far (reviewed in **Chapter 2**). Most likely they arise from the same common progenitor, and the CD1 locus was then shaped by deletions, insertions, and mutations probably under immunological selective pressure during the diversion of mammalian species. The result is that all mammals have different CD1 loci. The smallest locus that is known is in mice. Their locus consist of two CD1d homologues¹⁵. *In silico* prediction of the size of the CD1 locus in different mammals shows a large variation in the number of CD1 genes predicted¹⁶. These predicted genes can be either functional or pseudo genes (Figure 2). Humans have one functional gene of all CD1 isoforms. CD1 molecules are divided into two groups; group 1 (CD1a, CD1b, and CD1c) and group 2 (CD1d). CD1e does not present lipid antigens to T cells, but helps with lipid processing and loading of the lipid onto other CD1 molecules¹⁷⁻¹⁹.

An important difference between group 1 and group 2 molecules is their gene regulation. Group 1 CD1 molecules are exclusively found on professional antigen presenting cells (APCs) and thymocytes. Langerhans cells express extremely high levels of CD1a and low levels of CD1c. CD1b and CD1c is expressed on dermal dendritic cells (DC). Peripheral expression of CD1a and CD1b is limited to DCs. In addition to expression by DCs, CD1c is expressed by B cells²⁰. Furthermore, group 1 CD1 molecules are upregulated in response to GM-CSF or active IL-1 β ^{21,22}. The group 2 CD1 molecule, CD1d, is always present on the cell surfaces in blood and peripheral tissues²⁰.

The four CD1 molecules that present lipids to T cells have different cellular locations of antigen loading. In the Endoplasmic reticulum all four bind endogenous lipids and go to the cell surface via the Golgi apparatus. On the cell surface CD1a and CD1c exchange their lipid for an exogenous lipid. CD1b, CD1c and CD1d are internalized and transported from the cell surface to the early endosome and late endosome. CD1b molecules are further redirected to the lysosome. The compartment the CD1 molecules travel to correlates with the degree of acidification needed for the CD1 molecule to load its lipid. After loading the CD1b, CD1c, and CD1d molecules return to the cell surface to present their ligand to T cells (reviewed in Moody & Cotton 2017²³).

Lipid vaccines

Because CD1 molecules can present lipid antigens to the immune system, they should not be excluded from future vaccine development. Vaccines based on lipids can have numerous advantages. Pathogens frequently escape an immune reaction by antigen mutation. However, lipids are generated by an enzymatic cascade. Therefore, unlike peptides which are

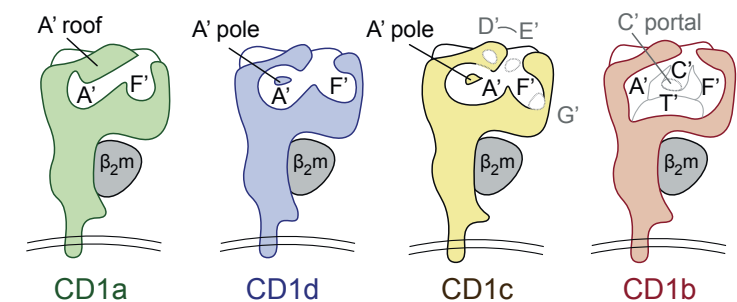
hardcoded in the genome, lipids are unlikely to mutate. Furthermore, pathogenic bacteria can share lipids, so a vaccine based on a lipid from one species can potentially give protection against multiple species. It is shown that the lipid glucose monomycolate (GMM) from related mycobacterial species is recognized by the same human CD1b specific T cells²⁴. Some studies report immunization experiments with mycobacterial lipids. When the GMM lipid was injected as vaccine in cattle, it raised a T cell response but not an antibody response²⁵. In a vaccination study in guinea pigs, the vaccine contained the mycobacterial lipids SL37 and PIM₂. Vaccinated animals had a better outcome of experimental tuberculosis than unvaccinated animals, although animals vaccinated with the commonly used BCG vaccine were better protected²⁶. Before a vaccine based on a specific lipid of pathogenic bacteria can be made, lipids need to be identified that bind to at least one CD1 isoform and the complex should stimulate a T cell repertoire in the host.

CD1 ligands

Initially, CD1 studies were aimed to find immunogenic lipids in mycobacteria²⁷. Currently, lipids from non-mycobacterial pathogenic bacteria and self-derived lipid and their reacting T cells are studied as well²⁸⁻³³.

Crystal structures are available for the four antigen presenting CD1 molecules and the shape and size of their clefts are known³⁴⁻³⁷. Whereas MHC molecules have an open binding groove, all CD1 molecules have a closed cleft. The so-called A' roof covers the cavity and defines the volume of the cleft (Figure 3). CD1a, CD1c, and CD1d have a somewhat similar structure: they all have an A' and F' pocket similar to the A and F pocket in MHC molecules. The cleft of CD1b is clearly different from the other molecules: the C' channel and T' tunnel connect the A' and F' pockets so it can accommodate larger lipids, up to C80 in total length^{34,38}. Although the clefts of CD1a, CD1c and CD1d molecules differ in their internal structure, all are able to present lipids with an overall chain length of C36-42 (Figure 3).

Just like MHC, CD1 can bind endogenous or “self” ligands as well as exogenous, often pathogen-derived ligands. Many endogenous lipid ligands are described to bind to CD1a^{37,39-41}. Furthermore, it is described that phospholipase A2 activates CD1a-restricted T cells by cleaving endogenous nonantigenic phospholipids into lysophospholipids which can be presented via CD1a^{42,43}. However, also some exogenous lipids are known to bind to CD1a, like urushiol derived from the Poison Ivy plant and the mycobacterial derived lipopeptide dideoxymycobactin⁴⁴⁻⁴⁶.



	CD1a	CD1d	CD1c	CD1b
Pockets	2	2	2	4
Accessory Portals	0	0	2 or 3 Depending on ligand	1
Approximate Capacity	C36	C42	C40	C72

Current Opinion in Immunology

Figure 3: Schematic overview of the binding clefts of the CD1 molecules.

CD1a, CD1b, CD1c, and CD1d all have an A' and F' pocket but the internal structure of the binding cleft of CD1b is considerably different due to the C' channel and T' tunnel.

Reprint from Moody & Cotton, Current Opinion in Immunology, 2017²³

The first and best known CD1b antigens were identified in studies of mycobacterial lipids. Lipids bound by CD1b of mycobacterial origin are glucose monomycolate (GMM), mycolic acid (MA), and diacylated sulfolipids (Ac₂SGL)^{24,27,47}. Like CD1a, CD1b is also known to present self-lipids to T cells^{33,48,49}.

CD1c is the only CD1 molecule that presents polyisoprenoids⁵⁰ and phosphoketides^{51,52}. Like CD1a, CD1c is also able to present lipopeptides to T cells. These lipopeptides can have peptide moieties up to 12 amino acids long^{53,54}. Similar to CD1a and CD1b, CD1c is also able to present self-lipids to T cells⁵⁵.

α -galactosylceramide (α -GalCer) is the most well-known ligand for CD1d⁵⁶. Beside α -GalCer, CD1d is capable to present a variety of other lipids to T cells. A number of lipids derived from microbial origin^{29,30,57,58} and self-lipids^{28,59-61} are described.

CD1-specific T cells

The first CD1 ligands and the first T cells that recognize CD1 were identified around the same time. Unlike MHC class I restricted T cells, which are CD8⁺ because CD8 is the co-receptor for MHC class I, and MHC class II-restricted cells, which are CD4⁺ because CD4 is the coreceptor for MHC class II, no coreceptors for CD1 are known and CD1 reactive T cells can be either CD4⁺, CD8⁺ or double negative for CD4 and CD8^{62,63}. Another difference is that due to the polymorphic nature of the CD1 molecules, interdonor conserved T cell receptors exist. Because of the enormous potential diversity of the TCR repertoire, most TCRs are unique to an individual. An interdonor conserved T cell is a TCR or TCR pattern that is present in all donors within one species and binds to the same antigen presenting molecule⁶⁴. The three known CD1 restricted interdonor conserved T cells are the NKT cell, the LDN5-like cell, and the GEM T cell. The CD1d restricted NKT cell is the most well-known of the three cells. In mice this T cell expresses the NK cell marker NK1.1 on its cell surface⁶⁵ while human NKT cells express CD161 (NKR-P1) on their cell surface⁶⁶. NKT cells can recognize CD1d with many different ligands with its invariant TCR. The human NKT TCR consist of an invariant α chain in which the V gene segment TRAV10 is always joined in an identical way to the J gene segment TRAJ18, while the β chain is biased towards using TRBV25-1. LDN5-like T cells⁶⁷ and GEM T cells⁶⁸ both recognize CD1b loaded with GMM. Although both cells have the same specificity, their TCRs are different. The TCR of GEM T cells consists of an invariant α chain of TRAV1-2 and TRAJ9 and a TRBV6-2-biased β chain. The TCRs of LDN5-like cells have a biased α chain (TRAV17) and β chain (TRBV4-1)⁶⁹, but because considerable sequence diversity is present where the V gene segments are joined to the J or D segments, they should not be considered invariant.

MHC molecules are known to be able to present endogenous peptides. However, after initial positive selection in the thymus, T cells that strongly recognize MHC presenting endogenous peptides are removed. This process provides some protection against autoimmune diseases and is called “central tolerance”. However, for all CD1 molecules, autoreactive T cells are easily and frequently identified^{33,40,49,55,70}. The presence of autoreactive CD1 restricted T cells in blood raises the question whether or not there negative selection in the thymus for these cells. Mice do not express group 1 CD1 molecules, so everything known concerns CD1d or originates from human group 1 CD1 transgenic mice. Regardless, it is clear that for positive selection hematopoietic cells are more important than thymic epithelial cells for CD1b⁷¹ and CD1d⁷². The mechanism that prevents these autoreactive T cells from causing autoimmune diseases is unknown. They may be unable to perform any inflammatory effector functions, or suppression by other cells might play a role. Lastly, with the exception of CD1a on Langerhans cells, CD1 molecules are not highly or ubiquitously expressed under steady state conditions and group 1 CD1 molecules are only upregulated during disease.

Scope of this thesis

This thesis centers around stimulation of the innate and adaptive immune system by pathogen-specific lipids, and mechanisms of “self” lipid presentation by CD1b to the TCR. Some chapters are highly interdisciplinary and contain chemical synthesis and NMR analysis of lipids (**chapter 6**) or protein crystallography (**chapter 4 and 5**) that helped address the immunological questions and that were contributed by collaborators as indicated.

Chapter 2 describes the expression and gene diversification of CD1 genes throughout the animal kingdom.

Chapter 3 describes that the lipid *Borrelia burgdorferi* glycolipid-II (BbGL-II) can be presented by CD1b to T cells. We identified a T cell clone that binds to CD1b molecules loaded with BbGL-II as well as the structurally related self-lipid diacylglycerol.

In **chapter 4** we try to understand the phenomenon of TCR bias in the CD1b system. We show that the CD1b is more likely to interact with a TCR when that TCR contains the β variable part TRBV4-1 and obtain insights based on amino acids critical for this interaction.

In **chapter 5** we distinguish narrow lipid specificity from broad cross-reactivity between highly diverse phospholipids in the CD1b system. A trimolecular structure of CD1b, phospholipid, and TCR shows how autoreactive TCRs can handle different lipid head group sizes.

Chapter 6 is the only chapter on the innate immune system. Here we describe the discovery of a novel lipid in *Salmonella*. The lipid stimulates the innate receptor Mincle so should be considered a candidate adjuvant. We identified a gene that is crucial for the synthesis of this lipid and its distribution among bacteria.

Chapter 7 is a summary of the main findings described in this thesis. These main findings are discussed in the context of current knowledge on CD1 antigen presentation, CD1b TCR interaction, and lipid biology.

References

1. Avery, O. T. & Goebel, W. F. Chemo-Immunological Studies on Conjugated Carbohydrate-Proteins: I. Immunological Specificity of Synthetic Sugar-Protein Antigens. *Journal of Experimental Medicine* **50**, 533–550 (1929).
2. Shah, R. R., Hassett, K. J. & Brito, L. A. Overview of Vaccine Adjuvants: Introduction, History, and Current Status. in *Vaccine Adjuvants: Methods and Protocols* (ed. Fox, C. B.) 1–13 (Springer New York, 2017). doi:10.1007/978-1-4939-6445-1_1
3. Hassett, K. J. *et al.* Stabilization of a Recombinant Ricin Toxin A Subunit Vaccine through Lyophilization. *Eur J Pharm Biopharm* **85**, 279–286 (2013).
4. Geisel, R. E., Sakamoto, K., Russell, D. G. & Rhoades, E. R. In Vivo Activity of Released Cell Wall Lipids of *Mycobacterium bovis* Bacillus Calmette-Guérin Is Due Principally to Trehalose Mycolates. *The Journal of Immunology* **174**, 5007–5015 (2005).
5. Ishikawa, E. *et al.* Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle. *Journal of Experimental Medicine* **206**, 2879–2888 (2009).
6. Decout, A. *et al.* Rational design of adjuvants targeting the C-type lectin Mincle. *PNAS* **114**, 2675–2680 (2017).
7. Hattori, Y. *et al.* Glycerol Monomycolate Is a Novel Ligand for the Human, but Not Mouse Macrophage Inducible C-type Lectin, Mincle. *J. Biol. Chem.* **289**, 15405–15412 (2014).
8. Davidsen, J. *et al.* Characterization of cationic liposomes based on dimethyldioctadecylammonium and synthetic cord factor from *M. tuberculosis* (trehalose 6,6'-dibehenate)—A novel adjuvant inducing both strong CMI and antibody responses. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1718**, 22–31 (2005).
9. Romagnani, S. TH1 and TH2 in Human Diseases. *Clinical Immunology and Immunopathology* **80**, 225–235 (1996).
10. Siegrist, C.-A. 2 - Vaccine immunology. in *Vaccines (Sixth Edition)* (eds. Plotkin, S. A., Orenstein, W. A. & Offit, P. A.) 14–32 (W.B. Saunders, 2013). doi:10.1016/B978-1-4557-0090-5.00004-5
11. Gorer, P. A. The Detection of Antigenic Differences in Mouse Erythrocytes by the Employment of Immune Sera. *Br J Exp Pathol* **17**, 42–50 (1936).
12. Gorer, P. A. The genetic and antigenic basis of tumour transplantation. *The Journal of Pathology and Bacteriology* **44**, 691–697 (1937).
13. Klein, J. Seeds of time: Fifty years ago Peter A. Gorer discovered the H-2 complex. *Immunogenetics* **24**, 331–338 (1986).
14. Snell, G. D. Methods for the study of histocompatibility genes. *Journ. of Genetics* **49**, 87–108 (1948).
15. Bradbury, A., Belt, K. T., Neri, T. M., Milstein, C. & Calabi, F. Mouse CD1 is distinct from and co-exists with TL in the same thymus. *EMBO J* **7**, 3081–3086 (1988).
16. Reinink, P. & Van Rhijn, I. In Silico method for identification of MHC class I-like molecules in whole genomes. *bioRxiv* (2016). doi:10.1101/046607
17. Facciotti, F. *et al.* Fine tuning by human CD1e of lipid-specific immune responses. *Proc Natl Acad Sci USA* **108**, 14228 (2011).
18. Angénioux, C. *et al.* Characterization of CD1e, a Third Type of CD1 Molecule Expressed in Dendritic Cells. *Journal of Biological Chemistry* **275**, 37757–37764 (2000).
19. Angénioux, C. *et al.* The Cellular Pathway of CD1e in Immature and Maturing Dendritic Cells. *Traffic* **6**, 1 (2005).
20. Dougan, S. K., Kaser, A. & Blumberg, R. S. CD1 Expression on Antigen-Presenting Cells. in *T Cell Activation by CD1 and Lipid Antigens* 113–141 (Springer, Berlin, Heidelberg, 2007). doi:10.1007/978-3-540-69511-0_5
21. Kasinrerk, W., Baumruker, T., Majdic, O., Knapp, W. & Stockinger, H. CD1 molecule expression on human monocytes induced by granulocyte-macrophage colony-stimulating factor. *The Journal of Immunology* **150**, 579–584 (1993).
22. Yakimchuk, K. *et al.* Borrelia burgdorferi infection regulates CD1 expression in human cells and tissues via IL1- β . *Eur. J. Immunol.* **41**, 694–705 (2011).
23. Moody, D. B. & Cotton, R. N. Four pathways of CD1 antigen presentation to T cells. *Current Opinion in Immunology* **46**, 127–133 (2017).
24. Moody, D. B. *et al.* Structural Requirements for Glycolipid Antigen Recognition by CD1b-Restricted T Cells. *Science* **278**, 283–286 (1997).
25. Nguyen, T. K. A. *et al.* Immune response of cattle immunized with a conjugate of the glycolipid glucose monomycolate and protein. *Veterinary Immunology and Immunopathology* **142**, 265–270 (2011).
26. Larrouy-Maumus, G. *et al.* Protective efficacy of a lipid antigen vaccine in a guinea pig model of tuberculosis. *Vaccine* **35**, (2017).
27. Beckman, E. M. *et al.* Recognition of a lipid antigen by CD1-restricted $\alpha\beta$ T cells. *Nature* **372**, 691–694 (1994).
28. Brennan, P. J. *et al.* Invariant natural killer T cells recognize lipid self antigen induced by microbial danger signals. *Nature Immunology* **12**, 1202–1211 (2011).
29. Kinjo, Y. *et al.* Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520–525 (2005).
30. Kinjo, Y. *et al.* Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nat Immunol* **7**, 978–986 (2006).
31. Shahine, A. *et al.* A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Science Immunology* **2**, eaao1384 (2017).
32. Tatituri, R. V. V. *et al.* Recognition of microbial and mammalian phospholipid antigens by NKT cells with diverse TCRs. *Proc Natl Acad Sci USA* **110**, 1827 (2013).
33. Van Rhijn, I. *et al.* Human autoreactive T cells recognize CD1b and phospholipids. *PNAS* **113**, 380–385 (2016).
34. Gadola, S. D. *et al.* Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. *Nature Immunology* **3**, 721–726 (2002).
35. Koch, M. *et al.* The crystal structure of human CD1d with and without α -galactosylceramide. *Nature Immunology* **6**, 819–826 (2005).
36. Scharf, L. *et al.* The 2.5 Å Structure of CD1c in Complex with a Mycobacterial Lipid Reveals an Open Groove Ideally Suited for Diverse Antigen Presentation. *Immunity* **33**, 853–862 (2010).
37. Zajonc, D. M., Elsliger, M. A., Teyton, L. & Wilson, I. A. Crystal structure of CD1a in complex with a sulfatide self antigen at a resolution of 2.15 Å. *Nature Immunology* **4**, 808–815 (2003).
38. Moody, D. B. *et al.* Lipid length controls antigen entry into endosomal and nonendosomal pathways for CD1b presentation. *Nature Immunology* **3**, 435–442 (2002).
39. Birkinshaw, R. W. *et al.* $\alpha\beta$ T cell antigen receptor recognition of CD1a presenting self lipid ligands. *Nature Immunology* **16**, 258–266 (2015).
40. de Jong, A. *et al.* CD1a-autoreactive T cells recognize natural skin oils that function as headless antigens. *Nature Immunology* **15**, 177–185 (2014).

41. Shamshiev, A. *et al.* Presentation of the Same Glycolipid by Different CD1 Molecules. *Journal of Experimental Medicine* **195**, 1013–1021 (2002).
42. Bourgeois, E. A. *et al.* Bee venom processes human skin lipids for presentation by CD1a. *Journal of Experimental Medicine* **212**, 149–163 (2015).
43. Cheung, K. L. *et al.* Psoriatic T cells recognize neolipid antigens generated by mast cell phospholipase delivered by exosomes and presented by CD1a. *Journal of Experimental Medicine* **213**, 2399–2412 (2016).
44. Zajonc, D. M. *et al.* Molecular Mechanism of Lipopeptide Presentation by CD1a. *Immunity* **22**, 209–219 (2005).
45. Kim, J. H. *et al.* CD1a on Langerhans cells controls inflammatory skin disease. *Nature Immunology* **17**, 1159–1166 (2016).
46. Kasmar, A. G. *et al.* Cutting Edge: CD1a Tetramers and Dextramers Identify Human Lipopeptide-Specific T Cells Ex Vivo. *J Immunol* **191**, 4499–4503 (2013).
47. Stenger, S. *et al.* Differential Effects of Cytolytic T Cell Subsets on Intracellular Infection. *Science* **276**, 1684–1687 (1997).
48. Shamshiev, A. *et al.* The $\alpha\beta$ T Cell Response to Self-Glycolipids Shows a Novel Mechanism of CD1b Loading and a Requirement for Complex Oligosaccharides. *Immunity* **13**, 255–264 (2000).
49. Vincent, M. S. *et al.* CD1-dependent dendritic cell instruction. *Nature Immunology* **3**, 1163–1168 (2002).
50. Moody, D. B. *et al.* CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. *Nature* **404**, 884–888 (2000).
51. Matsunaga, I. *et al.* *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *Journal of Experimental Medicine* **200**, 1559–1569 (2004).
52. de Jong, A. *et al.* CD1c Presentation of Synthetic Glycolipid Antigens with Foreign Alkyl Branching Motifs. *Chemistry & Biology* **14**, 1232–1242 (2007).
53. Moody, D. B. *et al.* T Cell Activation by Lipopeptide Antigens. *Science* **303**, 527–531 (2004).
54. Van Rhijn, I. *et al.* CD1c bypasses lysosomes to present a lipopeptide antigen with 12 amino acids. *Journal of Experimental Medicine* **206**, 1409–1422 (2009).
55. De Lalla, C. *et al.* High-frequency and adaptive-like dynamics of human CD1 self-reactive T cells. *European Journal of Immunology* **41**, 602–610 (2011).
56. Kawano, T. *et al.* CD1d-Restricted and TCR-Mediated Activation of Va14 NKT Cells by Glycosylceramides. *Science* **278**, 1626–1629 (1997).
57. Fischer, K. *et al.* *Mycobacterium* phosphatidylinositol mannoside is a natural antigen for CD1d-restricted T cells. *PNAS* **101**, 10685–10690 (2004).
58. Mattner, J. *et al.* Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature* **434**, 525–529 (2005).
59. Facciotti, F. *et al.* Peroxisome-derived lipids are self antigens that stimulate invariant natural killer T cells in the thymus. *Nature Immunology* **13**, 474–480 (2012).
60. Gumperz, J. E. *et al.* Murine CD1d-Restricted T Cell Recognition of Cellular Lipids. *Immunity* **12**, 211–221 (2000).
61. López-Sagaseta, J., Sibener, L. V., Kung, J. E., Gumperz, J. & Adams, E. J. Lysophospholipid presentation by CD1d and recognition by a human Natural Killer T-cell receptor. *The EMBO Journal* **31**, 2047–2059 (2012).
62. Cardell, S. *et al.* CD1-restricted CD4+ T cells in major histocompatibility complex class II-deficient mice. *Journal of Experimental Medicine* **182**, 993–1004 (1995).
63. Porcelli, S. *et al.* Recognition of cluster of differentiation 1 antigens by human CD4–CD8– cytolytic T lymphocyte. *Nature* **341**, 447–450 (1989).
64. Van Rhijn, I. & Moody, D. B. Donor Unrestricted T cells: a shared human T cell response. *J Immunol* **195**, 1927–1932 (2015).
65. Makino, Y., Kanno, R., Ito, T., Higashino, K. & Taniguchi, M. Predominant expression of invariant Va14+ TCR α chain in NK1.1+ T cell populations. *Int Immunol* **7**, 1157–1161 (1995).
66. Godfrey, D. I., Hammond, K. J. L., Poulton, L. D., Smyth, M. J. & Baxter, A. G. NKT cells: facts, functions and fallacies. *Immunology Today* **21**, 573–583 (2000).
67. Van Rhijn, I. *et al.* TCR Bias and Affinity Define Two Compartments of the CD1b–Glycolipid-Specific T Cell Repertoire. *J Immunol* **192**, 4054–4060 (2014).
68. Van Rhijn, I. *et al.* A conserved human T cell population targets mycobacterial antigens presented by CD1b. *Nat Immunol* **14**, 706–713 (2013).
69. Van Rhijn, I. & Moody, D. B. CD1 and mycobacterial lipids activate human T cells. *Immunological Reviews* **264**, 138–153 (2015).
70. Mallevaey, T. *et al.* A Molecular Basis for NKT Cell Recognition of CD1d-Self-Antigen. *Immunity* **34**, 315–326 (2011).
71. Zhao, J. *et al.* Mycolic acid-specific T cells protect against *Mycobacterium tuberculosis* infection in a humanized transgenic mouse model. *eLife Sciences* **4**, e08525 (2015).
72. Bendelac, A., Savage, P. B. & Teyton, L. The Biology of NKT Cells. *Annu. Rev. Immunol.* **25**, 297–336 (2007).



Chapter 2

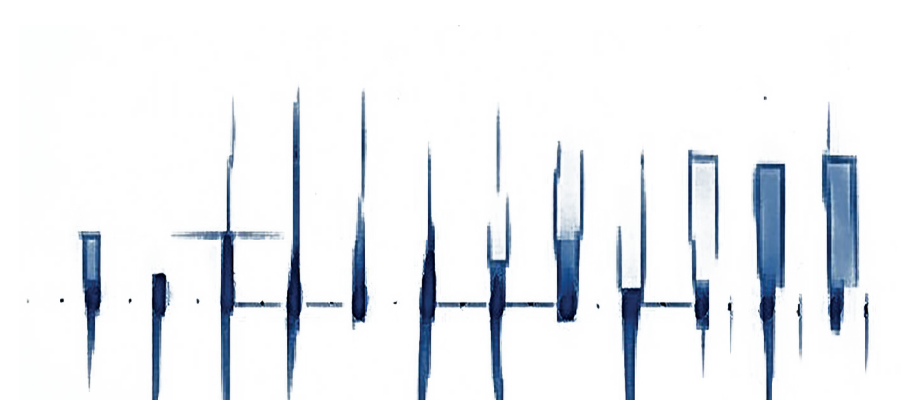
Mammalian CD1 and MR1 genes

Peter Reinink¹ and Ildiko Van Rhijn¹



¹Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA and Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands.

Immunogenetics, August 2016



Abstract

All higher vertebrates share the fundamental components of the adaptive immune system: the B cell receptor, the T cell receptor, and classical MHC proteins. At a more detailed level their immune systems vary considerably, especially with respect to the non-polymorphic MHC class I-like proteins. In mammals, the CD1 family of lipid-presenting proteins is encoded by clusters of genes of widely divergent size and composition. Another MHC class I-like protein, MR1, is typically encoded by a single gene that is highly conserved among species. Based on mammalian genomes and the available data on cellular expression profiles and protein structure, we review MR1 genes and families of CD1 genes in modern mammals from a genetic and functional perspective. Understanding the CD1 and MR1 systems across animal species provides insights into the specialized functions of the five types of CD1 proteins and facilitates careful consideration of animal models for human diseases in which immune responses to lipids and bacterial metabolites play a role.

Whereas MHC genes and the genes for T cell receptors appeared simultaneously during evolution and are present in all extant jawed vertebrates, the genes for CD1 and MR1 have a more limited distribution. Keeping the phylogenetic relationships among animal species in mind, we will describe the CD1 and MR1 genes in extant species. The evolution of these genes is beyond the scope of this review. CD1 and MR1 proteins bind lipids and vitamin B metabolites respectively, and present these to T cells, as opposed to the classical MHC proteins, which present peptide antigens to T cells. CD1 and MR1 genes have not been found in fish, while fish do contain MHC class I and II genes (Dascher 2007). Reptiles (Yang et al. 2015), birds (Miller et al. 2005; Salomonsen et al. 2005) and marsupials (Baker and Miller 2007; Cheng and Belov 2014) have CD1 genes that can clearly be distinguished from classical MHC genes and form an interspecies group with the mammalian CD1 genes. CD1a, CD1b, CD1c, CD1d, and CD1e proteins most likely arose in a common ancestor of placental mammals from a primordial form of CD1. Gene duplications, deletions, and gene inactivations shaped the composition of the CD1 family of genes, possibly under selective pressure associated with an immune function. MR1 genes are absent in fish and reptiles, but in marsupials and mammals there is typically a single functional MR1 gene. Of all the MHC Class I-like proteins, MR1 shows the highest conservation among species.

Chromosomal location and gene structure of CD1 and MR1

In mammals, the CD1 and MR1 genes are not part of the MHC locus. In humans, the CD1 and MR1 genes are located on chromosome 1 and the MHC locus lies on chromosome 6. However, in chicken, the CD1 and MHC loci are linked (Miller et al. 2005; Salomonsen et al. 2005). The genes that encode the classical MHC molecules are highly polymorphic with hundreds to thousands allelic variants, and this is thought to be closely related to their function of presenting peptides to T cells. No allelic variants for CD1b, CD1c, and MR1 are known. Allelic variants in the form of non-synonymous single nucleotide polymorphisms in human CD1a and CD1d are known, but the affected amino acids are not located in the antigen binding cleft or the TCR binding surface (Han et al. 1999; Oteo et al. 2001). Together, this justifies the common description of CD1 and MR1 genes as non-polymorphic.

Because of its non-polymorphic nature, shared structural features, and the fact that it presents small metabolite antigens rather than peptides, MR1 is often compared to CD1. However, MR1 genes are distinct from CD1 genes, and, based on sequence alignments, form their own interspecies group. Even though in humans MR1 and CD1 are located in the same MHC paralogous region on chromosome 1, the MR1 locus is separate from the CD1 cluster of genes, or CD1 locus (Hashimoto et al. 1995; Shiina et al. 2001). The CD1 locus is located between the KIRREL and olfactory receptor genes and consists of multiple genes, often in both orientations. The MR1 gene is embedded between STX6 and IER5 genes. In mice, a chromosomal rearrangement caused the separation of the CD1 and MR1

genes, which are located on chromosome 3 and chromosome 1, respectively (Dascher and Brenner 2003). In the available assembled primate data, MR1 and CD1 are located on the same chromosome, but in other mammals that were studied, like cow and pig, they are not (Goldfinch et al. 2010; Reinink and Van Rhijn 2016).

The human genome encodes one MR1 protein and five different CD1 proteins, called CD1a through CD1e. These five CD1 proteins and their orthologs are called CD1 isoforms. Even though for other proteins this term is used to define RNA splice variants derived from the same gene, in the CD1 field the word isoforms is used to indicate products of separate genes. The overall structure of CD1 and MR1 proteins resembles MHC class I molecules: a type I transmembrane protein, called the heavy chain, consisting of $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains associated with $\beta 2$ microglobulin. CD1 and MR1 genes have an intron-exon structure comparable to MHC class I genes: they contain 6 exons that encode 5' UTR and leader signal peptide, $\alpha 1$ domain, $\alpha 2$ domain, $\alpha 3$ domain, trans membrane domain, and cytoplasmic tail and 3' UTR combined (Yamaguchi et al. 1998). CD1 genes in mammals are named after the human CD1 isoform they group with, based on sequence comparison. Comparison based on overall alignment of the full coding sequence or only the $\alpha 1$ and $\alpha 2$ domains that form the antigen binding cleft, gives identical results. This is caused by the fact that $\alpha 3$ domains are highly conserved among all CD1 isoforms within one species (Balk et al. 1989), and cytoplasmic domains, even though they show considerable differences among the isoforms, are very short. Thus, effectively, CD1 isoforms are grouped and named according to resemblance of the sequence encoding their antigen binding cleft-forming $\alpha 1$ and $\alpha 2$ domains.

CD1 in humans and common research-, farm-, and companion animals

For nine mammalian species (human, rabbit, guinea pig, cow, pig, dog, horse, mouse, and rat), the CD1 genes have been studied extensively and, except for rabbit and guinea pig, their CD1 loci have been carefully mapped and curated. The functionality of many of these CD1 genes has been studied by cloning the transcripts from cDNA, sometimes followed by protein expression studies. Genomes from these nine mammalian species contain from one (rat (Katabami et al. 1998)) to thirteen (horse and dog (Dossa et al. 2014; Schjaerff et al. 2016)) CD1 genes. From rabbits, two CD1a, two CD1b, one CD1d, and one CD1e transcripts have been identified, but this study was not set up or intended to define the complete rabbit CD1 locus (Hayes and Knight 2001). Southern blots suggested the presence of at least eight CD1 genes in the rabbit (Calabi et al. 1989). Guinea pigs have been reported to contain four functional genes for CD1b, three for CD1c, one for CD1e gene and at least five CD1 pseudogenes (Dascher et al. 1999). Later, one functional gene for CD1d was described (Looringh van Beeck et al. 2009), and one gene of unknown functionality for CD1a (Van Rhijn and Moody 2015). The guinea pig CD1 locus has not been mapped, but while the

genome is being updated and assembled, even more CD1 genes have been identified in guinea pig (Reinink and Van Rhijn 2016). The first attempt to define the bovine CD1 locus was based on an early draft of the bovine genome and identified one CD1a gene, five CD1b genes, two CD1d genes, and a CD1e gene. Among these, the CD1a gene, three CD1b genes, and CD1e were identified as functional genes (Van Rhijn et al. 2006). A later version of the genome brought the number of bovine CD1 genes to 12 (Nguyen et al. 2015). The porcine CD1 locus has been described based on BAC sequencing and contains six CD1 genes: two genes for CD1a, and one for each of the other isoforms (Eguchi-Ogawa et al. 2007). One of the CD1a genes is a pseudogene. Twelve CD1 genes were mapped in the canine locus based on the available genome at the time, and a thirteenth gene was not mapped in the locus and was thought to be an allele (Looringh van Beeck et al. 2008). Subsequently, BAC sequencing placed the thirteenth gene in the locus, which is now known to contain nine CD1a genes, and one gene or each of the other isoforms (Schjaerff et al. 2016). Four of the canine CD1a genes are thought to be functional. The horse genome contains eighteen CD1 genes, among which five pseudogenes (Dossa et al. 2014). The functional equine CD1 genes are: seven CD1a genes, two CD1b genes, one CD1c gene, one CD1d gene, and two CD1e genes. The mouse genome, with its small locus consisting of only two CD1d genes, has been shown to have undergone a rearrangement that caused the loss of the other CD1 genes (Bradbury et al. 1988; Dascher and Brenner 2003). Rats have only one CD1 gene, which encodes CD1d (Ichimiya et al. 1994).

CD1 loci in less well studied mammalian genomes

Only a very small number of CD1 loci of the total of almost 4000 extant species of placental mammals has been studied. For species other than human, rabbit, guinea pig, cow, pig, dog, horse, mouse, and rat, the study of CD1 genes is now facilitated by the availability of multiple genomes. It should be noted however that many genomes, especially from species other than primates and rodents, are not yet finalized and contain multiple gaps. Also, not all sequence material may have been assigned to a chromosomal location yet, and artifacts like duplications may be present. Therefore, attempts to describe a CD1 locus based on genomic data available at a certain point in time may need to be adjusted later based on improved versions of the genome (Nguyen et al. 2015; Schjaerff et al. 2016). Because transcription and correct splicing of existing genes can not reliably be predicted, mRNA and cDNA-based sequences provide the most reliable data on functionality and the possible expression of individual CD1 proteins *in vivo*. However, RNA sequence-based data are unlikely to provide a complete overview of CD1 proteins in a species because expression of CD1 genes is often limited to a specific tissue and cell type. Therefore, to get an indication of the total number of CD1 genes in a species, searches using the basic local alignment search tool (BLAST) or collection of annotated genes in genomes provide more reliable information than expressed sequence tags or cDNA databases. Table 1 shows

numbers of CD1 genes obtained from whole genomes of 15 mammals, including less well studied species like alpaca, dolphin, elephant, two bat species, panda, and sloth (Reinink and Van Rhijn 2016). Of note, BLAST-based searches reveal functional genes and pseudogenes. Therefore, differences between results of de novo BLAST searches, automated annotation of open reading frames, and published data with regard to the total number of CD1 genes per species are expected.

Table 1: CD1 and MR1 gene numbers

Common name	Genome	Binomial species name	CD1a	CD1b	CD1c	CD1d	CD1e	Total CD1	MR1
Alpaca	vicPac2	<i>Vicugna pacos</i>	1	1	1	1	1	5	1
Bonobo	panPan1	<i>Pan paniscus</i>	1	1	1	1	1	5	2
Chimpanzee	panTro4	<i>Pan troglodytes</i>	1	1	1	1	1	5	2
Dog	CanFam3	<i>Canis lupus</i>	9	1	1	1	1	13	1
Dolphin ^a	turTru2	<i>Tursiops truncatus</i>	0	1	0	0	0	1	0
Elephant	loxAfr3	<i>Loxodonta africana</i>	1	2	1	1	1	6	1
Horse	equCab2	<i>Equus caballus</i>	9	2	2	1	2	16	1
Human	hg38	<i>Homo sapiens</i>	1	1	1	1	1	5	2
Megabat	pteVam1	<i>Pteropus vampyrus</i>	3	1	1	0	1	6	1
Microbat	myoLuc2	<i>Myotis lucifugus</i>	17	2	0	5	2	26	1
Mouse	mm10	<i>Mus musculus</i>	0	0	0	2	0	2	1
Panda	ailMel1	<i>Ailuropoda melanoleuca</i>	8	1	1	1	1	12	1
Pig	susScr3	<i>Sus scrofa</i>	2	1	1	1	2	7	1
Rabbit	oryCun2	<i>Oryctolagus cuniculus</i>	5	2	0	1	2	10	0
Rhesus macaque	rheMac3	<i>Macaca mulatta</i>	2	1	1	1	1	6	2
Sloth ^a	choHof1	<i>Choloepus hoffmanni</i>	1	0	0	1	0	2	1

For each of the indicated mammalian genomes, a list of CD1 and MR1 genes as determined by BLAST-based searches was merged with a list of Ensembl-annotated CD1 and MR1 genes when available (adapted from Reinink and Van Rhijn (Reinink and Van Rhijn 2016)). Redundancies (genes with identical genomic location) were removed. ^aThe dolphin and sloth genomes are not completely assembled and consist of relatively small contigs, which may have led to the fragmentation of CD1 or MR1 genes and subsequent failure of identification.

CD1 genes have undergone multiple duplications in most mammals, leading to extended multi gene families. The general picture that emerges is that CD1a has undergone the most extensive multiplications and CD1e the least or none. Automated annotations of CD1 isoforms of species other than mouse or human often indicate a certain degree of uncertainty, and are sometimes annotated as “CD1a-like”, for example. However, the

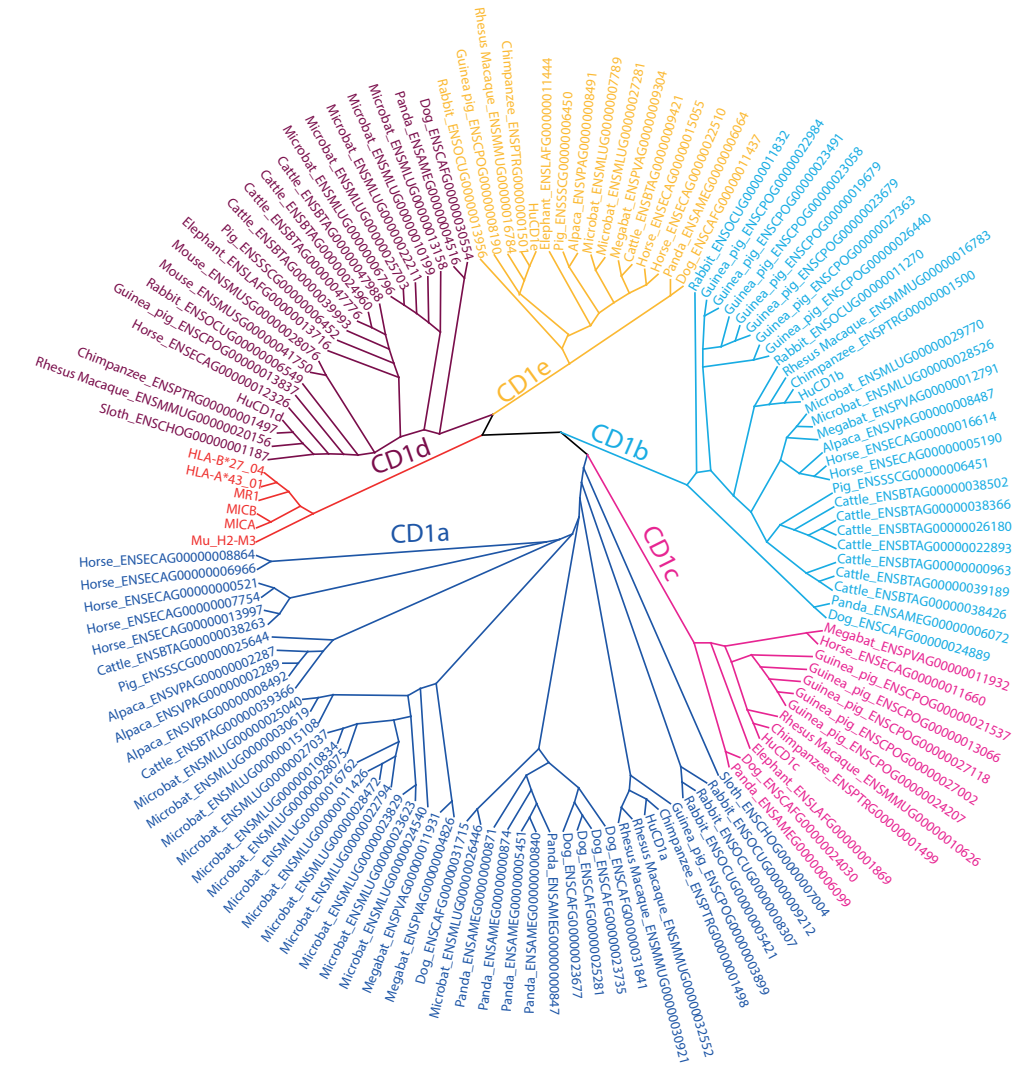


Figure 1: CD1 and MR1 genes in mammals

From 16 mammals, known CD1 genes and predicted CD1 paralog open reading frames were obtained from Ensembl (www.ensembl.org). An alignment of these sequences and human MICA, MICB, HLA-A, HLA-B and murine H2-M3 was generated by MUSCLE (Edgar 2004), clustered according to a neighbor joining algorithm, and shown as a radial cladogram. Groups were color-coded based on the clustering with human CD1 isoforms.

golden standard for isoform assignment is sequence alignment of the combined $\alpha 1$ and $\alpha 2$ domains. All CD1 genes from 15 mammals that were recovered, group with one of the five known CD1 isoforms, but were not always correctly named by automated annotation (Fig. 1). The clear grouping with the five isoforms suggests that the isoform nomenclature that was based on the human CD1 locus, appropriately describes all currently known mammalian CD1 isoforms. With more than 10 CD1 genes, the little brown bat, *Myotis lucifugus* can be considered a species with high numbers of CD1 genes. Intermediate numbers are found primates, alpaca, and elephant. These data confirm once more that mouse and rat are atypical with the lowest absolute number of CD1 genes (two and one, respectively). Though not identical, the gene numbers obtained by BLAST-based searches and automated annotation are largely consistent (Reinink and Van Rhijn 2016). Because misassemblies of repeated sequences and unmerged overlaps due to polymorphisms resulting in artificial duplications can occur in incompletely assembled genomes, we expect that in some species, the exact numbers of CD1 genes will be adjusted in the future. However, it is clear that humans should be considered to have an intermediate number of CD1 genes, while many other mammals have more extensive families of CD1 genes. Large CD1 families are often dominated by multiplied CD1a genes and to a lesser degree, CD1b, CD1c, and CD1d genes.

Mammalian MR1 genes

Searches in whole genomes (Table 1), as well as other published data show that primates have two MR1 genes. In humans and chimpanzees one of these genes is known to be functional and the other one is a pseudogene (Parra-Cuadrado et al. 2001; Parra-Cuadrado et al. 2000). No MR1 gene could be identified in the rabbit genome, which is in line with recently published data on lagomorphs (rabbit and pika) (Boudinot et al. 2016). The panda and dog MR1 genes that resulted from BLAST-based searches (Table 1) are most likely pseudogenes because it was shown that members of the order of Carnivora, including cats, dogs, ferrets and pandas have a MR1 pseudogene and lack a functional MR1 gene (Boudinot et al. 2016). Other mammals (alpaca, elephant, horse, bats) also have one MR1 gene. While in dolphin no MR1 gene was found, this does not conclusively prove the absence of MR1 in these species because especially the dolphin genome is still incompletely assembled. The sloth MR1 gene has a gap in the current version of the sloth genome (choHof1).

Functional specialization of CD1 isoforms

Interspecies comparisons between individual CD1 proteins have been made based on sequence alignment, protein models (canine versus human CD1a (Loringh van Beeck et al. 2008)), or crystal structures (bovine versus human CD1b (Girardi et al. 2010), bovine versus human CD1d (Wang et al. 2012), and murine versus human CD1d (Koch et al. 2005)). CD1 proteins within one species differ from each other in ways that suggest functional specialization. One aspect that determines the function of a CD1 protein is the

size and shape of the antigen binding cleft, which is formed by the $\alpha 1$ and $\alpha 2$ helices. Among the human CD1 proteins, the biggest size difference is observed between CD1a and CD1b. In addition, there are significant differences in shape with the CD1b cleft consisting of four pockets (Gadola et al. 2002) of which three are interconnected and the CD1a cleft consisting of one cavity (Zajonc et al. 2003). These differences translate into differences in size and shape of antigens that can be bound by the human CD1a and CD1b proteins. Because the CD1 isoforms form interspecies groups based on phylogenetic relationship based on the sequences of their $\alpha 1$ and $\alpha 2$ helices, functional specializations of CD1 molecules that relate to the antigen binding cleft are generally conserved across species. Especially the size of the cleft seems to be well conserved when human CD1 isoforms are compared with an ortholog. However, as seen in the bovine CD1d molecule, despite a very high homology between human and bovine CD1d, tryptophan 166 in the bovine protein, which is a cysteine in humans and mice, blocks part of the A' pocket so that long fatty acyl chains will not fit (Wang et al. 2012). A comparable situation exists when human CD1b and bovine CD1b3 are compared: the tunnel that is present in the human CD1b protein is closed in bovine CD1b3 by valine 98, which prevents the binding of extremely long ligands (Girardi et al. 2010). Structural aspects of these CD1 molecules will be reviewed by Zajonc et al. in this special issue of Immunogenetics. Thus, despite very high sequence homology and overall structural resemblance between orthologs, a single amino acid difference can have profound functional impact.

One specific feature of CD1 molecules that is difficult to predict from crude genomic data, is the cytoplasmic tail, which determines the transport and subcellular location of individual CD1 isoforms (Moody and Porcelli 2003). The cytoplasmic tail is encoded by the very small exon 5 that is located between two longer introns of variable length, and can not reliably be predicted based on the genomic sequence of the CD1 gene. However, the many cDNA sequences that are available have provided a number of cytoplasmic tails that allows for comparative analysis and predictions on intracellular trafficking of these CD1 proteins (Table 2).

Intracellular trafficking is determined by the CD1 cytoplasmic tail

Cellular lipids are bound in the antigen binding cleft during synthesis of the CD1 protein. When the CD1 protein recycles and encounters other lipids in an endosomal compartment, these may replace the initial endoplasmic reticulum-derived lipid. The type of endosomal compartment a lipid travels to is thus a determining factor in the kind of antigenic lipids it presents. The cytoplasmic tails of CD1 molecules may or may not contain short consensus sequences for adaptor proteins. CD1b, CD1c, or CD1d molecules often carry a dileucine motif and/or a tyrosine-based YXXZ motif, where X is any amino acid, and Z is a bulky hydrophobic amino acid. YXXZ motifs interact with adaptor proteins and are responsible for

recycling from the cell surface to intermediate and late endosomal compartments (Briken et al. 2002; Sugita et al. 2002). CD1e does not appear at the cell surface and does not recycle and has a function in lipid antigen loading and processing (Angenieux et al. 2005; de la Salle et al. 2005).

Table 2: Cytoplasmic tails of CD1 proteins in mammals

Gene name (Alias)	Cytoplasmic tail	Motif	Reference cdna
boCD1a2	RKSWSTYMSDA		(Nguyen et al. 2015)
boCD1a1	WKHWTHRESPSVLPLE		(Van Rhijn et al. 2006)
boCD1a			
canCD1a2	KAHWRPQCMDFPSEREPSSSSTYLNPAQH		(Schjaerff et al. 2016)
canCD1a6	KRWKTHNRPCQCTDFPLK		(Looringh van Beeck et al. 2008)
canCD1a9	KAHWRPQCTDFPSEQEPSSGSSTYLNPAQH		(Looringh van Beeck et al. 2008; Schjaerff et al. 2016)
canCD1a8.1			
canCD1a8	KRWKAH		(Looringh van Beeck et al. 2008; Schjaerff et al. 2016)
canCD1a8.2			
eqCD1a1	THCEACTIVPLK		(Dossa et al. 2014)
eqCD1a2	IRHQLQRLLPLD	dileucine	(Dossa et al. 2014)
eqCD1a3	IHSELPRTLLPLE	dileucine	(Dossa et al. 2014)
eqCD1a4	VISISVSILVRKPCATPRTPLPSQ		(Dossa et al. 2014)
eqCD1a5	RSCESASNLLWNEIPGAQDPGHI	dileucine	(Dossa et al. 2014)
eqCD1a6	WLRKRWRTRCEPPSNLISLE		(Dossa et al. 2014)
eqCD1a7	WLRKRGRTHCEFPRTCLPLE		(Dossa et al. 2014)
huCD1a	RKRFC		(Calabi and Milstein 1986; Martin et al. 1987)
pigCD1a1	WHRKHWKHCDPSSALHRL		(Chun et al. 1999; Eguchi-Ogawa et al. 2007)
pCD1.1			
rabCD1a1	RKCWIHHGPLETLLPLQ	dileucine	(Hayes and Knight 2001)
rabCD1a2	KKRWSHHGSPNSLLPLK	dileucine	(Hayes and Knight 2001)
boCD1b1	RFMGSHRVGHD		(Nguyen et al. 2015; Van Rhijn et al. 2006)
boCD1b3	RRWSYQNIL	YXXZ, dileucine	(Nguyen et al. 2015; Van Rhijn et al. 2006)
boCD1b5	RRWSYQNIL	YXXZ, dileucine	(Nguyen et al. 2015)
canCD1b	RRWSYQNSIS	YXXZ	(Looringh van Beeck et al. 2008)
eqCD1b1	SYQNIS	YXXZ	(Dossa et al. 2014)
eqCD1b2	SYLNIP	YXXZ	(Dossa et al. 2014)
gpCD1b1	RRWSYEDIL	YXXZ, dileucine	(Dascher et al. 1999; Hiromatsu et al. 2002)
gpCD1b2	KHWSYQDIL	YXXZ, dileucine	(Dascher et al. 1999; Dascher et al. 2002)
gpCD1b3	RRLRCEGIF		(Dascher et al. 1999; Dascher et al. 2002)
gpCD1b4	RRWSYEDIF	YXXZ	(Dascher et al. 1999)
huCD1b	RRRSYQNIP	YXXZ	(Martin et al. 1987)
ovCD1b	RRWSYQNIL	YXXZ, dileucine	(Ferguson et al. 1996)
scd1b42			

Table 2: Continued

ovCD1b	RRWSHRNIL	dileucine	(Ferguson et al. 1996)
scd1b52			
ovCD1b	RRWSYLTL	YXXZ, dileucine	(Ferguson et al. 1996)
scd1a25			
pigCD1b	RRWSYQSVL	YXXZ	(Eguchi-Ogawa et al. 2007)
rabCD1b	RRRSYQNIL	YXXZ, dileucine	(Calabi et al. 1989; Hayes and Knight 2001)
canCD1c	RKCCSYQDIP	YXXZ	(Looringh van Beeck et al. 2008)
eqCD1c	SYQNIQRDSSPCFPHCNENCTAQEHRTTE	YXXZ	(Dossa et al. 2014)
gpCD1c1	KRCTYQGIQ	YXXZ	(Dascher et al. 1999)
gpCD1c2	KRCTYQGIP	YXXZ	(Dascher et al. 1999)
gpCD1c3	KKCCTYQGIQ ^a	YXXZ	(Dascher et al. 1999)
huCD1c	KKHCSYQDIL	YXXZ, dileucine	(Martin et al. 1987)
eqCD1d	KKRSSYQDIL	YXXZ, dileucine	(Dossa et al. 2014; Looringh van Beeck et al. 2009)
gpcD1d	RRGRSYQDIL	YXXZ, dileucine	(Looringh van Beeck et al. 2009)
huCD1d	RFKQTSYQGVL	YXXZ, dileucine	(Balk et al. 1989)
lafCD1d	KRHCS		(Looringh van Beeck et al. 2009)
moCD1d1	RRRSAYQDIR	YXXZ	(Bradbury et al. 1988)
moCD1d2	RRRSAYQDIR	YXXZ	(Bradbury et al. 1988)
ovCD1d	RKHRRYQDIS	YXXZ, dileucine	(Rhind et al. 1999)
scd1d			
pigCD1d	RRRVYQNIQ	YXXZ	(Eguchi-Ogawa et al. 2007)
rabCD1d	RRRCSYQGIL	YXXZ, dileucine	(Calabi et al. 1989; Looringh van Beeck et al. 2009)
ratCD1d	RRRSYQDIM	YXXZ	(Ichimiya et al. 1994)

The cytoplasmic tails of CD1 proteins, grouped by isoform. Only tail sequences that have been confirmed by cDNA sequences are included. Species from which the sequences were derived are indicated as bo: bovine; can: canine; eq: equine; gp: guinea pig; hu: human; laf: African elephant; mo: mouse; ov: ovine; rab: rabbit. ^aThe sequence shown is based on GenBank sequence NM_001172855, but has been published as KKCCTYQGFP (Dascher et al. 1999).

Comparing the cytoplasmic tails of all cell surface-expressed isoforms across species, the following general picture emerges: CD1a tails are highly variable in length. Most CD1a tails contain histidine residues, with unknown function, or cysteines, which can be palmitoylated. The human CD1a cytoplasmic tail contains a cysteine residue that is palmitoylated and involved in incorporation in lipid rafts (Barral et al. 2008; Sloma et al. 2008), but the mechanism of recycling to the sorting or early endosomal compartment, which is so typical for human CD1a is unknown. None of the 17 known cytoplasmic tails of CD1a proteins has a dileucine and/or a tyrosine-based YXXZ motif, consistent with the absence from intermediate and late endosomes of the human CD1a protein, and strongly suggesting that this is an evolutionary conserved feature Table 2. CD1b, CD1c and CD1d tails are of

comparable length, typically contain a YXXZ motif, and sometimes an additional dileucine motif. In addition, all known CD1c cytoplasmic tails contain a cysteine of unknown function. There are exceptions to these general observations, like the bovine CD1b1 (Nguyen et al. 2015), guinea pig CD1b3 (Dascher et al. 1999), and African elephant CD1d (Looringh van Beeck et al. 2009) cytoplasmic tails, which lack any known motif. Indeed, guinea pig CD1b3 does not travel to late endosomes and rather has a CD1a-like subcellular distribution pattern (Dascher et al. 2002). However, the bovine CD1b3 and CD1b5 and guinea pig CD1b1, CD1b2, and CD1b4 have “typical” CD1b cytoplasmic tails. Thus, cows and guinea pigs have typical and atypical cytoplasmic tails among the proteins belonging to the CD1b isoform. For the African elephant no other cytoplasmic tails have been sequenced from cDNA yet. Generally speaking, it seems that when there are multiple proteins of one CD1 isoform in a species, there is functional diversification, but when a species has just one gene for a certain CD1 isoform, this one gene usually shows the “typical” combination of cleft and tail.

Expression patterns of CD1 isoforms

Human cortical thymocytes express high levels of CD1a, CD1b, CD1c, and CD1d on their cell surface. Broad CD1 expression in the thymus was confirmed in all other species where CD1 expression was studied (Hiromatsu et al. 2002; Howard et al. 1993; Looringh van Beeck et al. 2008; Van Rhijn et al. 2006). Outside the thymus, the general picture that emerges is that CD1d is widely expressed at a low level, with some specific high CD1d-expressing cells in certain tissues, like hepatic stellate cells in the liver (Winau et al. 2007), and marginal zone B cells in the spleen (Makowska et al. 1999). In humans, CD1b expression is limited to dendritic cells, CD1a is expressed by dendritic cells and at high levels by Langerhans cells in the skin, and CD1c is expressed by dendritic cells and subsets of B cells. Like humans, dogs express CD1a in the skin and thymus (Looringh van Beeck et al. 2008). However, only two of the three studied canine CD1a genes (out of the nine CD1a genes in the dog CD1 locus) were preferentially expressed in the skin. In rabbits, one of the two CD1a genes is preferentially expressed in the skin, while the other one has a more general expression pattern (Hayes and Knight 2001). In guinea pigs, different CD1b genes have different expression patterns in peripheral blood and spleen (Hiromatsu et al. 2002). Mice that are transgenic for the part of the human locus that encodes CD1a, CD1b, CD1c, and CD1e with their endogenous promoters, show a CD1 expression pattern that is surprisingly comparable to the human CD1 expression pattern in that all isoforms are expressed on thymocytes and lymph node dendritic cells, while CD1a stands out as highly expressed on Langerhans cells, and CD1c is expressed on B cells (Felio et al. 2009).

Closing remarks

A typical pattern of gene distribution has been described for immune function-related genes including genes that encode TCRs, immunoglobulins, classical MHC molecules and NK receptors. This pattern is the result of gene family expansion, diversification, and contraction or “birth and death” evolution, and usually leads to families of genes that include a considerable number of pseudogenes (Nei and Rooney 2005). Furthermore, complete loss of certain family members, and expansion of other family members are observed when contemporary animal species are compared. With its many duplications, deletions, and pseudogenes, the CD1 loci seem to follow this pattern. Although MR1 genes have been subject to inactivation in the carnivores, extensive gene family expansion and diversification was not observed in any of the species studied. The lack of functional diversification of MR1 genes is hard to explain but may be related to its seemingly exclusive interaction with the highly specialized MAIT cells.

Even though mammals vary widely in the numbers of CD1 genes they have, some general observations can be made and help understand the function of the different CD1 isoforms. The skin seems to be a preferred site for CD1a expression across species. While CD1a genes have undergone extensive multiplication in some species, even more so than CD1b and CD1c, none of the currently described CD1a proteins contains any known signal for endosomal location. Multiplication and diversification of genes may have occurred in response to changes in the environment, including pathogenic and non-pathogenic microbes, to which evolving mammals were exposed.

Acknowledgments

This work was supported by The Netherlands Organisation for Scientific Research (NWO) grant 824.02.002.

References

- Angenieux C, Fraissier V, Maitre B, Racine V, van der Wel N, Fricker D, Proamer F, Sachse M, Cazenave JP, Peters P, Goud B, Hanau D, Sibarita JB, Salamero J, de la Salle H (2005) The cellular pathway of CD1e in immature and maturing dendritic cells. *Traffic* 6:286-302
- Baker ML, Miller RD (2007) Evolution of mammalian CD1: marsupial CD1 is not orthologous to the eutherian isoforms and is a pseudogene in the opossum *Monodelphis domestica*. *Immunology* 121:113-21
- Balk SP, Bleicher PA, Terhorst C (1989) Isolation and characterization of a cDNA and gene coding for a fourth CD1 molecule. *Proceedings of the National Academy of Sciences of the United States of America* 86:252-6
- Barral DC, Cavallari M, McCormick PJ, Garg S, Magee AI, Bonifacino JS, De Libero G, Brenner MB (2008) CD1a and MHC Class I Follow a Similar Endocytic Recycling Pathway. *Traffic* 9:1146-57
- Boudinot P, Mondot S, Jouneau L, Teyton L, Lefranc MP, Lantz O (2016) Restricting nonclassical MHC genes coevolve with TRAV genes used by innate-like T cells in mammals. *Proceedings of the National Academy of Sciences of the United States of America*
- Bradbury A, Belt KT, Neri TM, Milstein C, Calabi F (1988) Mouse CD1 is distinct from and co-exists with TL in the same thymus. *Embo J* 7:3081-6
- Briken V, Jackman RM, Dasgupta S, Hoening S, Porcelli SA (2002) Intracellular trafficking pathway of newly synthesized CD1b molecules. *Embo J* 21:825-34
- Calabi F, Belt KT, Yu CY, Bradbury A, Mandy WJ, Milstein C (1989) The rabbit CD1 and the evolutionary conservation of the CD1 gene family. *Immunogenetics* 30:370-7
- Calabi F, Milstein C (1986) A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* 323:540-3
- Cheng Y, Belov K (2014) Characterisation of non-classical MHC class I genes in the Tasmanian devil (*Sarcophilus harrisi*). *Immunogenetics* 66:727-35
- Chun T, Wang K, Zuckermann FA, Gaskins HR (1999) Molecular cloning and characterization of a novel CD1 gene from the pig. *J Immunol* 162:6562-71
- Dascher CC (2007) Evolutionary biology of CD1. *Curr Top Microbiol Immunol* 314:3-26
- Dascher CC, Brenner MB (2003) Evolutionary constraints on CD1 structure: insights from comparative genomic analysis. *Trends Immunol* 24:412-8
- Dascher CC, Hiromatsu K, Naylor JW, Brauer PP, Brown KA, Storey JR, Behar SM, Kawasaki ES, Porcelli SA, Brenner MB, LeClair KP (1999) Conservation of a CD1 multigene family in the guinea pig. *J Immunol* 163:5478-88
- Dascher CC, Hiromatsu K, Xiong X, Sugita M, Buhlmann JE, Dodge IL, Lee SY, Roura-Mir C, Watts GF, Roy CJ, Behar SM, Clemens DL, Porcelli SA, Brenner MB (2002) Conservation of CD1 intracellular trafficking patterns between mammalian species. *J Immunol* 169:6951-8
- de la Salle H, Mariotti S, Angenieux C, Gilleron M, Garcia-Alles LF, Malm D, Berg T, Paoletti S, Maitre B, Mourey L, Salamero J, Cazenave JP, Hanau D, Mori L, Puzo G, De Libero G (2005) Assistance of microbial glycolipid antigen processing by CD1e. *Science* 310:1321-4
- Dossa RG, Alperin DC, Hines MT, Hines SA (2014) The equine CD1 gene family is the largest and most diverse yet identified. *Immunogenetics* 66:33-42
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-7
- Eguchi-Ogawa T, Morozumi T, Tanaka M, Shinkai H, Okumura N, Suzuki K, Awata T, Uenishi H (2007) Analysis of the genomic structure of the porcine CD1 gene cluster. *Genomics* 89:248-61
- Felio K, Nguyen H, Dascher CC, Choi HJ, Li S, Zimmer MI, Colmone A, Moody DB, Brenner MB, Wang CR (2009) CD1-restricted adaptive immune responses to Mycobacteria in human group 1 CD1 transgenic mice. *J Exp Med* 206:2497-509
- Ferguson ED, Dutia BM, Hein WR, Hopkins J (1996) The sheep CD1 gene family contains at least four CD1B homologues. *Immunogenetics* 44:86-96
- Gadola SD, Zaccari NR, Harlos K, Shepherd D, Castro-Palomino JC, Ritter G, Schmidt RR, Jones EY, Cerundolo V (2002) Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. *Nat Immunol* 3:721-726
- Girardi E, Wang J, Mac TT, Versluis C, Bhowruth V, Besra G, Heck AJ, Van Rhijn I, Zajonc DM (2010) Crystal structure of bovine CD1b3 with endogenously bound ligands. *J Immunol* 185:376-86
- Goldfinch N, Reinink P, Connelley T, Koets A, Morrison I, Van Rhijn I (2010) Conservation of mucosal associated invariant T (MAIT) cells and the MR1 restriction element in ruminants, and abundance of MAIT cells in spleen. *Vet Res* 41:62
- Han M, Hannick LI, DiBrino M, Robinson MA (1999) Polymorphism of human CD1 genes. *Tissue Antigens* 54:122-7
- Hashimoto K, Hirai M, Kurosawa Y (1995) A gene outside the human MHC related to classical HLA class I genes. *Science* 269:693-5
- Hayes SM, Knight KL (2001) Group 1 CD1 genes in rabbit. *J Immunol* 166:403-10
- Hiromatsu K, Dascher CC, Sugita M, Gingrich-Baker C, Behar SM, LeClair KP, Brenner MB, Porcelli SA (2002) Characterization of guinea-pig group 1 CD1 proteins. *Immunology* 106:159-72
- Howard CJ, Sopp P, Bembridge G, Young J, Parsons KR (1993) Comparison of CD1 monoclonal antibodies on bovine cells and tissues. *Vet Immunol Immunopathol* 39:77-83
- Ichimiya S, Kikuchi K, Matsuura A (1994) Structural analysis of the rat homologue of CD1. Evidence for evolutionary conservation of the CD1D class and widespread transcription by rat cells. *J Immunol* 153:1112-23
- Katabami S, Matsuura A, Chen HZ, Imai K, Kikuchi K (1998) Structural organization of rat CD1 typifies evolutionarily conserved CD1D class genes. *Immunogenetics* 48:22-31
- Koch M, Stronge VS, Shepherd D, Gadola SD, Mathew B, Ritter G, Fersht AR, Besra GS, Schmidt RR, Jones EY, Cerundolo V (2005) The crystal structure of human CD1d with and without alpha-galactosylceramide. *Nat Immunol* 6:819-26
- Loringh van Beeck FA, Reinink P, Hermesen R, Zajonc DM, Laven MJ, Fun A, Troskie M, Schoemaker NJ, Morar D, Lenstra JA, Vervelde L, Rutten VP, van Eden W, Van Rhijn I (2009) Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants. *Mol Immunol* 46:1424-31
- Loringh van Beeck FA, Zajonc DM, Moore PF, Schlotter YM, Broere F, Rutten VP, Willemse T, Van Rhijn I (2008) Two canine CD1a proteins are differentially expressed in skin. *Immunogenetics* 60:315-24
- Makowska A, Faizunnessa NN, Anderson P, Midtvedt T, Cardell S (1999) CD1high B cells: a population of mixed origin. *European Journal of Immunology* 29:3285-94
- Martin LH, Calabi F, Lefebvre FA, Bilsland CA, Milstein C (1987) Structure and expression of the human thymocyte antigens CD1a, CD1b, and CD1c. *Proc Natl Acad Sci U S A* 84:9189-93

- Miller MM, Wang C, Parisini E, Coletta RD, Goto RM, Lee SY, Barral DC, Townes M, Roura-Mir C, Ford HL, Brenner MB, Dascher CC (2005) Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. *Proc Natl Acad Sci U S A* 102:8674-9
- Moody DB, Porcelli SA (2003) Intracellular pathways of CD1 antigen presentation. *Nat.Rev.Immunol.* 3:11-22
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121-52
- Nguyen TK, Reinink P, El Messlaki C, Im JS, Ercan A, Porcelli SA, Van Rhijn I (2015) Expression patterns of bovine CD1 in vivo and assessment of the specificities of the anti-bovine CD1 antibodies. *PLoS One* 10:e0121923
- Oteo M, Arribas P, Setien F, Parra JF, Mirones I, Gomez del Moral M, Martinez-Naves E (2001) Structural characterization of two CD1A allelic variants. *Human Immunology* 62:1137-41
- Parra-Cuadrado JF, Gomez del Moral M, Garcia-Pavia P, Setien F, Martinez-Naves E (2001) Characterization of the MHC class I-related MR1 locus in nonhuman primates. *Immunogenetics* 53:643-8
- Parra-Cuadrado JF, Navarro P, Mirones I, Setien F, Oteo M, Martinez-Naves E (2000) A study on the polymorphism of human MHC class I-related MR1 gene and identification of an MR1-like pseudogene. *Tissue Antigens* 56:170-2
- Reinink P, Van Rhijn I (2016) In Silico method for identification of MHC class I-like molecules in whole genomes. *BioRxiv* 10.1101/046607
- Rhind SM, Hopkins J, Dutia BM (1999) Amino-terminal sequencing of sheep CD1 antigens and identification of a sheep CD1D gene. *Immunogenetics* 49:225-30
- Salomonsen J, Sorensen MR, Marston DA, Rogers SL, Collen T, van Hateren A, Smith AL, Beal RK, Skjodt K, Kaufman J (2005) Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proc Natl Acad Sci U S A* 102:8668-73
- Schjaerff M, Keller SM, Fass J, Froenicke L, Grahn RA, Lyons L, Affolter VK, Kristensen AT, Moore PF (2016) Refinement of the canine CD1 locus topology and investigation of antibody binding to recombinant canine CD1 isoforms. *Immunogenetics* 68:191-204
- Shiina T, Ando A, Suto Y, Kasai F, Shigenari A, Takishima N, Kikkawa E, Iwata K, Kuwano Y, Kitamura Y, Matsuzawa Y, Sano K, Nogami M, Kawata H, Li S, Fukuzumi Y, Yamazaki M, Tashiro H, Tamiya G, Kohda A, Okumura K, Ikemura T, Soeda E, Mizuki N, Kimura M, Bahram S, Inoko H (2001) Genomic anatomy of a premier major histocompatibility complex paralogous region on chromosome 1q21-q22. *Genome Res* 11:789-802
- Sloma I, Zilber MT, Vasselon T, Setterblad N, Cavallari M, Mori L, De Libero G, Charron D, Mooney N, Gelin C (2008) Regulation of CD1a surface expression and antigen presentation by invariant chain and lipid rafts. *Journal of Immunology* 180:980-7
- Sugita M, Cao X, Watts GF, Rogers RA, Bonifacino JS, Brenner MB (2002) Failure of trafficking and antigen presentation by CD1 in AP-3-deficient cells. *Immunity* 16:697-706
- Van Rhijn I, Koets AP, Im JS, Piebes D, Reddington F, Besra GS, Porcelli SA, van Eden W, Rutten VP (2006) The bovine CD1 family contains group 1 CD1 proteins, but no functional CD1d. *J Immunol* 176:4888-93
- Van Rhijn I, Moody DB (2015) CD1 and mycobacterial lipids activate human T cells. *Immunological Reviews* 264:138-53
- Wang J, Guillaume J, Pauwels N, Van Calenbergh S, Van Rhijn I, Zajonc DM (2012) Crystal structures of bovine CD1d reveal altered aGalCer presentation and a restricted A' pocket unable to bind very long-chain lipids. *Plos One* 7:1-10
- Winau F, Hegasy G, Weiskirchen R, Weber S, Cassan C, Sieling PA, Modlin RL, Liblau RS, Gressner AM, Kaufmann SH (2007) Ito cells are liver-resident antigen-presenting cells for activating T cell responses. *Immunity* 26:117-29
- Yamaguchi H, Kurosawa Y, Hashimoto K (1998) Expanded genomic organization of conserved mammalian MHC class I-related genes, human MR1 and its murine ortholog. *Biochem Biophys Res Commun* 250:558-64
- Yang Z, Wang C, Wang T, Bai J, Zhao Y, Liu X, Ma Q, Wu X, Guo Y, Ren L (2015) Analysis of the reptile CD1 genes: evolutionary implications. *Immunogenetics* 67:337-46
- Zajonc DM, Elsliger MA, Teyton L, Wilson IA (2003) Crystal structure of CD1a in complex with a sulfatide self antigen at a resolution of 2.15 Å. *Nat.Immunol.* 4:808-815



Chapter 3

CD1b presents self and *Borrelia burgdorferi* diacylglycerols to human T cells

3

Peter Reinink^{1,2}, Michael N. T. Souter^{3,4}, Tan-Yun Cheng², Tamara van Gorkom^{5,6}, Stefanie Lenz¹, Joanna Kubler-Kielb⁷, Klemen Strle⁸, Kristin Kremer⁶, Steven F.T. Thijsen⁵, Allen C. Steere⁸, Dale I. Godfrey^{3,4}, Daniel G. Pellicci^{3,4,9}, D. Branch Moody², Ildiko Van Rhijn^{1,2}

.....

¹Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584CL Utrecht, The Netherlands

²Brigham and Women's Hospital Division of Rheumatology, Immunology and Allergy and Harvard Medical School, Boston, MA 02115, USA

³Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC, 3000, Australia.

⁴ARC Centre of Excellence in Advanced Molecular Imaging, University of Melbourne, Parkville, VIC, 3010, Australia

⁵Department of Medical Microbiology and Immunology, Diaconessen Hospital, Utrecht, the Netherlands.

⁶Centre for Infectious Diseases Research, Diagnostics and Laboratory Surveillance, Centre for Infectious Diseases Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

⁷National Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

⁸Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

⁹Murdoch Children's Research Institute, Parkville, Victoria 3052, Australia

European journal of immunology: 10 March 2019

Abstract

Lyme disease is a common multi-system disease caused by infection with a tick-transmitted spirochete, *Borrelia burgdorferi* and related *Borrelia* species. The monoglycosylated diacylglycerol known as *B. burgdorferi* glycolipid II (BbGL-II) is a major target of antibodies in sera from infected individuals. Here we show that CD1b presents BbGL-II to human T cells and that the T cell receptor (TCR) mediates the recognition. However, we did not detect increased frequency of CD1b-BbGL-II binding T cells in the peripheral blood of Lyme disease patients compared to controls. Unexpectedly, mapping the T cell specificity for BbGL-II-like molecules using tetramers and activation assays revealed a concomitant response to CD1b-expressing antigen presenting cells in absence of BbGL-II. Further, among all major classes of self-lipid tested, BbGL-II responsive TCRs show strong cross-reactivity to diacylglycerol, a self-lipid antigen with structural similarities to BbGL-II. Extending prior work on MHC and CD1b, CD1c, and CD1d proteins, this study provides evidence for cross-reactive CD1b-restricted T cell responses to bacterial and self-antigens, and identifies chemically defined targets for future discovery of self and foreign antigen cross-reactive T cells.

Introduction

Major histocompatibility complex (MHC) molecules present peptides from self or foreign origin to T cells. Outside the MHC, the human genome encodes additional antigen presenting molecules that produce proteins that are structurally related to MHC class I proteins, including CD1a, CD1b, CD1c, and CD1d¹. CD1-β2microglobulin complexes are expressed on key antigen presenting cells (APCs) such as B cells, macrophages and dendritic cells², where they bind many types of lipid antigens for presentation to T cell receptors (TCRs). An important difference between MHC and CD1 genes is that CD1 genes show extremely low rates of polymorphism. Although single nucleotide polymorphisms in CD1 are known, they have not been shown to affect lipid antigen display or TCR recognition. Thus, unlike peptide antigens, where the pattern of response varies substantially among individuals, CD1 allows development of lipid antigens as immunogens for vaccines or targets of diagnostic studies that could work in any genetic background^{3,4}.

Among human CD1 molecules, CD1b was the first for which lipid antigen presentation was shown^{5,6}. High levels of CD1b are present on DC's or induced on monocytes differentiating into DCs^{7,8}, where it presents self and foreign lipids to T cells. Many foreign lipids that are known to be presented by CD1b are derived from mycobacteria or closely related actinobacteria, like glucose monomycolate, glycerol monomycolate⁹⁻¹¹, mycolic acid^{5,12,13}, and diacylated sulfoglycolipids^{14,15}. Also, some self-lipids from mammalian cells are presented by CD1b, including phosphatidylglycerol, phosphatidic acid, phosphatidylethanolamine, phosphatidylinositol, and gangliosides¹⁶⁻¹⁹. Interestingly, for some of these self-lipids, such as phosphatidylglycerol and phosphatidic acid, the T cells that recognize them cross-react with bacterial versions of the lipids, such as lipids found in *Salmonella* and *Staphylococcus* species¹⁸.

Here we tested the possibility that CD1b might play a role in Lyme disease, which is a tick-borne bacterial infection²⁰. The initial site of infection is the skin, where the spread of bacteria and immune response often begins as an expanding skin lesion, erythema migrans. In untreated patients, this skin lesion may be followed weeks later by neuroborreliosis or carditis, and months later, by arthritis. In North America *Borrelia burgdorferi* is the causative agent of Lyme disease, while in Europe and Asia *B. afzelii* and *B. garinii* are the predominant species²¹. Infections caused by *B. burgdorferi* in North America tend to be more arthritogenic, whereas *B. garinii* causes a more neurotropic disease. *B. afzelii* predominantly causes skin manifestations²². Most patients respond to sterilizing regimens of antibiotics. However, some patients have persistent or worsening, long lasting synovitis, called post-infectious or antibiotic-refractory Lyme arthritis where live *Borrelia* cannot be detected²³. The synovial lesion in these patients is associated with chronic inflammation, fibrosis, and autoimmune phenomena²⁴⁻²⁷.

B. burgdorferi has two major glycolipids, *Borrelia burgdorferi* glycolipid I (BbGL-I) and *Borrelia burgdorferi* glycolipid II (BbGL-II), which together comprise 36 % of the total lipid mass of the bacteria. These molecules belong to different lipid classes: BbGL-I is cholesterol-based, and BbGL-II is diacylglycerol-based. However, both lipids are components of both the inner and outer membrane^{28,29}, and both elicit antibody responses in infected people^{30,31}. In one study of North American patients, 52 % of patients with neuroborreliosis had weak IgG antibody responses to BbGL-II, and almost all patients with Lyme arthritis, had strong IgG reactivity with both BbGL-I and BbGL-II³². Although BbGL-I and BbGL-II were identified in *B. burgdorferi*, both lipids were also detected in *B. afzelii* and *B. garinii*³³. The gene named monogalactosyl-1,2-diacylglycerol synthase, *mgs*, which is thought to catalyze the synthesis of BbGL-II, is present in all three pathogenic *Borrelia* species³⁴. Thus, BbGL-II would be a potential glycolipid antigen for use in vaccines and diagnostics.

Several prior studies suggested the hypothesis that CD1b might present *Borrelia* lipids to human T cells. First, while modelling suggested that BbGL-I fits into the antigen binding cleft of human CD1c³⁵, we predicted that BbGL-II would fit in the antigen binding cleft of CD1b based on its 2-tailed glycolipid structure and size^{36,37}. Second, NKT cells are activated by CD1d and BbGL-II³⁸, and depletion of NKT cells or CD1d render mice more sensitive to infection^{39,40}. Third, human CD1b and CD1c transcripts and proteins are upregulated by TLR ligands or live *B. burgdorferi*. Studies of ex vivo *B. burgdorferi* infected human skin explants and erythema migrans biopsies show CD1b protein upregulation to high levels at the site of infection⁴¹. Because CD1b is rarely expressed in human tissues in the periphery, these findings suggested a model in which localized infection rapidly upregulates CD1b proteins at the site of infection, where bacterial lipids are being shed⁴². However, direct evidence for *Borrelia* lipid antigens for CD1b is lacking. In this study we investigated whether BbGL-II can be presented by CD1b to human T cells.

Results

Characterization of BbGL-II

To characterize in-house produced synthetic BbGL-II (1, 2 di-*O*-oleoyl-3-*O*- α -D-galactopyranosyl-*sn*-glycerol)³¹, we carried out positive mode, multi-stage CID-ESI-MS (Fig. 1). The MS² collision pattern of BbGL-II (C₄₅H₈₂O₁₀Na⁺, calculated *m/z* 805.6, detected *m/z* 805.9) shows loss of 282 u, representing the mass of any oleic acyl (C18:1) moiety. The MS³ collision pattern shows subsequent loss of the hexose (162 u), another oleic acid (282 u), or acylglycerol (338 u). These data confirm that the BbGL-II used in this study is intact and identifies key components, including two oleic acyl units.

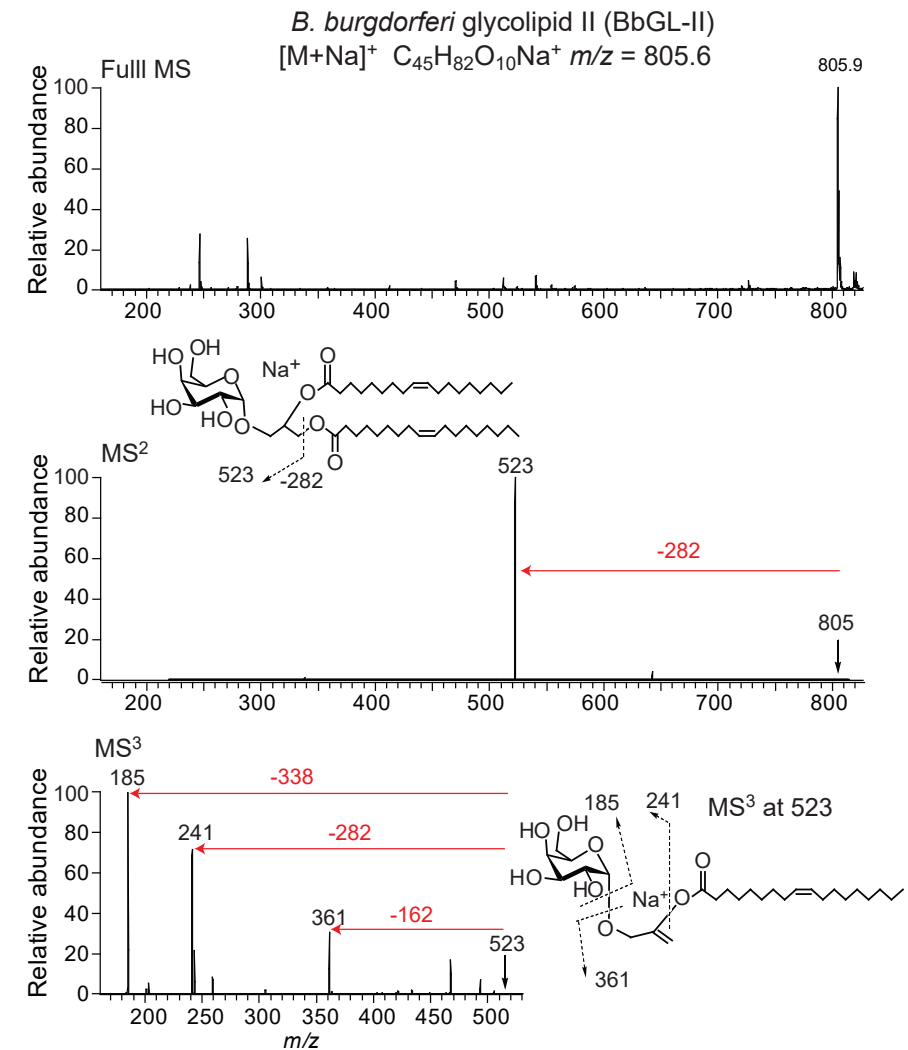


Figure 1. Mass spectrometry on synthetic BbGL-II lipid.

Positive ion mode ESI-MS of synthetic BbGL-II gave an expected major ion at *m/z* 805.9 (top panel), which was subjected to multistage CID-MS to yield fragment ions as indicated (middle and bottom panels).

BbGL-II specific T cells

To identify CD1b-BbGL-II specific T cells, we used peripheral blood mononuclear cells (PBMC) from a buffy coat from donor 24 (BC24). This donor was considered immune based on the presence of an erythema migrans lesion but did not have clinical signs of Lyme disease. After 14 days of expansion of T cells with anti-CD3, ~0.3 % of cells stained positively with CD1b-BbGL-II tetramers (Fig. 2A). These cells were sorted and expanded

to yield CD1b-BbGL-II tetramer positivity in 1.4 % of cells. After a second round of sorting and expansion with CD1b-BbGL-II tetramers, a clear CD1b-BbGL-II tetramer⁺ CD4⁻ population was identified. This large, homogeneously staining population represented 14 % of cells, which we named T cell line BC24A. The CD4⁺ cells contained CD1b-specific populations with a very broad reactivity pattern, including CD1b-phosphatidylglycerol¹⁹, and were not analyzed here because we were interested in discovering BbGL-II-specific T cells. Further analysis showed that BC24A T cells do not detectably bind to mock-treated CD1b tetramers (Fig. 2B), which are called CD1b-endo tetramers because they contain endogenous cellular lipids from the mammalian expression system for CD1b¹⁸. In contrast to nearly all previously reported CD1b reactive T cells lines, which are CD4⁺ or CD4⁻CD8⁻^{11,18,43}, BC24A cells uniformly express high levels of CD8.

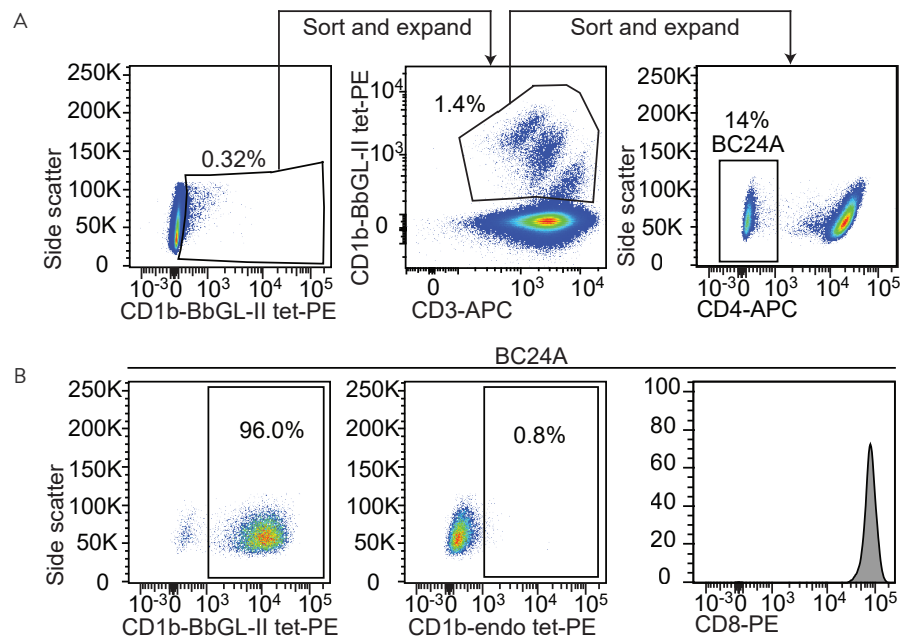


Figure 2. Isolation of CD1b-BbGL-II binding T cells.

A. PBMCs from blood donor BC24 (Table 1) with erythema migrans were sorted based on CD1b-BbGL-II tetramer binding. After three rounds of sorting and expansion, the 96 % of CD4⁻ T cells that stain with CD1b-BbGL-II tetramers were further tested for staining with CD1b-endo tetramers carrying endogenous (endo) lipids and an antibody against CD8 α . Positive CD1b-BbGL-II tetramer, negative CD1b-endo, and negative CD4 staining have been observed in five independent experiments. CD8 staining was performed once.

BC24A T cells were subjected to a single-cell TCR sequencing method⁴⁴, which yielded six separate interpretable TCR α and β chain pairs that were all identical. The BC24A TCR consists of an α chain containing the gene segments TRAV36/DV7*04 and TRAJ22*01 and a β chain containing the gene segments TRBV4-1*01 and TRBJ1-4*01 (Fig. 3A). Unlike the conserved α chain sequences of invariant T cells, such as NKT cells, the BC24A TCR contains seven non-template-encoded nucleotides. Invariant TCRs such as NKT cells and germline-encoded mycolyl reactive (GEM) T cells use predominantly or solely germline encoded sequences. So this finding suggests that BC24A is a representative of the private T cell repertoire unique to this individual. Whereas CD1 tetramers typically bind TCRs, there is some prior evidence for staining of non-T cells or binding to non-TCR proteins like ILT4^{45,46}. Therefore, to assess whether CD1b-BbGL-II tetramers bind this $\alpha\beta$ TCR, we performed a TCR transfer experiment. Jurkat 76 (J76) cells, which are deficient for both TCR α and β chains⁴⁷, were transduced with the BC24A TCR and CD3, both using green fluorescent protein (GFP) as a transduction marker. CD1b-BbGL-II tetramers selectively stained TCR-transduced, GFP⁺ cells but not untransduced cells, demonstrating that tetramer binding is TCR dependent (Fig. 3B). Like the primary T cell population BC24A, staining of TCR transduced J76 cells with CD1b-endo tetramers did not show detectable binding, demonstrating that binding between CD1b and the TCR is dependent on the BbGL-II glycolipid (Fig. 3B). Thus, BbGL-II is a CD1b-presented antigen for T cells that undergoes cognate recognition by an $\alpha\beta$ TCR.

T cell populations in random donors and *Borrelia*-infected people

Next, we sought to measure CD1b-BbGL-II and CD1b-endo tetramer staining in polyclonal T cell populations (Fig. 4, Table 1, Supplemental Figs. 1-5). We used anti-CD3-expanded T cells from PBMC from four Lyme arthritis, eight Lyme neuroborreliosis patients, and six random blood bank donors from the Netherlands (Fig. 4A). In contrast to the brightly staining cells seen after CD1b-BbGL-II tetramer sorting and expansion (Fig. 2), *ex vivo* staining did not yield cell populations that were well separated from the tetramer negative cells and did not result in significant differences between CD1b-endo and CD1b-BbGL-II tetramer staining in any of the groups. Prior studies^{18,46} suggest that T cell precursor frequencies of 10^{-5} to 10^{-4} above background are needed for detection of antigen-specific T cells. Thus, CD1b-BbGL-II tetramer staining was insufficiently sensitive to detect T cells these expanded polyclonal T cell populations. Because phospholipids are abundantly present in CD1b-endo¹⁹, another way of formulating this conclusion is to say that BbGL-II-specific T cells are not more abundant than phospholipid-specific T cells. T cell expansion was necessary from PBMC from Lyme disease patients because cell numbers directly *ex vivo* were insufficient for these experiments. However, we were able to perform experiments on PBMC from blood bank donors directly *ex vivo* (Fig. 4B). Also directly *ex vivo*, no significant differences between CD1b-endo and CD1b-BbGL-II tetramer staining were detected.

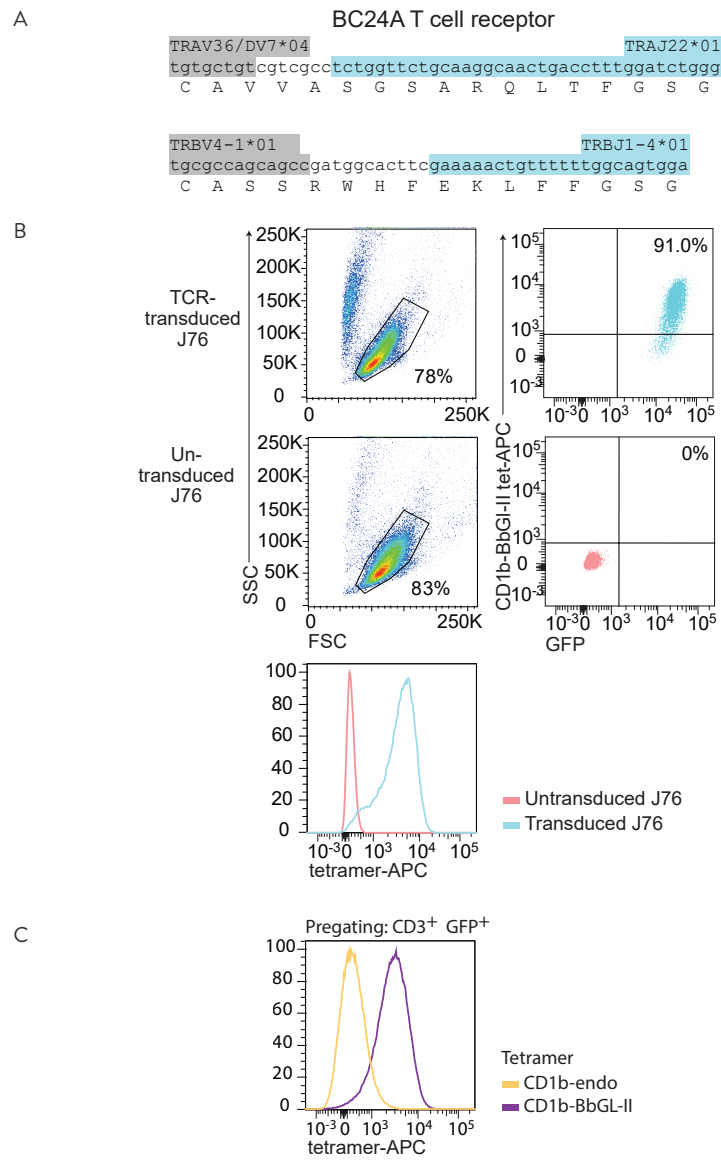


Figure 3. TCR dependent tetramer binding.

A. The CDR3 domain of the CD1b-BbGL-II binding TCR were determined by single cell TCR sequencing, where six independently obtained single cells showed the same CDR3 sequence. The nucleotides of the variable region (light grey) N-region additions (white) and joining segment (dark grey) are highlighted. B. J76 cells transduced with the BC24A TCR and untransduced J76 cells were stained with CD1b-BbGL-II tetramer. C: J76 cells transduced with the CD1b-BbGL-II specific TCR were stained with CD1b-BbGL-II and CD1b-endo tetramers. Four independent experiments were performed with comparable results. On representative experiment is shown.

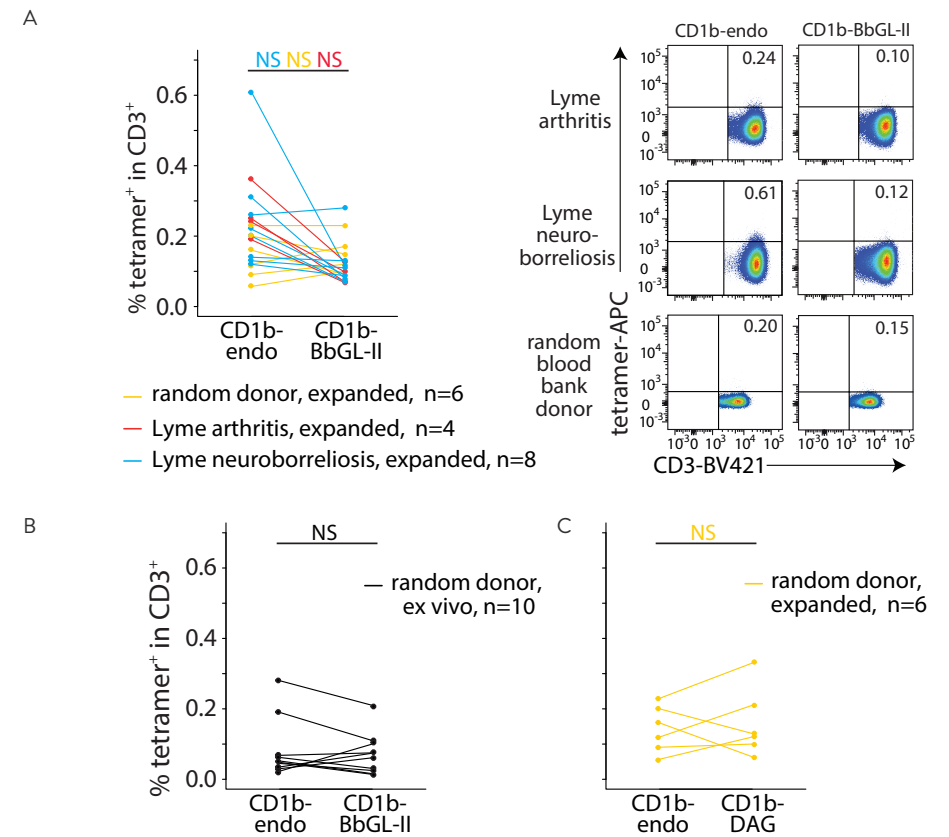


Figure 4. Tetramer staining of T cells from Lyme disease patients.

T cells that were expanded for two weeks (A, C) or PBMC that were used directly ex vivo (B) were stained with an α CD3 antibody and the indicated tetramers (CD1b-BbGL-II, CD1b-endo, or CD1b-DAG). The cells were pre-gated based on forward and side scatter and CD3. FACS data of three representative subjects (one Lyme arthritis, one Lyme neuroborreliosis, and one random blood bank donor) are shown in A. The complete set of flowcytometric data is shown in Supplemental Figs. 1-5. The percentages CD3⁺ tetramer⁺ cells are indicated. NS: not significant ($P > 0.05$) as calculated by paired Wilcoxon ranked sums tests and Benjamini & Hochberg correction for multiple testing.

Table 1. Overview of the blood donors used in this study.

Donor	Donor status	Time since diagnosis (months)
1	Lyme neuroborreliosis	6
2	Lyme arthritis	6
3	Lyme neuroborreliosis	0
4	Lyme neuroborreliosis	24
5	Lyme arthritis	12
6	Lyme arthritis	0
7	Lyme neuroborreliosis	12
8	Lyme arthritis	1
9	Lyme neuroborreliosis	0
11	Lyme neuroborreliosis	1
12	Lyme neuroborreliosis	3
14	Lyme neuroborreliosis	24
BC1	Random blood bank donor	NA
BC2	Random blood bank donor	NA
BC8	Random blood bank donor	NA
BC9	Random blood bank donor	NA
BC13	Random blood bank donor	NA
BC14	Random blood bank donor	NA
BC17	Random blood bank donor	NA
BC18	Random blood bank donor	NA
BC19	Random blood bank donor	NA
BC20	Random blood bank donor	NA
BC24	Erythema migrans	unknown

All patients and blood bank donors are from the Netherlands.

NA: not applicable; BC: buffy coat

T cell activation assays demonstrate CD1b and self-antigen reactivity

To determine whether BC24A T cells can be functionally activated by cellular CD1b-BbGL-II complexes, IFN- γ production was measured with an ELISPOT assay using CD1b-expressing C1R cells as APCs (Fig. 5A). No response was seen upon stimulation with C1R.CD1a cells or C1R.CD1b treated with anti-CD1b monoclonal antibodies (BCD1b3.1), demonstrating an absolute requirement for CD1b expression. However, we observed similarly high levels of IFN- γ production in response to C1R.CD1b with no added antigen, C1R.CD1b cells treated with BbGL-II lipid, or C1R.CD1b cells treated with sonicated *B. burgdorferi* bacteria. In this

system, CD1b-dependent reactivity to human APCs in the absence of exogenously added antigen is interpreted as CD1b autoreactivity. Together, the data suggested the existence of a self-antigen in C1R cells, which we sought to discover.

Diacylglycerol is a self-antigen

Recent studies have identified CD1b-presented lipid autoantigens, including phosphatidic acid (PA), phosphatidylglycerol, phosphatidylcholine, and related molecules^{16,18}. Structural studies of CD1b-phosphatidylglycerol-TCR show that CD1b binds the diacylglycerol (DAG) moiety and presents the phosphoglycerol anion to cationic cup in the TCR. Therefore, we tested staining with a panel of CD1b tetramers loaded with phospholipids, including phosphatidic acid, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylserine, phosphatidylcholine, as well as one smaller (lysophosphatidic acid) and larger (phosphatidylinositol mannoside) lipid with a similar core structure (Fig. 5B). Although some of these alternately loaded tetramers have been previously shown to bind TCRs¹⁸, BC24A did not stain, ruling out the many of the most abundant self-phospholipids as antigens in this case.

Next, we considered self-lipids that more closely resemble the structure of BbGL-II, which has a simple DAG backbone coupled to a hexose sugar (Fig. 1). Hexose-DAGs are not known to exist in human cells. However, free DAG has been reported in low amounts in human cells as a second messenger and as an intermediate of triglyceride synthesis. Therefore, we tested whether the C1R cells, which were used as APCs in our functional assay, as well as two other commonly used APCs, K562 and 293T, expressed detectable amounts of DAG. To detect DAG, we used reverse phase HPLC-quadrupole time-of flight mass spectrometry (Q-ToF), which provides high mass accuracy and retention time information for the detected ions (Fig. 5C). In all cell lines we detected ions with a mass and retention time matching synthetic dioleoyl-DAG standard. A mass spectrum at that retention time (13.5-14 min) showed ammoniated ions of the chain length analogs C34:1 (m/z 612.5567) and C36:2 (m/z 638.5723) DAGs. (Fig. 5D).

Next, we found that dioleoyl DAG-treated CD1b tetramers brightly stained BC24A cells. This result was confirmed in BC24A TCR-transduced J76 cells (Fig. 5E). Polyclonal T cells from six random blood bank donors did not show significant differences between CD1b-endo and CD1b-DAG tetramer staining (Fig. 4C). So, even though BC24A staining provides proof of principle that this reactivity pattern exists, detection of such T cells is not possible in peripheral blood at precursor frequencies of $>10^{-4}$, which is needed for detection above background staining levels. Whereas prior work has identified sphingolipid¹⁷ and phospholipid autoantigens for CD1b^{16,18}, these data identify a neutral lipid self-antigen: simple diacylglycerols mediate CD1b autoreactivity.

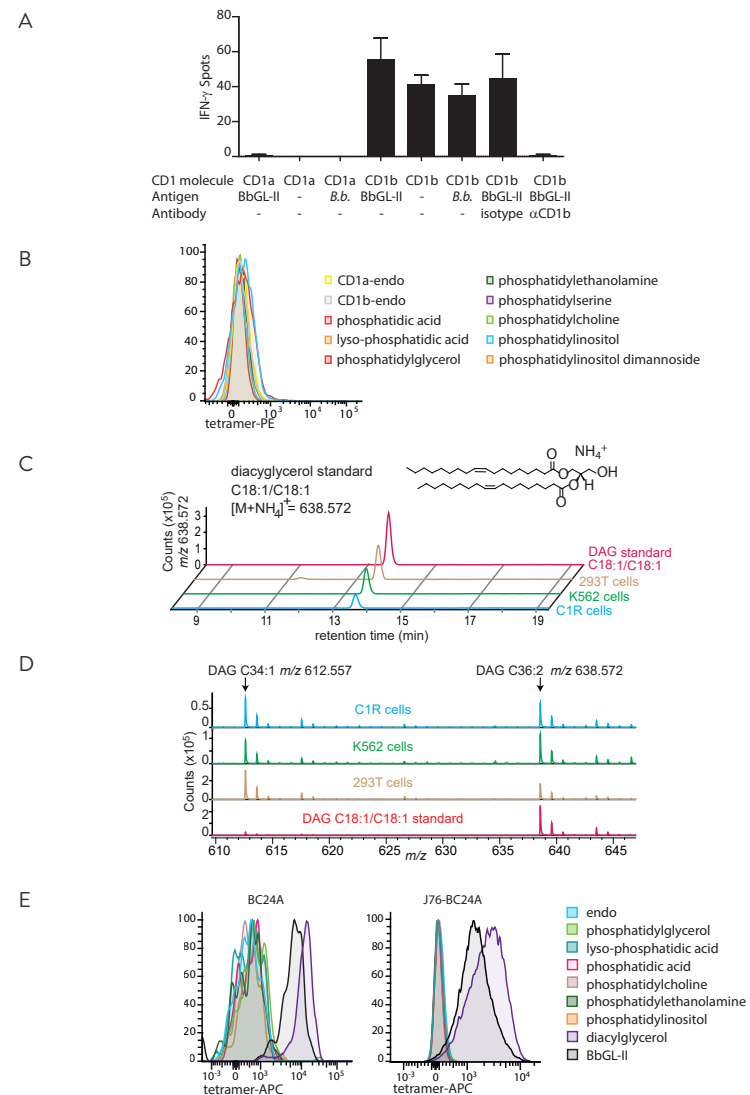


Figure 5. CD1b-BbGL-II-binding T cells are autoreactive and recognize diacylglycerol

A: IFN- γ production was measured by co-culture of BC24A T cells and C1R cells transduced to express human CD1a or CD1b. The cells were incubated in the presence of BbGL-II lipid, sonicated *Borrelia burgdorferi* (B.b.), or without antigen (-). A blocking antibody or isotype control was added. Data is mean value of triplicates with error bars showing standard deviation. Data shown is representative of 3 independent experiments. B: Staining of BC24A with the indicated tetramers. C: Detection of C36:2 DAG standard or lipid extracts of human cell lines C1R, K562, and 293T cells by high mass resolution mass high performance liquid chromatography time of flight mass spectrometry. D: Mass spectrum between 13.5 minutes and 13.7 minutes shows the presence of C34:1 DAG and C36:2 DAG in cells. E: Cultured primary T cells or J76 cells transduced with the BC24A TCR were stained with CD1b tetramers loaded with the indicated antigens.

BC24A staining was performed twice independently, J76-BC24A staining was performed three times independently. One representative staining is shown.

Discussion

Whereas most studies of human $\alpha\beta$ T cell responses focus on MHC and peptide antigens, new information on the CD1 system increasingly points to a parallel system of antigen response involving recognition of foreign lipids. Direct evidence for T cell responses to *B. burgdorferi* lipids started with Kinjo's seminal work, demonstrating that CD1d-BbGL-II complexes activate invariant NKT cells³⁸. NKT cells form an interdonor-conserved innate T cell population. Based on the non-template encoded nucleotide additions in its TCR α chain, the BC24A TCR most likely represents one example of the private, diverse repertoire. A further difference between the CD1d and CD1b systems is that CD1d is constitutively expressed on most cell types, whereas CD1a, CD1b and CD1c protein expression is normally undetectable on myeloid cells in the periphery until Toll-like receptors, cytokines or other stimuli induce CD1b transcription⁸. Because *B. burgdorferi* infection stimulates CD1b expression in human skin, the existence of *Borrelia* lipid-specific T cells was predicted⁴¹. Here we provide proof of principle for CD1b-mediated T cell recognition of a major glycolipid that is expressed in the three major disease-causing species of *Borrelia*.

Whether or not *Borrelia* lipid recognition plays a role in Lyme disease outcomes will require further study, given the lack of increased *ex vivo* CD1b-BbGL-II tetramer staining seen in our pilot study. However, BbGL-II and DAG provide specific molecular targets for future clinical studies in Lyme disease patients. Given the importance of intradermal T cell responses in the erythema migrans lesion, prior evidence for local upregulation of CD1b in skin and the expectation that lipids would be shed locally, the next phase of studies will take advantage of skin to focus on tissue-specific T cells in human responses to *Borrelia* lipids. Further, the pattern of recognition observed here matches a broader theme of immune response in Lyme disease. At the molecular level, B cell and MHC-restricted T cell responses in Lyme disease show demonstrable responses to self-antigens^{25, 26, 48} and cross-reactivities among bacterial and self-antigens²⁷. Thus *Borrelia*-initiated, self-propagating immune responses may contribute to post-infectious Lyme arthritis.

From a molecular perspective, these studies identify DAG as a previously unknown autoantigen in the CD1b system. Although recognition of a moderately abundant self-lipid was unexpected, the data demonstrate antigen recognition using several types of assays. A cloned TCR binds DAG lipids loaded onto CD1b complexes, and a CD1b-mediated autoreactive response to antigen presenting cells that produce DAG was found.

This study reports a T cell line with combined self and foreign lipid antigen specificity. Given the low staining of BC24A T cells with CD1b-endo tetramers, and the high staining with CD1b-BbGL-II tetramers, the strong response of this T cell line to CD1b expressed on C1R cells was unexpected, but can be readily explained by the different valency of TCR ligands on tetramers and cells. DAG constitutes a low percentage of cellular lipids compared to phospholipids. Tetramers typically require that two or more arms be loaded with a TCR binding ligand to achieve adequate avidity to bind T cells. In contrast, the surface of a CD1b-expressing cell has a much larger number of CD1b proteins from which to generate binding partners for TCRs, so rare lipids can generate multivalent TCR ligands. Our data directly document DAG presence in C1R cells, competence of CD1b-DAG tetramers to stain BC24A brightly, and maximal cytokine response to CD1b-expressing C1R cells. These results highlight the ability of cells to present relatively rare antigens to T cells, as contrasted with tetramers, which require higher absolute occupancy.

While the precise molecular mechanism for the observed cross-reactivity between BbGL-II and DAG remains to be established through structural studies, recent work on CD1b-phosphatidylglycerol-TCR complexes suggest a likely model. DAG, BbGL-II, and phosphatidylglycerol have essentially the same diacylglycerol anchor, and this shared chemical moiety also matches the lipid-anchoring moiety seen in the CD1b-phosphatidylglycerol-TCR structure¹⁶. Here the DAG anchor sits atop a lipid spacer and positions the glycerol unit just at the opening of the F' portal of CD1b. If the shared chemical moieties of DAG and BbGL-II are positioned similarly inside CD1b, the hexose sugar unique to BbGL-II would protrude through the F' portal to rest on the outer surface of CD1b in some way. Our data indicate that the glucose is not required for TCR binding, and it does not inhibit TCR response. There are several known mechanisms by which the distal moiety on lipid antigen can be 'ignored' by an approaching TCR. The simplest is that the BbGL-II molecule sits completely inside CD1b, but this is less favored given the lack of clear precedent for carbohydrates to be seated inside the hydrophobic cleft. A second, more favored possibility is that the hexose sugar protrudes to the surface of CD1 through the F' portal, whereas the TCR binds near the center of CD1. This mechanism is known as left-right mismatch⁴⁹. Last, CD1b-reactive TCRs can have an escape channel whereby larger head groups escape sideways between the TCR α and β chains¹⁹. Overall, these findings support and extend key emerging concepts in CD1b biology: TCR cross-reactivity for foreign and self-antigens.

Material and methods

Mass spectrometry

For nanospray analysis, a 1 μ M solution of lipid in methanol was loaded onto a nanospray tip and analyzed by electrospray ionization mass spectrometry (ESI-MS) and collision-induced dissociation spectrometry (CID-MS) on the LXQ Ion Trap Mass Spectrometer (Thermo Fisher Scientific) in positive ion mode. Collision energy was 20 to 30 % of the maximum, and product ions were trapped with a q value of 0.25. Chloroform/methanol extracted total lipids from cell lines were prepared at 0.5 mg/ml and 10 μ l was injected for HPLC-ESI-MS analysis (Agilent 6520 Accurate-Mass Q-TOF and 1200 series HPLC system using a reverse phase Eclipse Plus-C18 column (3.5 μ M, 2.1 mm \times 30 mm, Agilent Technologies) according to the published method⁴⁶.

Tetramers and flow cytometry.

Biotinylated CD1b monomers were obtained from the National Institute Health (NIH) and loaded with 32 μ g phosphatidylglycerol (#841188P, Avanti polar lipids), phosphatidic acid (#840857, Avanti polar lipids), lyso-phosphatidic acid (#857130, Avanti polar lipids), phosphatidylethanolamine (#850757, Avanti polar lipids), phosphatidylserine (#840032, Avanti polar lipids), phosphatidylcholine (#850475, Avanti polar lipids), phosphatidylinositol (#840042, Avanti polar lipids), phosphatidylinositol dimannoside (Bill and Melinda Gates Foundation lipid bank) or in-house produced synthetic BbGL-II³¹. Lipids were dissolved in citrate buffer at pH 4.5 with 0.5 % CHAPS (Sigma) by sonication for 10 min at 37°C, followed by incubation for 2 hours at 65°C and repeat sonication for 10 min at 37°C, after which the CD1b monomers were added. This mixture was incubated overnight at 37°C. After incubation, the pH was neutralized by adding TRIS pH 8.5. As control CD1b monomers were treated in the same way without adding lipid (CD1b-endo). Tetramers were generated by linking monomer together with streptavidin (SA)-allophycocyanin (APC) (Life Technologies) or SA-phycoerythrin (PE) (Life Technologies). Experimental setup of flow cytometric experiments adhered to published guidelines⁵⁰. Cells were stained with tetramers and α CD3 monoclonal antibody (clone SK7, BD biosciences). Cells were first incubated for 10 min with tetramers at room temperature, then the α CD3 antibody was added and incubated again for 10 min at room temperature, followed by an incubation of 20 minutes at 4°C. The cells were either sorted on the FACSAria (BD biosciences) or analyzed on the FACSCanto II (BD biosciences).

Human subjects

PBMCs of random blood donors were derived from the blood bank (Sanquin, Amsterdam). Lyme disease patient-derived blood was obtained as part of an ongoing Lyme disease study, which is approved by the regional Medical Research Ethics Committees United (Nieuwegein, Netherlands; MEC-U: NL36407.100.11).

T cell culture

PBMC were isolated from buffy coats or whole blood by density gradient centrifugation. PBMCs or sorted T cells were cultured in T cell medium⁵¹. Cells were cultured in either 25 mL (1×10^5 - 1×10^6 T cells) with 25×10^6 irradiated PBMCs and 5×10^6 irradiated Epstein-Barr virus transformed B cells as feeder cells or in 200 μ l (800 - $8,000$ T cells) with 2×10^5 irradiated PBMCs and 4×10^4 irradiated Epstein-Barr virus transformed B cells as feeder cells. Cells were stimulated with 30 ng/mL α CD3 antibody (OKT3) in the absence of IL-2 followed by adding 1 ng/mL IL-2 after 24 hours and thereafter. After 14-21 days cells were resorted and cultured.

Single cell TCR sequencing.

96 well Eppendorf plates were coated with Vapor-Lock (Qiagen). 0.5 μ l 5x Iscript buffer, 0.5 μ l reverse transcriptase (Iscrip, Bio-Rad), 0.1 % triton X-100, and 1.25 μ l H₂O were added per well for cDNA synthesis. Single cells were sorted in these 96 wells plate on the FACSaria followed by centrifugation at 3000 RPM at 4°C for 10 minutes. CDNA synthesis and nested PCRs were performed as described by Wang *et al.*⁴⁴ with small modifications. For the first PCR reaction, the following conditions were used: 95°C for 2 min, 35 cycles of 95°C for 20sec, 50°C for 20 sec, 72°C for 45 sec, followed by 1 cycle of 72°C for 7 min. For the second PCR the following conditions were used: 95°C for 2 min, 35 cycles of 95°C for 20sec, 56°C for 20 sec, 72°C for 45 sec, followed by 1 cycle of 72°C for 7 min. Samples were loaded on 1.5 % (α chain) or 2 % (β chain) agarose gel for electrophoresis. Multiple wells that showed a result for both α chain and β chain on gel electrophoresis underwent Sanger sequencing.

ELISPOT

96 wells 0.45 μ m Hydrophobic Multiscreen plates (Milipore) were coated with α IFN- γ mAb (clone 1-D1K, Mabtech) overnight at 4°C. The plate was blocked for 2 hr after which 20,000 C1R cells transfected with CD1a or CD1b were co-cultured with 200 T cells in the present or absence of a blocking antibody (clone BCD1b.3) and different antigens. The antigens used in this assay are BbGL-II lipid (5 μ g/ml), sonicated *B. burgdorferi* (200,000 bacteria per well) and media as control. After incubation overnight at 37°C cells were lysed and washed away with PBS-Tween, and the plates were incubated for 2 hr with a biotinylated α IFN- γ antibody (clone 7-B6-1, Mabtech). The wells were washed with PBS-

Tween and incubated with Extravidin-ALP (Sigma-Aldrich) for 1 hr. After washing with PBS-Tween followed by washing with PBS, 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (SigmaFAST tablets, Sigma) was added to visualize the spots. Spots were counted by the Immunospot reader (C.T.L Technologies).

Acknowledgements

This work was supported by funding from Nederlands Wetenschappelijk Onderzoek grant 824.02.002 and the NIH (AI116604, AI04393). PR was supported by a Boehringer Ingelheim Fonds travel grant. MJS, DIG and DGP are supported by the National Health and Medical Research Council (NHMRC, 1063587) and the Australian Research Council (ARC; CE140100011). DIG is also supported by NHMRC Senior Principal Research Fellowship (1020770) and DGP is supported by an NHMRC Career Development Fellowship (APP1144308).

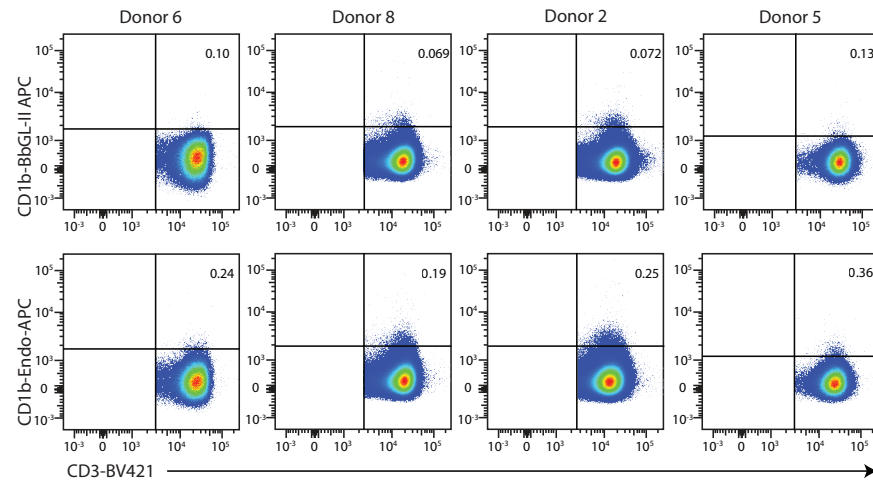
Conflict of interest

The authors declare no commercial or financial conflict of interest.

References

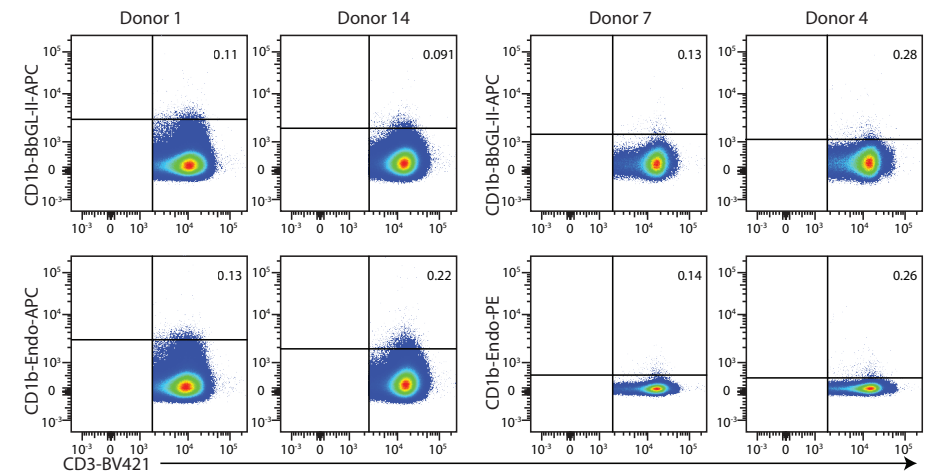
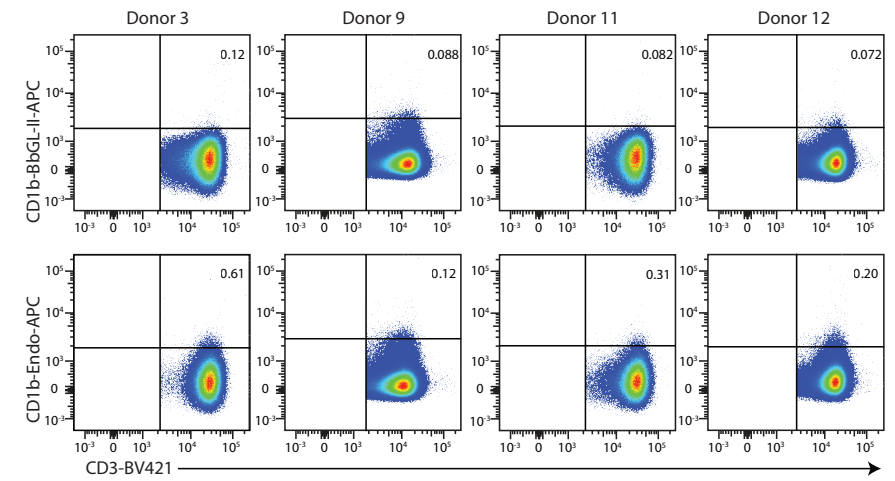
- 1 Calabi, F., Jarvis, J. M., Martin, L. and Milstein, C., Two classes of CD1 genes. *Eur J Immunol* 1989. **19**: 285-292.
- 2 Dougan, S. K., Kaser, A. and Blumberg, R. S., CD1 expression on antigen-presenting cells. *Curr Top Microbiol Immunol* 2007. **314**: 113-141.
- 3 Rossjohn, J., Gras, S., Miles, J. J., Turner, S. J., Godfrey, D. I. and McCluskey, J., T cell antigen receptor recognition of antigen-presenting molecules. *Annual Review of Immunology* 2015. **33**: 169-200.
- 4 Van Rhijn, I. and Moody, D. B., Donor Unrestricted T Cells: A Shared Human T Cell Response. *Journal of Immunology* 2015. **195**: 1927-1932.
- 5 Beckman, E. M., Porcelli, S. A., Morita, C. T., Behar, S. M., Furlong, S. T. and Brenner, M. B., Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature* 1994. **372**: 691-694.
- 6 Porcelli, S., Brenner, M. B., Greenstein, J. L., Balk, S. P., Terhorst, C. and Bleicher, P. A., Recognition of cluster of differentiation 1 antigens by human CD4-CD8-cytolytic T lymphocytes. *Nature* 1989. **341**: 447-450.
- 7 Krutzik, S. R., Tan, B., Li, H., Ochoa, M. T., Liu, P. T., Sharfstein, S. E., Graeber, T. G., Sieling, P. A., Liu, Y. J., Rea, T. H., Bloom, B. R. and Modlin, R. L., TLR activation triggers the rapid differentiation of monocytes into macrophages and dendritic cells. *Nature Medicine* 2005. **11**: 653-660.
- 8 Roura-Mir, C., Wang, L., Cheng, T. Y., Matsunaga, I., Dascher, C. C., Peng, S. L., Fenton, M. J., Kirschning, C. and Moody, D. B., Mycobacterium tuberculosis regulates CD1 antigen presentation pathways through TLR-2. *J Immunol* 2005. **175**: 1758-1766.
- 9 Layre, E., Collmann, A., Bastian, M., Mariotti, S., Czaplicki, J., Prandi, J., Mori, L., Stenger, S., De Libero, G., Puzo, G. and Gilleron, M., Mycolic acids constitute a scaffold for mycobacterial lipid antigens stimulating CD1-restricted T cells. *Chem Biol* 2009. **16**: 82-92.
- 10 Moody, D. B., Reinhold, B. B., Guy, M. R., Beckman, E. M., Frederique, D. E., Furlong, S. T., Ye, S., Reinhold, V. N., Sieling, P. A., Modlin, R. L., Besra, G. S. and Porcelli, S. A., Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* 1997. **278**: 283-286.
- 11 Van Rhijn, I., Kasmar, A., de Jong, A., Gras, S., Bhati, M., Doorenspleet, M. E., de Vries, N., Godfrey, D. I., Altman, J. D., de Jager, W., Rossjohn, J. and Moody, D. B., A conserved human T cell population targets mycobacterial antigens presented by CD1b. *Nature Immunology* 2013. **14**: 706-713.
- 12 Chancellor, A., Tocheva, A. S., Cave-Ayland, C., Tezera, L., White, A., Al Dulayymi, J. R., Bridgeman, J. S., Tews, I., Wilson, S., Lissin, N. M., Tebruegge, M., Marshall, B., Sharpe, S., Elliott, T., Skylaris, C. K., Essex, J. W., Baird, M. S., Gadola, S., Elkington, P. and Mansour, S., CD1b-restricted GEM T cell responses are modulated by Mycobacterium tuberculosis mycolic acid meromycolate chains. *Proceedings of the National Academy of Sciences of the United States of America* 2017. **114**: E10956-E10964.
- 13 Van Rhijn, I., Iwany, S. K., Fodran, P., Cheng, T. Y., Gapin, L., Minnaard, A. J. and Moody, D. B., CD1b-mycolic acid tetramers demonstrate T-cell fine specificity for mycobacterial lipid tails. *European Journal of Immunology* 2017. **47**: 1525-1534.
- 14 Gilleron, M., Stenger, S., Mazorra, Z., Wittke, F., Mariotti, S., Bohmer, G., Prandi, J., Mori, L., Puzo, G. and De Libero, G., Diacylated Sulfoglycolipids Are Novel Mycobacterial Antigens Stimulating CD1-restricted T Cells during Infection with Mycobacterium tuberculosis. *J Exp Med* 2004. **199**: 649-659.
- 15 James, C. A., Yu, K. K. Q., Gilleron, M., Prandi, J., Yedulla, V. R., Moleda, Z. Z., Diamanti, E., Khan, M., Aggarwal, V. K., Reijneveld, J. F., Reinink, P., Lenz, S., Emerson, R. O., Scriba, T. J., Souter, M. N. T., Godfrey, D. I., Pellicci, D. G., Moody, D. B., Minnaard, A. J., Seshadri, C. and Van Rhijn, I., CD1b Tetramers Identify T Cells that Recognize Natural and Synthetic Diacylated Sulfoglycolipids from Mycobacterium tuberculosis. *Cell Chem Biol* 2018. **25**: 392-402 e314.
- 16 Shahine, A., Van Rhijn, I., Cheng, T. Y., Iwany, S., Gras, S., Moody, D. B. and Rossjohn, J., A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Sci Immunol* 2017. **2**.
- 17 Shamshiev, A., Donda, A., Prigozy, T. I., Mori, L., Chigorno, V., Benedict, C. A., Kappos, L., Sonnino, S., Kronenberg, M. and De Libero, G., The alphabeta T cell response to self-glycolipids shows a novel mechanism of CD1b loading and a requirement for complex oligosaccharides. *Immunity*. 2000. **13**: 255-264.
- 18 Van Rhijn, I., van Berlo, T., Hilmennyuk, T., Cheng, T. Y., Wolf, B. J., Tatituri, R. V., Uldrich, A. P., Napolitani, G., Cerundolo, V., Altman, J. D., Willemsen, P., Huang, S., Rossjohn, J., Besra, G. S., Brenner, M. B., Godfrey, D. I. and Moody, D. B., Human autoreactive T cells recognize CD1b and phospholipids. *Proceedings of the National Academy of Sciences of the United States of America* 2016. **113**: 380-385.
- 19 Shahine, A., Reinink, P., Reijneveld, J. F., Gras, S., Holzheimer, M., Cheng, T. Y., Minnaard, A. J., Altman, J. D., Lenz, S., Prandi, J., Kubler-Kielb, J., Moody, D. B., Rossjohn, J. and Van Rhijn, I., A T-cell receptor escape channel allows broad T-cell response to CD1b and membrane phospholipids. *Nature communications* 2019. **10**: 56.
- 20 Pfister, H. W., Wilske, B. and Weber, K., Lyme borreliosis: basic science and clinical aspects. *Lancet* 1994. **343**: 1013-1016.
- 21 Steere, A. C., Strle, F., Wormser, G. P., Hu, L. T., Branda, J. A., Hovius, J. W., Li, X. and Mead, P. S., Lyme borreliosis. *Nat Rev Dis Primers* 2016. **2**: 16090.
- 22 Stanek, G., Wormser, G. P., Gray, J. and Strle, F., Lyme borreliosis. *Lancet* 2012. **379**: 461-473.
- 23 Arvikar, S. L. and Steere, A. C., Diagnosis and treatment of Lyme arthritis. *Infectious Disease Clinics of North America* 2015. **29**: 269-280.
- 24 Collins, C., Shi, C., Russell, J. Q., Fortner, K. A. and Budd, R. C., Activation of gamma delta T cells by *Borrelia burgdorferi* is indirect via a TLR- and caspase-dependent pathway. *Journal of Immunology* 2008. **181**: 2392-2398.
- 25 Crowley, J. T., Drouin, E. E., Pianta, A., Strle, K., Wang, Q., Costello, C. E. and Steere, A. C., A Highly Expressed Human Protein, Apolipoprotein B-100, Serves as an Autoantigen in a Subgroup of Patients With Lyme Disease. *Journal of Infectious Diseases* 2015. **212**: 1841-1850.
- 26 Crowley, J. T., Strle, K., Drouin, E. E., Pianta, A., Arvikar, S. L., Wang, Q., Costello, C. E. and Steere, A. C., Matrix metalloproteinase-10 is a target of T and B cell responses that correlate with synovial pathology in patients with antibiotic-refractory Lyme arthritis. *Journal of Autoimmunity* 2016. **69**: 24-37.
- 27 Drouin, E. E., Seward, R. J., Strle, K., McHugh, G., Katchar, K., Londono, D., Yao, C., Costello, C. E. and Steere, A. C., A novel human autoantigen, endothelial cell growth factor, is a target of T and B cell responses in patients with Lyme disease. *Arthritis and Rheumatism* 2013. **65**: 186-196.
- 28 Ben-Menachem, G., Kubler-Kielb, J., Coxon, B., Yergely, A. and Schneerson, R., A newly discovered cholesteryl galactoside from *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences of the United States of America* 2003. **100**: 7913-7918.

- 29 Toledo, A., Huang, Z., Coleman, J. L., London, E. and Benach, J. L., Lipid rafts can form in the inner and outer membranes of *Borrelia burgdorferi* and have different properties and associated proteins. *Molecular Microbiology* 2018. **108**: 63-76.
- 30 Hossain, H., Wellensiek, H. J., Geyer, R. and Lochnit, G., Structural analysis of glycolipids from *Borrelia burgdorferi*. *Biochimie* 2001. **83**: 683-692.
- 31 Pozsgay, V., Kubler-Kielb, J., Coxon, B., Marques, A., Robbins, J. B. and Schneerson, R., Synthesis and antigenicity of BBGL-2 glycolipids of *Borrelia burgdorferi*, the causative agent of Lyme disease. *Carbohydrate Research* 2011. **346**: 1551-1563.
- 32 Jones, K. L., Seward, R. J., Ben-Menachem, G., Glickstein, L. J., Costello, C. E. and Steere, A. C., Strong IgG antibody responses to *Borrelia burgdorferi* glycolipids in patients with Lyme arthritis, a late manifestation of the infection. *Clinical Immunology* 2009. **132**: 93-102.
- 33 Stubs, G., Fingerle, V., Wilske, B., Gobel, U. B., Zahringer, U., Schumann, R. R. and Schroder, N. W., Acylated cholesteryl galactosides are specific antigens of *borrelia* causing lyme disease and frequently induce antibodies in late stages of disease. *Journal of Biological Chemistry* 2009. **284**: 13326-13334.
- 34 Ostberg, Y., Berg, S., Comstedt, P., Wieslander, A. and Bergstrom, S., Functional analysis of a lipid galactosyltransferase synthesizing the major envelope lipid in the Lyme disease spirochete *Borrelia burgdorferi*. *FEMS Microbiology Letters* 2007. **272**: 22-29.
- 35 Mansour, S., Tocheva, A. S., Cave-Ayland, C., Machelett, M. M., Sander, B., Lissin, N. M., Molloy, P. E., Baird, M. S., Stubs, G., Schroder, N. W., Schumann, R. R., Rademann, J., Postle, A. D., Jakobsen, B. K., Marshall, B. G., Gosain, R., Elkington, P. T., Elliott, T., Skylaris, C. K., Essex, J. W., Tews, I. and Gadola, S. D., Cholesteryl esters stabilize human CD1c conformations for recognition by self-reactive T cells. *Proceedings of the National Academy of Sciences of the United States of America* 2016. **113**: E1266-1275.
- 36 Gadola, S. D., Zaccai, N. R., Harlos, K., Shepherd, D., Castro-Palomino, J. C., Ritter, G., Schmidt, R. R., Jones, E. Y. and Cerundolo, V., Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. *Nat.Immunol.* 2002. **3**: 721-726.
- 37 Huang, S., Cheng, T. Y., Young, D. C., Layre, E., Madigan, C. A., Shires, J., Cerundolo, V., Altman, J. D. and Moody, D. B., Discovery of deoxyceramides and diacylglycerols as CD1b scaffold lipids among diverse groove-blocking lipids of the human CD1 system. *Proc Natl Acad Sci U S A* 2011. **108**: 19335-19340.
- 38 Kinjo, Y., Tupin, E., Wu, D., Fujio, M., Garcia-Navarro, R., Benhnia, M. R., Zajonc, D. M., Ben-Menachem, G., Ainge, G. D., Painter, G. F., Khurana, A., Hoebe, K., Behar, S. M., Beutler, B., Wilson, I. A., Tsuji, M., Sellati, T. J., Wong, C. H. and Kronenberg, M., Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nat Immunol* 2006. **7**: 978-986.
- 39 Kumar, H., Belperron, A., Barthold, S. W. and Bockenstedt, L. K., Cutting edge: CD1d deficiency impairs murine host defense against the spirochete, *Borrelia burgdorferi*. *Journal of Immunology* 2000. **165**: 4797-4801.
- 40 Tupin, E., Benhnia, M. R., Kinjo, Y., Patsey, R., Lena, C. J., Haller, M. C., Caimano, M. J., Imamura, M., Wong, C. H., Crotty, S., Radolf, J. D., Sellati, T. J. and Kronenberg, M., NKT cells prevent chronic joint inflammation after infection with *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences of the United States of America* 2008. **105**: 19863-19868.
- 41 Yakimchuk, K., Roura-Mir, C., Magalhaes, K. G., de Jong, A., Kasmar, A. G., Granter, S. R., Budd, R., Steere, A., Pena-Cruz, V., Kirschning, C., Cheng, T. Y. and Moody, D. B., *Borrelia burgdorferi* infection regulates CD1 expression in human cells and tissues via IL1-beta. *European Journal of Immunology* 2011. **41**: 694-705.
- 42 Moody, D. B., TLR gateways to CD1 function. *Nat Immunol* 2006. **7**: 811-817.
- 43 Van Rhijn, I., Gherardin, N. A., Kasmar, A., de Jager, W., Pellicci, D. G., Kostenko, L., Tan, L. L., Bhati, M., Gras, S., Godfrey, D. I., Rossjohn, J. and Moody, D. B., TCR Bias and Affinity Define Two Compartments of the CD1b-Glycolipid-Specific T Cell Repertoire. *Journal of Immunology* 2014. **193**: 5338-5344.
- 44 Wang, G. C., Dash, P., McCullers, J. A., Doherty, P. C. and Thomas, P. G., T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Sci Transl Med* 2012. **4**: 128ra142.
- 45 Li, D., Hong, A., Lu, Q., Gao, G. F., Jin, B., Sreaton, G. R. and Xu, X. N., A novel role of CD1c in regulating CD1d-mediated NKT cell recognition by competitive binding to Ig-like transcript 4. *Int Immunol* 2012.
- 46 Wun, K. S., Reijneveld, J. F., Cheng, T. Y., Ladell, K., Uldrich, A. P., Le Nours, J., Miners, K. L., McLaren, J. E., Grant, E. J., Haigh, O. L., Watkins, T. S., Suliman, S., Iwany, S., Jimenez, J., Calderon, R., Tamara, K. L., Leon, S. R., Murray, M. B., Mayfield, J. A., Altman, J. D., Purcell, A. W., Miles, J. J., Godfrey, D. I., Gras, S., Price, D. A., Van Rhijn, I., Moody, D. B. and Rossjohn, J., T cell autoreactivity directed toward CD1c itself rather than toward carried self lipids. *Nat Immunol* 2018. **19**: 397-406.
- 47 Heemskerck, M. H., Hooijboom, M., de Paus, R. A., Kester, M. G., van der Hoorn, M. A., Goulmy, E., Willemze, R. and Falkenburg, J. H., Redirection of antileukemic reactivity of peripheral T lymphocytes using gene transfer of minor histocompatibility antigen HA-2-specific T-cell receptor complexes expressing a conserved alpha joining region. *Blood* 2003. **102**: 3530-3540.
- 48 Pianta, A., Drouin, E. E., Crowley, J. T., Arvikar, S., Strle, K., Costello, C. E. and Steere, A. C., Annexin A2 is a target of autoimmune T and B cell responses associated with synovial fibroblast proliferation in patients with antibiotic-refractory Lyme arthritis. *Clinical Immunology* 2015. **160**: 336-341.
- 49 Cotton, R. N., Shahine, A., Rossjohn, J. and Moody, D. B., Lipids hide or step aside for CD1-autoreactive T cell receptors. *Current Opinion in Immunology* 2018. **52**: 93-99.
- 50 Cossarizza, A., Chang, H. D., Radbruch, A., Akdis, M., Andra, I., Annunziato, F., Bacher, P., Barnaba, V., Battistini, L., Bauer, W. M., Baumgart, S., Becher, B., Beisker, W., Berek, C., Blanco, A., Borsellino, G., Boulais, P. E., Brinkman, R. R., Buscher, M., Busch, D. H., Bushnell, T. P., Cao, X., Cavani, A., Chattopadhyay, P. K., Cheng, Q., Chow, S., Clerici, M., Cooke, A., Cosma, A., Cosmi, L., Cumano, A., Dang, V. D., Davies, D., De Biasi, S., Del Zotto, G., Della Bella, S., Dellabona, P., Deniz, G., Dessing, M., Diefenbach, A., Di Santo, J., Dieli, F., Dolf, A., Donnemeyer, V. S., Dorner, T., Ehrhardt, G. R. A., Endl, E., Engel, P., Engelhardt, B., Esser, C., Everts, B., Dreher, A., Falk, C. S., Fehniger, T. A., Filby, A., Fillatreau, S., Follo, M., Forster, I., Foster, J., Foulds, G. A., Frenette, P. S., Galbraith, D., Garbi, N., Garcia-Godoy, M. D., Geginat, J., Ghoreschi, K., Gibellini, L., Goettlinger, C., Goodyear, C. S., Gori, A., Grogan, J., Gross, M., Grutzkau, A., Grummitt, D., Hahn, J., Hammer, Q., Hauser, A. E., Haviland, D. L., Hedley, D., Herrera, G., Herrmann, M., Hiepe, F., Holland, T., Hombrink, P., Houston, J. P., Hoyer, B. F., Huang, B., Hunter, C. A., Iannone, A., Jack, H. M., Javega, B., Jonjic, S., Juelke, K., Jung, S., Kaiser, T., Kalina, T., Keller, B., Khan, S., Kienhofer, D., Kroneis, T., Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur J Immunol* 2017. **47**: 1584-1797.
- 51 Ly, D., Kasmar, A. G., Cheng, T. Y., de Jong, A., Huang, S., Roy, S., Bhatt, A., van Summeren, R. P., Altman, J. D., Jacobs, W. R., Jr., Adams, E. J., Minnaard, A. J., Porcelli, S. A. and Moody, D. B., CD1c tetramers detect ex vivo T cell responses to processed phosphomycolketide antigens. *J Exp Med* 2013. **210**: 729-741.



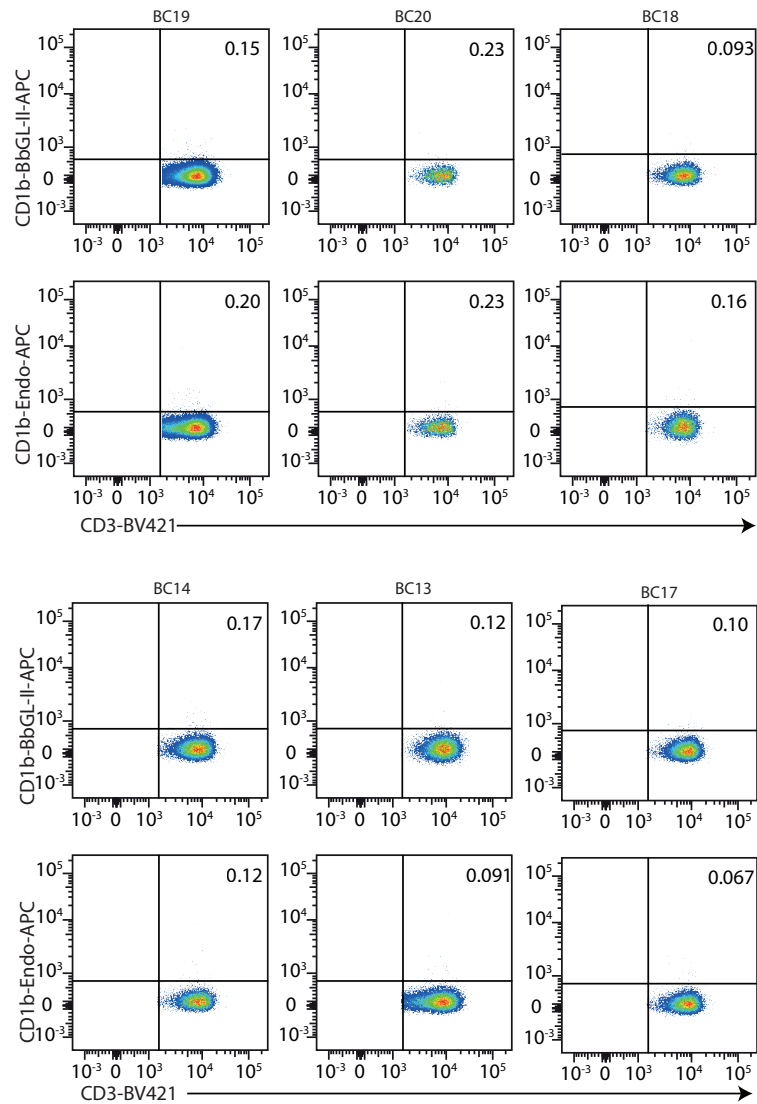
Supplemental figure 1. Tetramer staining on Lyme arthritis patients

Raw Facs plots of data summarized in Fig. 4A. PBMCs of four Lyme arthritis patients in different disease states (Table 1) were expanded via α CD3 stimulation for two weeks and stained with α CD3 antibody and CD1b-BbGL-II or CD1b-endo tetramers. Cells are pre-gated for CD3.



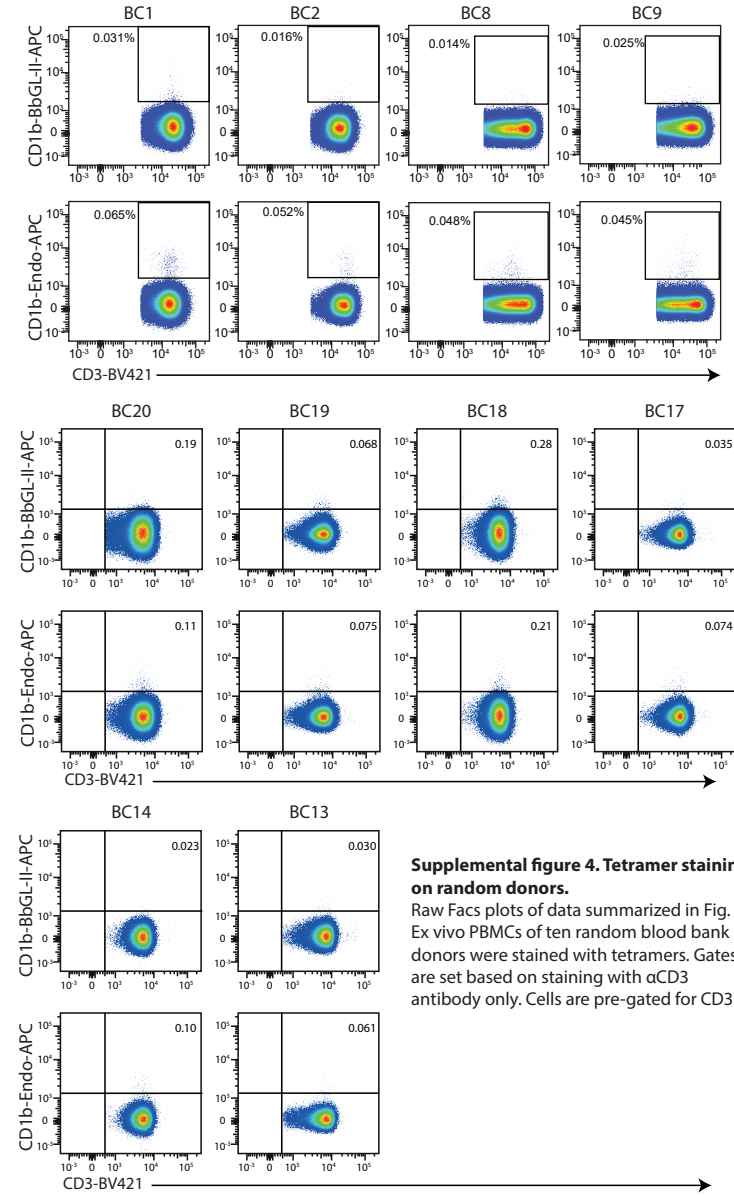
Supplemental figure 2. Tetramer staining on Lyme neuroborreliosis patients

Raw Facs plots of data summarized in Fig. 4A. PBMCs of eight Lyme neuroborreliosis patients in different disease states (Table 1) were expanded via α CD3 stimulation for two weeks and stained with an α CD3 antibody and CD1b-BbGL-II or CD1b-endo tetramers. Cells are pre-gated for CD3.



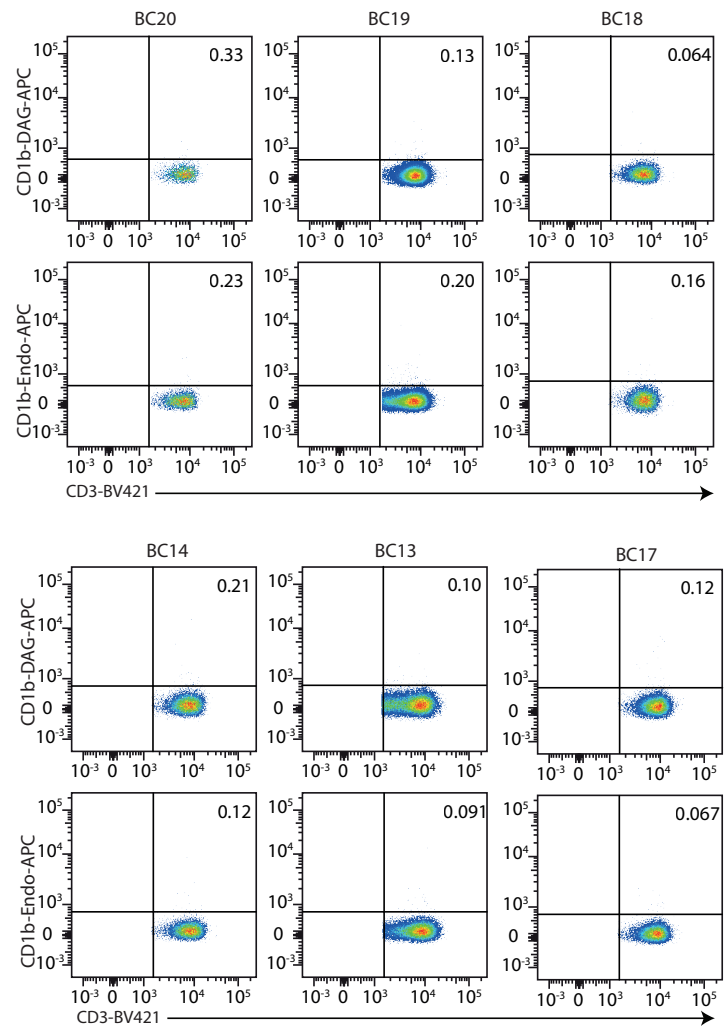
Supplemental figure 3. Tetramer staining on random blood donors

Raw Facs plots of data summarized in Fig. 4A. PBMCs of six random blood bank donors were expanded via α CD3 stimulation for two weeks and stained with an α CD3 antibody and CD1b-BbGL-II or CD1b-endo tetramers. Cells are pre-gated for CD3.



Supplemental figure 4. Tetramer staining on random donors.

Raw Facs plots of data summarized in Fig. 4B. Ex vivo PBMCs of ten random blood bank donors were stained with tetramers. Gates are set based on staining with α CD3 antibody only. Cells are pre-gated for CD3.



Supplemental figure 5. Tetramer staining on random blood donors

Raw FACS plots of data summarized in Fig. 4C. PBMCs of six random blood bank donors were expanded via α CD3 stimulation for two weeks and stained with an α CD3 antibody and CD1b-DAG or CD1b-endo tetramers. Cells are pre-gated for CD3.



Chapter 4

TRBV4-1 (V β 7.1) bias among T cells recognizing CD1b regardless of the presented lipid antigen

4

Peter Reinink^{a,b}, Adam Shahine^{c,d}, Stephanie Gras^{c,d}, Tan-Yun Cheng^b, Jamie Rossjohn^{c,d,e}, D. Branch Moody^b, Ildiko Van Rhijn^{a,b}

.....

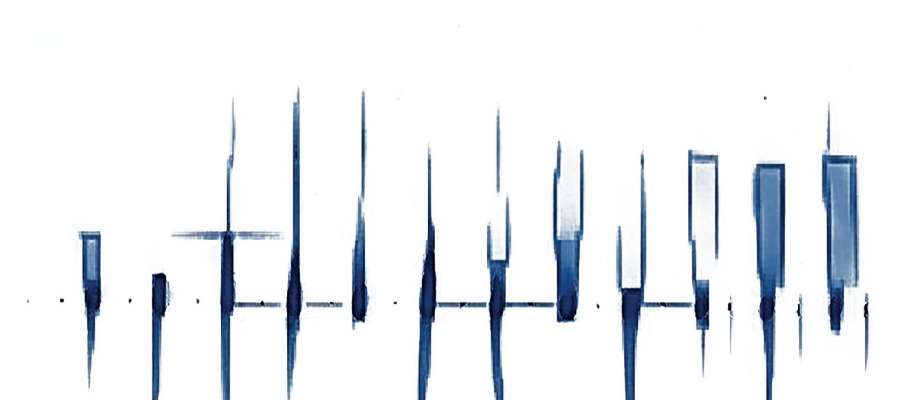
^a Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584CL Utrecht, The Netherlands

^b Brigham and Women's Hospital Division of Rheumatology, Immunology and Allergy and Harvard Medical School, Boston, MA 02115, USA

^c Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

^d Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia

^e Institute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, UK



Abstract

An important step towards meaningful interpretation of the human TCR repertoire is to predict interactions between T cell receptors and their antigenic targets, which has been proven to be difficult. Unlike classical MHC molecules, the CD1 family of antigen presenting molecules is non-polymorphic. T cell types with a species-wide distribution, or “interdonor-conserved” T cells, can exist as a consequence of the non-polymorphic nature of CD1 molecules. Examples are invariant NKT cells that recognize CD1d bound to α -galactosylceramide, and GEM T cells and LDN5-like cells that both recognize CD1b bound to glucose monomycolate. Of the many possible gene segments that can be used to form TCRs, LDN5-like cells use TRBV4-1. In fact, binding of glucose monomycolate-loaded CD1b tetramers in combination with an antibody against TRBV4-1 defines LDN5-like cells. Here we describe a much broader involvement of TRBV4-1 in the recognition of CD1b regardless of the chemical nature of the lipid antigen. Crystallography of two TCRs shows how germline-encoded amino acids of CDR1 and CDR3 of TRBV4-1 interact and consolidate the CD1b-facing surface of the TCR. Together with recently discovered bias of TRBV4-1 towards CD1c recognition, these findings suggest co-evolution of CD1 genes and specific TCR gene segments.

Introduction

Antigens are presented to T cells by major histocompatibility complex (MHC) molecules or by the structurally related CD1 and MR1 molecules. In humans, the CD1 family consists of four cell surface-expressed molecules (CD1a, CD1b, CD1c, CD1d) that present lipids to T cells, rather than peptides. In contrast to classical MHC molecules, CD1 molecules are non-polymorphic. Humans express different MHC alleles and use a wide variety of T-cell receptors (TCRs) to recognize the many occurring MHC-peptide complexes. An overly simplified view of TCR-MHC interaction is that the variable (V) gene-encoded complementarity determining region (CDR) 1 and CDR2 interact with the MHC molecule and the CDR3, which is formed by the V-J junction and incorporates random, non-template-encoded nucleotides, interacts with the peptide. This model led to the prediction that certain MHC types or allomorphs are preferentially recognized by certain V segments, and that an individual’s MHC haplotype biases the selected TCR repertoire. However, there is only limited data available to support this, and the existing data only account for a minor association between MHC haplotype and TCR repertoire¹. Thus, the diversity in the MHC system, the diversity of the bound peptides, and the fact that each MHC molecule or allomorph can interact with many TCR V segments, all contribute to high TCR repertoire diversity and unpredictability in the system.

Despite the lack of strong association between a particular MHC molecule and V genes used by the TCRs that recognize it, for CD1 and other non-polymorphic molecules such associations, driven by coevolved protein-protein interactions², might exist. In the CD1 system there are some examples of identical TCR chains that recognize the same CD1-lipid complex in all humans, which is referred to as type III bias³. The best-known example is the invariant NKT TCR that recognizes the CD1d- α -galactosylceramide complex in all humans studied, and another example is the GEM TCR that recognizes CD1b-glucose monomycolate. Outside the CD1 system, the non-polymorphic MR1 molecule stimulates invariant MAIT cells. Co-evolution between MR1 and the Va segment that is used by MAIT cells has been suggested⁴. However, in all three of these T cell populations, TCR recognition is also dependent on the bound antigen, which makes direct contacts with the invariant α chain.

The β chains of these three invariant T cells show type I bias (shared V gene usage but unrelated CDR3s). Type I-biased β chains, not in combination with invariant α chains, are found among CD1c-restricted autoreactive T cells that preferentially express a TCR containing a TRBV4-1 segment⁵. Another example of type I bias in the CD1 system is formed by CD1b-glucose monomycolate-specific cells: LDN5-like cells preferentially use TRAV17 or TRBV4-1 but have unrelated CDR3 sequences⁶. But, like MAIT, NKT, and

GEM T cells, LDN5-like cells also share specificity for a specific lipid ligand, so whether the TCR bias is driven by CD1b or by the antigen is not known.

Thus, TCR bias is common among T cells that recognize a specific CD1-lipid complex. If coevolution between CD1 and TCR V gene(s) is a driving force for such a bias, we hypothesize that CD1 isoform-specific bias, independent of the bound lipid antigen, also exists. To study this, we looked at the TCRs of known CD1b restricted T cells specific for several different antigens and show that there is a strong overrepresentation of TRBV4-1⁺ T cells and vice versa: among TRBV4-1⁺ cells there are more CD1b restricted cells than among cells that express other Vβ segments. Despite the strong focus on invariant α chains and antigen specificity of invariant TCRs in literature, we discovered that among CD1b-specific T cells, Vβ usage is highly biased, but independent from lipid antigen specificity.

Results

Overrepresentation of TRBV4-1 among CD1b-specific T cell clones

While generating CD1b-specific T cells in the past we noticed that TRBV4-1 was used several times. To study this phenomenon in a systematic way, we collected all TCR sequences from confirmed CD1b-restricted T cell clones from the literature, regardless of their antigen specificity and TCR gene segment usage. It is clear that TRBV4-1 is overrepresented among these TCRs (Fig. 1a): 12 out of the 27 CD1b restricted TCRs use the TRBV4-1 segment, while in PBMC, TRBV4-1 is typically used by 0.5-4% of the T cells⁷⁻⁹. Among the TRBV4-1⁺ CD1b restricted T cells, specificity for at least four lipid antigens was found: phosphatidylglycerol (PG), glucose monomycolate (GMM), mycolic acid (MA), and 1, 2 di-O-oleyl-3-O-α-D-galactopyranosyl-sn-glycerol (BbGL-II), and sulfoglycolipid 37 (SL37). These lipids differ in size and charge of their headgroups and include lipids with a hexose sugar, a dihexose, a phosphate-bearing headgroup, and a small negatively charged headgroup (Fig. 1b).

CD1b-specific T cells among polyclonal T cell lines

To further study the occurrence of CD1b-specific T cells amongst TRBV4-1-expressing T cells, we generated a TRBV4-1⁺ polyclonal T cell line from PBMC of a random blood bank donor 43 (D43). As control, we generated a T cell line that does not contain any TRBV4-1-expressing cells, but otherwise represents the natural diversity of all other TCR sequences in that donor. Equal numbers of TRBV4-1⁺ and TRBV4-1⁻ T cells were sorted based on staining with antibodies against CD3 and TRBV4-1, and both populations were expanded by stimulation with αCD3 antibody. After this initial round of expansion, the uniform expression or absence of TRBV4-1 was confirmed (Supplemental Fig. 1). Both T

cell lines were stained with CD1b tetramers that were loaded with PG or GMM, or not loaded with an exogenously added antigen (called CD1b-endo). In the TRBV4-1⁺ population there is an increased frequency of cells that bind to CD1b-PG or CD1b-GMM compared to the TRBV4-1⁻ population (Fig. 2a). This indicates that TRBV4-1⁺ is more likely to confer CD1b-reactivity than any other Va segment and that the reactivity is not limited to CD1b bound to one specific antigen. In fact, PG and GMM are very different from each other with a glucose versus a phosphoglycerol on the TCR-facing surface of CD1b^{10,11}.

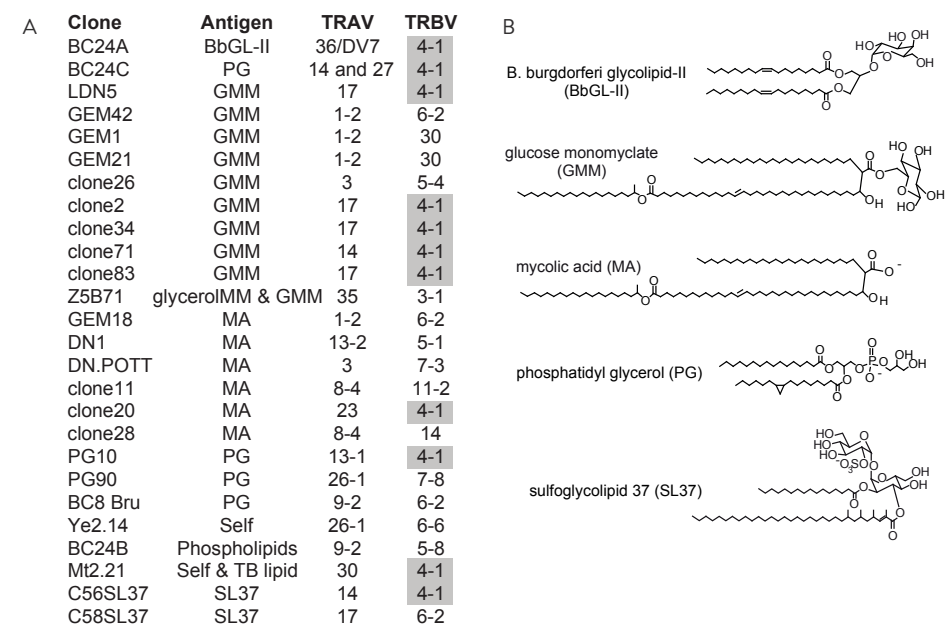


Figure 1: TCRs of established CD1b-restricted clones

(A) Established and new CD1b-restricted T cell clones are listed with their TCR α and β chain variable gene and the antigen they recognize. The TRBV4-1 gene segment is marked in grey. T cell clones were previously published: LDN5³⁶; GEM1, GEM18, GEM21, GEM42⁶; clone 2, clone26, clone34, clone71, clone83⁶; Z5B71³⁹; DN1¹⁶; DN.POTT^{17,40}; PG10, PG90, BC8 Bru¹⁵; YE2.14⁴¹; MT2.21³¹; C56SL37, C58SL37⁴²; clone 11, clone 20, clone 28¹⁸; BC24A, BC24B, and BC24C [Shahine in press]. BC24C expresses two different α chains. (B) Structures of the lipid antigens BbGL-II: 1,2-di-oleyl-α-galactopyranosyl-sn-glycerol; PG: phosphatidylglycerol; GMM: glucose monomycolate; MM: monomycolate; SL37: a synthetic di-acylated sulfoglycolipid analog⁴³.

A representative TCR sequence was obtained from the CD1b-PG or CD1b-GMM tetramer⁺ populations by single cell TCR sequencing. Among the CD1b-PG binding cells were several single cells that expressed identical α chains composed of TRAV8-2 and TRAJ38 and

identical β chains composed of TRBV4-1 and TRBJ1-6, suggesting a clonally expanded population, possibly caused by the in vitro expansion after stimulation with α CD3 antibody. Likewise, among cells that bind to CD1b-GMM were cells that expressed identical α chains composed of TRAV13-2 and TRAJ32 and identical β chains composed of TRBV4-1 and TRBJ1-4 (Fig. 2b). Even though in healthy blood bank-derived donors the frequency of GMM- and PG-specific cells is typically below or at the detection limit of 10^{-4} , our data show that pre-enrichment for TRBV4-1 apparently enriches for CD1b-specific cells to a level that they can now be detected.

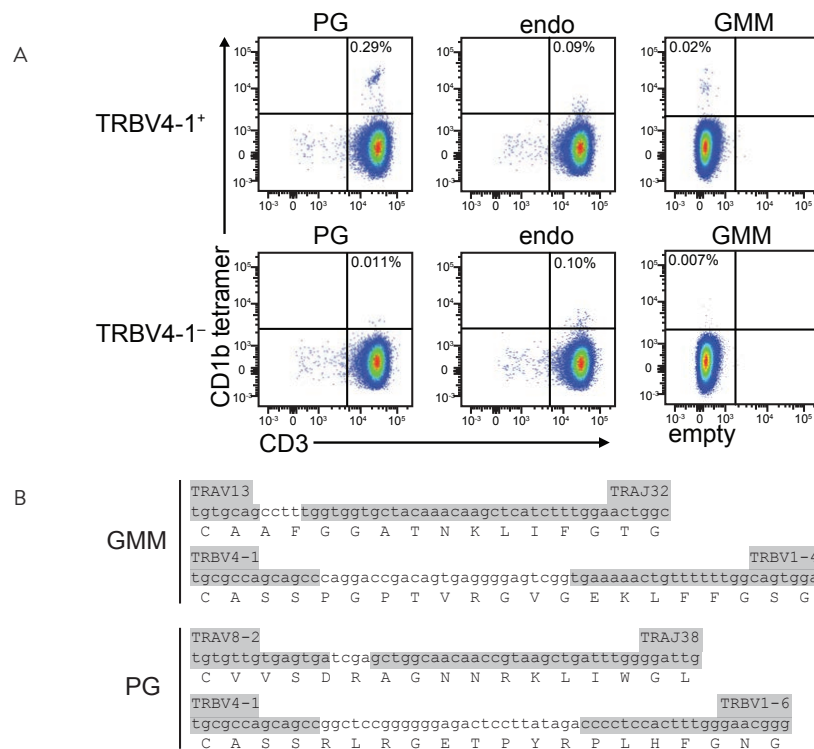


Figure 2: CD1b recognition by TRBV4-1⁺ T cells

(A) Sorted TRBV4-1⁺ (top) and TRBV4-1⁻ (bottom) T cell populations from random blood bank donor D43 (Supplemental Fig. 1) were stained with PG loaded CD1b tetramers, mock loaded CD1b tetramers, and GMM loaded CD1b tetramers. (B) TCR sequences obtained by single cell TCR sequencing of CD1b-GMM (4 out of 5 sequenced cells) and CD1b-PG (5 out of 6 sequenced cells) tetramer positive cells. In grey: corresponding gene segment.

CD1b-specific T cells among TRBV4-1⁺ T cells ex vivo

Knowing that CD1b-specific cells are enriched among established CD1b-specific T cell clones (Fig. 1), and TRBV4-1⁺ cell lines (Fig. 2), we now asked the question whether CD1b-

specific cells are enriched among TRBV4-1⁺ cells directly ex vivo. We used freshly isolated PBMCs from three random blood bank donors and stained them with anti-CD3, anti-TRBV4-1, CD1b-PG tetramer, and CD1b-GMM tetramer. In all three donors we detected a higher percentage of cells that stain with CD1b-PG or CD1b-GMM tetramer in the TRBV4-1⁺ population compared to the TRBV4-1⁻ population (Fig. 3a). CD1a tetramers do not show this difference (Supplemental Fig. 2). Thus, the enrichment of CD1b tetramer⁺ cells among TRBV4-1⁺ T cells and vice versa is not an artefact caused by culturing T cells because the same phenomenon was detected in freshly isolated, non-manipulated PBMC.

Several studies have demonstrated that the tetramers generally underestimate the frequency of functionally antigen-reactive T cells¹²⁻¹⁴. We were wondering whether many more TRBV4-1⁺ T cells recognize CD1b, regardless of whether they bind CD1b tetramer. All known CD1b-restricted phospholipid-specific T cell clones are autoreactive: they respond to antigen presenting cells (APCs) that express CD1b but do not need exogenously added antigen^{11,15}. To determine activation of TRBV4-1⁺ and TRBV4-1⁻ T cells by CD1b directly ex vivo, we tested for CD1b autoreactivity of sorted T cells in an ELISPOT for IFN- γ . As APCs we used the K562 myelomonocytic cell line expressing CD1a as negative control or CD1b. APCs were incubated with TRBV4-1⁺ T cells, TRBV4-1⁻ T cells, or an established CD1b-autoreactive T cell line as a positive control. This cell line, A25Salmonella, is known to produce IFN- γ after stimulation with CD1b-expressing cells in the absence of exogenously added antigen¹⁵ (Fig. 3b). Although the autoreactive positive control cell line did show a response to CD1b-expressing APCs, none of the ex vivo sorted cells showed increased IFN- γ production using CD1b-expressing APCs compared to using CD1a-expressing APCs or CD1b expressing APCs in the presence of a CD1b blocking antibody control. From this we conclude that even though a low frequency of CD1b tetramer⁺ T cells is detected among TRBV4-1⁺ T cells, we cannot detect an CD1b-autoreactive immunophenotype among the general TRBV4-1⁺ T cell population.

TRBV4-1⁺ expression by CD1b tetramer-positive and -negative T cells

Next, we determined the expression of all V β 's among CD1b tetramer⁺ and CD1b tetramer⁻ populations among polyclonal T cells (Fig. 4). T cells from two random blood bank donors (CX and C63) and one person who was latently infected with *M. tuberculosis* (C58) were stimulated with autologous monocyte-derived dendritic cells and mycolic acid, which is a CD1b-presented lipid¹⁶⁻¹⁸. At day 17 after antigen stimulation, mycolic acid tetramer⁺ were enriched by cell sorting (Supplemental Fig. 3) and the junctional regions of their T cell receptors were sequenced (Adaptive technology). When compared to tetramer⁻ sorted cells from the same polyclonal population, tetramer⁺ cells expressed TRBV4-1 much more frequently in CX and C58 (Fig. 4a). In addition, publicly available datasets¹⁹ containing junctional regions of CD1b-glucose monomycolate tetramer⁺ T cell population sorted

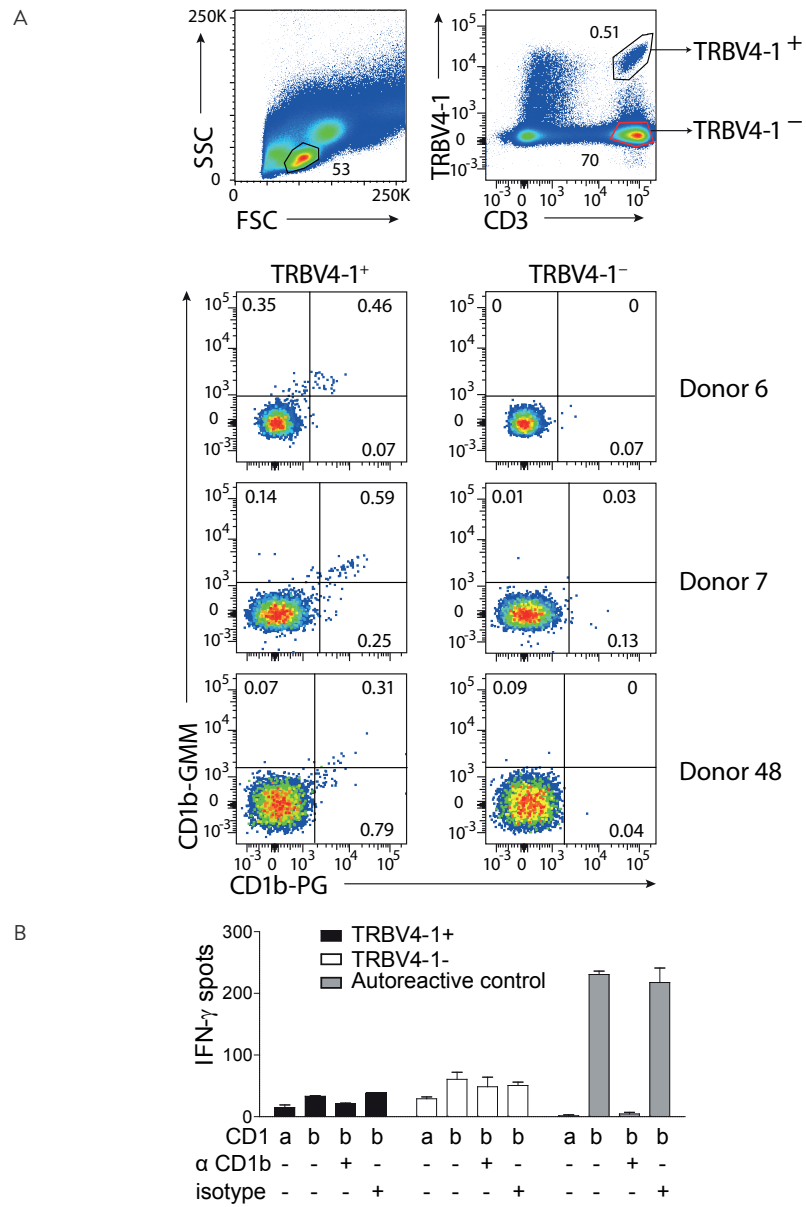


Figure 3: CD1b tetramer binding by PBMCs

After pre-gating using anti-CD3 and anti-TRBV4-1 antibodies (top), TRBV4-1⁺ and TRBV4-1⁻ T cells from PBMC from three random blood bank donors are analyzed for binding of CD1b-PG and CD1b-GMM tetramers directly ex vivo. Equal numbers of cells are shown in each plot. All acquired TRBV4-1⁺ cells are shown in Supplemental Figure 2A. (B) IFN- γ ELISPOT using TRBV4-1⁺ and TRBV4-1⁻ T cell populations sorted directly ex vivo (3000 cells/well), and an autoreactive T cell line named A25Salmonella (15) (1000 cells/well), in the presence (+) or absence (-) of a blocking antibody against CD1b.

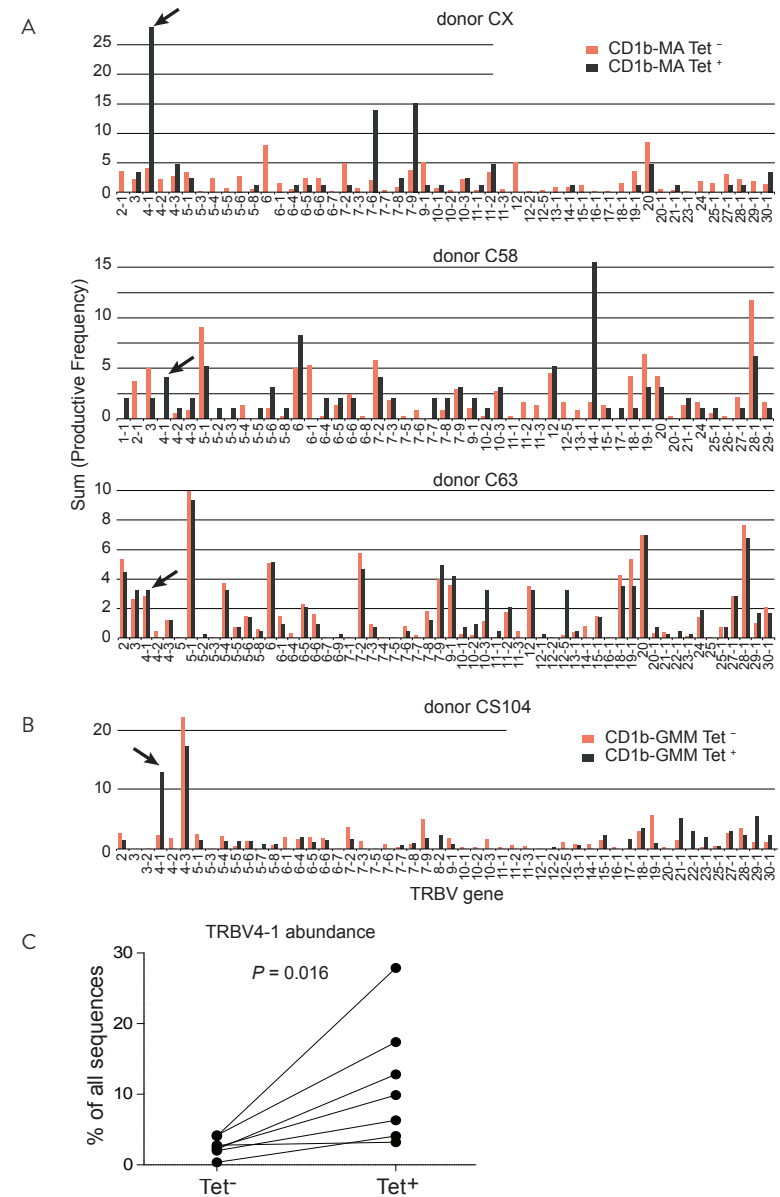


Figure 4: TRBV4-1⁺ T cells are enriched among CD1b-mycolic acid tetramer⁺ T cells

(A) PBMC from three blood donors were stimulated with autologous monocyte-derived dendritic cells and mycolic acid (MA) for 18 days. The resulting cells were stained with CD1b-MA tetramers and an antibody against CD3 followed by sorting of tetramer⁺ and tetramer⁻ cells (Supplemental Figure 3) and TCR sequencing (Adaptive). The percentages of TRBV gene usage are shown. (B) Percentage of TRBV usage of donor CS104 of a publicly available dataset¹⁹ of CD1b-glucose monomycolate (GMM) tetramer⁺ and tetramer⁻ cells. (C) TRBV4-1 percentage among tetramer⁺ and tetramer⁻ cells of the three donors presented here and four donors of a publicly available dataset.

from blood from four blood donors were enriched for TRBV4-1 compared to negative control populations from the same donors (Fig. 4b). Together, these datasets from seven donors, generated by two different labs, show a highly significant enrichment of TRBV4-1 among CD1b tetramer⁺ cells (Fig. 4c). Because glucose monomycolate and mycolic acid are chemically distinct antigens and TCRs specific for one do not cross-react with the other, preferred expansion of TRBV4-1⁺ cells by both antigens suggests that TRBV4-1 is driving CD1b recognition rather than antigen recognition.

A	Clone	Donor	TRBV	TRBJ	Antigen	CDR3 β
	BC24A	BC24	4-1	1-4	BbGL-II	CASSRWHFEKLF
	BC24C	BC24	4-1	2-7	BbGL-II & PG	CASSQRQGRGEQYF
	LDN5	Leprosy	4-1	2-1	GMM	CASSQPIGGGEQFF
	clone2	A14	4-1	1-1	GMM	CASSPLGREGLNTEAFF
	clone34	C58	4-1	2-2	GMM	CASSQLQVTGELFF
	clone71	C58	4-1	2-6	GMM	CASSLIGLADPSGANVLTFF
	clone83	C39	4-1	1-3	GMM	CASSPLTAAKVGNTIYF
	D43 - GMM	BC43	4-1	1-4	GMM	CASSPGPTVRGVGEKLF
	PG10	A25	4-1	2-3	PG	CASSQPPLGVGTDQYF
	D43 - PG	BC43	4-1	1-6	PG	CASSRLRGETPYRPLHF
	clone20	C58	4-1	1-1	MA	CASSHPLGSSKREAFF
	C56SL37	C56	4-1	2-7	SL37	CASSQALLTGSYEQYF

B	TRBV4-1 (clone2, PG10)	CDR1 β	CDR2 β	CDR3 β
	TRBV4-1 (clone2, PG10)	MGHRA	YSYEKL	CASSPLGREGLNTEAFF
	TRBV6-2 (GEM42)	MNHEY	SVGEGT	CASSQPPLGVGTDQYF
	TRBV7-8 (PG90)	SGHVS	FQNEAQ	CAVRNTGGFKTIF
		27	56	104

Figure 5: TCRs of CD1b-reactive TRBV4-1⁺ clones

(A) Of all TRBV4-1⁺ CD1b clones listed in Fig. 1 and the additional TRBV4-1⁺ clones described here, the amino acid sequence of the CDR3 β region is shown. BbGL-II: 1,2-di-oleyl-a-galactopyranosyl-sn-glycerol; PG: phosphatidylglycerol; GMM: glucose monomycolate; MM: monomycolate; SL37: a synthetic diacylated sulfoglycolipid analog (43). (B) CDR1 β , CDR2 β , and CDR3 β sequences of the clone2, PG10, GEM42 and PG90 TCRs.

TCR β chain CDR3 sequences

TRBV4-1 directly encodes CDR1 β and CDR2 β and the first few amino acids of CDR3 β . The C-terminal part the CDR3 β is encoded by N nucleotides, D segments and part of the J segment. It is possible that shared patterns in the CDR3 β that are not derived from TRBV4-1 contribute to recognition of CD1b. When looking in more detail at all known CD1b-reactive, TRBV4-1⁺ β chain sequences, including TCRs that recognize several different antigens, we observed no enrichment for a single TRBJ segment or a preference for TRBJ

segments that belong to the TRBC1 or TRBC2 group (Fig. 5a). No obvious CDR3 β patterns were found beyond the germline-encoded residues 104-107 CASS, suggesting that if conserved interactions between CD1b and TCR exist, they are most likely conferred by the germline-encoded CDR1 β , CDR2 β , or residue 104-107 of CDR3 β (Fig. 5b).

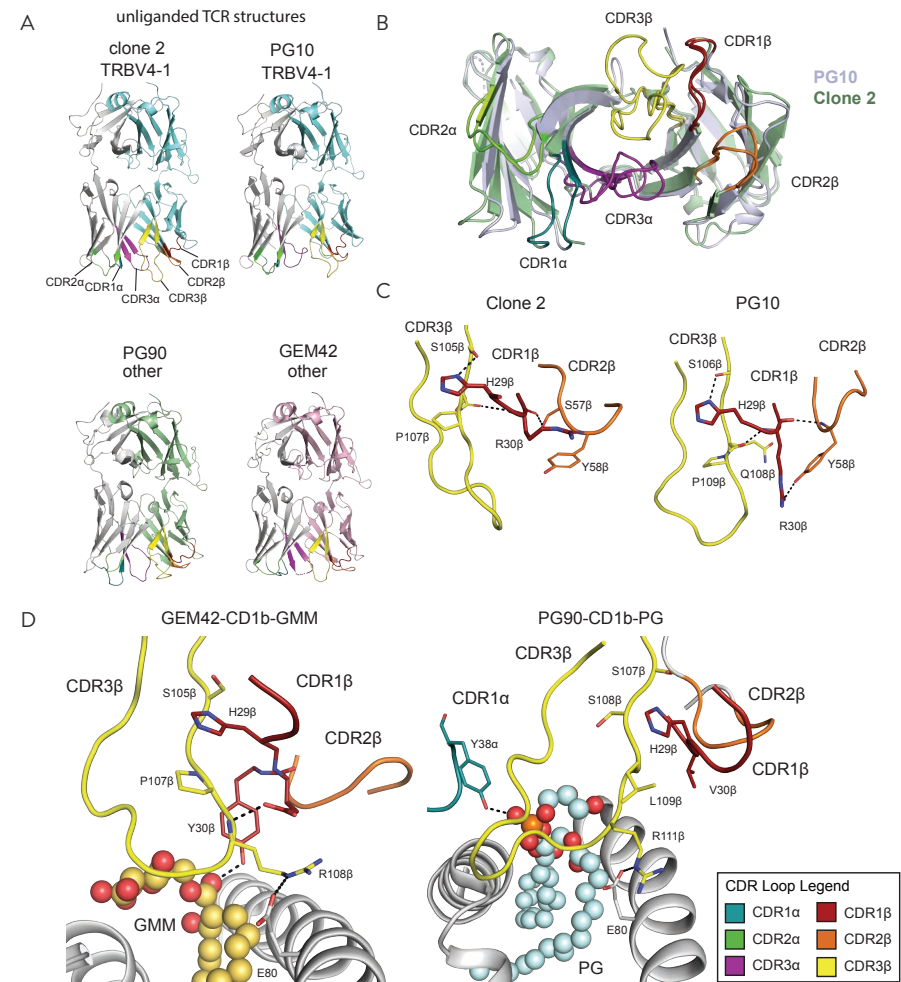


Figure 6: Structural analysis of CD1b-specific TRBV4-1⁺ TCRs

(A) Structure of unliganded TRBV4-1⁺ TCRs of clone 2 and PG10 and of the previously published structures of the TRBV4-1⁺ PG90 and GEM42 TCRs. (B) Overlay of the clone 2 (green) and PG10 (grey) TCRs, seen from the V domain side. The CDR loops are color-coded. (C) Overview of the interactions between the CDR β loops of the unliganded PG10 and clone 2 TCRs. (D) Overview of the interactions between the CDR β loops of the liganded PG90 and GEM42 TCRs.

Conserved TCR structure of clone 2 (GMM) and PG10 (PG)

To study how the germline-encoded CDR1 β , CDR2 β , and residue 104-107 of CDR3 β are exposed at the CD1b-facing surface of the TCR, we crystallized two unliganded TRBV4-1⁺ TCRs: the clone 2 TCR, which recognizes CD1b-GMM⁶, and the PG10 TCR, which recognizes CD1b-PG¹⁵ (Fig. 6a). Despite the different lipid ligand these TCRs recognize, the location of CDR1 β and CDR2 β was highly conserved among these two TCRs. In addition, the shape and orientation of CDR1 β was essentially identical (Fig. 6b). In both TCRs, CDR1 β residue H29 appears to play a stabilizing role, anchoring the germline-encoded N-terminal end of the CDR3 β loop by forming a hydrogen bond with S106 β . In turn, the backbone hydroxyl group of H29 is stabilized by the CDR2 β residues S57 β or Y58 β (Fig. 6c). Interestingly, in the GEM42 and PG90 TCRs, despite using TRBV6-2 and TRBV7-8, respectively, an H is also present on position 29 (Fig. 5b), and it performs a highly similar role (Fig. 6d).

In the unliganded structures, R30 appears to be flexible, hence the two different positions observed in the Clone 2 and PG10 structures (Fig. 6c). Its position in these structures depends on whether or it forms a hydrogen bond between the R30 headgroup and the Y58 β headgroup. When this hydrogen bond is formed (PG10), R30 is orientated downwards, presumably to where the TCR docks onto CD1b.

Mapping of essential residues in the CD1b-clone 2 interaction

To test if CDR1 β residues are crucial for recognition of CD1b-GMM complexes by the TRBV4-1⁺ clone 2, we made recombinant, mutant clone 2 TCR proteins. We mutated CDR1 β amino acids H29, and R30, and, as a negative control, framework region 1 amino acid T16 by alanine substitution. Surface plasmon resonance (SPR) experiments were performed to measure binding affinity between these TCR mutants and GMM-loaded CD1b monomers coupled to a chip. The K_D of wild type TCR and T16A-mutated TCR were comparable: $6.9 \pm 1.0 \mu\text{M}$ and $6.1 \pm 0.6 \mu\text{M}$ respectively. However, both mutations in the CDR1 β region showed a strong decrease in binding affinity, each resulting in a K_D of $>200 \mu\text{M}$ (Fig. 7a), indicating that these residues in the CDR1 β region are essential for CD1b-GMM binding. A converse experiment testing the ability of wild type clone 2 TCR protein to bind to mutated CD1b proteins was performed to determine the CD1b residues involved in the CD1b-clone 2 interaction (Fig. 7b). Among the ten CD1b mutants tested, only E80A, Y151A, and I154A caused greatly reduced or loss of clone 2 TCR binding.

CD1b E80 is essential for binding of all tested TRBV4-1⁺ TCRs

Having established the importance of certain CDR1 β and CD1b residues for the CD1b-GMM-clone 2 interaction, we wanted to identify candidate CD1b residues that are important for the interaction with a broader selection of TRBV4-1⁺ TCRs. We generated

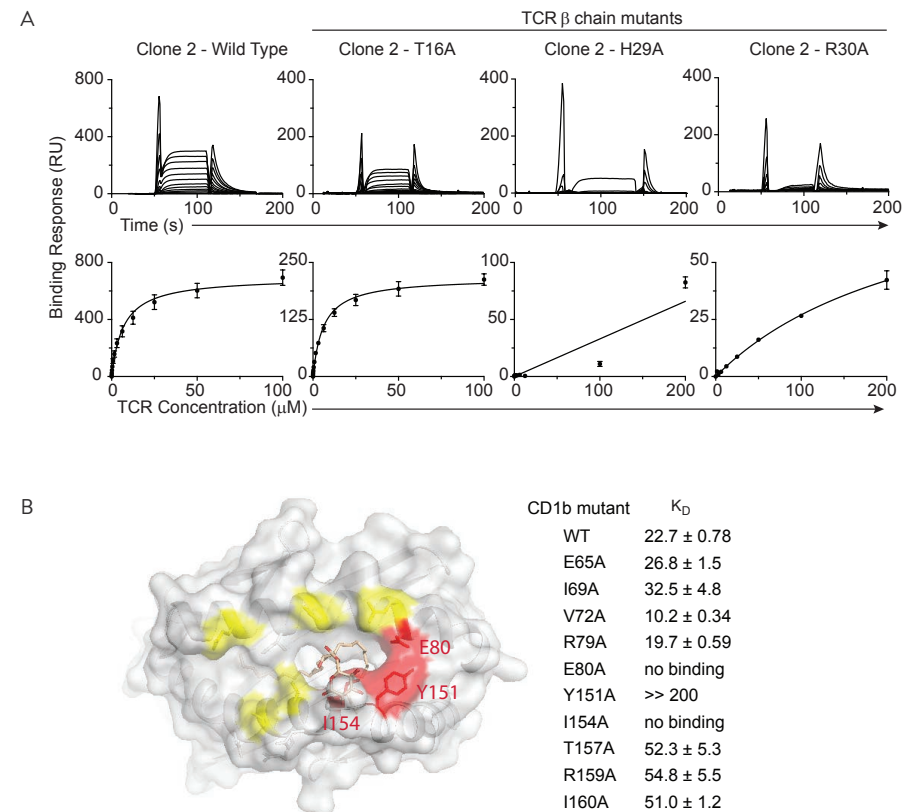
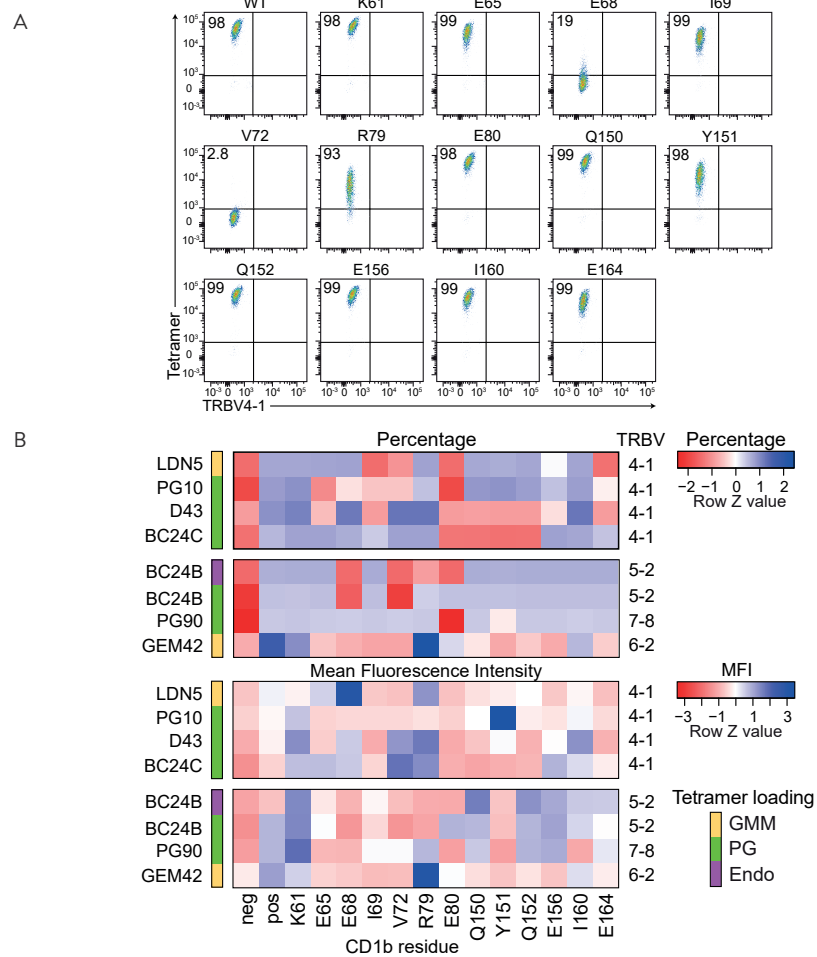


Figure 7: Mutational analysis of the clone 2-CD1b interaction by surface plasmon resonance

(A) The steady state affinity of the wild type TCR and TCR β chain point mutants of clone 2 for wild type CD1b loaded with GMM is measured. (B) The steady state affinity of the wild type clone 2 TCR for mutant CD1b proteins loaded with GMM was determined. Residues that, when mutated, greatly diminish the ability of the clone 2 TCR to bind, are shown in red. Residues that can be mutated without loss of clone 2 TCR affinity are shown in yellow.

mutated CD1b proteins and used these to make tetramers. In each mutant protein, one amino acid was replaced by an alanine. The mutated amino acids were all located at the surface of the CD1b protein that faces the TCR and were not predicted to be interfering with antigen binding^{10,20}. Four TRBV4-1⁺ and four TRBV4-1⁻ clones were analyzed using wild type and 13 mutant tetramers loaded with relevant antigen (Fig. 8 and Supplemental Figs. 4-10). With the exception of the K61A mutation, which did not have any effect compared to wild type CD1b, most mutations had a different effect on tetramer binding, depending on which TRBV4-1⁺ T cell line was tested, suggesting non-conserved interactions. However, all TRBV4-1⁺ T cells lost the capacity to bind to the E80A mutant CD1b-lipid complex. This was true for GMM- and for PG-loaded E80A mutant CD1b protein. The

E80A CD1b tetramers were not non-functional in a general way because they bound normally to two TRBV4-1⁻ T cells (Fig. 8b). These data demonstrate that E80 in CD1b is an important amino acid for interaction with TRBV4-1, independently of the antigen, but do not demonstrate direct interactions with TRBV4-1.



Discussion

In this paper we show that CD1b has a propensity to bind to TRBV4-1⁺ T cells and that among all T cells, TRBV4-1⁺ T cells are the most likely to bind to CD1b. CD1b genes are widely present among mammals²¹⁻²⁷ but despite the presence of two CD1D genes, no CD1a, CD1b, or CD1c homolog is present in mice²⁸. Interestingly, mice also lack a TRBV4-1 homolog^{29,30}. Analogous to the discovery of co-evolution between MR1 and TRAV1-2⁴, our findings suggest that a comparable relationship might exist between the presence of a TRBV4-1 and a CD1b ortholog in a species. Taking the recent discovery into account that CD1c restricted T cells also preferentially express TRBV4-1⁵, this may be true for CD1c as well.

The TRBV4-1⁺ T cells that we studied here recognize CD1b with several different ligands, including phosphatidylglycerol, glucose monomycolate, and mycolic acid. This is consistent with the hypothesis that genome-encoded parts of TRBV4-1 common to all TRBV4-1⁺ cells interact with CD1b, while other TCR segments, like the non-TRBV4-1-encoded part of CDR3 β , and the TCR α chain confer antigen dependence. However, crystal structures of two CD1b-specific TRBV4-1⁺ TCRs show that genome-encoded parts of the β chain interact with each other, creating a shaped surface and providing support of non-TRBV4-1-encoded parts.

For the amino acid E80 of CD1b, which was described by us and others as being essential for TRBV4-1 TCR recognition of CD1b³¹, different types of interactions have been described. It was previously suggested that E80 is part of a “basic triad” of amino acids around the CD1b portal, which was hypothesized to be necessary for antigen loading³². In our study, mutating E80 does not cause a lack of lipid loading *in vitro* because E80A mutant CD1b protein is still recognized by an antigen-dependent TRBV4-1⁻ T cell. Furthermore, in a binary structure of CD1b with SGL12 E80 interacts with Y151 of the CD1b molecule and mutation of E80 or Y151 in an alanine resulted in loss of TCR recognition³³. This E80-Y151 interaction is not observed in the bimolecular structure of CD1b with PC^{33,34}. Lastly, E80 can be involved in CD1b TCR interaction. In all CD1b-lipid antigen-TCR crystal structures currently available, E80 interacts with non-germline-encoded CDR3 β residues. This is true for the TRBV6-2⁺ GEM42 TCR¹⁰, the TRBV7-8⁺ PG90 TCR¹¹, and the autoreactive TRBV6-2⁺ BC8B TCR [shahine in press], and involves arginine, alanine, or leucine. However, even though these interactions illustrate the possibility of direct interactions between E80 and the TCR, they do not point to specific TRBV4-1-encoded residues.

In a published model of the docking of the TRBV5-1⁺ DN1 TCR on CD1b-mycolic acid, E80 interacts with CDR1 β residue R30³⁵. The CDR1 region of TRBV4-1 also contains an R30

and both TRBV5-1⁺ and TRBV4-1⁺ TCRs are affected by mutating E80 of CD1b³¹. For the clone 2 TCR, in addition to the importance of an intact CD1b E80, we demonstrated that the CDR1 β R30A mutation results in a loss of TCR binding. Remarkably, TRBV4-1⁺ TCRs of CD1c autoreactive T cells with an R30A mutation lost their CD1c reactivity⁵, and both CD1c and CD1b have a glutamic acid at position 80 (E80). Only the trimolecular structure of TRBV4-1⁺ TCRs with CD1b or CD1c can give a definitive answer about the TRBV4-1 TCR docking onto and molecular interactions with these molecules, but our and previously published data suggest that the TRBV4-1 H29 supports and shapes the overall TCR surface and R30 is flexible and can be available for contacts with CD1b or CD1c E80.

Material and methods

Tetramers

20 μ g CD1b monomers obtained from the National Institute Health (NIH) tetramer facility, were loaded with 32 μ g PG (Avanti lipids), GMM (Bill and Melinda Gates Foundation lipid bank) or synthetic BbGL-II (a gift from J. Kubler-Kielb, NIH). Lipids were dissolved in citrate buffer pH 4.5 with 0.5% CHAPS (Sigma) and incubated overnight at 37°C. After incubation the pH was neutralized by adding tris pH 8.5-9. As negative control CD1b monomers were treated in the same way without adding a lipid (CD1b-Endo).

T cell lines

The previously published cell lines A25Salmonella, which contains the clones PG10 and PG90¹⁵, LDN5³⁶, GEM42³⁷, BC24A, BC24B, and BC24C [Shahine, in press] were grown by stimulation with 30ng/mL α CD3 antibody and 25 x 10⁶ irradiated allogeneic peripheral blood mononuclear cells and 5 x 10⁶ irradiated Epstein-Barr virus-transformed B cells, and 1 ng/mL IL-2, which was added on day 2 of the culture. To generate a TRBV4-1⁺ and TRBV4-1⁻ T cell line, PBMC from random blood bank donor D43 were stained with α CD3 (555342, BD) and α TRBV4-1 (IM2287, Beckman Coulter). Cells were sorted on a BD FACSAria (BD Biosciences), lymphocytes were selected based on forward scatter and side scatter. 1x10⁶ CD3⁺TRBV4-1⁺ and 1x10⁶ CD3⁺TRBV4-1⁻ cells were sorted. After 2 weeks of stimulation as described above both cells were stained with α CD3 and α TRBV4-1 antibodies to check purity.

Single cell TCR sequencing

To each well of a Vapor-Lock (Qiagen)-coated 96 wells plate (Eppendorf) a mixture of 0.5 μ l 5x Iscript buffer, 0.5 μ l reverse transcriptase (Iscrip, Bio-Rad), and 1.25 μ l H₂O was added per well, with a final concentration of 0.1% Triton X-100. Single cells were sorted in this 96 wells plate on the FACSAria (BD Biosciences). After sorting, the plate was spun

down at 3000 RPM at 4°C for 10 min. For cDNA synthesis, the plate was incubated at 25°C for 5 min, 42°C for 30 min, and 80°C for 5 min. TCR transcripts were amplified in two subsequent, nested PCRs. The primary reaction consisted of 2.5 μ l cDNA mixture as a template, 0.75 U Taq polymerase (Denville), 2.5 μ l 10X PCR buffer (Denville), 0.5 μ l 10 mM dNTPs, 2.5 pMol of each external TRAV and TRBV primers and 10 pMol of both external antisense TRAC and TRBC primers as described by Wang (38) in a total volume of 25 μ l. The following PCR conditions were used: 95°C for 2 min, 35°C times 95°C for 20 sec, 50°C for 20 sec, 72°C for 45 sec followed by 1 cycle of 72°C for 7 min. 2.5 μ l of this product was used as a template in two separate secondary PCR reactions. The mixtures are identical as the primary PCR except that in one reaction the internal TRAV and TRAC primers were used and in the other reaction the internal TRBV and TRBC primers as described by Wang³⁸. The following PCR conditions were used for the secondary PCR: 95°C for 2 min, 35 cycles of 95°C for 20 sec, 56°C for 20 sec, 72°C for 45 sec, followed by 1 cycle of 72°C for 7 min.

PCR products were analyzed by gel electrophoresis and paired α and β chain PCR products that resulted from the same single cell-containing well were sequenced by Sanger sequencing.

Flowcytometry on TRBV4-1⁺ and TRBV4-1⁻ cell lines.

Equal cell numbers of the TRBV4-1⁺ and TRBV4-1⁻ population were incubated for 10 min at room temperature in PBS 1% BSA with CD1b tetramers. Subsequently anti-CD3 (349201, BD) antibody was added and incubated for 10 min at room temperature and 20 min at 4°C. Cells were washed with PBS with 1% BSA and analyzed on LSRFortessa (BD Biosciences). PBMC were isolated by ficoll gradient centrifuge and washed twice with PBS. 0.2 μ g CD1b tetramers loaded with PG and CD1b tetramers loaded with GMM were mixed and added to 3x10⁶ PBMC for 10 min at room temperature, an anti-CD3 antibody (BD Biosciences) was added and incubated for 10 min at room temperature. An anti-TRBV4-1 antibody was added and incubated at 4°C for 20 min. Cells were washed with PBS 1% BSA and measured on LSRFortessa (BD Biosciences).

ELISPOT

96 wells 0.45 μ m Hydrophobic Multiscreen plates (Milipore) were coated with α IFN- γ antibody 1-D1K (Mabtech)(o/n, 4°C). The plate was blocked for 2hr after which 20,000 K562 transfected with CD1a or CD1b were co-cultured with 200 T cells in the presence or absence of blocking antibodies. After incubation overnight at 37°C cells were washed away and the plates were incubated for 2hr with a biotinylated α IFN- γ antibody (7-B6-1, Mabtech). The wells were washed with PBS/Tween and incubated with Extravidin-ALP (Sigma-Aldrich) for 1 hr. After washing with PBS/Tween followed by washing with PBS,

BCIP/NBT (Sigma-Aldrich) was added to visualize the spots. Spots were counted by the Immunospot reader (C.T.L).

Flowcytometry with CD1b point mutants

T cells were stained with all 13 mutants and wild type CD1b tetramers loaded with their specific antigen. The CD1b tetramers were added for 10 min at room temperature, an α CD3 antibody (555342, BD) was added and incubated for 10 min a room temperature. An α TRBV4-1 (IM2287, Beckman Coulter) antibody was added and incubated at 4°C for 20 min. Cells were washed with PBS/BSA (1%) and measured on LSRFortessa (BD Biosciences). Both percentage positive cell and MFI of the positive cells were Z score normalized and a heatmap was created using the heatmap.2 function of R (<https://cran.r-project.org/web/packages/gplots/index.html>).

Surface plasmon resonance

Soluble wild type and mutated TCR proteins were expressed and purified and passed in increasing concentrations over wild type or mutated biotinylated CD1b proteins loaded with the indicated lipid antigen, coupled to streptavidin-coated chips in a Biacore 3000. BIAevaluation version 3.1 software (Biacore) was used to fit the data to the 1:1 Langmuir binding model, and the equilibrium data were analyzed with the Prism program for biostatistics, curve fitting, and scientific graphing (GraphPad Software).

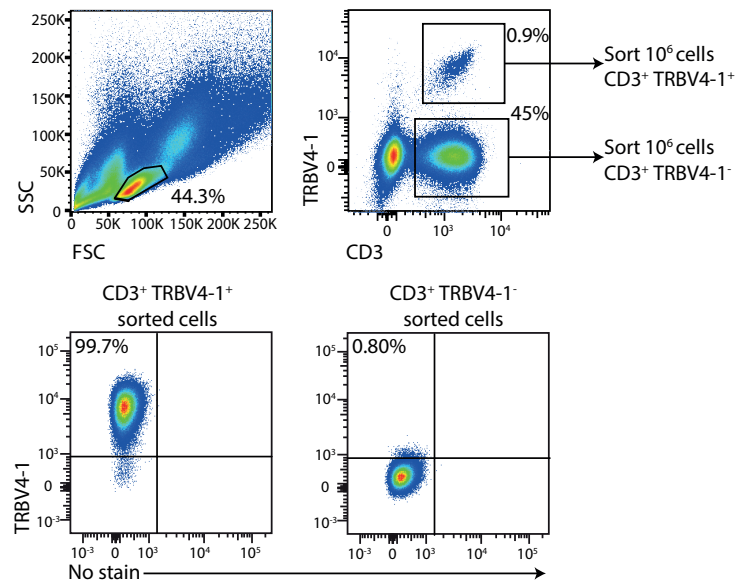
Statistics

On the data shown in Figure 4C, significance was calculated by a paired Wilcoxon ranked sums test.

References

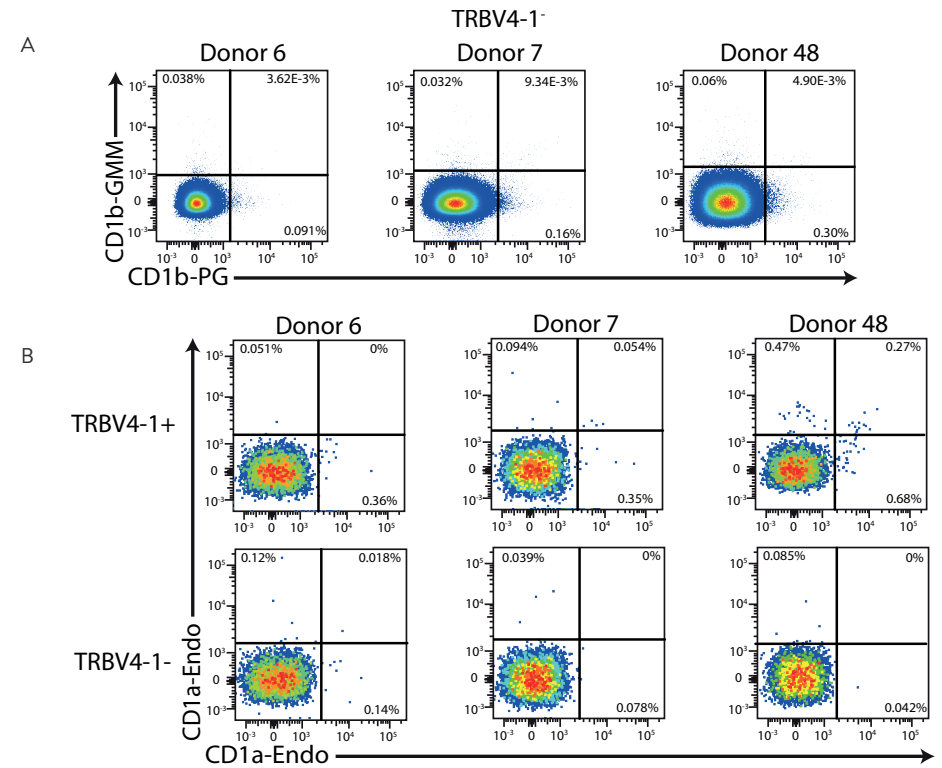
1. Sharon, E., L. V. Sibener, A. Battle, H. B. Fraser, K. C. Garcia, and J. K. Pritchard. 2016. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* 48:995-1002.
2. de Juan, D., F. Pazos, and A. Valencia. 2013. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 14:249-261.
3. Turner, S. J., P. C. Doherty, J. McCluskey, and J. Rossjohn. 2006. Structural determinants of T-cell receptor bias in immunity. *Nat. Rev. Immunol.* 6:883-894.
4. Boudinot, P., S. Mondot, L. Jouneau, L. Teyton, M. P. Lefranc, and O. Lantz. 2016. Restricting nonclassical MHC genes coevolve with TRAV genes used by innate-like T cells in mammals. *Proc. Natl. Acad. Sci. U. S. A.*
5. Guo, T., M. Y. Koo, Y. Kagoya, M. Anczurowski, C. H. Wang, K. Saso, M. O. Butler, and N. Hirano. 2018. A Subset of Human Autoreactive CD1c-Restricted T Cells Preferentially Expresses TRBV4-1(+) TCRs. *J. Immunol.* 200:500-511.
6. Van Rhijn, I., N. A. Gherardin, A. Kasmar, W. de Jager, D. G. Pellicci, L. Kostenko, L. L. Tan, M. Bhati, S. Gras, D. I. Godfrey, J. Rossjohn, and D. B. Moody. 2014. TCR Bias and Affinity Define Two Compartments of the CD1b-Glycolipid-Specific T Cell Repertoire. *J. Immunol.* 193:5338-5344.
7. Emerson, R., A. Sherwood, C. Desmarais, S. Malhotra, D. Phippard, and H. Robins. 2013. Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *J. Immunol. Methods* 391:14-21.
8. Freeman, J. D., R. L. Warren, J. R. Webb, B. H. Nelson, and R. A. Holt. 2009. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 19:1817-1824.
9. Warren, R. L., J. D. Freeman, T. Zeng, G. Choe, S. Munro, R. Moore, J. R. Webb, and R. A. Holt. 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* 21:790-797.
10. Gras, S., I. Van Rhijn, A. Shahine, T. Y. Cheng, M. Bhati, L. L. Tan, H. Halim, K. D. Tuttle, L. Gapin, J. Le Nours, D. B. Moody, and J. Rossjohn. 2016. T cell receptor recognition of CD1b presenting a mycobacterial glycolipid. *Nat Commun* 7:13257.
11. Shahine, A., I. Van Rhijn, T. Y. Cheng, S. Iwany, S. Gras, D. B. Moody, and J. Rossjohn. 2017. A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Science immunology* 2.
12. Laugel, B., H. A. van den Berg, E. Gostick, D. K. Cole, L. Wooldridge, J. Boulter, A. Milicic, D. A. Price, and A. K. Sewell. 2007. Different T cell receptor affinity thresholds and CD8 coreceptor dependence govern cytotoxic T lymphocyte activation and tetramer binding properties. *J. Biol. Chem.* 282:23799-23810.
13. Gallegos, A. M., H. Xiong, I. M. Leiner, B. Susac, M. S. Glickman, E. G. Pamer, and J. W. van Heijst. 2016. Control of T cell antigen reactivity via programmed TCR downregulation. *Nat. Immunol.* 17:379-386.
14. Rius, C., M. Attaf, K. Tungatt, V. Bianchi, M. Legut, A. Bovay, M. Donia, P. Thor Straten, M. Peakman, I. M. Svane, S. Ott, T. Connor, B. Szomolay, G. Dolton, and A. K. Sewell. 2018. Peptide-MHC Class I Tetramers Can Fail To Detect Relevant Functional T Cell Clonotypes and Underestimate Antigen-Reactive T Cell Populations. *J. Immunol.* 200:2263-2279.

15. Van Rhijn, I., T. van Berlo, T. Hilmenyuk, T. Y. Cheng, B. J. Wolf, R. V. Tatituri, A. P. Uldrich, G. Napolitani, V. Cerundolo, J. D. Altman, P. Willemsen, S. Huang, J. Rossjohn, G. S. Besra, M. B. Brenner, D. I. Godfrey, and D. B. Moody. 2016. Human autoreactive T cells recognize CD1b and phospholipids. *Proc. Natl. Acad. Sci. U. S. A.* 113:380-385.
16. Beckman, E. M., S. A. Porcelli, C. T. Morita, S. M. Behar, S. T. Furlong, and M. B. Brenner. 1994. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature* 372:691-694.
17. Grant, E. P., M. Degano, J. P. Rosat, S. Stenger, R. L. Modlin, I. A. Wilson, S. A. Porcelli, and M. B. Brenner. 1999. Molecular recognition of lipid antigens by T cell receptors. *J. Exp. Med.* 189:195-205.
18. Van Rhijn, I., S. K. Iwany, P. Fodran, T. Y. Cheng, L. Gapin, A. J. Minnaard, and D. B. Moody. 2017. CD1b-mycolic acid tetramers demonstrate T-cell fine specificity for mycobacterial lipid tails. *Eur. J. Immunol.* 47:1525-1534.
19. DeWitt, W. S., K. K. Q. Yu, D. B. Wilburn, A. Sherwood, M. Vignali, C. L. Day, T. J. Scriba, H. S. Robins, W. J. Swanson, R. O. Emerson, P. H. Bradley, and C. Seshadri. 2018. A Diverse Lipid Antigen-Specific TCR Repertoire Is Clonally Expanded during Active Tuberculosis. *J. Immunol.* 201:888-896.
20. Relloso, M., T. Y. Cheng, J. S. Im, E. Parisini, C. Roura-Mir, C. DeBono, D. M. Zajonc, L. F. Murga, M. J. Ondrechen, I. A. Wilson, S. A. Porcelli, and D. B. Moody. 2008. pH-dependent interdomain tethers of CD1b regulate its antigen capture. *Immunity* 28:774-786.
21. Dascher, C. C., K. Hiromatsu, J. W. Naylor, P. P. Brauer, K. A. Brown, J. R. Storey, S. M. Behar, E. S. Kawasaki, S. A. Porcelli, M. B. Brenner, and K. P. LeClair. 1999. Conservation of a CD1 multigene family in the guinea pig. *J. Immunol.* 163:5478-5488.
22. Dossa, R. G., D. C. Alperin, M. T. Hines, and S. A. Hines. 2014. The equine CD1 gene family is the largest and most diverse yet identified. *Immunogenetics* 66:33-42.
23. Eguchi-Ogawa, T., T. Morozumi, M. Tanaka, H. Shinkai, N. Okumura, K. Suzuki, T. Awata, and H. Uenishi. 2007. Analysis of the genomic structure of the porcine CD1 gene cluster. *Genomics* 89:248-261.
24. Hayes, S. M., and K. L. Knight. 2001. Group 1 CD1 genes in rabbit. *J. Immunol.* 166:403-410.
25. Loring van Beeck, F. A., D. M. Zajonc, P. F. Moore, Y. M. Schlotter, F. Broere, V. P. Rutten, T. Willemsen, and I. Van Rhijn. 2008. Two canine CD1a proteins are differentially expressed in skin. *Immunogenetics* 60:315-324.
26. Reinink, P., and I. Van Rhijn. 2016. Mammalian CD1 and MR1 genes. *Immunogenetics* 68:515-523.
27. Van Rhijn, I., A. P. Koets, J. S. Im, D. Piebes, F. Reddington, G. S. Besra, S. A. Porcelli, W. van Eden, and V. P. Rutten. 2006. The bovine CD1 family contains group 1 CD1 proteins, but no functional CD1d. *J. Immunol.* 176:4888-4893.
28. Bradbury, A., K. T. Belt, T. M. Neri, C. Milstein, and F. Calabi. 1988. Mouse CD1 is distinct from and co-exists with TL in the same thymus. *EMBO J.* 7:3081-3086.
29. Bosc, N., and M. P. Lefranc. 2000. The mouse (*Mus musculus*) T cell receptor beta variable (TRBV), diversity (TRBD) and joining (TRBJ) genes. *Exp. Clin. Immunogenet.* 17:216-228.
30. Koop, B. F., and L. Hood. 1994. Striking sequence similarity over almost 100 kilobases of human and mouse T-cell receptor DNA. *Nat. Genet.* 7:48-53.
31. Spada, F. M., E. P. Grant, P. J. Peters, M. Sugita, A. Melian, D. S. Leslie, H. K. Lee, E. van Donselaar, D. A. Hanson, A. M. Krensky, O. Majdic, S. A. Porcelli, C. T. Morita, and M. B. Brenner. 2000. Self-recognition of CD1 by gamma/delta T cells: implications for innate immunity. *J. Exp. Med.* 191:937-948.
32. Moody, D. B., D. M. Zajonc, and I. A. Wilson. 2005. Anatomy of CD1-lipid antigen complexes. *Nat. Rev. Immunol.* 5:387-399.
33. Garcia-Alles, L. F., A. Collmann, C. Versluis, B. Lindner, J. Guiard, L. Maveyraud, E. Huc, J. S. Im, S. Sansano, T. Brando, S. Julien, J. Prandi, M. Gilleron, S. A. Porcelli, H. de la Salle, A. J. Heck, L. Mori, G. Puzo, L. Mourey, and G. De Libero. 2011. Structural reorganization of the antigen-binding groove of human CD1b for presentation of mycobacterial sulfolipids. *Proc. Natl. Acad. Sci. U. S. A.* 108:17755-17760.
34. Garcia-Alles, L. F., K. Versluis, L. Maveyraud, A. T. Vallina, S. Sansano, N. F. Bello, H. J. Gober, V. Guillet, H. de la Salle, G. Puzo, L. Mori, A. J. Heck, G. De Libero, and L. Mourey. 2006. Endogenous phosphatidylcholine and a long spacer ligand stabilize the lipid-binding groove of CD1b. *EMBO J.* 25:3684-3692.
35. Grant, E. P., E. M. Beckman, S. M. Behar, M. Degano, D. Frederique, G. S. Besra, I. A. Wilson, S. A. Porcelli, S. T. Furlong, and M. B. Brenner. 2002. Fine specificity of TCR complementarity-determining region residues and lipid antigen hydrophilic moieties in the recognition of a CD1-lipid complex. *J. Immunol.* 168:3933-3940.
36. Moody, D. B., B. B. Reinhold, M. R. Guy, E. M. Beckman, D. E. Frederique, S. T. Furlong, S. Ye, V. N. Reinhold, P. A. Sieling, R. L. Modlin, G. S. Besra, and S. A. Porcelli. 1997. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* 278:283-286.
37. Van Rhijn, I., A. Kasmar, A. de Jong, S. Gras, M. Bhati, M. E. Doorenspleet, N. de Vries, D. I. Godfrey, J. D. Altman, W. de Jager, J. Rossjohn, and D. B. Moody. 2013. A conserved human T cell population targets mycobacterial antigens presented by CD1b. *Nat. Immunol.* 14:706-713.
38. Wang, G. C., P. Dash, J. A. McCullers, P. C. Doherty, and P. G. Thomas. 2012. T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Science translational medicine* 4:128ra142.
39. Layre, E., A. Collmann, M. Bastian, S. Mariotti, J. Czaplicki, J. Prandi, L. Mori, S. Stenger, G. De Libero, G. Puzo, and M. Gilleron. 2009. Mycolic acids constitute a scaffold for mycobacterial lipid antigens stimulating CD1-restricted T cells. *Chem. Biol.* 16:82-92.
40. Stenger, S., R. J. Mazzaccaro, K. Uyemura, S. Cho, P. F. Barnes, J. P. Rosat, A. Sette, M. B. Brenner, S. A. Porcelli, B. R. Bloom, and R. L. Modlin. 1997. Differential effects of cytolytic T cell subsets on intracellular infection. *Science* 276:1684-1687.
41. Vincent, M. S., D. S. Leslie, J. E. Gumperz, X. Xiong, E. P. Grant, and M. B. Brenner. 2002. CD1-dependent dendritic cell instruction. *Nat. Immunol.* 3:1163-1168.
42. James, C. A., K. K. Q. Yu, M. Gilleron, J. Prandi, V. R. Yedulla, Z. Z. Moleda, E. Diamanti, M. Khan, V. K. Aggarwal, J. F. Reijneveld, P. Reinink, S. Lenz, R. O. Emerson, T. J. Scriba, M. N. T. Souter, D. I. Godfrey, D. G. Pellicci, D. B. Moody, A. J. Minnaard, C. Seshadri, and I. Van Rhijn. 2018. CD1b Tetramers Identify T Cells that Recognize Natural and Synthetic Diacylated Sulfolipids from *Mycobacterium tuberculosis*. *Cell chemical biology* 25:392-402 e314.
43. Larrouy-Maumus, G., E. Layre, S. Clark, J. Prandi, E. Rayner, M. Lepore, G. de Libero, A. Williams, G. Puzo, and M. Gilleron. 2017. Protective efficacy of a lipid antigen vaccine in a guinea pig model of tuberculosis. *Vaccine* 35:1395-1402.



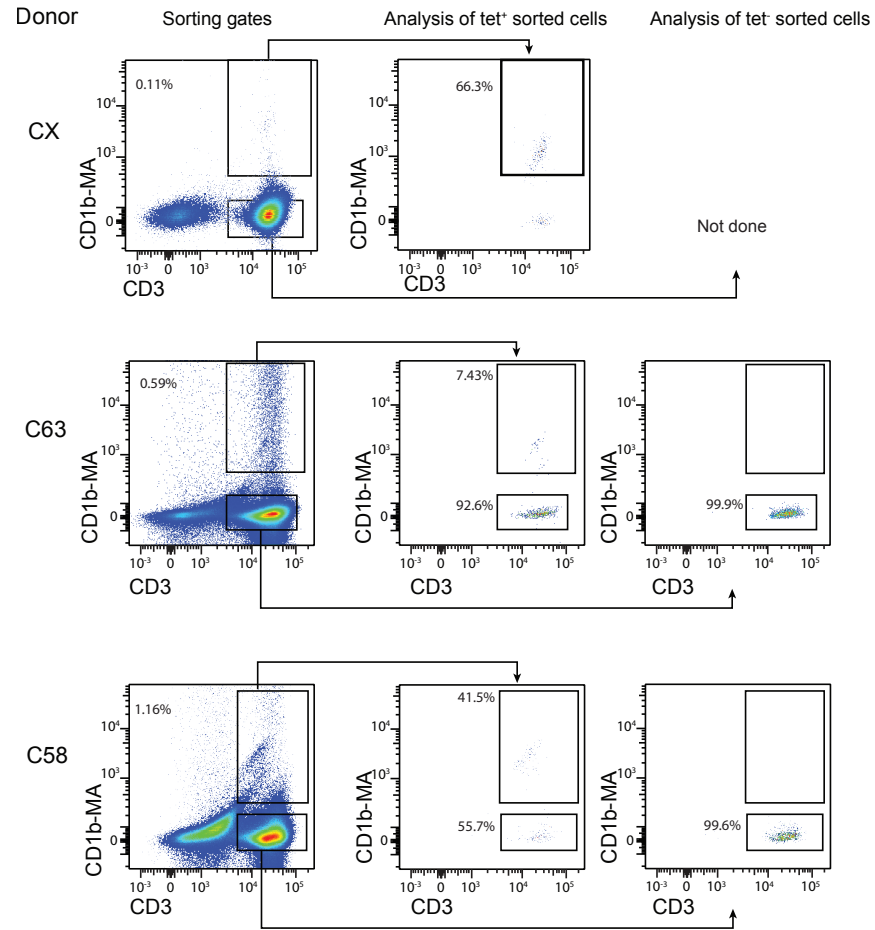
Supplemental Figure 1 Generation and validation of a TRBV4-1+ and a TRBV4-1- cell line

10⁶ TRBV4-1⁺ and 10⁶ TRBV4-1⁻ T cells were sorted from random blood bank donor D43 and stimulated with an anti-CD3 antibody (OKT3) and irradiated feeder cells. After two weeks of expansion their TRBV4-1 expression was checked by staining with an antibody against TRBV4-1.

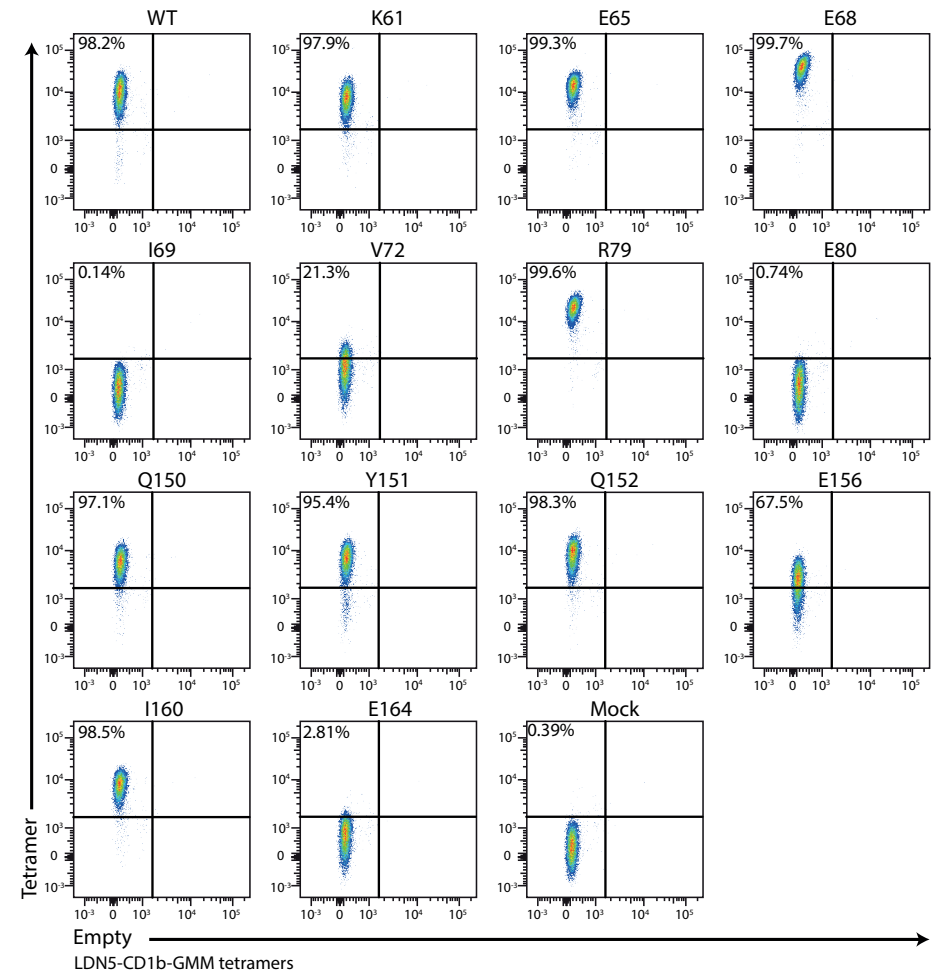


Supplemental Figure 2 CD1b tetramer binding by ex vivo T cells

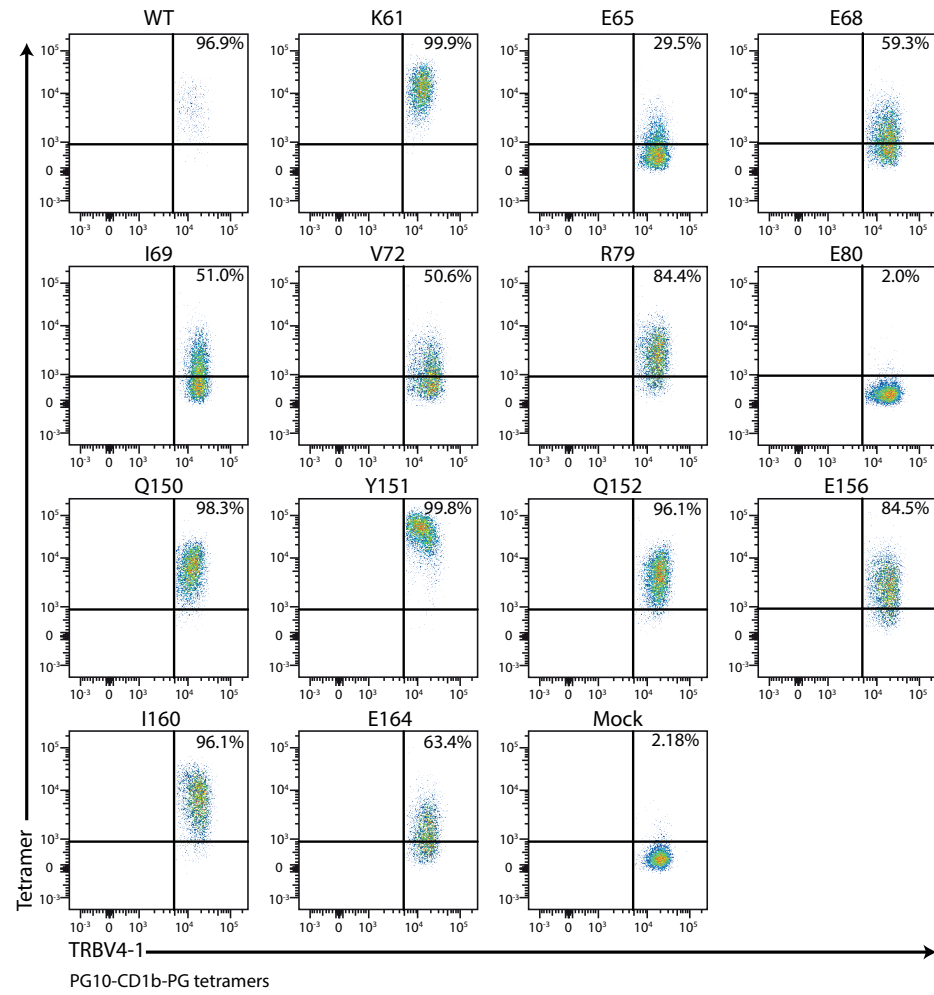
A: In Figure 3, some TRBV4-1⁻ cellular events were removed to be able to show equal cell numbers in all panels. Here we show all TRBV4-1⁻ cells stained for CD1b-PG and CD1b-GMM. B: Negative control staining with CD1a-endo.



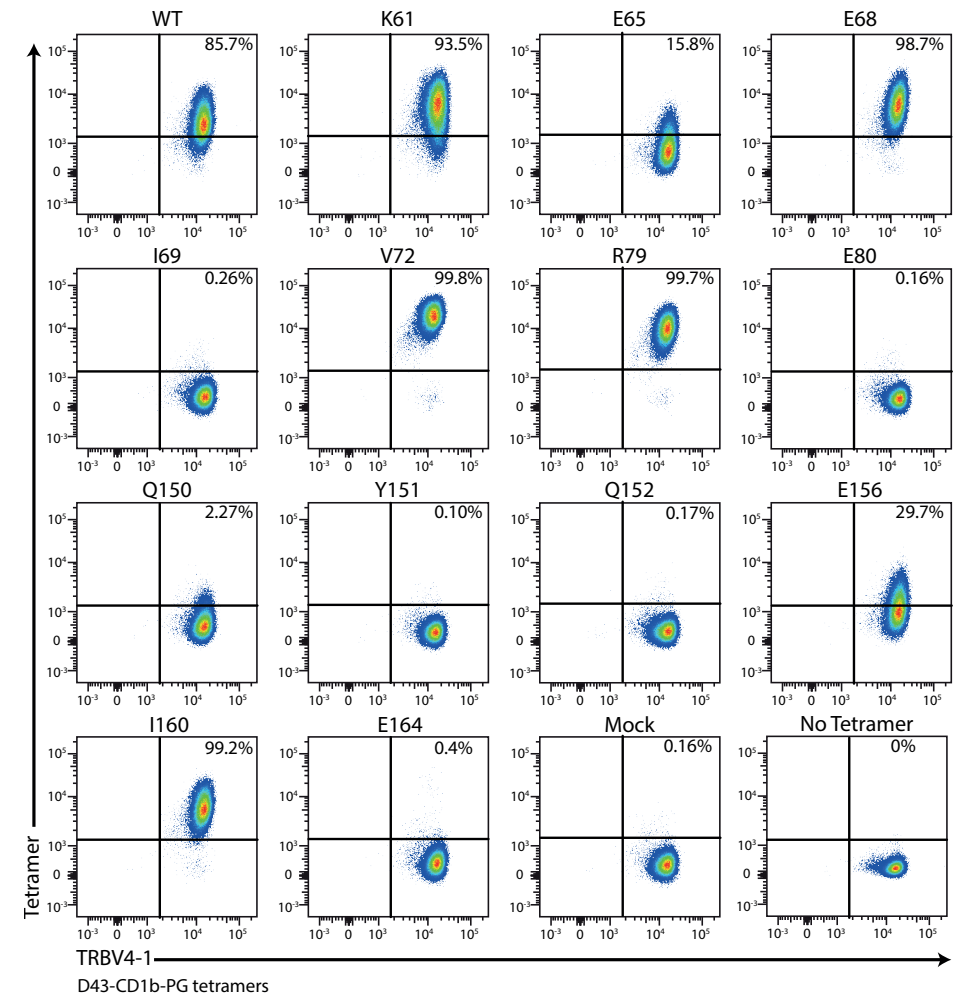
Supplemental Figure 3 Sorting strategy for CD1b-MA tetramer enriched cell populations
 From donors CX, C63 and C58, CD1b-MA positive cells were enriched by cell sorting. An aliquot of the sorted cells was immediately re-analyzed to determine the percentage of CD1b-MA tetramer+ cells.



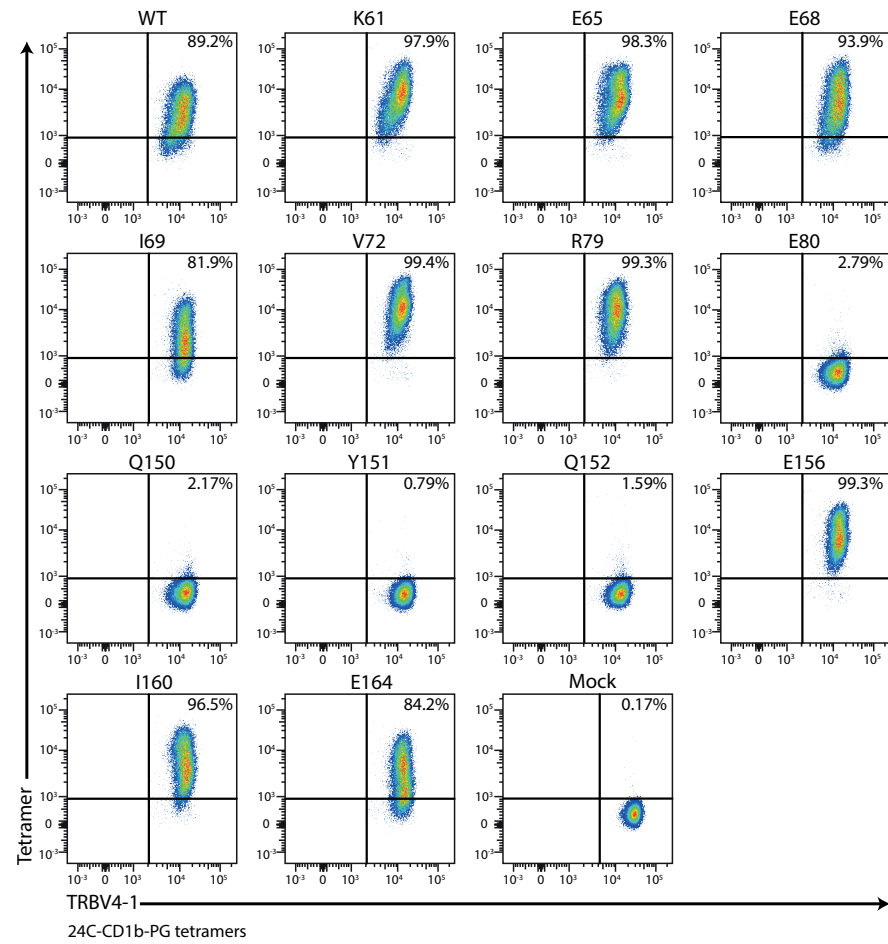
Supplemental Figure 4: Flow cytometry data using point mutant CD1b tetramers
 Raw flow cytometry data used for the heatmaps in Fig 8.: LDNS stained with CD1b-GMM.



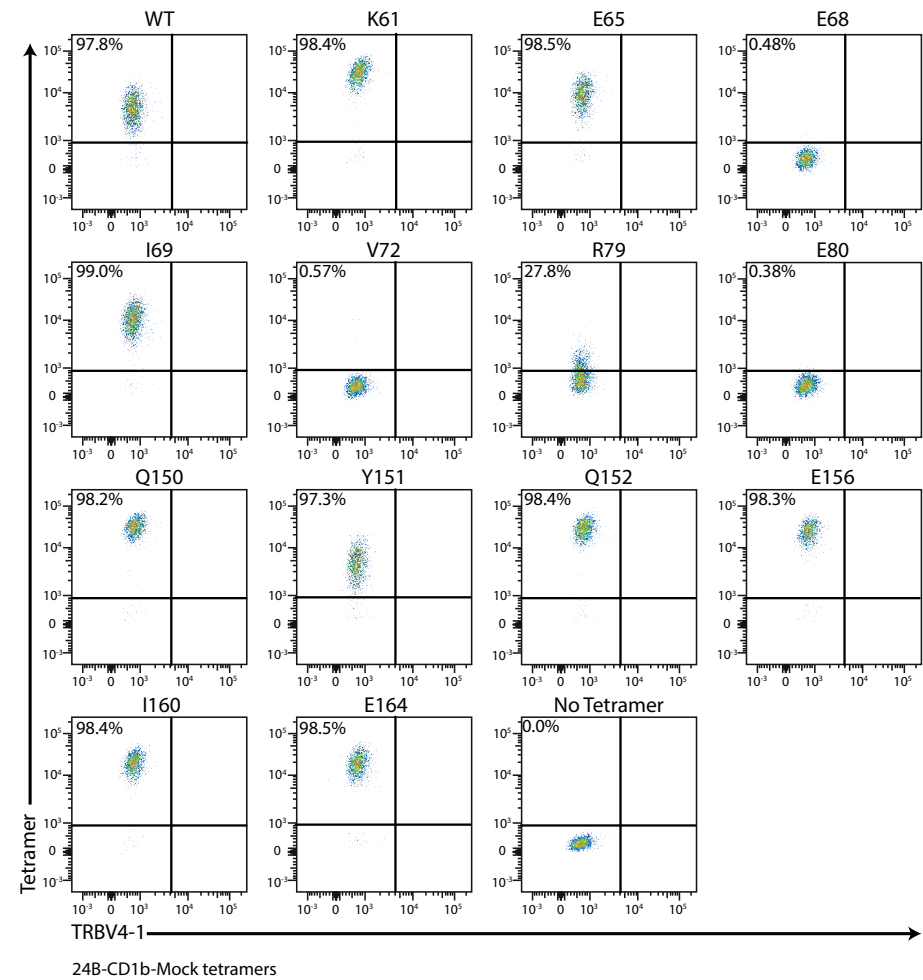
Supplemental Figure 5: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: PG10 stained with CD1b-PG.



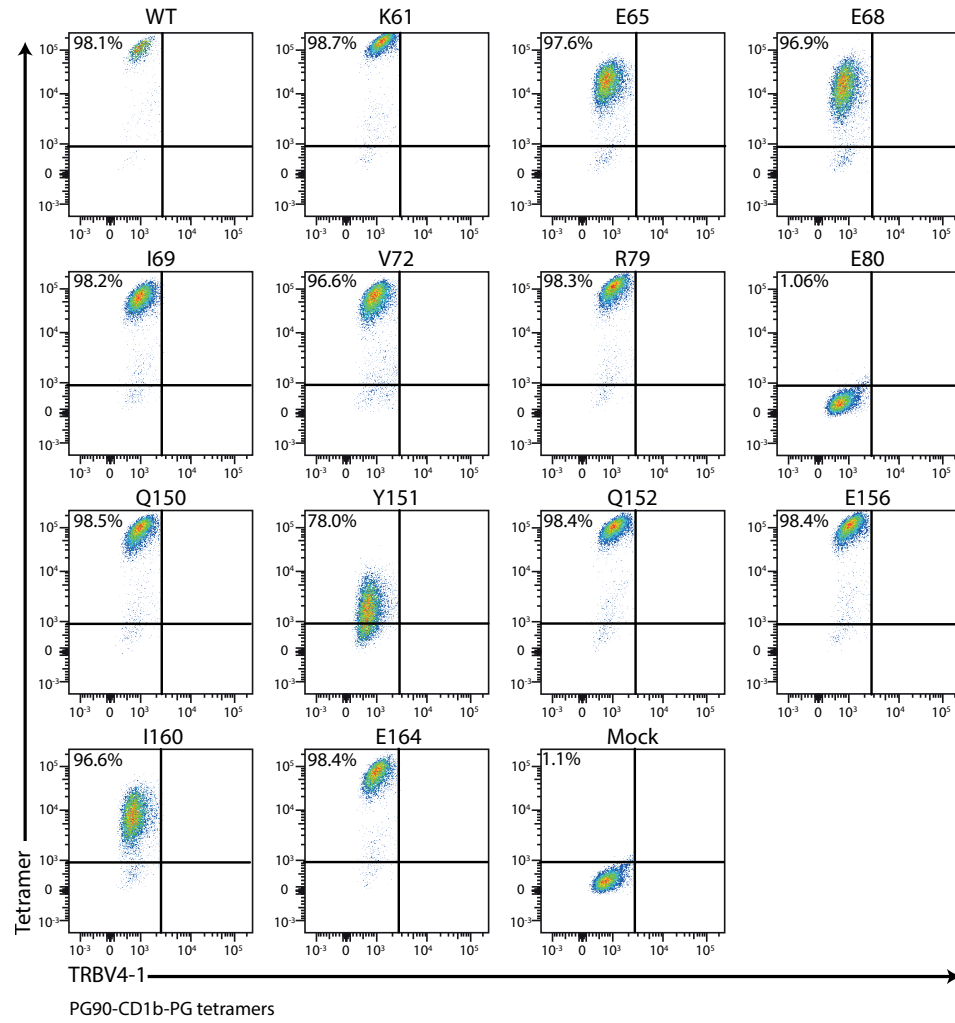
Supplemental Figure 6: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: D43 stained with CD1b-PG.



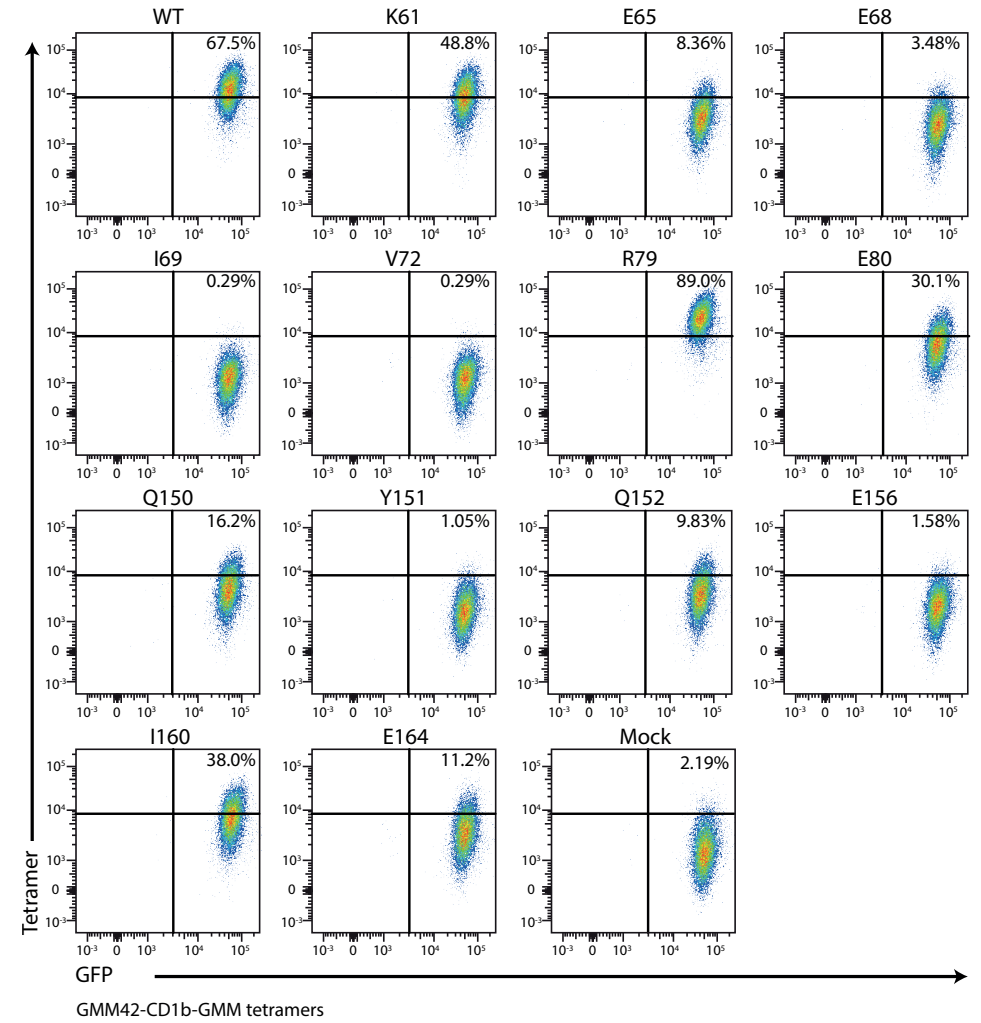
Supplemental Figure 7: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: BC24C stained with CD1b-PG.



Supplemental Figure 8: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: BC24B stained with CD1b-endo.



Supplemental Figure 9: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: PG90 stained with CD1b-PG.



Supplemental Figure 10: Flow cytometry data using point mutant CDb tetramers.
Raw flow cytometry data used for the heatmaps in Fig 8.: GEM42 stained with CD1b-GMM.

Chapter 5

A T-cell receptor escape channel allows broad T-cell response to CD1b and membrane phospholipids

5

Adam Shahine^{a, b}, Peter Reinink^{c, d}, Josephine F. Reijneveld^{c, d, e}, Stephanie Gras^{a, b}, Mira Holzheimer^e, Tan-Yun Cheng^d, Adriaan J. Minnaard^e, John Altman^f, Steffi Lenz^c, Jacques Prandi^g, Joanna Kubler-Kielb^b, D. Branch Moody^d, Jamie Rossjohn^{a, b, h, i} & Ildiko Van Rhijn^{c, d, *}

.....

^aInfection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

^bAustralian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia

^cDepartment of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584CL Utrecht, The Netherlands

^dBrigham and Women's Hospital Division of Rheumatology, Immunology and Allergy and Harvard Medical School, Boston, MA 02115, USA

^eStratingh Institute for Chemistry, University of Groningen, 9747AG Groningen, The Netherlands

^fDepartment of Microbiology and Immunology, Emory University School of Medicine, 1510 Clifton Road Atlanta, GA 30322, USA

^gInstitut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, Université Paul Sabatier, F-31077 Toulouse, France

^hNational Institute of Child Health and Human Development, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

ⁱInstitute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, UK

*Joint senior *Nature Communications*, 4 January 2019

Abstract

CD1 proteins are expressed on dendritic cells, where they display lipid antigens to T-cell receptors (TCRs). Here we describe T-cell autoreactivity towards ubiquitous human membrane phospholipids presented by CD1b. These T-cells discriminate between two major types of lipids, sphingolipids and phospholipids, but were broadly cross-reactive towards diverse phospholipids including phosphatidylcholine, phosphatidylinositol and phosphatidylethanolamine. The crystal structure of a representative TCR bound to CD1b-phosphatidylcholine provides a molecular mechanism for this promiscuous recognition. We observe a lateral escape channel in the TCR, which shunted phospholipid head groups sideways along the CD1b-TCR interface, without contacting the TCR. Instead the TCR recognition site involved the neck region phosphate that is common to all major self-phospholipids but absent in sphingolipids. Whereas prior studies have focused on foreign lipids or rare self-lipids, we define a new molecular mechanism of promiscuous recognition of common self-phospholipids including those that are known targets in human autoimmune disease.

Introduction

Human $\alpha\beta$ T-cells recognize antigen complexes formed from MHC proteins bound to diverse peptides, CD1 proteins bound to diverse lipids, and MR1 presenting small molecules¹. When an $\alpha\beta$ T-cell receptor (TCR) recognizes a self-peptide presented by an MHC protein, the sequence and structure of the peptide bound controls whether the T-cells respond^{2,3}. T-cells with high affinity for MHC-self-peptide are rarely isolated in the periphery because the negative selection mechanism in the thymus prevents most MHC-restricted autoreactive T-cells from entering the circulation. Nevertheless, T-cell autoreactivity towards self-peptide-MHC complexes occurs, which can manifest as autoimmune disease⁴. With increasing evidence that CD1 proteins present highly diverse self and foreign lipids to $\alpha\beta$ T-cells³, the interrelated questions of negative selection, T-cell fine specificity for lipids and ratios of self and foreign reactive T-cells in the periphery are likewise coming to the fore for lipids.

In this study we measured T-cell responses to lipid antigens using CD1b tetramers, which bind to antigen specific TCRs⁵. Although the four types of human CD1 antigen presenting molecules (CD1a, CD1b, CD1c, CD1d) are related in structure³, CD1b has distinct cellular and immunological functions. As compared to other human CD1 proteins, CD1b shows particularly strong recycling through the most acidic endosomes^{6,7}, use of a particularly capacious cleft to capture large lipids⁸, and it is uniquely expressed on a subset of activated macrophages in the periphery⁹. CD1b was the first of the human CD1 proteins discovered to present exogenous antigens¹⁰. There is extensive evidence for its presentation of mycobacterial lipids to T-cells, including mycolic acids¹¹, lipoarabinomannan¹², glucose monomycolate¹³, phosphatidylinositol dimannoside (PIM2)¹⁴, glycerol monomycolate¹⁵ and sulfoglycolipids^{16,17}. More recent results show that CD1 protein expression¹⁸ or CD1-restricted T-cells, can respond to lipids from other human pathogens, including *Borrelia burgdorferi*¹⁹ and Gram negative bacteria²⁰.

To investigate potentially new targets of CD1b-mediated T-cell response, we undertook an effort to detect human T-cells responding to mycobacterial diacyltrehalose, *Borrelia burgdorferi* glycolipid 2 and bacterial lipid extracts. The goal of the study was to select for foreign antigen-specific T-cells, and in so doing, identify new bacterial antigens. The approach was based on the premise that mechanisms of negative selection likely bias the peripheral T-cell repertoire toward TCRs with specificity for foreign lipids. We selected for foreign lipid reactive T-cells by using CD1b tetramers treated with pure bacterial ligands or complex lipid extracts from bacterial pathogens. Unexpectedly, we derived a series of CD1b autoreactive T-cells lines with broad responses to common self-phospholipids that are widely expressed in human cellular membranes. After identifying a distinct pattern whereby

T-cells respond broadly to phospholipids in preference to sphingolipids, we solve the basis of this response via a ternary crystal structure of a CD1b-phosphatidylcholine-TCR complex.

Results

CD1b-autoreactive T-cell lines are frequently isolated

To isolate bacteria reactive T-cells, we loaded CD1b tetramers with foreign lipid antigens or lipid extract (Table 1). After several rounds of T-cell sorting and expansion from blood bank-derived buffy coats (BC) or peripheral blood mononuclear cells (PBMC) from healthy donors (HD), we obtained oligoclonal T-cell populations that were > 85 % CD1b tetramer⁺ (Figure 1a and Supplementary Figs. 1-5). Prior approaches using activation assays largely failed to detect CD1b-reactive cells in unfractionated PBMCs^{21,22}, but these tetramer-based enrichment procedures succeeded in recovering CD1b tetramer-positive T-cells in each of five unrelated donors with a measured precursor frequency of about 1 in 10⁴ T-cells, ranging between 0.018 to 0.33 % (Supplementary Figs. 1-5 and Table 1). These numbers are somewhat higher than naive MHC-restricted T-cells and lower than human NKT-cells²³. The initial CD1b tetramer⁺ cell populations (HD1, BC8, BC10, BC13, BC24) were named according to the donor, and sublines A, B or C were designated when the line could be sorted into clearly distinct subpopulations (Supplementary Figs. 1-5).

Table 1: T-cell lines and subpopulations

Human subject	Selecting CD1b tetramer	Name of T-cell subpopulation	CD1b tetramer staining
BC10	BBGL1	BC10A	PL brighter than SL
BC8	<i>Brucella</i> lipid extract	BC8A	LPA
		BC8B	PL but not SL
		BC8C	PL but not SL
BC24	BBGL2	BC24A	BBGL2
		BC24B	PL but not SL
		BC24C	PG
		BC24D	PG
HD1	diacyltrehalose	HD1A	diacyltrehalose
		HD1B	PL but not SL
BC13	<i>Salmonella</i> lipid extract	BC13A	PL and SL

Summary of the T-cell lines used, including the human subject number from which the original blood PBMC were derived, the tetramer that was used to isolate the T-cells, the antigen, the names of the subpopulations, and their reactivity pattern. PL: phospholipids; SL: sphingolipids; LPA: lysophosphatidic acid; PG: phosphatidylglycerol. Subject numbers preceded by BC or HD refer to blood bank-derived buffy coats or healthy donors, respectively.

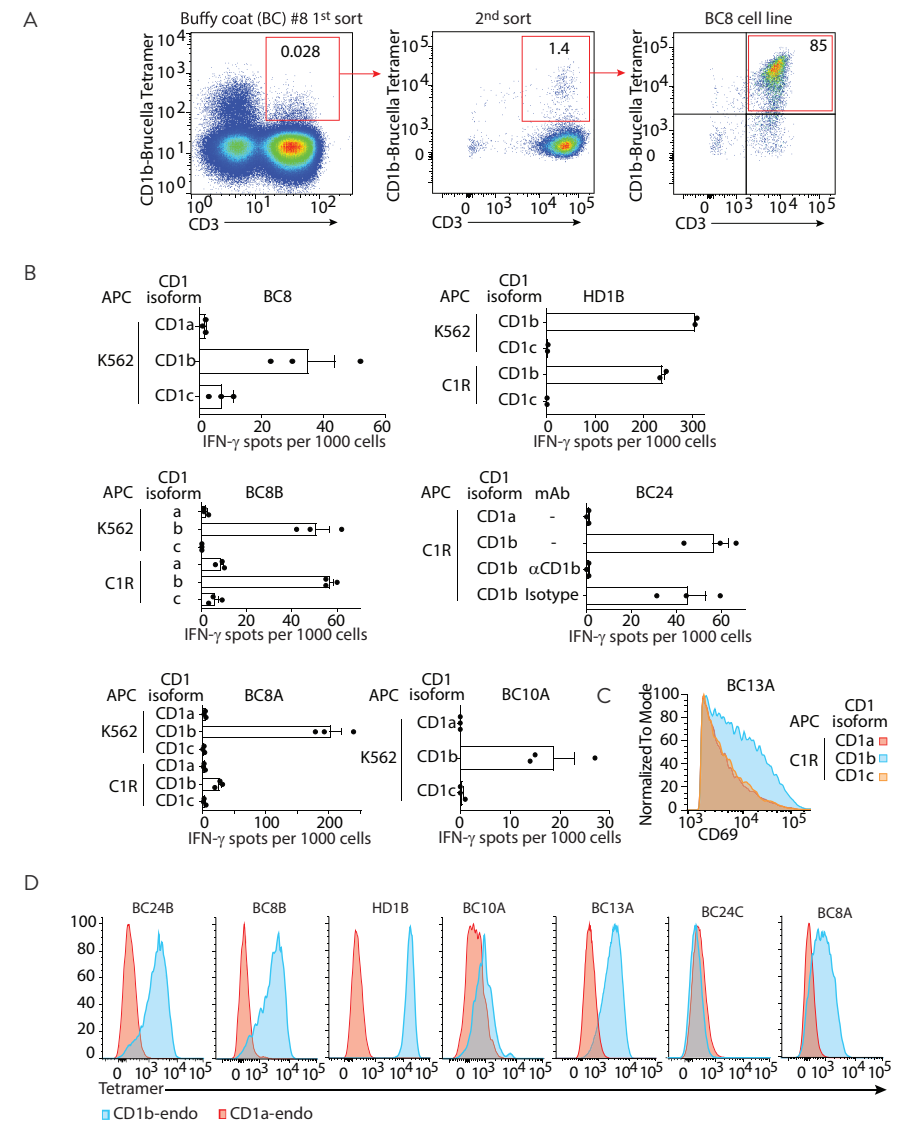


Figure 1: CD1b tetramer selection generates autoreactive T-cell lines.

(A) Peripheral blood mononuclear cells from blood bank-derived buffy coat 8 (BC8) were sorted based on expression of CD3 and binding to CD1b tetramers treated with *Brucella* lipid extract, followed by expansion in vitro. Numbers in the outlined red areas indicate percent cells in gate. The generation of additional T-cell lines is shown in Supplementary Figures 1-5. (B) IFN- γ ELISPOT assay of T-cell lines and subpopulations stimulated with K562 or C1R cells transfected to express CD1a, CD1b or CD1c in the absence of exogenously added antigen. Error bars represent standard error of the mean (SEM) of triplicate wells. (C) Jurkat cells transfected to express the BC13A T-cell receptor were stimulated with C1R cells transfected to express CD1a, CD1b, CD1c in the absence of exogenously added antigen. After 24 hours, CD69 expression was determined by flow cytometry. (D) Polyclonal T-cell populations from the indicated donor were stained with CD1a-endo or CD1b-endo tetramers. T-cell line derivation (A) and screening (B, D) was performed once on each of five donors. The experiments in C were performed twice with comparable results. APC: antigen presenting cell.

As expected, two cell lines recognized bacterial lipids: BC24A and HD1A (Supplementary Figs. 3-4 and Table 1). However, all other lines showed functional autoreactivity to CD1b proteins in the absence of added foreign lipid. For example, lines HD1B, BC24, BC10A, as well as line BC8 and two lines derived thereof, BC8A and BC8B, all showed interferon- γ release in response to CD1b-transfected but not CD1a- or CD1c-transfected human (K562 or C1R) cells (Fig 1b). Similarly, line BC13A upregulates the activation marker in response to CD1b-transfected C1R cells (Figure 1c). Further, CD1b-dependent autoreactive T-cell responses could be seen against monocyte-derived dendritic cells, which represent physiological APCs (Supplementary Figure 6). Thus, despite using methods to select foreign lipid reactive T-cells, all new lines showed substantial autoreactivity to CD1b.

Although somewhat counterintuitive, the recovery of CD1b autoreactive T-cells has been observed previously²⁰ and could be explained by extremely cross-reactive TCRs that do not discriminate lipid structure, retention of self-lipids from the expression system used to make CD1b tetramers, or antigenic lipids expressed both by mammalian cells and bacteria. This functional autoreactivity observed in activation assays was confirmed using untreated CD1b tetramers that carried endogenous self-lipids from the human expression system (CD1b-endo) and CD1a-endo as a negative control (Figure 1d). For five CD1b autoreactive sub-lines (HD1B, BC10A, BC13A, BC8B, BC8A), the pattern of CD1b-specific tetramer staining matched the observed CD1b-specific functional response. For donor BC24, functional reactivity to CD1b (Figure 1b) was likely explained by the response of the sub-line BC24B rather than BC24C (Figure 1d). Overall, these results confirmed CD1b autoreactivity using a separate method that relies on TCR binding to CD1b-endo complexes.

Self-lipids in CD1b complexes

Because few chemically identified self-antigens are known for the human CD1b system, we used these CD1b-endo complexes and the panel of CD1b autoreactive lines as a route to discover lipid autoantigens. Using a recently reported mass spectrometry approach²⁴, we identified several lipids present in the CD1b-endo monomers used to make CD1b tetramers. After monomers were treated with chloroform, methanol and water, we captured lipids in the organic phase and precipitated proteins into the aqueous phase. Lipid eluents were subjected to negative mode nano-electrospray ionization (nanoESI) ion trapping MS, identifying ions corresponding to deprotonated⁻ ions of phosphatidic acid (PA, m/z 673.5), phosphatidylethanolamine (PE, m/z 742.5), phosphatidylserine (PS, m/z 760.5), and phosphatidylinositol (PI, m/z 885.7, 887.7, 909.8). Further we detected chloride adducts of phosphatidylcholine (PC, m/z 794.5, 808.5, 820.5, 822.5, 842.5, 844.5) and sphingomyelin (SM, m/z 737.5, 765.5, 849.6). For some ions, these tentative identifications based on mass alone could not be formalized due to limited MS intensity or mass accuracy (Figure 2a). However, for all six lipid classes, CID mass

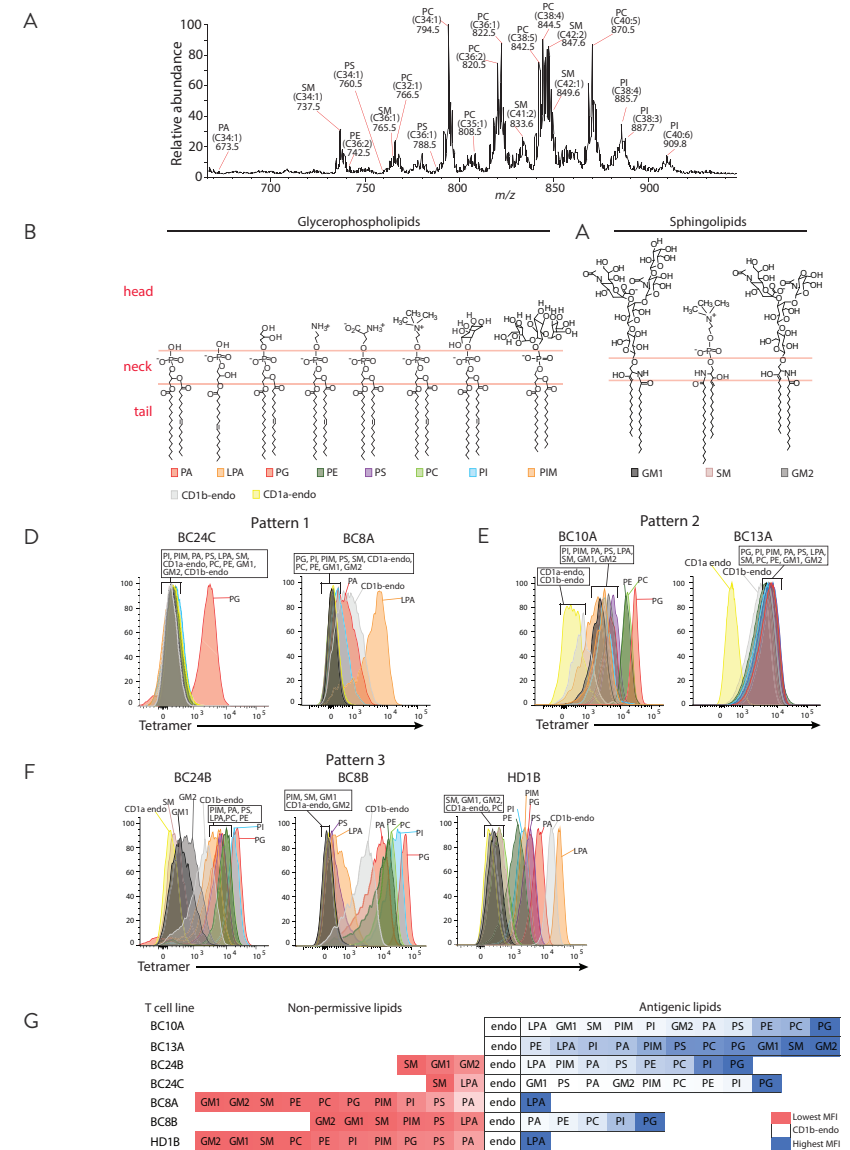


Figure 2: Patterns of antigen specificity among CD1b-restricted T-cell lines.

(A) Lipids were eluted from untreated CD1b proteins and analyzed by mass spectrometry. The identities of the indicated lipids were initially established by based on m/z value matching to masses of known compounds. All six lipid classes were confirmed by collisional mass spectrometry based on diagnostic fragmentation patterns (Supplementary Figure 6). (B-F) A panel consisting of 12 allophycocyanin-labeled CD1b tetramers and one CD1a tetramer, treated with the indicated phospholipid (B) or sphingolipid (C) or mock treated (endo), was used to stain the indicated T-cell lines. (G) Tetramer binding data are arranged according to mean fluorescence intensity (MFI) and aligned relative to the MFI of CD1b-endo staining. An experiment using the full panel of 13 tetramers and all T-cell lines was performed once. Replicate experiments with a smaller subset of the tetramer panel were conducted at least once for each T-cell line and tetramer.

spectrometry identified diagnostic ions leading to their confirmed structural identification (Supplementary Figure 7). For example, cleavage of PA ²⁵⁻ *m/z* 673.5 generated fragments identifiable as fatty acids (*m/z* 255, 281), and monoacylglycerols (*m/z* 391, 409, 417) (Supplementary Figure 7a). Considering the combined length of acylglycerol or sphingosine units, the lipid anchors of ligands ranged from C30-C42, with strongest signals seen for C38 and C40 lipids (Figure 2a). Similar results were obtained using HPLC-TOF-MS methods (Supplementary Figure 8).

Discovery of self-lipid antigens

In addition to lipids directly identified in CD1b-endo complexes, we designed a larger panel of lipids for screening in CD1b tetramer assays. These additional lipids included lysophosphatidic acid (LPA), ganglioside M1 (GM1) and GM2 and synthetic phosphatidylinositol dimannoside (PIM2) (Figure 2b-c). These lipids were chosen because they are naturally occurring, available in pure forms, and were known as CD1 ligands in other contexts ^{14, 26, 27}. Recent studies on human CD1a ^{28, 29} and CD1c ³⁰ have shown how lipids are more likely to act as autoantigens if they have small head groups or fewer lipid tails ^{20, 24, 31, 32}. Smaller CD1 ligands can remain sequestered within CD1 clefts, where head groups do not interfere with CD1-TCR contacts. Two decades after the discovery of neural tissue-derived glycosphingolipid presentation by CD1b ³³, ubiquitous antigen-presenting cell-intrinsic CD1b-presented autoantigens have only been recently discovered ^{20, 31} and general size constraints are unknown. Accordingly, lipids in this panel spanned a large range of head group size, from PA, which has no head group, to lipids with small head groups (PS, PE, PC, SM) up to glycolipids with one polyalcohol (inositol), or three (PIM2), four (GM2) or five (GM1) carbohydrates. Also, this panel sampled the two main types of lipids in mammalian membranes: glycerophospholipids and sphingolipids. Glycerophospholipids differ from sphingolipids in their ‘neck region,’ where glycerophospholipids contain an anionic phosphate ester, and sphingolipids have an amide linkage (Figure 2b-c). Thus, in addition to acting as probes to define new antigenic molecules, this panel tested lipid tails, neck regions and head groups as determinants of CD1b-TCR contact.

We treated CD1b-endo tetramers with vehicle alone and separately with each of the 11 lipids. These antigen-loaded CD1b tetramers were screened for staining of 7 CD1b autoreactive T-cell populations. Tetramers avoid certain false positive results seen in activation assays because tetramers rely on direct physical interactions of TCRs with antigen complexes, but tetramers also give false positive or negative results in some circumstances. Therefore, it is notable that for every lipid tested, alternatively loaded tetramers generated differential up or down regulated binding, based on the T-cell line used (Figure 2d-f). Visualization of the data as a matrix of 77 staining conditions makes clear that the seven lines provided positive and negative controls for one another (Figure 2g). No lipid generally failed to bind

tetramers, generally inactivated tetramers or generally caused tetramers to bind non-specifically to T-cells. Instead differing patterns within the matrix were most consistent with differential interactions with clonally distributed TCRs and the type of lipid loaded onto CD1b. This matrix identified many lipids that could block (non-permissive ligands) or enhance (antigens) CD1b-endo staining of T-cells, demonstrating broad recognition of self-lipids by human CD1b-reactive T-cells.

Three patterns of self-antigen reactivity

Further, by observing larger patterns in which few or many lipids affected T-cell staining, as well as the general classes of lipids that controlled staining, we could identify 3 general patterns of T-cell specificity for CD1b lipid-complexes. Pattern 1 was observed for alternatively loaded tetramer staining of BC24C and BC8A T-cell populations. Here T-cells recognized phospholipids in preference to sphingolipids, and they specifically stained with phospholipids carrying small head groups, such as PG and LPA (Figure 2d). The staining was highly specific for the lipid loaded, as there was no cross-reactive recognition of membrane phospholipids with larger head groups such as PC and PI. These results confirmed a known antigen response pattern previously observed in clones PG90 and PG10 ^{20, 24} and broadened the result to additional human donors.

In addition, this matrix identified two previously unknown lipid antigen reactivity patterns for CD1b. Two T-cell populations, BC10A and BC13A, constituted pattern 2. They recognized CD1b tetramers treated with many phospholipids or sphingolipids, (Figure 2e). Despite this extreme promiscuity for lipid structure, both lines were still specific for the CD1b isoform in T-cell activation (Figure 1) and tetramer staining (Figure 2e) assays. Last, pattern 3 was exemplified by BC24B, BC8B, and HD1B (Figure 2f). Here T-cells showed broadly cross-reactive recognition of many classes of self-phospholipids, including PG, PA, PI, PC, PS, and PE. Unlike pattern 1, binding was highly promiscuous among phospholipid types, with T-cells cross-reacting with phospholipids lacking a head group (PA) and those with varying head group size and structure. Like pattern 1, these lines were not cross-reactive to any tested sphingolipid, including gangliosides (GM1, GM2) and SM (Figure 2f). Of note, the phospholipid PC and the sphingolipid SM have the same phosphocholine head group (Figure 2b) but are differentially recognized. Thus, pattern 3 suggests that the T-cell response might be in some way specific for the phosphoglycerol neck region that differentiates sphingolipids from phospholipids.

Thus, screening with major classes of cellular self-lipids was highly informative regarding the identity of individual antigens identified, including PS, PE, PI and SM. Pattern 3, with apparent ‘neck region’ specificity that translated into preference for phospholipids over sphingolipids, was particularly intriguing. To our knowledge, this pattern lacked precedent.

Yet, it could potentially explain how T-cells respond differentially to the two major classes of membrane lipids present in human cells. Also from a structural perspective, pattern 3 was difficult to understand. The only prior structural study of an autoreactive TCR bound to CD1b, CD1b-PG-PG90 TCR, shows that head group protrudes ‘upward’ to contact the TCR²⁴. If head groups usually dominate TCR contact with phospholipids, it was unclear how pattern 3 TCRs could somehow equally recognize large (PI, PC), small (PS, PE, PG) or absent (PA) head groups, yet still discriminate neck region structures, which are likely positioned further from the TCR and nearer to CD1b. Therefore, we chose the BC8B TCR (Figure 2f) as a representative pattern 3 T-cell for detailed structural studies.

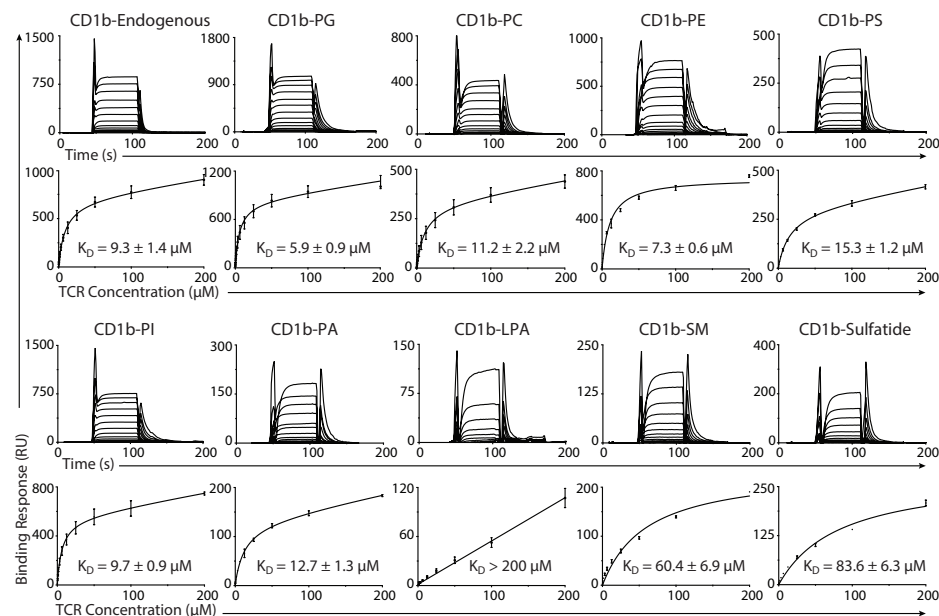


Figure 3: Steady-state affinity measurements of BC8B TCR against CD1b by SPR.

Steady state affinity measurements by SPR of the BC8B TCR against CD1b presenting a range of phospholipid and non-phospholipid antigens. Three (CD1b-endo, CD1b-PS, and CD1b-PC) or two independent (all other antigens) SPR experiments were conducted. Steady state (K_D) values and error bars represent mean \pm SEM in μM of duplicate measurements. Sensograms (upper) and equilibrium curves (lower) were generated in Graphpad Prism 7.0.

Binding of a broadly cross-reactive TCR to CD1b-phospholipid

First, we cloned and sequenced the TCR of BC8B. The TCR α chain complementarity determining region (CDR3 α) sequence CALTPSGGYQKVTF was formed by the gene segments TRAV9-2 and TRAJ13, and a CDR3 β sequence CASSMPGLRSSYEQYF was formed by TRBV6-2 and TRBJ2-7. TCR heterodimers were generated and purified

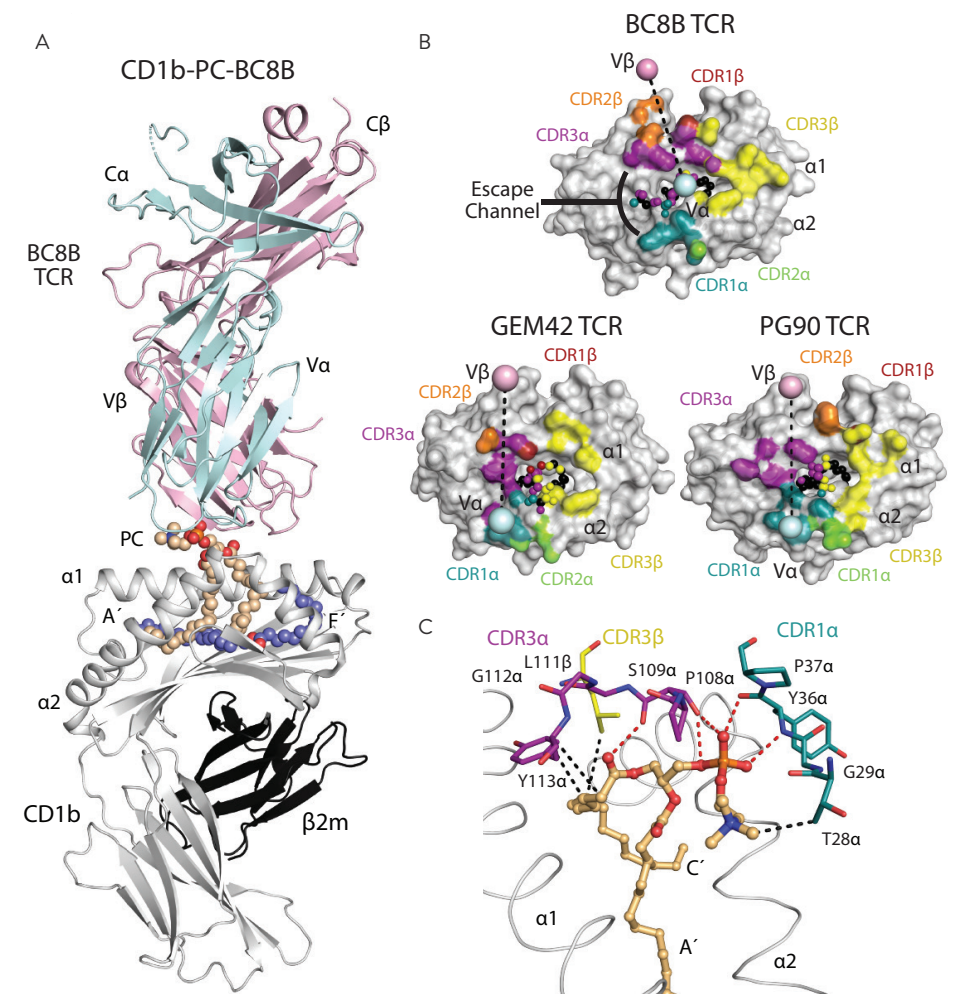


Figure 4: Overview of the CD1b-PC-BC8B crystal structure. (A) Overview of crystal structure of CD1b presenting PC in complex with the BC8B TCR. CD1b (grey), $\beta 2m$ (black), BC8B TCR α chain (light blue), and β chain (light pink) are represented as ribbons, with PC (light orange) and the scaffold lipid (dark blue) represented as spheres. Dashed lines indicate missing residues in the structure due to lack of electron density. (B) Footprint of BC8B CDR loops on CD1b-PC, in comparison to the PG90 TCR and GEM42 TCR CDR loops onto CD1b-PG and CD1b-GMM respectively. CD1b and antigens are represented as grey surface and spheres respectively, with points of contacts with TCR CDRs indicated as follows: CDR1 α (teal), CDR2 α (green), CDR3 α (purple), CDR1 β (red), CDR2 β (orange), and CDR3 β (yellow). Ligand atoms not contacted by the BC8B TCR are indicated in black. TCR α and β centers of mass are indicated by spheres and colored as indicated in Fig4a. Escape channel formed by BC8B TCR CDR1 α and CDR3 β regions is labelled. (C) Contacts between the BC8B TCR CDR1 α (teal), CDR3 α (purple), and CDR3 β (yellow) regions, CD1b (grey), and the PC antigen (light orange). Amino acid residues involved in contacts are represented as sticks, with the CDR regions, and $\alpha 1$ and $\alpha 2$ helices of CD1b represented as ribbons. Hydrogen bonds and hydrophobic interactions are indicated as red and black dashes respectively. Oxygen, nitrogen and phosphate ions are colored red, blue and orange respectively.

according to published protocols²⁴. In surface plasmon resonance (SPR) assays, CD1b was loaded with various lipid antigens, coupled to the surface, and recombinant BC8B TCR was the analyte (Figure 3). For CD1b loaded with PC, PE, PA, PI, PG or PS, steady state affinities were measured in the range of 5.9-15.3 μM , which represents a high affinity interaction for any self-ligand². Very weak binding was detected to CD1b carrying the single chain lipid LPA ($K_D > 200 \mu\text{M}$), consistent with the absence of staining with LPA-treated tetramers (Figure 2f, Supplementary Figure 2). The BC8B TCR detectably interacted with CD1b presenting the sphingolipid antigens, SM and sulfatide, but did so with extremely low affinity, 60.4 μM and 83.6 μM , respectively, as compared with CD1b-phospholipids. Although there were some differences in avidity or affinity among individual phospholipids, both tetramer and SPR-based assays showed a clear preference for self-phospholipids over sphingolipids and promiscuity among phospholipids, implying neck region specificity.

BC8B TCR docking onto CD1b-phospholipid

Next, we determined the crystal structure of the CD1b-PC-BC8B TCR complex to a resolution of 2.4 Å (Supplementary Table 1, Figure 4a). The entire PC molecule was clearly defined in the electron density map (Supplementary Figure 9), so that its interaction with the TCR could be analyzed in detail. Similar to prior CD1b complexes with self-lipids^{24, 34, 35} and the detection of CD1b ligands whose lipids are smaller than the CD1b cleft (Figs. 2a, Supplementary Figure 9), the large cleft of CD1b contained a scaffold lipid in addition to the acyl chains of PC (Supplementary Figure 9). The TCR docked centrally with an approximate 110° angle across the antigen-binding cleft, where the TCR contacted the lipid and CD1b (Figure 4a-b). The total buried surface area (BSA) of the interaction was 1990 Å², with equivalent contributions by the TCR α (53%) and β (47%) chain (Figure 4b).

The TCR α -chain dominated contacts with PC, where the CDR1 α and CDR3 α regions (20% and 30% BSA, respectively) interact with the phosphate, sn1 acyl tail, and both the $\alpha 1$ and $\alpha 2$ helices CD1b (Supplementary Table 2, Figure 4b-c). Specifically, the main chain of Y36 α and P37 α in the CDR1 α loop and P108 α in the CDR3 α loop form hydrogen bonds with the phosphate. S109 α of the CDR3 α loop hydrogen bonds with the hydroxyl group on the sn1 lipid tail (Figure 4c). The hydroxyl group of Y36 α of the CDR1 α forms a hydrogen bond with S109 α in the CDR3 α , aiding in stabilizing the TCR interaction with PC. Furthermore, Y36 α and K58 α of the CDR1 α and CDR2 α loops, respectively, make contacts proximally to the CD1b $\alpha 2$ hinge regions (Supplementary Table 2, Supplementary Figure 10a). Whereas CDR3 α loop contacts the PC neck region and sn1 tail regions (Figure 4c), Y113 α contacts R79 and E80 of the $\alpha 1$ helix of CD1b to further anchor CD1b-PC docking (Supplementary Table 2, Supplementary Figure 10b).

In contrast, TCR β lacks interactions with PC, and instead contacts the F' portal of CD1b (32% BSA) (Figure 4b, Supplementary Table 2, Supplementary Figure 10c-d). CDR3 β docks across the $\alpha 1$ helix, with the sulfur side chain and hydroxyl backbone of M108 β and P109 β , respectively, bonding with R79 (Supplementary Figure 10d). In fact, the CDR3 β loop penetrates the F' portal to gain access to the CD1b cleft, where L111 β contacts the buried L154 on the $\alpha 2$ helix and L111 β amine backbone and E80 on the $\alpha 1$ helix (Supplementary Figure 10d). TCR docking onto CD1b is further solidified by R112.1 β forming salt bridges and a hydrogen bond with D87 and D83 respectively (Supplementary Figs. 10d and 11). While the CDR1 β and CDR2 β loops play lesser roles, hydrogen bond formation between Y31 β and Y113 α of the CDR1 β and CDR3 α loops respectively (Supplementary Figure 10c), stabilize the PC antigen binding region architecture. Overall, the BC8B TCR binds near the center of CD1b with both chains providing defined contacts that explain CD1b and PC specificity

Conserved TCR docking modes on CD1b

Comparison of the BC8B TCR with the GEM42 and PG90 TCRs, the only ternary TCR structures currently solved with CD1b^{24, 36}, provides insight into key commonalities that underpin antigen recognition (Figure 4b, Supplementary Figure 11). All three TCRs exhibit a β chain center of mass above the $\alpha 1$ helix of CD1b, with contacts spreading across the $\alpha 1$ helix, and bridging towards the $\alpha 2$ helix (Figure 4b)^{24, 36}. The CDR3 β region contributes to significant contacts with CD1b in each case. TCR stabilization on CD1b is partially driven by salt bridge formation between key acidic residues on the $\alpha 1$ helix of CD1b, and a non-germline encoded arginine residue present on the N-region of each CDR3 β loop. Specifically, R112.1 β of the BC8B TCR binds D87 of CD1b (Supplementary Figs. 9d and 11), whereas salt bridges form between R110 β of the PG90 TCR and E80 and D83, respectively, and R109 β of GEM42 TCR binds E80 of CD1b (Supplementary Figure 11). These polar interactions contribute to anchoring the respective TCRs to allow for optimal lipid antigen co-recognition. These similarities are observed despite the lack of significant conservation of TCR gene usages^{24, 36}. Conversely, both the BC8B and GEM42 TCRs share TRBV6-2 genes, but their respective CD1-TCR structures reveal the CDR1 β and CDR2 β regions play minimal roles in CD1b and antigen contacts (Figure 4b, Supplementary Table 2). Instead, the CDR1 β region plays an indirect role in stabilizing the antigen binding architecture of the CDR3 α and CDR3 β regions of the BC8B and GEM42 TCRs, respectively (Supplementary Figure 10c). Considering the contacts between the TCRs and their respective lipid antigens, all three TCR footprints span across the F' portal, allowing contact with elements of the glycolipid (GEM42) or phospholipid (BC8B and PG90) that protrude through the F' portal onto the surface of CD1b (Figure 4b, Fig 5a-c).

A TCR escape channel limits antigen contact

Despite these commonalities, the key difference in antigen recognition is that the PG90 and GEM42 TCRs are highly specific for antigen head groups, whereas the BC8 TCR is not. For the GEM42 and PG90 TCRs, the TCRs make a tight seal on CD1b, forming a gasket-like, O-ring seal that fully surrounds the F' portal and protruding antigen (Figure 4b, Figure 5a-b). Surrounded on all sides, glucose or phospholipid head groups are forced upward towards the interface of the TCR α and β chains. Here the two chains act like two arms of tweezers that extensively surround and broadly contact head groups^{24,36} (Figure 5a-b). In contrast, the contacts between the BC8B TCR and lipid antigen are predominantly limited to the TCR α -chain, where the CDR1 α and CDR3 α loops contact both the phosphocholine neck region and the sn1 lipid tail (Figure 4c, Figure 5c). This alternative binding mechanism propagates a differing docking angle and TCR-CD1b contact interface creating a reversed C-shaped TCR contact footprint (Figure 4b, Figure 5d).

This forms a wide channel between the CDR1 α and CDR3 α loops that allows the choline head group of PC to escape laterally across the surface of CD1b, where it rests on the A' roof of CD1b (Figure 4b, Figure 6a). The 'bent' conformation of the antigen allows the anionic phosphate unit of the neck region to cradle within a cationic cup-like surface in the TCR (Figure 5d), with the choline group escaping TCR contact (Figure 4c). Overall, the BC8B TCR escape channel provides a straightforward explanation for both the neck region specificity for phosphoglycerol lipids over sphingolipids, as well as the lack of head group specificity for phospholipid subclasses: the head groups protrude laterally through a channel in the TCR.

Two mechanisms of cationic cup formation

The electronegative phosphocholine head group fits into an electropositive pocket of the BC8 TCR formed by the amine backbones of the TCR α chain (Figure 5d). This situation is reminiscent of the "cationic cup" that was recently described in the PG90 TCR, which mediates the phosphate unit of PG²⁴. In both cases, the CDR loops that form the cationic cups sequester the anionic phospholipid antigen head group, either partially or completely for the BC8B and PG90 TCRs, respectively. To characterize the extent of antigen and TCR remodeling during BC8B TCR docking, we determined the crystal structures of the CD1b-PC complex to 1.7 Å and the unliganded BC8B TCR to 2.0 Å (Supplementary Table 1). Next these three structures were compared with the CD1b-PG-PG90 TCR, CD1b-PG and PG90 TCR²⁴, demonstrating distinct mechanisms of cationic cup formation (Figure 5e-f). Comparison of CD1b-PG and CD1b-PG-PG90 TCR complexes shows that TCR binding resulted in extensive repositioning of the lipid, with the TCR β -chain moving 7 Å toward the PG head group and vice versa, pulling the antigen up slightly from its fully seated position in the cleft (Figure 5e)²⁴. Similarly, upon BC8B TCR docking, the PC

antigen is lifted from the CD1b antigen binding cleft in both the A' and C' pockets, allowing for significant TCR contacts with the sn1 acyl tail. The PC head group is lifted upwards by approximately ~6 Å and bent, allowing for favorable contacts as the phosphate moiety is inserted into the BC8B TCR cationic cup (Figure 5f).

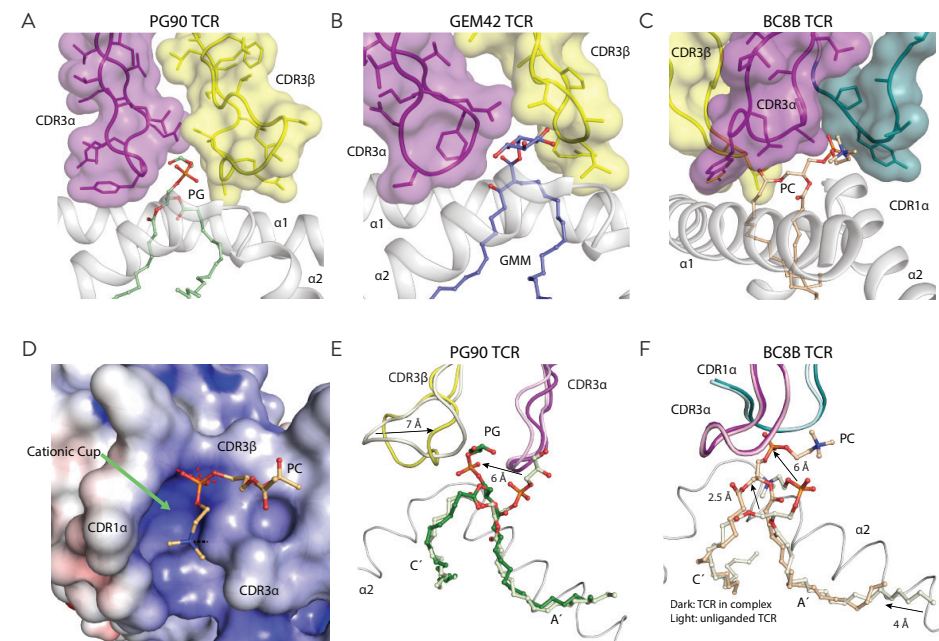


Figure 5. Cationic cup formation and rearrangement upon TCR docking.

(A-C) CDR loop contacts with antigen head group. Surface representation of the CDR1 α (teal), CDR3 α (purple), and CDR3 β (yellow) loops of (A) the PG90 TCR in contact with PG (green), (B) the GEM42 TCR in contact with GMM (blue), and (C) the BC8B TCR in contact with PC (light orange). (D) Electrostatic potential of the cationic cup (green arrow) formed by the BC8B TCR viewed bottom up from CD1b to the TCR. Hydrogen bonds and hydrophobic interactions between residues within the cationic cup and PC are indicated in red and black dashes respectively. The potential contours are shown on a scale from +10.0 (positive charge, blue) to -10.0 $k_B T e^{-1}$ (negative charge, red); white indicates a value close to 0 $k_B T e^{-1}$ (neutral charge). Cationic cup and direction of lateral escape are indicated by green and black arrows respectively. (E) Cationic cup and lipid antigen remodeling upon docking of (E) the PG90 TCR, and (F) the BC8B TCR. CDR loops of the unliganded TCR and TCR in complex with CD1b, as well as lipid antigens presented in CD1b binary and in complex with TCR structures, are represented in lighter and darker colors respectively as per Figure 4A. Directions of movement are indicated by black arrows.

Otherwise the formation mechanisms are highly distinct with regard to the cooperating TCR chains and the role of CD1b-phospholipid in driving formation. Whereas the BC8B TCR cationic cup is formed almost exclusively via the TCR α -chain (Figure 5d-f), the PG90 TCR cationic cup was formed from interchain interactions among the CDR3 α and CDR3 β loops (Figure 5a, 5e). For BC8B, the unliganded and CD1b-complexed BC8B TCR structures overlaid closely, with a Root Mean Square Deviation (R.M.S.D) of 0.65 Å. No significant movement of the CDR1 α and CDR3 α loops was observed between the BC8B TCR structures, revealing that the antigen interaction surface that contributes to the cationic cup is pre-formed (Figure 5f). In the case of the PG90 TCR, the CDR3 β loop undergoes a major conformational change upon CD1b-PG docking, demonstrating that the PG90 cationic cup is only formed after ligation (Figure 5e). Thus, both autoantigen-reactive CD1b-restricted TCRs solved to date form a cationic cup, which appears to be a phosphate-selective feature within these TCRs. However, this structure can be pre-formed within TCR- α or generated by inter-loop cooperation during binding of the negatively charged antigen.

Cross-reactivity explained by head group escape channel

Both PG90 and BC8B TCRs use cationic cups to position the phosphate at the base of the TCR. Whereas PG90 seals tightly over the phospholipid antigen head group²⁴, an escape channel forms between the CDR1 α and CDR3 α loops of the BC8B TCR. A lateral view shows that this escape channel is 10 Å at its maximum width between the side chains of T28 α and Q114 α of the CDR1 α and CDR3 α loops (Fig 6a-b). This gap is large enough to accommodate the choline head group that defines PC (Figure 6a-b). Although other glycerophospholipids were not co-crystallized with CD1b, their highly analogous structures (Figure 2b) are consistent with a general mechanism in which their shared phosphoglycerol groups lie in the cationic cup. The 5 glycerophospholipids cross-reactively recognized by this TCR (PE, PG, PS, PI, PA) (Figure 2f-g) would be expected to present ethanolamine, glycerol, serine, inositol, or no head group, respectively, towards the escape channel. Comparison of the molecular radii of these phospholipid head groups versus the size of the TCR channel predicts that all 5 could pass through the lateral gap in the TCR (Figure 6c). Conversely, PIM2, which has a core PI structure that carries two additional mannose units, is predicted to be sterically hindered, and this ligand blocks CD1b tetramer binding (Figure 2f-g). Last, this model plausibly explains why diversely loaded CD1b-endo tetramers stain the BC8B TCR. CD1b presents several abundant T-cellular phospholipids, including PG²⁴, PC, PS, PE and PI (Figure 2a). If each of these lipids act as antigens via trans-TCR escape as proposed here, their combined occupancy of a population CD1b monomers increase the probability that two or more TCR binding CD1b proteins are present within any tetramer, as multimeric interactions are generally needed for T-cell staining.

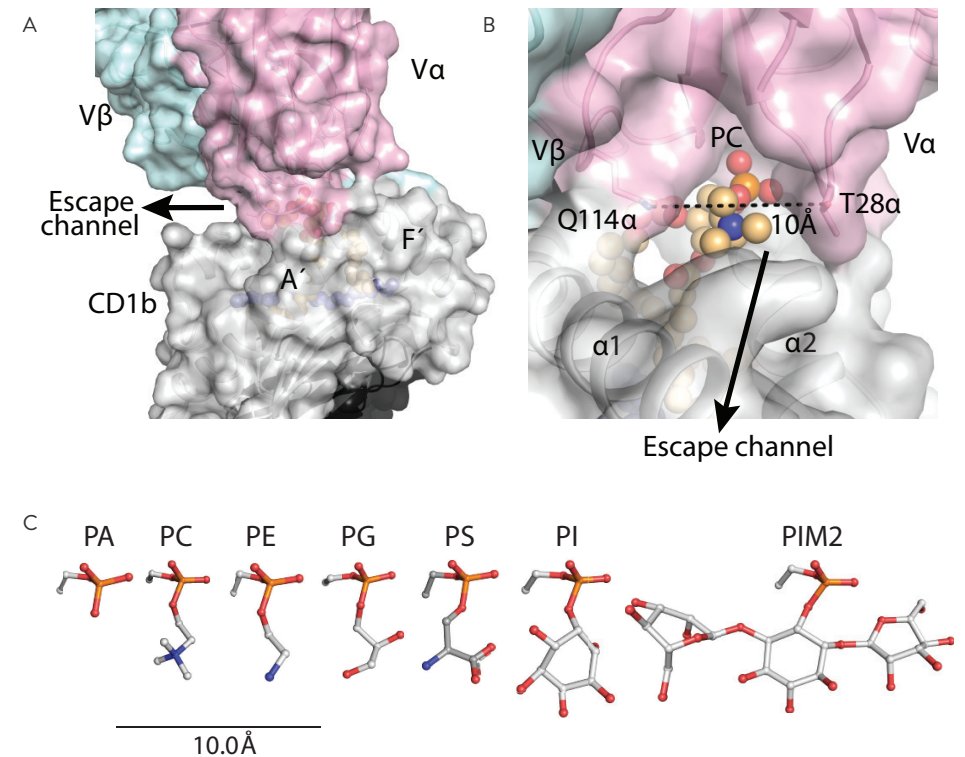


Figure 6: The lateral escape channel of the BC8B TCR.

(A) BC8B docking onto CD1b creates a lateral escape channel above the A' roof, to accommodate lipid antigen. The red arrow indicates the direction of the lateral channel opening. (B) The 10 Å escape channel is flanked by T28 α and Q114 α of the CDR1 α and CDR3 α loops respectively. Distance measurement is indicated by the black dash. Color coding for BC8B, CD1b, and the PC antigen are as in Figure 4A. (C) Chemical structures of the phospholipid head groups of phosphatidic acid (PA), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylserine (PS), phosphatidylinositol (PI), and phosphatidylinositol dimannoside (PIM2) are shown in comparison to an atomic ruler.

Discussion

In follow up to early evidence for phospholipid presentation by CD1d^{37, 38} these results provide new evidence for human T-cell response to CD1b and a broad range of common self-phospholipid autoantigens from unrelated human blood donors. From each of five donors tested we describe a set of CD1b-restricted T-cells present in the blood that recognize ubiquitous self-lipids and are functionally autoreactive to CD1b-expressing APCs. Considering the identities of the antigenic lipids seen by these T-cell lines, they respond to phospholipids that form the majority of the biomass of every mammalian cell membrane. Further, we found that individual lines showed broad, cross-reactive recognition of many

antigens^{24,34,35}, representing a new specificity pattern that is distinct from previous studies of CD1b-reactive T-cell clones that specifically recognize rare self-phospholipids associated with mitochondrial stress^{20,24}. Whereas human B cell and immunoglobulin responses to phospholipids have been studied for decades as manifestation of autoimmune disease such as systemic lupus erythematosus^{39,40}, these data provide broader evidence for human T-cell response to anionic phospholipids.

Further, the ternary TCR-PC-CD1b structure contributes a previously unknown mode of self-antigen recognition. The first autoreactive TCR footprint in the CD1a system bound to the A¹-roof of CD1a itself rather than lipid antigen protruding from the F¹ portal²⁹. The first autoreactive TCR footprint in the CD1c system docked centrally over the CD1c antigen portal, which can occur only when a small ligand is present that is sequestered within CD1c and does not interfere with the TCR approach⁴¹. These two related mechanisms emphasize TCR binding to CD1 itself, where the carried lipid essentially does not contribute to the TCR epitope. This shared 'absence of interference' mechanism favors small lipids and can account for the promiscuous recognition of structurally diverse lipids, so long as the lipids do not protrude from the cleft to substantially cover the CD1 surface⁴². Here we describe a new mechanism for CD1b, where lipid head groups take a bent conformation, positioning the anionic phosphate into a cationic cup in the TCR. The more distal head group exits laterally through a 10 Å gap, which can act as a sizing mechanism. The TCR does not significantly interact with the parts of the head group that enter the escape channel, which explains its cross-reactivity to phosphatidic acid-based antigens with absent, small or large head groups. Larger head group, like the trisaccharide in PIM cannot likely cannot traverse escape channel based on its measured width. Sphingolipids, including sphingomyelin, which otherwise mimics the head group of PC, lack an anionic neck region for the cationic cup. This bent-neck, head group escape mechanism of BC8B contrasts with the only known TCR and CD1b-autoantigen interaction, where the PG90 TCR is highly specific for a phosphoglycerol head group²⁴.

In MHC-peptide-TCR structures, key interactions occur between the TCR, MHC and upward facing residues of the peptide bound within the MHC. These data contribute to the view that CD1-antigen-TCR interactions operate via the similar mechanisms. Indeed, CD1-antigen-TCR structures typically show that the distal, upward protruding head group of the lipid dominates antigen specificity. One counterexample is CD1b-presentation of GM1 by CD1b²⁷. Here, additional sugar moieties could be added to a pentasaccharide core epitope without abrogating the T-cell response, leading to the speculation that the distal moieties might escape laterally below or within the TCR. Here we prove that the phospholipid neck region extends upward within the crevice between the TCR chains, makes specific interactions with a cationic cup and the head group exits laterally through an

escape channel in the TCR. Because many types of head groups fit inside the channel and are ignored, the broad response to membrane phospholipids is mechanistically explained.

Early functional evidence indicates that certain aspects of the detailed mechanism observed for BC8B may be more broadly conserved. The TCR cationic cup mechanism of recognition of anionic phospholipids is seen in both CD1b-autoantigen-TCR complexes solved to date, though the cup forms by two clearly distinct mechanisms. The cationic cup observed here is pre-formed with the TCR α chain, whereas the PG90 cationic cup is formed by three variable loops in TCR α and β , and becomes a recognizable structure only after contact with CD1b-phospholipid²⁴. Also, we isolated additional T-cell lines that show high specificity for small head groups (pattern 1) and lines with general phospholipid versus sphingolipid specificity (pattern 3) as examples of human T-cells recognizing common membrane phospholipids. CD1a, CD1b and CD1d also bind and present membrane phospholipids, including phosphatidylcholines that are similar or identical to lipids studied here^{28,41,43}. For CD1d, there is some evidence for promiscuous responses of clones to phospholipids³⁷, so plausibly use mechanisms similar to those described here.

Increasing evidence for promiscuous T-cell responses to cellular CD1 proteins and self-lipid antigens is beginning to raise questions about negative regulation of such responses. For CD1a and CD1c autoreactive T-cells can be directly detected in unfractionated blood^{22,28}, and our data suggest that CD1b autoreactive cells can be readily detected with a simple selection procedure. The T-cell lines that we describe here, recognizing broadly distributed, abundant phospholipids were isolated from healthy blood bank donors. If such CD1b-specific T-cells are generally present in human donors, the question arises as to whether they would cause autoreactivity or autoimmunity upon contact with any CD1b-expressing cell. Known mechanisms for limiting autoreactivity may be in play. T-cells may be suppressed by other cells or are inherently unable to perform inflammatory functions *in vivo*. They may not see the proper stimulus under steady state conditions because the antigen, the antigen presenting molecule, or costimulatory factors are in limited supply. Even though phospholipid antigens are widely available, CD1b expression in healthy tissues in the periphery is rare⁴⁴, but can be induced by Toll-like receptors, interleukin-1, cytokines and other primary inflammatory signals^{8,9,45,46}. Thus, in contrast to the ubiquitous expression of MHC class I proteins, activation and potentially harmful consequences of the T-cells described in this paper may be limited by tightly regulated CD1b expression in healthy tissues, and not by antigen availability.

On the other hand, such cells have the potential to cause autoimmunity, as was recently demonstrated in mice that expressed human CD1b and phospholipid specific TCR. These mice developed an autoimmune skin disease, when ApoE deletion causes marked

overexpression of phospholipids in the skin³¹. This study raises the possibility that CD1b and phospholipids might play an unrecognized role in autoimmune disease. Also, clear evidence for immune response to phospholipids via CD1b and CD1d invites consideration of possible B cell-T-cell interactions in the anti-phospholipid antibody syndrome^{47,48}, an autoimmune state associated with systemic lupus erythematosus and other autoimmune diseases.

Methods

Tetramers and analytical flow cytometry

CD1 monomers were obtained from the NIH tetramer facility. For loading of monomers, 32 µg of lipid was sonicated at 37°C for 2 hours in 90 µl of 0.5% CHAPS 50 mM sodium citrate buffer pH7.4 in 10 mm diameter glass tubes. Subsequently, 20 µg CD1b monomer was added to the tubes and incubated overnight at 37°C. Molecular Probes streptavidin-APC or streptavidin-PE was used for tetramerization. Human PBMC and T-cell lines were stained with tetramers at 2 µg ml⁻¹ in PBS containing 1% BSA and 0.01% sodium azide. Cells and tetramer were incubated for 10 minutes at room temperature, followed by addition of 3 µl of anti CD3 (clone SK7-Fitc from BD biosciences) per 50 µl staining and another incubation for 10 minutes at room temperature and 20 minutes at 4°C. Cells were analyzed using the BD LSRFortessa flow cytometer and FlowJo software.

Lipids

Phosphatidylcholine (PC, #850475; C18:1/C16:0), sphingomyelin (SM, #860584; C18:1/C16:0), phosphatidylinositol (PI, #840042; mixture from bovine liver with a range of fatty acids), phosphatidylserine (PS, #840032; mixture from porcine brain with a range of fatty acids), phosphatidic acid (PA, #840857; C18:1/C16:0), lyso-PA (#857130; C18:1), phosphatidylethanolamine (PE, #850757 C18:1/C16:0), phosphatidylglycerol (PG, #840503; C18:1/C18:0), and sulfatides (#131305P; mixture from porcine brain with a range of fatty acids) were purchased from Avanti polar lipids. Gangliosides GM1 (G7641; mixture from bovine brain with a range of fatty acids) and GM2 (G8397; mixture from bovine brain with a range of fatty acids) were purchased from Sigma. Phosphatidylinositol dimannoside (PIM) and diacyltrehalose (DAT), previously described as DAT2a⁴⁹, were provided by the Bill and Melinda Gated Foundation lipid bank. Cholesteryl 6-O-palmitoyl-β-galactopyranoside, also known as *Borrelia burgdorferi* glycolipid 1 (BBGL1) and 1,2-dioleoyl-α-galactosyl-glycerol, also known as *Borrelia burgdorferi* glycolipid 2 (BBGL2) were synthesized as described previously^{50,51}. *Salmonella* Typhi was cultured in Luria Broth. *Brucella melitensis* human isolate, strain X10017283-001 (CVI) was cultured in Tryptic Soy Broth. Lipid extracts were prepared by extracting bacterial pellet with chloroform/methanol 1:2 (V:V) for two hours at room temperature, followed by chloroform/methanol 2:1.

CD1b-lipid complex analysis

Human CD1b proteins made in U293 cells at the NIH tetramer facility as described^{5,24} were transferred into a 15-ml glass tube and extracted with chloroform, methanol and water according to the Bligh and Dyer method⁵². The lipids were recovered from the organic layer and redissolved in chloroform/methanol (1:2) to a concentration equivalent to 5 µM of the input protein. The extracted lipids (5-10 µl) were loaded onto a nanospray tip for negative mode ESI-MS and multistage collision-induced dissociation tandem mass (CID-MS) using linear ion trap mass spectrometer (LXQ, Thermo Scientific). Collision energy was 20–35% of maximum and product ions were trapped with a *q* value of 0.25. The negative ions were calibrated with external reference Hexakis (1H, 1H, 3H-tetrafluoropropoxy) phosphazine ([M+Cl]⁻ at *m/z* 955.97 (Agilent part # 18720241).

Generation of T-cell lines

PBMC were obtained from leukoreduction collars provided by the Brigham and Women's Hospital Specimen Bank, as approved by the Partners Healthcare IRB. Human PBMC and PBMC derived T-cells were stained with tetramers at 2 µg ml⁻¹ in PBS containing 1% BSA and 0.01% sodium azide. Cells and tetramer were incubated for 10 minutes at room temperature, followed by addition of 3 µl of anti-CD3 (clone SK7-Fitc from BD biosciences) per 50 µl staining and another incubation for 10 minutes at room temperature and 20 minutes at 4 °C. PBMC were sorted for positive staining with anti-CD3 and the indicated tetramers (Supplementary Figure 1-5). Expansion of sorted cells was performed using anti-CD3 at 30 ng ml⁻¹ (clone OKT3), irradiated feeder cells, and IL-2. After two weeks, the sorting and expansion procedure was repeated once or multiple times. When the resulting cell lines consisted mainly of tetramer⁺ T-cells, they were screened for binding to a panel of antibodies against TCR Vβ and 2 µl of anti-CD4 (clone RPA-T4 from Biolegend) per 50 µl staining. Suitable combinations of antibodies were used to isolate subsets of the original T-cell lines (Supplementary Figure 1-5). IFN-γ ELISPOT was performed using the 1D1K and GB-11-biotin antibodies (Mabtech), according to the manufacturer's instructions.

TCR sequencing and transduction

The TCR sequences were determined using RNA isolated with an RNeasy kit (Qiagen), and cDNA synthesized with a Quantitect reverse transcription kit (Qiagen). V segment usage was determined by PCR using primerset IPS000030 as described at www.imgt.org (Lefranc MP, 1989), as well as by a multiplex approach⁵³, followed by direct Sanger sequencing of the PCR product. Full length TCRα and TCRβ chains were cloned into a self-cleaving 2A peptide based (MSCV)-IRES-GFP (pMIG) vector⁵⁴ and co-transfected into HEK293T-cells in the presence of the retroviral packaging vectors pPAM-E and pVSV-g. The supernatant from the transfected HEK293T-cells was harvested and used to stably transduce TCR-deficient Jurkat clone 76 cells⁵⁵ to generate the BC13A T-cell

line. After 5 days, cells that had the highest expression of GFP and CD3 were enriched by FACS sorting.

Protein expression and purification

Recombinant BC8B TCR were cloned into the pET30a vector and expressed, refolded and purified from *E. coli* inclusion bodies^{24,56}. Inclusion bodies were resuspended in 8 M urea, 20 mM Tris-HCl (pH 8.0), 0.5 mM Na-EDTA, and 1 mM DTT. The TCR was refolded by flash dilution in a solution containing 5 M urea, 100 mM Tris (pH 8.0), 2 mM Na-EDTA, 400 mM L-arginine-HCl, 0.5 mM oxidized glutathione, 5 mM reduced glutathione, and EDTA-free anti-protease cocktail. The refolding solution was then dialyzed to eliminate urea. The resulting protein solution was then purified by size exclusion chromatography and HiTrap-Q anion exchange chromatography. Soluble CD1b was expressed from either insect *Trichoplusia ni* High Five cell lines, or HEK 293S GnT1⁻ (American Type Culture Collection), and purified by HisTrap Ni²⁺-affinity chromatography and size exclusion chromatography^{24,36}. For SPR and crystallography purposes, purified CD1b was loaded with a molar excess of target lipid in the presence of tyloxapol (Sigma) over a 16 h incubation period at 20°C, then purified via anion exchange chromatography to homogeneity^{24,36}.

Crystallization and structure determination

Crystals of the BC8B TCR were grown via the hanging drop vapour diffusion method, using a protein-reservoir drop ratio of 1:1, at a protein concentration of 5 mg mL⁻¹ in 10 mM Tris-HCl (pH 8.0), with a crystallisation condition of 20% (v/v) PEG 3350, 0.2M Potassium sodium tartrate tetrahydrate. Crystals of the CD1b-PC-BC8B TCR ternary complex were grown at protein concentrations of 5 mg/mL, with a crystallisation condition of 22% to 24% (v/v) PEG 3350, 0.01 M Tris-HCl pH 7.0 – 7.4. Crystals of the CD1b-PC complex were grown at protein concentrations between 5–8 mg mL⁻¹, with a crystallisation condition of 22 – 26 % PEG 3350, 0.2 M Sodium Iodide. Crystals of BC8B TCR, CD1b-PC-BC8B TCR, and Cd1b-PC were soaked in a cryoprotectant comprised of reservoir solution containing 10% (v/v) ethylene glycol, before being flash-frozen in liquid nitrogen. Data were collected at the Australian Synchrotron at the MX2 beamline for the BC8B TCR and CD1b-PC-BC8B TCR crystals, and at the MX1 beamline for the CD1b-PC crystals⁵⁷. Data was processed using the iMosflm software, and scaled using Aimless as part of the CCP4i program suite⁵⁸. Crystal structures were solved via Molecular replacement using Phaser as part of the phenix program suite⁵⁹, with the structures of CD1b-PG (PDB accession code: 5WL1), and GEM42 TCR (PDB accession code: 4G8F) used as models for solving CD1b and TCR structures respectively. Manual adjustments of the models were conducted in the coot graphics program⁶⁰, following maximum-likelihood refinement with Buster 2.10⁶¹. Electron density for ligands were well defined in the CD1b-PC-BC8B TCR structure. Electron density for regions of the PC molecule in the CD1b-PC structure were less

defined, indicative of flexibility. All molecular representations were generated in PyMOL. BSA values were calculated using areaimol, and contacts generated using the CONTACT program in the CCP4i program suite⁵⁸.

Surface plasmon resonance

SPR analysis on the BC8B TCR against CD1b-phospholipids was conducted on the BIACore 3000 instrument at 25°C in 10 mM tris-HCl (pH 8.0), 150 mM NaCl, and 1% (w/v) bovine serum albumin²⁴. Binding of the recombinant, refolded TCR in solution to CD1b, loaded with exogenous lipids and amine coupled to a CM5 chip, was analyzed. All experiments were conducted as two or three independent experiments in duplicate. Data analysis and visualization were generated using Graphpad Prism 7.0, using the 1:1 Langmuir binding model.

Data availability

Structural data were deposited in the Protein Data Bank, with the following accession codes: BC8B TCR (6CUH) [www.ncbi.nlm.nih.gov/Structure/pdb/6CUH], CD1b-PC-BC8B (6CUG) [www.ncbi.nlm.nih.gov/Structure/pdb/6CUG], and CD1b-PC (6D64) www.ncbi.nlm.nih.gov/Structure/pdb/6D64]. All remaining data are available within the article and its supplementary information files and from the corresponding authors on request. A reporting summary for this article is available as a Supplementary Information file.

Acknowledgements

This work was supported by the Australian Research Council and National Health and Medical Research Council, the National Institutes of Health (AR048632 and AI049313), the Bill and Melinda Gates Foundation, and Netherlands Organization for Scientific Research (NWO). SG is a Monash Senior Research Fellow. J.R. is supported by an ARC Laureate Fellowship.

Author contributions

A.S., P.R., J.F.R., and S.L. undertook the research and analyzed the data. A.S., P.R., S.G., T-Y C., D.B.M., I.V.R., and J.R. designed the research and analyzed the data. M.H., A.J.M., J.A., J.P., J.K.-K. provided unique reagents. D.B.M., J.R., I.V.R. wrote the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Godfrey, D.I., Uldrich, A.P., McCluskey, J., Rossjohn, J. & Moody, D.B. The burgeoning family of unconventional T cells. *Nat. Immunol.* **16**, 1114-1123 (2015).
2. Rossjohn, J. et al. T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169-200 (2015).
3. Salio, M., Silk, J.D., Jones, E.Y. & Cerundolo, V. Biology of CD1- and MR1-restricted T cells. *Annu. Rev. Immunol.* **32**, 323-366 (2014).
4. Dendrou, C.A., Petersen, J., Rossjohn, J. & Fugger, L. HLA variation and disease. *Nat. Rev. Immunol.* **18**, 325-339 (2018).
5. Kasmar, A.G. et al. CD1b tetramers bind {alpha}{beta} T cell receptors to identify a mycobacterial glycolipid-reactive T cell repertoire in humans. *J. Exp. Med.* **208**, 1741-1747 (2011).
6. Jackman, R.M. et al. The tyrosine-containing cytoplasmic tail of CD1b is essential for its efficient presentation of bacterial lipid antigens. *Immunity* **8**, 341-351 (1998).
7. Sugita, M. et al. Cytoplasmic tail-dependent localization of CD1b antigen-presenting molecules to MHCs. *Science* **19;273**, 349-352 (1996).
8. Chancellor, A. et al. CD1b-restricted GEM T cell responses are modulated by Mycobacterium tuberculosis mycolic acid meromycolate chains. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10956-E10964 (2017).
9. Krutzik, S.R. et al. TLR activation triggers the rapid differentiation of monocytes into macrophages and dendritic cells. *Nat. Med.* **11**, 653-660 (2005).
10. Porcelli, S., Morita, C.T. & Brenner, M.B. CD1b restricts the response of human CD4-8- T lymphocytes to a microbial antigen. *Nature* **360**, 593-597 (1992).
11. Beckman, E.M. et al. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature* **372**, 691-694 (1994).
12. Sieling, P.A. et al. CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science* **269**, 227-230 (1995).
13. Moody, D.B. et al. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* **278**, 283-286 (1997).
14. de la Salle, H. et al. Assistance of microbial glycolipid antigen processing by CD1e. *Science* **310**, 1321-1324 (2005).
15. Layre, E. et al. Mycolic acids constitute a scaffold for mycobacterial lipid antigens stimulating CD1-restricted T cells. *Chem. Biol.* **16**, 82-92 (2009).
16. Gilleron, M. et al. Diacylated Sulfoglycolipids Are Novel Mycobacterial Antigens Stimulating CD1-restricted T Cells during Infection with Mycobacterium tuberculosis. *J. Exp. Med.* **199**, 649-659 (2004).
17. James, C.A. et al. CD1b Tetramers Identify T Cells that Recognize Natural and Synthetic Diacylated Sulfoglycolipids from Mycobacterium tuberculosis. *Cell chemical biology* **25**, 392-402 e314 (2018).
18. Yakimchuk, K. et al. Borrelia burgdorferi infection regulates CD1 expression in human cells and tissues via IL1-beta. *Eur. J. Immunol.* **41**, 694-705 (2011).
19. Kinjo, Y. et al. Natural killer T cells recognize diacylglycerol antigens from pathogenic bacteria. *Nat. Immunol.* **7**, 978-986 (2006).
20. Van Rhijn, I. et al. Human autoreactive T cells recognize CD1b and phospholipids. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 380-385 (2016).
21. de Jong, A. et al. CD1a-autoreactive T cells are a normal component of the human alphabeta T cell repertoire. *Nat. Immunol.* **11**, 1102-1109 (2010).
22. de Lalla, C. et al. High-frequency and adaptive-like dynamics of human CD1 self-reactive T cells. *Eur. J. Immunol.* **41**, 602-610 (2011).
23. Gherardin, N.A. et al. Human blood MAIT cell subsets defined using MR1 tetramers. *Immunol. Cell Biol.* **96**, 507-525 (2018).
24. Shahine, A. et al. A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Science immunology* **2** (2017).
25. Cosma, C.L., Sherman, D.R. & Ramakrishnan, L. The secret lives of the pathogenic mycobacteria. *Annu. Rev. Microbiol.* **57**, 641-676 (2003).
26. Lepore, M. et al. A novel self-lipid antigen targets human T cells against CD1c(+) leukemias. *J. Exp. Med.* **211**, 1363-1377 (2014).
27. Shamshiev, A. et al. The alphabeta T cell response to self-glycolipids shows a novel mechanism of CD1b loading and a requirement for complex oligosaccharides. *Immunity* **13**, 255-264 (2000).
28. de Jong, A. et al. CD1a-autoreactive T cells recognize natural skin oils that function as headless antigens. *Nat. Immunol.* **15**, 177-185 (2014).
29. Birkinshaw, R.W. et al. alphabeta T cell antigen receptor recognition of CD1a presenting self lipid ligands. *Nat. Immunol.* **16**, 258-266 (2015).
30. Wun, K.S. et al. T cell autoreactivity directed toward CD1c itself rather than toward carried self lipids. *Nat. Immunol.* **19**, 397-406 (2018).
31. Bagchi, S. et al. CD1b-autoreactive T cells contribute to hyperlipidemia-induced skin inflammation in mice. *J. Clin. Invest.* **127**, 2339-2352 (2017).
32. Bourgeois, E.A. et al. Bee venom processes human skin lipids for presentation by CD1a. *J. Exp. Med.* **212**, 149-163 (2015).
33. Shamshiev, A. et al. Self glycolipids as T-cell autoantigens. *Eur. J. Immunol.* **29**, 1667-1675 (1999).
34. Gadola, S.D. et al. Structure of human CD1b with bound ligands at 2.3 Å, a maze for alkyl chains. *Nat. Immunol.* **3**, 721-726 (2002).
35. Garcia-Alles, L.F. et al. Endogenous phosphatidylcholine and a long spacer ligand stabilize the lipid-binding groove of CD1b. *EMBO J.* **25**, 3684-3692 (2006).
36. Gras, S. et al. T cell receptor recognition of CD1b presenting a mycobacterial glycolipid. *Nat Commun* **7**, 13257 (2016).
37. Gumperz, J.E. et al. Murine CD1d-restricted T cell recognition of cellular lipids. *Immunity* **12**, 211-221 (2000).
38. Mallevaey, T. et al. A molecular basis for NKT cell recognition of CD1d-self-antigen. *Immunity* **34**, 315-326 (2011).
39. Giannakopoulos, B. & Krilis, S.A. The pathogenesis of the antiphospholipid syndrome. *N Engl J Med* **368**, 1033-1044 (2013).
40. Roubey, R.A. & Hoffman, M. From antiphospholipid syndrome to antibody-mediated thrombosis. *Lancet* **350**, 1491-1493 (1997).
41. Borg, N.A. et al. CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* **448**, 44-49 (2007).
42. Cotton, R.N., Shahine, A., Rossjohn, J. & Moody, D.B. Lipids hide or step aside for CD1-autoreactive T cell receptors. *Curr Opin Immunol* **52**, 93-99 (2018).

43. Haig, N.A. et al. Identification of self-lipids presented by CD1c and CD1d proteins. *The Journal of biological chemistry* **286**, 37692-37701 (2011).
44. Dougan, S.K., Kaser, A. & Blumberg, R.S. CD1 expression on antigen-presenting cells. *Curr. Top. Microbiol. Immunol.* **314**, 113-141 (2007).
45. Sieling, P.A. et al. CD1 expression by dendritic cells in human leprosy lesions: correlation with effective host immunity. *J. Immunol.* **162**, 1851-1858 (1999).
46. Van Rhijn, I., Van den Berg, L.H., Bosboom, W.M., Otten, H.G. & Logtenberg, T. Expression of accessory molecules for T-cell activation in peripheral nerve of patients with CIDP and vasculitic neuropathy. *Brain* **123 (Pt 10)**, 2020-2029 (2000).
47. Leadbetter, E.A. et al. NK T cells provide lipid antigen-specific cognate help for B cells. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8339-8344 (2008).
48. Barral, P. et al. B cell receptor-mediated uptake of CD1d-restricted antigen augments antibody responses by recruiting invariant NKT cell help in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8345-8350 (2008).
49. Besra, G.S. et al. Structural elucidation of a novel family of acyltrehaloses from *Mycobacterium tuberculosis*. *Biochemistry (Mosc.)* **31**, 9832-9837 (1992).
50. Pozsgay, V. & Kubler-Kielb, J. Synthesis of an experimental glycolipoprotein vaccine against Lyme disease. *Carbohydr. Res.* **342**, 621-626 (2007).
51. Pozsgay, V. et al. Synthesis and antigenicity of BBGL-2 glycolipids of *Borrelia burgdorferi*, the causative agent of Lyme disease. *Carbohydr. Res.* **346**, 1551-1563 (2011).
52. Bligh, E.G. & Dyer, W.J. A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology* **37**, 911-917 (1959).
53. Wang, G.C., Dash, P., McCullers, J.A., Doherty, P.C. & Thomas, P.G. T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Science translational medicine* **4**, 128ra142 (2012).
54. Szymczak, A.L. et al. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22**, 589-594 (2004).
55. Heemskerck, M.H. et al. Redirection of antileukemic reactivity of peripheral T lymphocytes using gene transfer of minor histocompatibility antigen HA-2-specific T-cell receptor complexes expressing a conserved alpha joining region. *Blood* **102**, 3530-3540 (2003).
56. Gras, S. et al. The shaping of T cell receptor recognition by self-tolerance. *Immunity* **30**, 193-203 (2009).
57. Cowieson, N.P., Aragao, D., Clift, M., Ericsson, D.J. MX1: A bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation* **22**, 187-190 (2015).
58. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235-242 (2011).
59. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221 (2010).
60. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).
61. Bricogne G., B.E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, S.A., Smart O.S., Vornrhein C., Womack T.O. BUSTER version X.Y.Z. Cambridge, United Kingdom: Global Phasing Ltd.; 2017.

Chapter 6

Discovery of immunogenic trehalose phospholipids in *Salmonella* species reveals functional convergence with mycobacteria

Peter Reinink^{1,2,10}, Jeffrey Buter^{2,10}, Vivek K. Mishra³, Eri Ishikawa⁴, Tan-Yun Cheng², Peter T. J. Willemsen⁵, Steffen Porwollik⁶, Patrick J. Brennan², Eva Heinz⁷, Gordon Dougan⁷, Cécile A. van Els⁸, Vincenzo Cerundolo⁹, Giorgio Napolitani⁹, Sho Yamasaki⁴, Adriaan J. Minnaard³, Michael McClelland⁶, D. Branch Moody², Ildiko Van Rhijn^{2,1*}

¹ Department of Infectious Diseases and Immunology, School of Veterinary Medicine, Utrecht University, 3584CL Utrecht, the Netherlands

² Department of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, 02115, Massachusetts, USA

³ Stratingh Institute for Chemistry, University of Groningen, 9747AG Groningen, The Netherlands

⁴ Department of Molecular Immunology, Immunology Frontier Research Center, Osaka University, Osaka

⁵ Wageningen Bioveterinary Research, Department of Infection Biology, Lelystad, The Netherlands

⁶ Department of Microbiology and Molecular Genetics, University of California, Irvine, USA

⁷ Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

⁸ Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands.

⁹ MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

¹⁰ These authors contributed equally

*Corresponding author

Abstract

Salmonella species are among the world's most prevalent pathogens. The bacterial cell wall interfaces with the host, so we designed a lipidomics approach to reveal pathogen-specific cell wall compounds. Among the molecules differentially expressed between *Salmonella* Paratyphi and *S. Typhi*, we focused on lipids that are enriched in *S. Typhi* because it causes typhoid fever. We discovered a previously unknown family of trehalose phospholipids, 6,6'-diphosphatidyltrehalose (diPT) and 6-phosphatidyltrehalose (PT). Cardiolipin synthase B (ClsB) is essential for PT and diPT but not for cardiolipin biosynthesis. Chemotyping outperformed *clsB* homology analysis in evaluating synthesis of diPT. DiPT is restricted to a subset of Gram-negative bacteria: large amounts are produced by *S. Typhi*, lower amounts by other pathogens, and variable amounts by *E. coli* strains, but it is undetectable in intestinal microbiota. DiPT potently activates Mincle, a macrophage activating receptor that also recognizes mycobacterial cord factor (6,6'-trehalose dimycolate). These previously unknown innate stimuli in pathogenic Gram-negative bacteria show convergent function with mycobacteria. Overall, we discovered a previously unknown immunostimulant that is selectively expressed among medically important bacterial species.

Introduction

Salmonella Typhi is the etiologic agent of typhoid fever. This enteric fever has a high mortality rate if untreated and is responsible for an estimated 200,000 deaths annually¹. More generally, diarrheal diseases from Gram-negative bacteria are among the most prevalent and most deadly bacterial infectious diseases in the world, comparable to *M. tuberculosis* which causes tuberculosis². *Salmonella enterica enterica Typhi* (*S. Typhi*) lives in a membrane bound vacuole, which does not fuse with lysosomes and permits the intracellular replication of the bacteria. Gram-negative bacterial cell walls are comprised of an inner membrane dominated by phospholipids, a thin layer of peptidoglycan polymer, and an outer membrane with more complex lipids, which directly interface with the host.

Like other Gram-negative bacteria, *Salmonella* species synthesize lipopolysaccharide (LPS), which resides in the outer leaflet of the outer membrane and strongly stimulates the innate immune system by triggering Toll-like receptor 4 (TLR4)³. LPS is among the most studied molecules in infectious diseases⁴, where it unquestionably controls fever and sepsis as key manifestations of *S. Typhi* and other Gram-negative bacterial syndromes. Yet, anti-LPS therapies have had limited success in treating sepsis, and differing LPS chemotypes do not fully explain the markedly different immunogenicity, fever inducing capacity and pathogenicity of diverse Gram-negative bacterial species and strains. Further, given the focus on LPS as a strong stimulant for the mammalian immune system, the many other foreign molecules in Gram-negative cell walls have received less attention. To our knowledge comparative lipidomic analyses of important pathogenic and non-pathogenic Gram-negative species have not been reported, raising the possibility that undiscovered virulence-associated lipids exist.

To test this hypothesis, we recently developed a rapid method that takes advantage of normal phase chromatography to separate and analyze dozens of classes and thousands of molecular species of bacterial lipids by mass spectrometry (MS)⁵. Combined with manual methods of thin layer chromatography (TLC), collision induced dissociation-mass spectrometry (CID-MS), and nuclear magnetic resonance (NMR) spectroscopy, this system has been proven to offer an approach to discovery of previously unknown or pathogen-specific lipids, such as phosphomycoketides, dideoxymycobactins, and tuberculosinyladenosines⁶⁻⁸. Using *S. Typhi* as an example of a major Gram-negative pathogen, we compared its lipids to those of less pathogenic serovars including *S. Paratyphi*, *S. Enteritidis* and *S. Typhimurium*, generating clear evidence for strain-specific differences in lipid synthesis.

Focusing on the most abundant lipids that are selectively expressed by key pathogenic species, we discovered the products and defined the key genes of a new glycolipid

biosynthesis pathway in a subset of pathogenic Gram-negative bacteria. We found that a gene annotated as a cardiolipin synthase B functions as the essential enzyme for an abundant family of previously unknown immunogenic trehalose-containing phospholipids. Whereas most phylogenetic analyses and clinical strain typing rely on genetic methods, these studies illustrate the unique information available through systematic analysis of bacterial lipids. Similarities among mycobacterial and salmonella trehalose-containing lipids suggest functional convergence in their activation of human immune response, highlighting a new pathway to adjuvant development.

Results

Lipidomic analysis of pathogenic *Salmonella* serovars

We used high-performance liquid chromatography-mass spectrometry (HPLC-MS)-based comparative lipidomics⁵ to study a major pathogenic *Salmonella enterica* serovar, *S. Typhi*, and compare its lipid profile with the less virulent, but closely related serovar, *S. Paratyphi* A. Each molecular species of lipid isolated from the bacteria or its adduct is detected as a three-component data point known as a molecular event. A molecular event consists of a retention time on the HPLC column, a mass-to-charge ratio (m/z), and an intensity value. The number of molecular events estimates the total number of lipids present, including chain length variants, altered adducts, and isotopes of each molecule. The two *Salmonella* serovars combined generated 4569 molecular events (Figure 1a). As expected, molecular diversity is lower than the 6,000 to 10,000 events in the highly complex cell wall of *Mycobacterium tuberculosis*⁵. However, this number still represents substantial lipid diversity in a Gram-negative pathogen. Comparison of *S. Typhi* and *S. Paratyphi* A resulted in 865 lipids that differed in intensity by 2-fold with a corrected p value of < 0.05 , documenting substantial divergence of the two lipidomes. This number exceeded our experimental throughput for compound identification, so we designed strategies to prioritize the unknowns for identification.

Two previously unidentified, abundant lipids

To focus on lipids of high biological interest, including previously unknown compounds or candidate virulence factors, we prioritized molecular events that: 1) were enriched in *S. Typhi*, 2) had high absolute intensity, and 3) had m/z values that did not match known compounds in LIPID MAPS or other databases^{9,10}. Strain-specific enrichment of two abundant lipids was evident even by relatively insensitive normal phase TLC method (Figure 1b). Two spots with retardation factors (R_f) of 0.26 and 0.22 showed much denser spots in *S. Typhi* as compared to *S. Paratyphi* A. Positive mode nanoelectrospray ionization (nano-ESI) MS analysis of TLC scrapings of these spots yielded spectra that were dominated

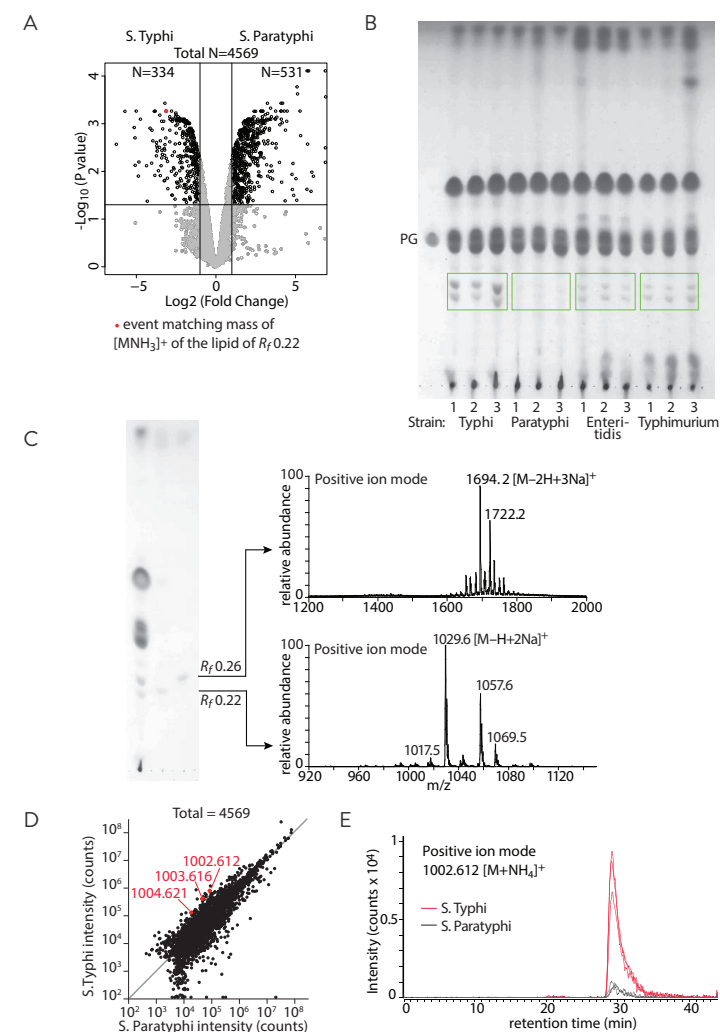


Figure 1: Detection of two unknown lipids in *Salmonella*

a. Three independent cultures of *S. Typhi* (Quailes) and *S. Paratyphi* A (NVGH308) underwent lipid extractions and comparative lipidomics. Among all 4569 molecular events (black and grey circles), 865 (black circles) show two-fold intensity change ($p < 0.05$, corrected for multiple comparisons). **b.** Total lipid extracts (300 $\mu\text{g}/\text{lane}$) of 12 independently derived clinical isolates belonging to four different strains of *Salmonella enterica enterica* were analyzed by TLC. Phosphatidylglycerol (PG, 40 μg) was used as a standard. Green boxes are drawn around two unknown lipids with R_f 0.22 and 0.26. **c.** After isolation of the two unknown lipids from *S. Typhi* from a water-sprayed TLC plate (not shown), part of the isolated material was re-applied on a TLC plate and charred for analysis, and part was analyzed by positive mode nanoESI-MS. **d.** Mean intensities (integrated counts) of 4569 ions detected in 3 replicates of *S. Typhi* (Quailes) and *S. Paratyphi* A (NVGH308) in a lipidomics experiment are shown in a scatterplot with red dots corresponding to ions matching the ammonium adduct of M corresponding to the lipid with R_f 0.22 (m/z 1002.612) and its isotopes. **e.** Extracted ion chromatograms of the R_f 0.22 lipid for three independent cultures of *S. Typhi* and three *S. Paratyphi* A measured at m/z 1002.6.

by ions of m/z 1694.2 and 1029.6, respectively. These were seen along with ions corresponding to chain length and saturation variants that differed by 14 (CH_2) or 12 (C) u, including m/z 1722.2, m/z 1057.6 and m/z 1069.5 (Figure 1c). Initial discovery efforts emphasized HPLC-TOF-MS over TLC because the former is $\approx 10^6$ -fold more sensitive. That these two lipids were visible by the less sensitive technique suggests high abundance in *S. Typhi* (Figure 1b). Other lipids identified by the HPLC-lipidomics system are produced at < 10 part per million of total cellular lipid^{11,12}. The limit of detection of charring on TLC is $\sim 1 \mu\text{g}$, so any clearly visible spot in a profile from 300 μg of total bacterial lipid suggest production in the low parts per hundred range. If the combined density of the two unknowns is one-tenth or more of that of the phosphatidylglycerol (PG) standard (40 μg), these two unknown lipids would comprise more than 1 % of total extractable lipids of *S. Typhi* (Figure 1b).

As a first step to identification, we sought to match low mass accuracy values from TLC-nano-ESI MS (Figure 1c) to high mass resolution data from the HPLC-TOF-MS lipidomic data expressed as a volcano plot (Figure 1a) or scatterplot (Figure 1d). This approach identified ions corresponding to ammoniated adducts ($[\text{M}+\text{NH}_4]^+$) of M in the spectrum from the lipid with R_f 0.22 on TLC (Figure 1c). Using the higher mass accuracy data (m/z 1002.612), we deduced M as likely being $\text{C}_{48}\text{H}_{89}\text{O}_{18}\text{P}$. This suggested that the compound was a phospholipid, consistent with its observed retention time in HPLC-MS (Figure 1e, 29.5 min) matching known phosphoglycolipids⁵. Importantly, high mass accuracy mass value and molecular formula allowed searching of lipid databases, which returned no matches, suggesting that the target was previously unknown. Given that *S. Typhi* is a widely studied pathogen of world-wide significance, finding a previously unidentified compound among the most abundant lipids in this organism was unexpected. These data provided a strong rationale for focused analysis of the compounds of R_f 0.22 and 0.26 to determine their complete structures.

Serovar-specific patterns of expression

Differential analysis of four pathogenic *Salmonella enterica* serovars (*S. Typhi*, *S. Paratyphi*, *S. Enterica*, *S. Typhimurium*) using three separately obtained clinical isolates within each serovar, suggested uniform abundance of the two unknown lipids among independent clinical isolates (Figure 1b). Looking across the serovars, the two lipids showed the same ranking with *S. Typhi* $>$ *S. Typhimurium* \approx *S. Enteritidis* $>$ *S. Paratyphi*. These findings were consistent with the possibility of co-regulation of two structurally related lipids in the same pathway. TLC (Figure 1b) resulted in barely discernable bands in *S. Paratyphi*. However, using the established strains *S. Typhi* Quailles and *S. Paratyphi* A strain NVGH308, ion chromatograms reproducibly demonstrated a co-eluting lipid with a mass of 1002.612 in *S. Paratyphi* A, confirming the presence of trace amounts of the unknown in this serovar (Figure 1e). Based on ion chromatogram intensity calculated as peak area, *S. Paratyphi*

produces ≈ 9 -fold less product than *S. Typhi*. Overall, these data demonstrated serovar-specific synthesis of many lipids in *Salmonella* and identified two previously unknown phospholipids enriched in *S. Typhi*.

Discovery of two trehalose phospholipids

Nearly all MS signals from TLC-purified *S. Typhi* material corresponded to m/z 1694.2 or m/z 1029.6 and the identifiable isotopes and chain length variants thereof, suggesting each spot contained only one major product (Figure 1b). Initial characterization by ion trap CID-MS tentatively identified the lipids as two structurally related dihexose phospholipids that contained either one or two phosphatidyl groups (Figure 2a, Supplementary Figure 1). The lower band consisted of phosphatidyl dihexose and the upper band of diphosphatidyl dihexose. One of the fatty acids in the phosphatidyl group was a palmitic acid (C16). The other one was a C17:1, suggesting the presence of either an unsaturation or a cyclopropyl group.

Because CID-MS cannot unequivocally identify the particular hexoses or differentiate an unsaturation from a cyclopropyl group in the C17 fatty acid, we undertook 1- and 2-dimensional NMR spectroscopy analysis (Figure 2b, Supplementary Figure 2). This approach showed that the lower migrating unknown compound possesses two anomeric protons (doublets at δ 5.10 and 5.11 ppm) both with a coupling constant (J) of 3.7 Hz, suggesting an α,α -linked dihexose structure¹³. The signals between δ 3.3-4.1 ppm were complex, necessitating correlation spectroscopy (COSY) analysis, which allowed assignment of the protons at C1-C5, showing two C5 signals at δ 3.82 and δ 3.95 (Supplementary Figure 2b). 2D-total correlation spectroscopy (TOCSY) NMR¹⁴ showed a one spin-system for the dihexose unit with five cross-peaks and the anomeric center, revealing the C6 signal at δ 4.08 (Supplementary Figure 2c). Distortionless Enhancement by Polarization Transfer Heteronuclear single quantum coherence spectroscopy (135DEPT-HSQC) (Supplementary Figure 2d) showed clear C6 signals at δ 4.08 + δ 4.02 (with $^{13}\text{C} = \delta$ 65.6), and C6' signals at δ 3.79 + δ 3.67 (with $^{13}\text{C} = \delta$ 62.3). Also, an additional C4 signal could be assigned at δ 3.36-71.7. The significant difference of the C6-signals in both the ^1H and ^{13}C -NMR resulted from a phosphate moiety at C6, as shown in the ^{31}P -decoupled ^1H -NMR (Supplementary Figure 2e) and the CID-MS spectrum (Figure 2a). Coupling constants for the dihexose CH units were established to be in the 8-10 Hz interval, consistent with J -values for *trans* stereochemistry. The only hexose with all-*trans* stereochemistry between the H-atoms is glucose.

Finally, the α,α -trehalose-6-phosphate moiety was unambiguously verified by comparing the ^1H - and ^{13}C -NMR signals with that of α,α -trehalose-6-phosphate¹⁵. The signals at δ 5.24 (m, 1H), δ 4.00 (t, $J = 5.6$, 2H) and the diastereotopic protons at δ 4.45 (dd, $J = 3.0$, 12.0, 1H) and δ 4.20 (dd, $J = 6.9$, 12.0, 1H) were part of one spin-system in the 2D-TOCSY,

These considerations provided a strong rationale to test for immune responses to these newly-discovered trehalose-based compounds. However, even though MS and NMR spectroscopy suggested high purity of TLC-purified trehalose phospholipids, *S. Typhi* synthesizes LPS, which is bioactive at picomolar concentrations. As LPS or other minor contaminants could create false positive results in cellular assays, we undertook the complete chemical synthesis of diPT. Based on prior reports describing naturally occurring Gram-negative bacterial fatty acids^{18,19} we first synthesized 9*R*,10*S* cyclopropyl fatty acid. This compound, together with palmitic acid, was used to prepare the required diacylglycerol, which along with hexabenzyl trehalose were used to assemble diPT via phosphoramidite coupling and subsequent deprotection. The 9-step synthesis was carried out with an overall yield of 69%, resulting in 15 mg of the final product. The final product was validated by HPLC-MS and NMR spectroscopy showing good correspondence of the chemical shifts between the two samples, thereby verifying the proposed chemical structure of diPT (Supplementary Figure 5 versus Supplementary Figure 2f-i).

Access to milligram quantities of a synthetic diPT standard allowed more quantitative assessments of PT and diPT as a component of the *S. Typhi* lipidome (Supplementary Figure 2k). Using synthetic diPT as an internal standard for HPLC-MS and applying the method of standard additions to the tree brightest natural diPT ions, we estimate that diPT is 2.5% of *S. Typhi* total lipid extract. This is a conservative estimate because more than 6 ions are seen, and it generally matches the estimate for diPT of 1.7 percent derived from TLC (Figure 1b). Both results indicate the diPT and PT are among the most abundant lipids in *S. Typhi*.

S. Typhi trehalose phospholipids are potent Mincle ligands

The overall chemical resemblance between diPT and TDM is that both are symmetrical molecules with a trehalose core that is substituted with lipids attached at the two 6'-positions (Figure 3b). The key difference is that diPT is substituted with phospholipids whereas TDM is substituted with mycolic acids. To determine whether diPT is recognized by Mincle, we used a reporter cell line stably transduced with murine Mincle, its signaling partner, FcR γ and an NF κ B-driven green fluorescent protein (GFP) reporter construct. Because hydrophobic glycolipids like TDM are integrated in membranes and do not act in solution, bioassays to mimic this interaction rely on first coating wells with lipid and then adding cells in aqueous media. Despite this technical limitation, dose responses are measurable (Figure 3c). The dose responses to TDM and natural diPT were similar and highly sensitive in absolute terms, with cells responding to 5 ng of coated lipid. Pure synthetic diPT was less potent but still sensitive with responses seen at 20 ng of coated lipid. Although the potent stimulation of reporter cells by synthetic diPT rules out effects of LPS or other trace bacterial contaminants as mediating Mincle response, the reduced GFP signal at high doses

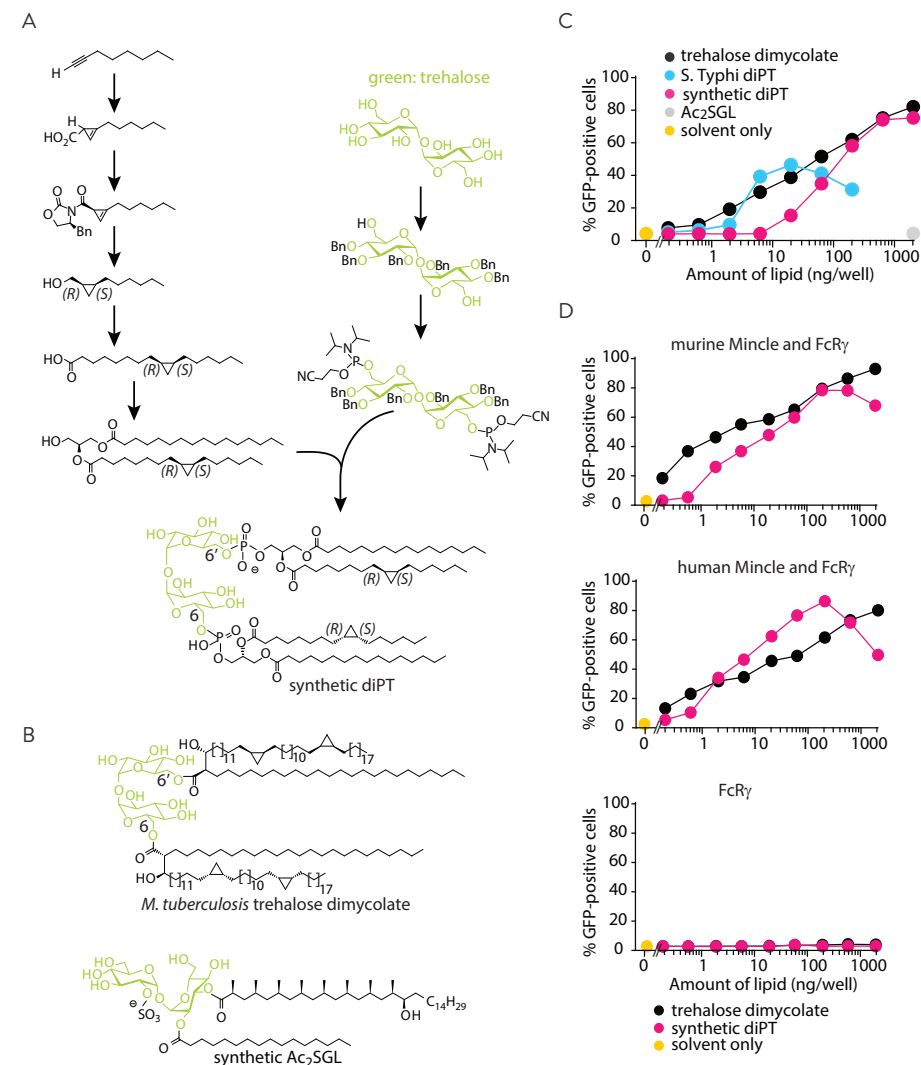


Figure 3: Synthetic diphosphatidyltrehalose (diPT) is a high potency Mincle agonist.

a. Schematic representation of the chemical synthesis of diPT containing C17:1 and C16 fatty acyl units such as found in natural diPT. **b.** For comparison, chemical structures of the two natural trehalose based lipids, the known Mincle ligand, *M. tuberculosis* trehalose dimycolate (TDM) and diacylated sulfolipid (Ac₂SGL). **c.** GFP expression was measured using a murine Mincle-FcR γ reporter cell line and the indicated plate bound lipids (ng/well). Data are representative of five experiments using different combinations and dose of antigens with similar results. **d.** GFP expression of reporter cell lines expressing murine Mincle-FcR γ , human Mincle-FcR γ , or FcR γ alone, after stimulation with trehalose dimycolate or synthetic diPT.

of natural diPT could be LPS-mediated. Synthetic sulfoglycolipid²⁷, a glycolipid substituted at the 2 and 3 positions of trehalose, did not stimulate the Mincle reporter cell line at high doses. Subsequently, we compared murine and human Mincle reporter cell lines and again found highly sensitive responses to synthetic diPT, with activation present at nanogram levels. Human Mincle is marginally more sensitive to diPT than to TDM, while for murine Mincle the opposite is true (Figure 3d). DiPT failed to activate a negative control cell line expressing FcRγ only. Thus, diPT efficiently stimulates Mincle.

Candidate gene approach to biosynthesis

To determine the mechanism of biosynthesis of trehalose phospholipids, we used a candidate gene approach. DiPT was previously unknown, and we could not identify substantially similar lipids in other bacteria that could point us towards candidate synthetic enzymes involved in phospholipid transfer onto trehalose. However, genes for biosynthesis of the two components of diPT, trehalose and phosphatidylglycerol, are well known, as are those enzymes that produce ‘substituted phosphatidylglycerols’ like phosphatidylinositol and cardiolipin. Considering PT and diPT as carbohydrate-substituted phosphatidylglycerol and mining the *S. Typhi* genome, we identified 12 candidate genes for biosynthesis. Among these 12 candidates are four biosynthetic enzymes that produce trehalose from either glucose 6-phosphate and UDP-glucose (OtsA, OtsB) or from α(1-4)-linked glucose polymers (TreY, TreZ), and a fifth enzyme that shares a domain with PapA3, which is responsible for lipid transfer onto mycobacterial trehalose phosphate (EntF)²⁸. In addition, we tested four enzymes involved in synthesis of phospholipids (PagP, PgpA, PgpB, PgpC) and three putative cardiolipin synthases (ClsA, ClsB, ClsC), which are thought to transfer phosphatidyl units onto phosphatidylglycerol.

Cardiolipin synthase B is essential for trehalose phospholipids, but not cardiolipin

We tested 12 candidate diPT and PT biosynthetic genes using two sets of single gene knockouts in *S. Typhimurium*, which were generated independently as kanamycin or chloramphenicol selected mutants²⁹. Because diPT did not resolve well on normal phase HPLC-MS, we developed a suitable reverse phase HPLC method that could reliably detect both PT and diPT. An equivalent mass of total lipid, as determined by weighing dry lipid, from each single gene knockout (Figure 4a) was analyzed using phosphatidylethanolamine (PE) as a secondary loading control (Supplementary Figure 5). This screen provided a clear result that was reproducible in both sets of mutants and for both lipids, with essentially all or nothing effects for each gene studied. Of 12 candidates, 11 genes (*aas*, *clsA*, *entF*, *otsA*, *otsB*, *pagP*, *pgpA*, *pgpB*, *pgpC*, *treY*, *treZ*) were non-essential for biosynthesis (Figure 4a). There was no source of trehalose in the culture media, so the likely reason for production of trehalose phospholipids after single gene deletion of trehalose synthases likely relates to redundancy

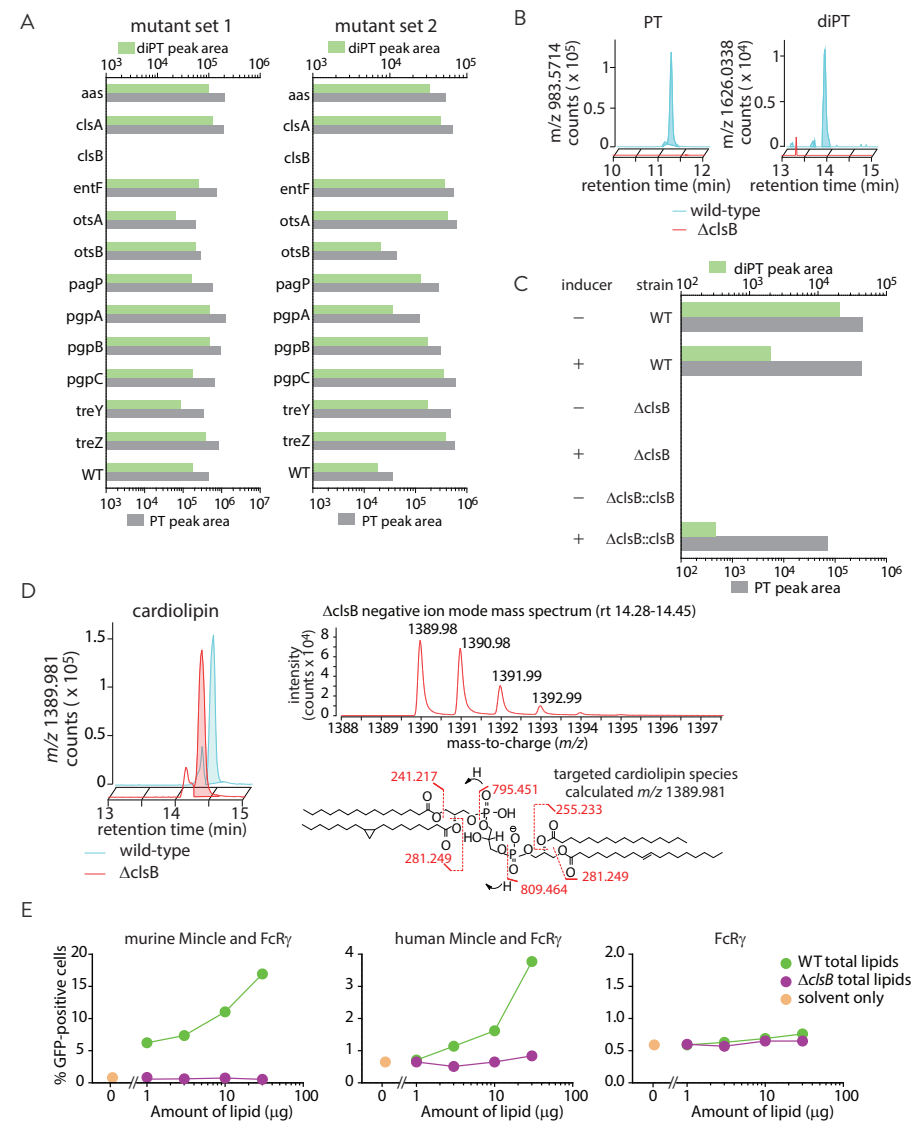


Figure 4: ClsB is necessary for the biosynthesis of PT and diPT.

a. Abundance of PT and diPT was determined in total lipid extracts from two independently generated sets of *S. Typhimurium* single gene knockouts generated by selection under kanamycin (set 1) or chloramphenicol (set 2). **b.** Extracted ion chromatograms of PT (m/z 983.6) and diPT (m/z 1626) in lipid extracts from wild-type *S. Typhimurium* and in *S. Typhimurium* $\Delta clsB$. **c.** PT and diPT production in wild-type *S. Typhimurium* (WT), *S. Typhimurium* $\Delta clsB$ mutant ($\Delta clsB$) and genetically complemented *S. Typhimurium* $\Delta clsB$ ($\Delta clsB::clsB$) in the present or absence of the inducer, arabinose. **d.** Extracted ion chromatogram of cardiolipin in lipid extracts from wild-type *S. Typhimurium* and *S. Typhimurium* $\Delta clsB$. The identity of cardiolipin was confirmed by high resolution MS, where the observed m/z 1389.98 matches the expected mass of 1389.981, and CID-MS, which identified fragments corresponding to phosphatidyl units and fatty acids. **e.** Total lipids from wild type or $\Delta clsB$ *S. Typhimurium* were used to stimulate reporter cell lines expressing murine Mincle-FcR γ , human Mincle-FcR γ , or FcR γ alone.

of biosynthetic pathways. OtsA OtsB act via a glucose 6-phosphate intermediate. TreY and TreZ convert maltodextrin to trehalose, and other enzymes, including TreT and TreS, convert maltose into trehalose. Only the *clsB* single gene knockout showed loss of PT and diPT in both mutant sets, while PE levels were unchanged (Figure 4a-b and Supplementary Figure 5). That this loss of PT and diPT was observed in two independently generated single gene knockout sets reduces the likelihood that lipid loss was due to an unrelated second hit occurring elsewhere in the genome of *clsB* mutants. Even so, we genetically complemented a *clsB* knockout using an arabinose-inducible system. Only when treated with arabinose, the reconstituted *clsB* knockout ($\Delta clsB::clsB$) produced PT and diPT, demonstrating that *clsB* is necessary and sufficient for its production (Figure 4c). In the wild type strain, arabinose treatment partially diminished diPT synthesis for unknown reasons, which might partly explain why the arabinose-treated, genetically complemented strains did not fully restore lipid production to wild type levels. Taken together, these data point to ClsB as an essential enzyme for the synthesis of trehalose phospholipids.

The acronym *clsB* (previously known as *ybhO* or *f413*) stands for cardiolipin (CL) synthase B³⁰, an assignment based on sequence homology rather than direct demonstration of this enzymatic role. Using HPLC-MS to detect signals (*m/z* 1390) at the retention time of a CL standard, we observed no significant change in the intensity or shape of the biphasic peak assigned to CL in the *clsB* knockouts (Figure 4d and Supplementary Figure 5). We conclude that *clsB* is non-essential, and under the conditions tested, does not affect CL concentration. Although an adjunctive role in CL biosynthesis might be masked by parallel functions of ClsA and ClsC, the simplest interpretation is that ClsB is, despite its name, not a CL synthase. Instead of coupling phosphatidic acid to phosphatidylglycerol, ClsB could couple phosphatidic acid to trehalose. Consistent with this interpretation, previous work has also suggested the likely existence of an unknown substrate for ClsB^{31,32}, and that ClsA or ClsC is sufficient for cardiolipin synthesis.

The *clsB* knockout strain represented a new tool to determine if diPT and PT, considered among all cell wall lipids, were essential to stimulate Mincle. First, we found that total lipid extract from wild type *S. Typhimurium* stimulated Mincle (Figure 4e). Although lower Mincle reporter response was seen compared to using pure diPT, this outcome was expected based on the much lower concentration of diPT among all lipids, the plate-bound nature of the lipid presentation in this assay, and the possible existence of antagonists or toxic factors among total cell wall lipids. However, the total lipid extract from $\Delta clsB$ bacteria did not stimulate human or murine Mincle reporter lines (Figure 4e), suggesting that among all lipids in *S. Typhimurium*, the diPT pathway is the dominant or sole source of Mincle ligands.

Genetic and chemical phylogeny of *clsB* and diPT

After the initial identification of trehalose phospholipids in *S. Typhi*, *S. Typhimurium*, *S. Enteritidis*, and *S. Paratyphi*, we next asked how broadly distributed these compounds are among bacteria. Also, given that PT and diPT were discovered in enteric bacteria, the specific question arises as to whether humans are continuously exposed to trehalose phospholipids via intestinal microbiota or through infection with enteric pathogens. First, we used a basic local alignment search tool (BLAST)-reverse BLAST-based approach to ask whether a *S. Typhi* ClsB ortholog is present in genomes of closely and distantly related bacterial genera (Figure 5a-b). The four studied *Salmonella enterica enterica* strains had *clsB* genes predicted to encode proteins that were 100% identical, and *S. bongori* encoded a 99% identical protein. The genomes of species belonging to β - and γ -proteobacterial *Escherichia*, *Shigella*, *Pseudomonas*, and *Bordetella* encoded ClsB proteins that ranged from 51-87% identity to *S. Typhi* ClsB protein. We found no identifiable ClsB orthologs among β - and γ -proteobacteria, nor among more distant groups such as Gram-positive bacteria and actinobacteria. In all bacterial species found to have a ClsB ortholog, a ClsA ortholog, which has an established role in cardiolipin synthesis, was also present. Furthermore, phylogenetic analysis showed that ClsA, ClsB and ClsC form distinct monophyletic branches clearly separating the three enzymes (Figure 5b).

We next determined the extent to which such phylogenetic comparisons using gene homology analysis correctly predict the diPT and PT chemotypes. Therefore, we used HPLC-MS to chemotype 16 broadly divergent species highlighted in the bioinformatic analysis (Figure 5c-d). Despite high (99%) similarity with the *S. Typhi* *clsB* gene, we did not detect diPT or PT in *S. bongori*, or in other species in which we found a *clsB* ortholog, including *Pseudomonas aeruginosa*, *Bordetella pertussis*, and *Shigella* species. Whether this is due to functional differences of the enzyme itself or instead to differences in its species-specific expression or interaction with other enzymes in the pathway will require further detailed studies beyond the scope of this analysis. A rather unexpected finding, pointing towards unknown contribution of expression level or other interacting proteins, was the distribution of diPT and PT among the six *E. coli* strains tested. Despite identical culture conditions and identical coding regions of their *clsB* genes, we found that four of the six strains synthesized the compounds (clinical isolates CVI-7, CVI-19; laboratory strains DH10B, BW25113), while two did not (ATCC25922, DH5 α). We conclude that having a *clsB* gene is required for the synthesis of PT and diPT, but as yet unknown additional requirements, like gene transcription or production of precursor molecules or the lack of suppressors, must be met. In addition, genes annotated as *clsB*, based on the full-length sequence may not have equivalent functionality because of small changes in key residues at the catalytic site or other critical sites in the protein.

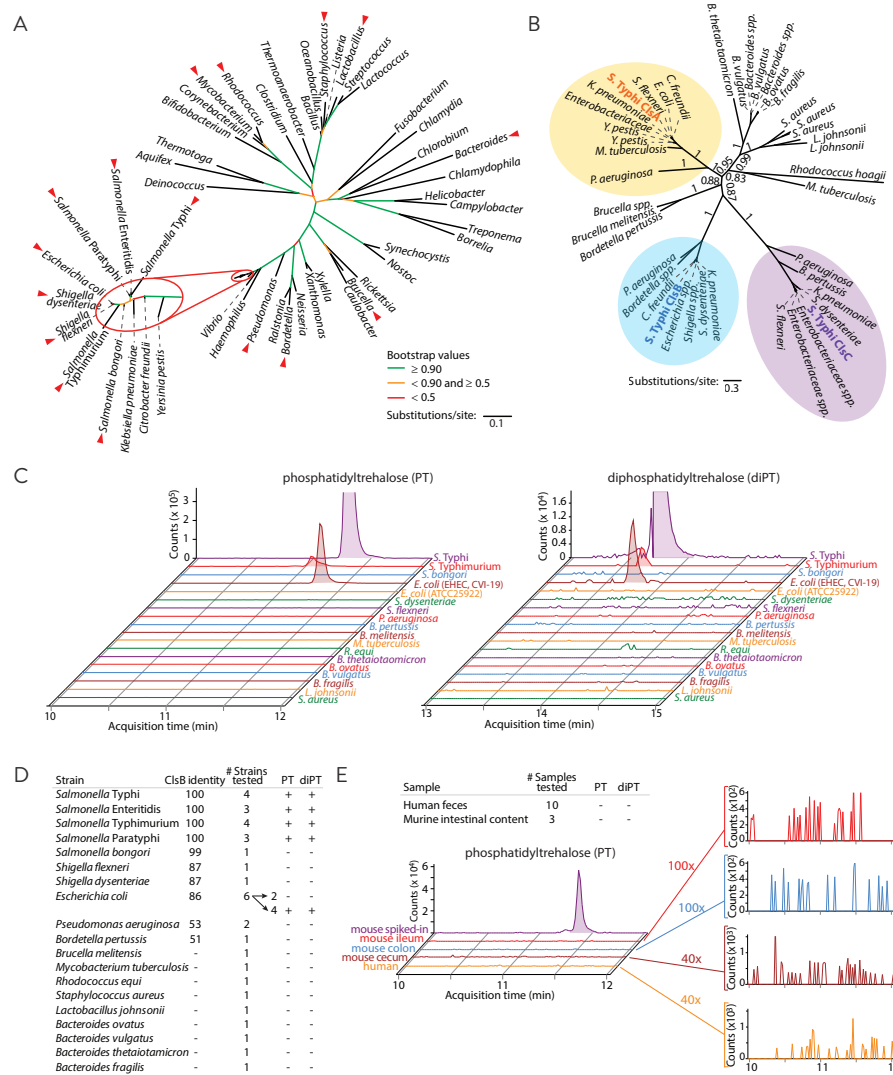


Figure 5: PT and diPT in other bacterial strains.

a. Phylogenetic tree based on 16S RNA sequences of a broad range of bacterial species, focusing on close relatives of *Salmonella* (inset). Species that underwent genomic analysis and lipid chemotyping are indicated (red triangles). **b.** Phylogenetic tree of Cls protein sequences. **c.** HPLC-TOF-MS yielded extracted ion chromatograms matching the mass of PT and diPT for the indicated bacterial strains. **d.** The amino acid identity of the translated *clsB* genes to *S. Typhi* ClsB and results of HPLC-MS-based chemotyping for PT and diPT are summarized (+: detected; -: not detected). **e.** Extracted ion chromatograms corresponding to the mass of PT and diPT in lipid extracts of human fecal samples, samples of murine intestinal content, and a positive control sample consisting of murine intestinal content pre-spiked with *S. Typhi* total lipid extract before processing. Results are representative of 10 human and 3 mouse fecal samples, and positive signals would not be detected after increasing gain by 40-100 fold (insets).

Our results pointed to diPT as a strain-specific phenotype present in certain gastrointestinal pathogens, but the unexpected detection of diPT in *E. coli* strains raised questions about its possible role in the intestinal microbiome. Although feces represents a complex mixture of bacteria, *E. coli* is a common and abundant species in human feces, and in vitro evidence for trehalose phospholipid production by this and other enteric bacteria raise the question of whether this compound can be detected in stool. Therefore, we looked for trehalose phospholipids in mouse and human stool samples by HPLC-MS. Neither PT nor diPT was detected in any of 10 human fecal samples, nor in murine ileum, colon, or cecum contents. This result held true for standard HPLC-MS readouts and also when signals were amplified 40-100-fold to assess diPT signal at the limit of detection by MS (Figure 5e). Because stool contains a multitude of bacteria and fungi as well as partly digested food and host-derived material, we considered possible false negative results by cross-suppression from unrelated compounds. However, separation by HPLC prior to ionization greatly reduces cross-suppression. A spike in positive control of an *S. Typhi* lipid extract containing an estimated amount of 20 µg PT and diPT generated strong trehalose phospholipid signals (Figure 5e). Thus, trehalose phospholipids appear to be undetectable or absent from the normal intestinal microbiota.

Overall, these data indicate that production of diPT is a strain- or serovar-specific phenotype and varies among isolates of the same species. ClsB ortholog searches are not sufficient to predict the presence of diPT, highlighting the usefulness of lipid chemotyping assays. Further, these results reveal a general pattern that diPT is restricted to a subset of Gram-negative bacteria: it is found in pathogens but not gastrointestinal flora of healthy donors. Considering its role as an agonist for MinCLE, a widely expressed innate immune pattern recognition receptor in humans, diPT becomes a candidate for adjuvant development, control of virulence, and a new marker of clinical interest to physicians.

Discussion

These data identify ClsB as essential for production of 6-phosphatidyltrehalose (PT) and 6,6'-diphosphatidyltrehalose (diPT) in the major disease-causing serovars of *Salmonella enterica enterica*. Although trehalose is common in plants and bacteria, trehalose phospholipids were previously unknown. Thus, the discovery and detection of diPT and PT was entirely unexpected and prompts the reconsideration of the function of the enzyme known as ClsB. ClsA condenses two PG molecules to form CL, whereas ClsC condenses PE and PG^{31,33}. Despite its name and homology to ClsA and ClsC, the function of ClsB has been less clear. *E. coli* ClsB mediates PG turnover to CL outside of cells, but in other experiments it does not affect cardiolipin content in *E. coli*^{30,31} or *Shigella flexneri*³³. Thus it

was previously suggested ClsB might actually use different, unknown, substrates or produce lipids other than cardiolipin³¹, a hypothesis that is directly ruled in by our data. Other members of the phospholipase D superfamily cleave phosphates to transfer phosphatidic acid to an acceptor, and here we show that *S. Typhi* *clsB* is absolutely required for PT and diPT biosynthesis, suggesting that trehalose could be an acceptor for phosphatidic acid transfer. While *in vitro* studies of recombinant ClsB are required to test this hypothesis, our findings cast further doubt on a gene function corresponding to its name as a cardiolipin synthase.

Here we identified murine and human Mincle as cellular receptors for diPT. Members of the C-type lectin family of receptors, which also includes Dectin-1, Dectin-2 and DC-SIGN, can activate monocytes, macrophages and myeloid dendritic cells via ITAM or ITAM-like domains that act as kinases to turn on phospho-Syk signaling pathways³⁴. The crystal structure of human Mincle identifies a calcium binding site positioned near a proposed carbohydrate binding site, which in turn is located adjacent to a hydrophobic patch^{35,36}. Mincle activation occurs in response to 1-linked and 6-linked glucose containing lipids³⁷⁻³⁹. Comparative analysis with other C-type lectins, mutational analysis, and molecular modeling suggest that the glucose or trehalose sugars common to natural Mincle ligands bind in the carbohydrate binding site^{35,36,40}. The 6,6'-linked C20 fatty acids in trehalose dibehenate and C80 mycolic acids in TDM are proposed to bind in the hydrophobic patch. Thus, diPT contains the two key chemical elements previously proposed to mediate Mincle binding. However, because there is no precedent for Mincle activation by phospholipids, prior models have not accounted for putative interactions with an anionic phosphoglycerol unit. There is no identifiable cationic binding site between the carbohydrate binding site and the hydrophobic patch on Mincle. However, docking simulations suggest that the most proximal part of 6-linked alkyl chains, corresponding to the two phosphoglycerol units in diPT, do not contact Mincle⁴⁰. Thus, existing models predict diPT docking in a manner in which the phosphate moieties act as a bridge between the sugar- and lipid-binding epitopes, rather than directly contacting Mincle.

These structural considerations and the precedent of TDM provide a rationale for adjuvant development using synthetic modifications of the natural diPT structure. Freund's adjuvant is widely used in experimental biology and provided the basis for use of pure TDM as an adjuvant. Subsequently, the chemically simplified and more hydrophilic adjuvant TDB was developed as a further refinement of the TDM natural structure^{23,24}. Unlike these molecules, diPT contains two phospholipid units, which further promote water solubility, a favorable feature in adjuvant development. This chemical consideration, along with the high potency of synthetic diPT for human Mincle agonism, support further modification of the diPT chemical scaffold for adjuvant development. In particular, the chemical syntheses

reported here for diPT can be modified to include simplified lipid moieties to further increase water solubility without affecting the Mincle binding regions of the molecule.

Results from reporter lines rule in both human and mouse Mincle receptors for PT and diPT. Conversely, the loss of Mincle agonism in the *clsB* knockout strongly suggests other unrelated compounds in *S. typhimurium* do not redundantly agonize Mincle. These data establish key receptor-ligand interactions, raising new questions about the broader cell biology of macrophage response. For example, whether PT and diPT activate other innate immune receptors is unknown, and the specific subtypes of myeloid cells or macrophages activated by this system is not yet understood. Although they are structurally related and nearly equipotent, the extent to which Gram-negative bacterial and mycobacterial agonists have equivalent function is not yet known. Study of these new natural and synthetic Mincle agonists can be used to test certain questions and controversies that have arisen through study of mycobacterial cord factor, 6-, 6'-trehalose dimycolate²⁵. In addition to activating Mincle, cord factor has been reported to activate Toll-like receptor 4 and related MyD88-based signaling pathways^{41,42}. The phosphate moieties of PT and DiPT render these compounds as 'phospholipid' variants of cord factor. This and the mono-versus divalent nature of PT and diPT might influence the final outcomes of cellular activation.

TDM synthesis in *M. tuberculosis* takes place by transfer of mycolic acids to the 6 and 6' positions of trehalose by the enzymes proteins Ag85A, Ag85B, and Ag85C^{43,44}. ClsB is not related to Ag85 proteins but is instead a member of the phospholipase D family of proteins. Therefore, the TDM and diPT-generating enzymes and the genes that encode these enzymes are unrelated, but their products are similar with regard to structure and function. Thus, the separate evolution of differing enzymes that lead to molecules with highly similar structure that trigger the same receptor might be a previously unrecognized example of evolved functional convergence of Gram-negative bacteria and mycobacteria.

Whereas prior work on Mincle has emphasized mycobacterial and fungal ligands, the data presented here generate a strong link with Gram-negative enteric bacteria. We demonstrate that five serovars or strains of common Gram-negative bacteria produce Mincle ligands. Further, although we do not yet know if diPT has other functions, other lipids in *S. Typhimurium* cannot substitute for ClsB-dependent lipids for Mincle activation. The emerging serovar- and species-specific patterns of diPT production suggest that diPT is likely restricted to Gram-negative enteric bacteria. We failed to detect diPT in the enteric flora of healthy humans and mice, but instead detect the compound in some *E. coli* strains and serovars of *Salmonella enterica enterica*, with highest expression in *S. Typhi*, which is the cause of typhoid fever. These correlations between virulence and diPT expression support directed studies of diPT as a virulence factor. Given the differential high, low, or absent

levels of diPT within strains or serovars of the same species, as well as the low predictive value of *ClfB* gene identity for diPT, lipid chemotyping of clinical samples will be needed to understand the role of diPT in infection. The high production of diPT by *S. Typhi*, its potent agonism of a major activating receptor on macrophages, and the presence of diPT among bacteria that cause enteric fever syndromes now raise a key question for future studies: does diPT contribute to fever and sepsis that define enteric fever syndromes?

The World Health Organization estimates that diarrheal diseases, most commonly caused by enteric Gram-negative pathogens, remain the ninth leading cause of death worldwide, the fourth leading cause of death in developing nations and a major cause of death among children². Accordingly, *S. Typhi* and related serovars are priority pathogens that have spurred development of diagnostics, drug treatments, anti-sepsis regimens and vaccines⁴⁵. Both clinical and basic research on Gram-negative bacterial lipid endotoxins focus on LPS-TLR4 interactions, which is one of the most extensively studied and widely recognized receptor-ligand pairs in immunology⁴. Given this high investment in LPS and extensive attention to disease-causing serovars, it is striking that two immunogenic lipids that are abundant in the most virulent serovar could go undiscovered in decades of research on *Salmonella* species. Here, we describe diPT as a immunogen hiding in plain sight in bacterial membranes. This example underscores that the membrane content of Gram-negative bacterial cell walls, particularly the complex glycolipids in the outer membrane interfacing with the human host, remain an understudied resource for immunogens and virulence factors. The lipidomic profiles reported here identify hundreds of molecular species that differ by serovar, pointing to specific future paths for detecting serovar-specific lipid markers, including other molecules that could control host response.

Methods

Bacterial cultures and total lipid extraction

Species or strain	Number	Culture medium
<i>Salmonella</i> Enteritidis	SAL RIVM 423.78	Luria broth
<i>Salmonella</i> Enteritidis	SAL RIVM 423.79	Luria broth
<i>Salmonella</i> Enteritidis	SAL RIVM 423.68	Luria broth
<i>Salmonella</i> Paratyphi B	SAL RIVM 423.37	Luria broth
<i>Salmonella</i> Paratyphi A	SAL RIVM 422.26	Luria broth
<i>Salmonella</i> Paratyphi A	SAL RIVM 421.73	Luria broth
<i>Salmonella</i> Paratyphi A	NVGH308	Luria broth
<i>Salmonella</i> Typhi	SAL RIVM 424.47	Luria broth
<i>Salmonella</i> Typhi	SAL RIVM 422.41	Luria broth
<i>Salmonella</i> Typhi	SAL RIVM 421.56	Luria broth
<i>Salmonella</i> Typhi	Quailes	Luria broth
<i>Salmonella</i> Typhimurium	SAL RIVM 423.77	Luria broth
<i>Salmonella</i> Typhimurium	SAL RIVM 423.57	Luria broth
<i>Salmonella</i> Typhimurium	SAL RIVM 423.62	Luria broth
<i>Salmonella</i> Typhimurium	14028s	Luria broth
<i>Salmonella bongori</i>	RIVM 66:Z65	Luria broth
<i>Shigella dysenteriae</i>	BD09-00284, serovar 3	Luria broth
<i>Shigella flexneri</i>	BD09-0271	Luria broth
<i>Escherichia coli</i>	DH10B	Luria broth
<i>Escherichia coli</i>	DH5a	Luria broth
<i>Escherichia coli</i>	bw25113	Luria broth
<i>Escherichia coli</i>	ATCC25922	Luria broth
<i>Escherichia coli</i>	EHEC CVI-7, O150:H2	Luria broth
<i>Escherichia coli</i>	EHEC CVI-19, O177:H25	Luria broth
<i>Pseudomonas aeruginosa</i>	ATCC27853	Luria broth
<i>Pseudomonas aeruginosa</i>	NCTC 10662	Luria broth
<i>Brucella melitensis</i>	x10017283-001 (WBVR)	Tryptic soy broth
<i>Bordetella pertussis</i>	B1917	THUS ⁴⁶
<i>Mycobacterium tuberculosis</i>	H37Ra	Obtained dry (Difco)
<i>Rhodococcus equi</i>	ATCC 33701	Brain Heart Infusion broth
<i>Staphylococcus aureus</i>	Coagulase+ clinical isolate	Brain Heart Infusion broth
<i>Lactobacillus johnsonii</i>	Cecum of BALB/cJ mouse	Basal medium, anaerobic ⁴⁷
<i>Bacteroides ovatus</i>	ATCC 8483	Basal medium, anaerobic ⁴⁷
<i>Bacteroides vulgatus</i>	ATCC 8482	Basal medium, anaerobic ⁴⁷
<i>Bacteroides thetaiotamicron</i>	VPI-5482	Basal medium, anaerobic ⁴⁷
<i>Bacteroides fragilis</i>	NCTC 9343	Basal medium, anaerobic ⁴⁷

In addition, we used *S. Typhimurium* 14028s strains in which single genes are replaced by a cassette containing a kanamycin resistance gene oriented in the sense direction, or a chloramphenicol resistance gene oriented in the antisense direction²⁹. A single colony was picked from a plate, transferred to a 3 ml starter culture and incubated overnight at 37°C while shaking. One ml of a starter culture was added to 500 ml of medium and incubated overnight at 37°C while shaking. Bacteria were centrifuged for 15 minutes at 3500 rpm and washed twice with PBS. Lipid were extracted by rocking the pellet in organic solvent for 1 hr at 20°C, centrifugation for 10 minutes at 3500 rpm and collection of the supernatant. Solvents used for extraction were HPLC grade 2:1 Chloroform (Merck): Methanol (Merck) (C:M), 1:1 C:M, and 1:2 C:M. The three supernatants were pooled and dried and lipids were dissolved and stored in 1:1 C:M. Murine colon, cecum or small intestinal content (0.5-1.0 gram) or fresh human stool (1.0-2.0 gram) were suspended in 10:1 CH₃OH:0.3% NaCl in water and subjected to a series of extractions against petroleum ether, 9:10:3 CHCl₃:CH₃OH:0.3% NaCl in water, and 5:10:4 CHCl₃:CH₃OH:0.3% NaCl in water. Extracts were combined, dried, and weighed. Spiked samples were spiked directly after initial suspension in 10:1 Methanol:0.3% NaCl in water.

Comparative lipidomics

Total lipid extracts of triplicate cultures of *S. Typhi* (Quailes) and *S. Paratyphi A* (NVGH308) were separated using GL-Sciences Inertsil® Diol 3 μm 2.1x150 mm normal phase HPLC column equipped with Varian Monochrom 3 μm x 4.6 mm diol guard column. Lipids were measured on an Agilent Quadrupole Time-of-Flight (Q-TOF) Accurate-Mass QTOF LC/MS G6520B instrument in positive mode as described⁵. Data were analyzed using Mass Hunter (Agilent), LIMMA⁴⁸, and XCMS⁴⁹.

Thin layer chromatography

DURASIL-25 TLC plates (Macherey-Nagel) were pre-cleared with C:M:H₂O (60:30:6) (V:V:V) and dried. Bacterial total lipid extract (300 μg) or purified standard (40 μg) were applied and resolved with C:M:H₂O (60:30:6)(V:V:V). The plates were dried and either stained for analytical purposes (Figure 1b) or used for specific lipid isolation. Staining was performed by spraying 3% copper acetate monohydrate (Sigma Aldrich) in 8% phosphoric acid (Merck) on the plate and baking at 140°C. Isolation of specific lipids was performed after spraying the plate with water, which makes lipids bands temporarily visible. Lipids of interest were marked, the plate was dried, and the silica layer containing the lipid was scraped of the glass plate. The lipid was isolated from the silica by rocking the silica for 1 hr in 1:1 C:M. After rocking the sample was centrifuged and the supernatant containing the lipid was stored. PE (850758P) and PG (840503P) standards were from Avanti Lipids.

Analytical mass spectrometry

For mass determination and higher order CID-MS, purified lipids were dissolved in methanol and measured by nano-ESI-MS in the positive mode (LXQ Linear Ion Trap Mass Spectrometer with MSⁿ software; Thermo Fisher Scientific). For optimal detection and quantification of PT, diPT, PE, and CL with high mass resolution, reversed phase HPLC-MS was performed on an accurate-mass QTOF LC/MS G6520 instrument. In 15 minutes, a gradient from 100% solvent A (5% H₂O in MeOH with 2mM ammonium formate) to 100% solvent B (10% cyclohexane in 1-propanol with 3 mM ammonium formate) was run on an Agilent Poroshell 120Å, EC-C18, 1.9 μm column, equipped with an Agilent EC-C18, 3.0x5mm, 2.7 μm guard column. The elution time of cardiolipin was determined using a synthetic standard (Avanti # 750332).

Nuclear magnetic resonance

NMR spectra were recorded on an Oxford 600 MHz Magnet (600 MHz for ¹H, 151 MHz for ¹³C), Bruker AVANCE II Console system, equipped with 5mm Prodigy HCN TXI cold probe and an Agilent 700 MHz magnet (700 MHz for ¹H, 176 MHz for ¹³C) with a DDR2 console equipped with an Agilent Triple resonance Helium cold probe. The 31P-decoupling experiments were performed on a Varian 400 MHz magnet equipped with 5mm AutoX OneProbe. Chemical shifts are reported in ppm and coupling constants (J) in Hz. The samples were measured in MeOH-d₄ of which the residual solvent resonance was used as an internal standard (δ 3.31 for ¹H, δ 49.00 for ¹³C). The ¹H-NMR spectra were assigned using COSY, TOCSY, multiplicity-edited HSQC (135DEPT-HSQC) and HMBC. The TOCSY experiments were performed with a MLEV17 mixing scheme with a 100 ms spin lock¹⁴. The data are reported as follows: chemical shifts (δ), multiplicity (s = singlet, d = doublet, dd = double doublet, ddd = double double doublet, td = triple doublet, t = triplet, dt = double triplet, q = quartet, m = multiplet), coupling constants J (Hz) and integration (Supplementary Figure 2).

Reconstitution of *ClsB* in *S. Typhimurium* Δ*clsB* mutant

ClsB was isolated via PCR using forward primer 5'CGCGGATCCGATACGGTAACGCGGTTCTTTCT and reverse primer 5'GATGAATCCGGCCGCAATAAAGCCGTCCAAG. The sequence of the PCR product was verified by Sanger sequencing. A second PCR was performed using forward primer 5'TAGAGGAATAATAAATGATGAAATGCGGCTGGCGTGAAGGTAATCAA and reverse primer 5'GTTAGGGCTTCACTCCTGTATTTTC to make it suitable for cloning into the pBAD-TOPO vector (Life-Technologies). Correct insertion was confirmed by restriction digestion with BsmI (New England Biolabs). A 2ml *S. Typhimurium* Δ*clsB* culture was grown overnight in LB media with chloramphenicol (25 μg/ml). 0.5 ml of the starter culture was diluted in 50ml of LB with CM. This culture was incubated for 2.5 hr at 37°C while shaking.

The culture was washed three times with ice-cold 10 % glycerol and used for electroporation by adding 50 ng of pBAD-TOPO-CLSB to 40 µl of bacteria. A 0.1 cm cuvette was pulsed with 1.6 kV, capacitance of 2.5 µF, for ~5 µs. Cells were grown for 4 hr in 1 ml of LB without antibiotics at 37°C before plating on LB plates with 100 µg/ml ampicillin. After overnight incubation colonies were picked and grown overnight in 3 ml of LB with chloramphenicol and ampicillin. Insertion of the *clsB* gene was confirmed by PCR. Total lipid extraction was performed on cultures of a reconstituted clone grown with or without induction with 0.2 % arabinose.

Mincle activation assay

Lipids were diluted in 20 µl isopropanol per well of a flat bottom 96 well plate. Isopropanol was evaporated, after which 3×10^4 reporter cells in 100 µl medium were added per well. Reporter cell lines were FcRγ-, murine Mincle and FcRγ-, or human Mincle and FcRγ-expressing 2B4-NFAT-GFP cells⁵⁰. After 18hr at 37C, cells were washed with PBS 1% BSA and GFP expression was measured on the FACSCanto II (BD). Data were analyzed using FlowJo software.

Identification of ClsB in other bacterial species

For a selection of bacterial species, the 16S sequence was downloaded from the SILVA database (<https://www.arb-silva.de>). Alignment, Classification and Tree (ACT) service from the SILVA server was used with default settings to align the sequences and generate a phylogenetic tree. The unrooted tree was visualized using the ITOL server (<http://itol.embl.de/>). Protein sequences of ClsA, ClsB, and ClsC of *S. Typhi* were used for a BLASTP search in bacterial genomes using the NCBI server. For each Cls protein, the percentage identity to the top hit in each bacterial species was determined. A BLASTP search of this top hit in the *S. Typhi* genome was performed to determine whether the closest hit was the same Cls protein used in the initial BLASTP search. For the phylogenetic analysis of Cls proteins, the amino acid sequences were aligned using muscle⁵¹, informative sites selected with trimAl⁵², and the tree as shown in Figure 5b was calculated with MrBayes⁵³ for 1 million generations under the mixed amino acid model and a 25% burning for consensus tree generation; a separate calculation using iqtree⁵⁴ under the automatic model selection resulted in the same topology. Both tree files as well as the input alignment are available at [<https://figshare.com/s/4d65ca000a63275f56fa>].

Synthesis of diPT

We developed a stereoselective chemical synthesis of diPT, including *de novo* synthesis of a fatty acid with a 9*R*,10*S* *cis*-cyclopropyl ring. The cyclopropyl fatty acid was prepared as described by Spencer et al.⁵⁵. Briefly, rhodium-catalyzed cyclopropanation of octyne with ethyl diazoacetate was followed by resolution of the enantiomers via diastereomer formation and chromatography⁵⁶. The resulting enantiopure cyclopropenes were then

converted to the desired enantiomeric fatty acids over several synthetic steps including a Wittig reaction⁵⁷ and a diimide reduction⁵⁸. Diacylglycerol was prepared by epoxide ring-opening⁵⁹ of protected *S*-glycidol with palmitic acid, followed by esterification of the resulting hydroxyl group with the 9*R*,10*S* cyclopropyl fatty acid. Careful deprotection, to avoid acyl shift, produced the free diacylglycerol⁶⁰. Suitably protected trehalose⁶¹ was converted to bis-2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite and coupled with diacylglycerol, mediated by dicyanoimidazole, and then immediately followed by oxidation to the phosphate. Finally, deprotection of the phosphates and removal of the benzyl protecting groups provided the desired product.

Acknowledgements

This work was supported by funding from Nederlands Wetenschappelijk Onderzoek (NWO) grant 824.02.002 and the NIH (AI116604 to DBM). P.R. was supported by an EMBO Short Term Fellowship (16-2015). A.J.M. and V.K.M. are supported by the University of Groningen. S.Y. was supported by MEXT (26293099 and 26110009) and AMED (JP17gm0910010 and JP17ak0101070). We thank B. Kuipers and E. Lambert for assistance with bacterial culture, J. Kemmink for NMR spectroscopy studies, and R. Cotton for critically reading the manuscript and advice. V.C., G.N. and D.B.M. were supported by the UK Medical Research Council (MRC) (MR/K021222/1). V.C. was also supported by Cancer Research UK (CRUK) (C399/A2291) and by the NIHR Oxford Biomedical Research Centre. S.P. and M.M. were supported by NIH contract HHSN272200900040C. The authors declare no competing financial interests.

Author contributions

I.V.R., A.J.M., D.B.M., S.Y., V.C. and M.M. conceived and designed the experiments. V.K.M. carried out chemical synthesis. P.R. performed bacteriology experiments and genomic searches. P.R. and E.I. performed immunology experiments. J.B., T-Y.C., V.K.M., A.J.M., and P.R. performed chemical analyses. S.Y., M.M., P.W., E.H., G.D., P.J.B., V.C., G.N., and C.A.V.E. contributed reagents or original ideas. I.V.R. and D.B.M. wrote the paper.

Competing interests

The authors declare no competing interests.

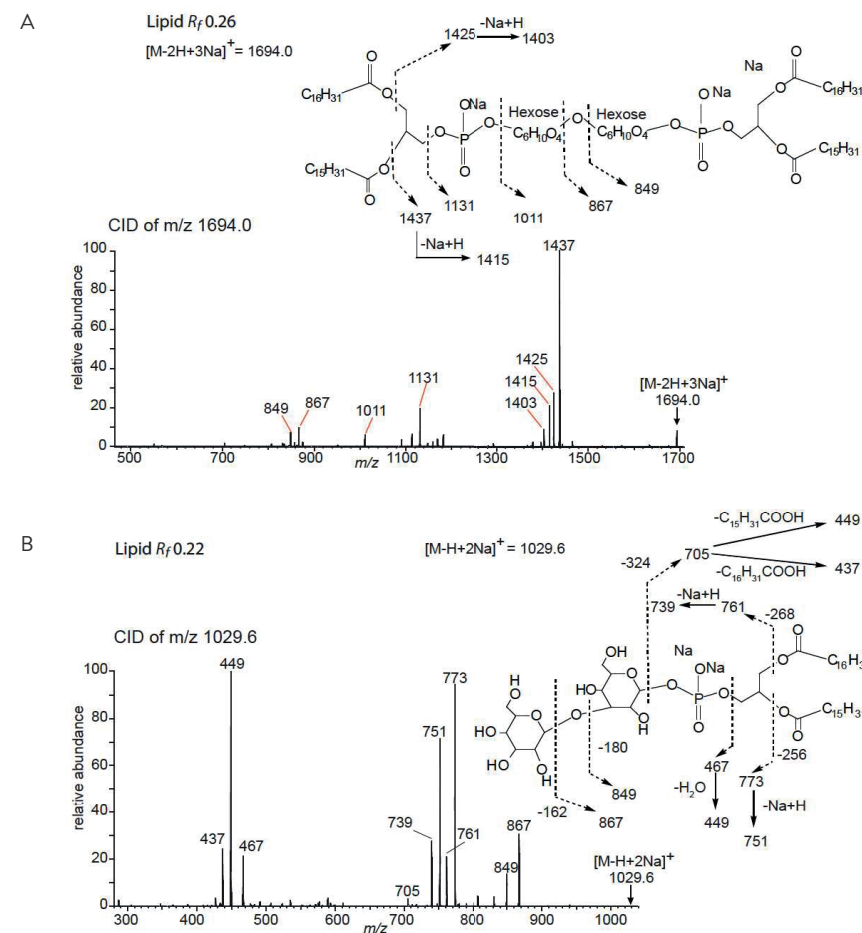
Correspondence and requests for materials should be addressed to I.V.R.: i.vanhijn@uu.nl

References

- Mogasale, V., et al. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. *The Lancet. Global health* **2**, e570-580 (2014).
- Organization, W.H. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, WHO (2018).
- Poltorak, A., et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085-2088 (1998).
- Nobel Prize to immunology. *Nature Reviews Immunology* **11**, 714 (2011).
- Layre, E., et al. A comparative lipidomics platform for chemotaxonomic analysis of *Mycobacterium tuberculosis*. *Chem. Biol.* **18**, 1537-1549 (2011).
- Matsunaga, I., et al. *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J Exp Med* **200**, 1559-1569 (2004).
- Madigan, C.A., et al. Lipidomic discovery of deoxysiderophores reveals a revised mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1257-1262 (2012).
- Layre, E., de Jong, A. & Moody, D.B. Human T cells use CD1 and MR1 to recognize lipids and small molecules. *Curr. Opin. Chem. Biol.* **23**, 31-38 (2014).
- Layre, E. & Moody, D.B. Lipidomic profiling of model organisms and the world's major pathogens. *Biochimie* **95**, 109-115 (2013).
- Fahy, E., et al. Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **50** Suppl, S9-14 (2009).
- Moody, D.B., et al. T cell activation by lipopeptide antigens. *Science* **303**, 527-531 (2004).
- Moody, D.B., et al. CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. *Nature* **404**, 884-888 (2000).
- Roslund, M.U., Tahtinen, P., Niemitz, M. & Sjöholm, R. Complete assignments of the (1)H and (13)C chemical shifts and J(H,H) coupling constants in NMR spectra of D-glucopyranose and all D-glucopyranosyl-D-glucopyranosides. *Carbohydr. Res.* **343**, 101-112 (2008).
- Gheysen, K., Mihai, C., Conrath, K. & Martins, J.C. Rapid identification of common hexapyranose monosaccharide units by a simple TOCSY matching approach. *Chemistry* **14**, 8869-8878 (2008).
- Wang, Y. & Hollingsworth, R.I. A solvent system for the high-resolution proton nuclear magnetic resonance spectroscopy of membrane lipids. *Anal. Biochem.* **225**, 242-251 (1995).
- Perez-Victoria, I., et al. Saturation transfer difference NMR reveals functionally essential kinetic differences for a sugar-binding repressor protein. *Chem Commun (Camb)*, 5862-5864 (2009).
- Knothe, G. NMR characterization of dihydrostercularic acid and its methyl ester. *Lipids* **41**, 393-396 (2006).
- Grogan, D.W. & Cronan, J.E., Jr. Cyclopropane ring formation in membrane lipids of bacteria. *Microbiology and molecular biology reviews : MMBR* **61**, 429-441 (1997).
- Hildebrand, J.G. & Law, J.H. Fatty Acid Distribution in Bacterial Phospholipids. The Specificity of the Cyclopropane Synthetase Reaction. *Biochemistry (Mosc.)* **3**, 1304-1308 (1964).
- Adam, A., Senn, M., Vilkas, E. & Lederer, E. [Mass spectrometry of glycolipids. 2. Natural and synthetic diesters of trehalose]. *Eur. J. Biochem.* **2**, 460-468 (1967).
- Geisel, R.E., Sakamoto, K., Russell, D.G. & Rhoades, E.R. In vivo activity of released cell wall lipids of *Mycobacterium bovis* bacillus Calmette-Guerin is due principally to trehalose mycolates. *J. Immunol.* **174**, 5007-5015 (2005).
- Shenderov, K., et al. Cord factor and peptidoglycan recapitulate the Th17-promoting adjuvant activity of mycobacteria through mincle/CARD9 signaling and the inflammasome. *J. Immunol.* **190**, 5722-5730 (2013).
- Pimm, M.V., Baldwin, R.W., Polonsky, J. & Lederer, E. Immunotherapy of an ascitic rat hepatoma with cord factor (trehalose-6, 6'-dimycolate) and synthetic analogues. *Int. J. Cancer* **24**, 780-785 (1979).
- Holten-Andersen, L., Doherty, T.M., Korsholm, K.S. & Andersen, P. Combination of the cationic surfactant dimethyl dioctadecyl ammonium bromide and synthetic mycobacterial cord factor as an efficient adjuvant for tuberculosis subunit vaccines. *Infect. Immun.* **72**, 1608-1617 (2004).
- Ishikawa, E., et al. Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle. *Journal of Experimental Medicine* **206**, 2879-2888 (2009).
- Werninghaus, K., et al. Adjuvanticity of a synthetic cord factor analogue for subunit *Mycobacterium tuberculosis* vaccination requires FcRgamma-Syk-Card9-dependent innate immune activation. *J Exp Med* **206**, 89-97 (2009).
- James, C.A., et al. CD1b Tetramers Identify T Cells that Recognize Natural and Synthetic Diacylated Sulfoglycolipids from *Mycobacterium tuberculosis*. *Cell chemical biology* **25**, 392-402 e314 (2018).
- Burbaud, S., et al. Trehalose Polyphleates Are Produced by a Glycolipid Biosynthetic Pathway Conserved across Phylogenetically Distant Mycobacteria. *Cell chemical biology* **23**, 278-289 (2016).
- Porwollik, S., et al. Defined single-gene and multi-gene deletion mutant collections in *Salmonella enterica* sv Typhimurium. *PLoS one* **9**, e99820 (2014).
- Guo, D. & Tropp, B.E. A second *Escherichia coli* protein with CL synthase activity. *Biochim. Biophys. Acta* **1483**, 263-274 (2000).
- Tan, B.K., et al. Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16504-16509 (2012).
- Li, C., Tan, B.K., Zhao, J. & Guan, Z. In Vivo and in Vitro Synthesis of Phosphatidylglycerol by an *Escherichia coli* Cardiolipin Synthase. *J. Biol. Chem.* **291**, 25144-25153 (2016).
- Rossi, R.M., Yum, L., Agaisse, H. & Payne, S.M. Cardiolipin Synthesis and Outer Membrane Localization Are Required for *Shigella flexneri* Virulence. *mBio* **8**(2017).
- Del Fresno, C., Iborra, S., Saz-Leal, P., Martinez-Lopez, M. & Sancho, D. Flexible Signaling of Myeloid C-Type Lectin Receptors in Immunity and Inflammation. *Frontiers in immunology* **9**, 804 (2018).
- Furukawa, A., et al. Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17438-17443 (2013).
- Feinberg, H., et al. Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor mincle. *J. Biol. Chem.* **288**, 28457-28465 (2013).
- Nagata, M., et al. Intracellular metabolite beta-glucosylceramide is an endogenous Mincle ligand possessing immunostimulatory activity. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E3285-E3294 (2017).
- Behler-Janbeck, F., et al. C-type Lectin Mincle Recognizes Glucosyl-diacylglycerol of *Streptococcus pneumoniae* and Plays a Protective Role in Pneumococcal Pneumonia. *PLoS Pathog.* **12**, e1006038 (2016).
- Decout, A., et al. Rational design of adjuvants targeting the C-type lectin Mincle. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2675-2680 (2017).
- Soldner, C.A., Horn, A.H.C. & Sticht, H. Interaction of Glycolipids with the Macrophage Surface Receptor Mincle - a Systematic Molecular Dynamics Study. *Sci Rep* **8**, 5374 (2018).
- Geisel, R.E., Sakamoto, K., Russell, D.G. & Rhoades, E.R. In vivo activity of released cell wall lipids of *Mycobacterium bovis* bacillus Calmette-Guerin is due principally to trehalose mycolates. *J. Immunol.* **174**, 5007-5015 (2005).

42. Oda, M., et al. Vizantin inhibits endotoxin-mediated immune responses via the TLR 4/MD-2 complex. *J Immunol* **193**, 4507-4514 (2014).
43. Belisle, J.T., et al. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**, 1420-1422 (1997).
44. Backus, K.M., et al. The three *Mycobacterium tuberculosis* antigen 85 isoforms have unique substrates and activities determined by non-active site regions. *J. Biol. Chem.* **289**, 25041-25053 (2014).
45. Andrews, J.R., et al. Typhoid conjugate vaccines: a new tool in the fight against antimicrobial resistance. *Lancet Infect. Dis.* (2018).
46. Thalen, M., et al. Rational medium design for *Bordetella pertussis*: basic metabolism. *J. Biotechnol.* **75**, 147-159 (1999).
47. Zitomersky, N.L., Coyne, M.J. & Comstock, L.E. Longitudinal analysis of the prevalence, maintenance, and IgA response to species of the order Bacteroidales in the human gut. *Infect. Immun.* **79**, 2012-2020 (2011).
48. Ritchie, M.E., et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
49. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787 (2006).
50. Yamasaki, S., et al. Mincle is an ITAM-coupled activating receptor that senses damaged cells. *Nat. Immunol.* **9**, 1179-1188 (2008).
51. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
52. Capella-Gutierrez, S., Silla-Martinez, J.M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
53. Ronquist, F., et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539-542 (2012).
54. Nguyen, L.T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).
55. Shah, S., White, J.M. & Williams, S.J. Total syntheses of cis-cyclopropane fatty acids: dihydromalvalic acid, dihydrosterculic acid, lactobacillic acid, and 9,10-methylenehexadecanoic acid. *Org. Biomol. Chem.* **12**, 9427-9438 (2014).
56. Liao, L., Zhang, F., Yan, N., Golen, J.A. & Fox, J.M. An efficient and general method for resolving cyclopropene carboxylic acids. *Tetrahedron* **60**, 1803-1816 (2004).
57. Coxon, G., et al. The synthesis of both enantiomers of lactobacillic acid and mycolic acid analogues. *Tetrahedron Lett.* **40**, 6689-6692 (1999).
58. Smit, C., Fraaije, M.W. & Minnaard, A.J. Reduction of carbon-carbon double bonds using organocatalytically generated diimide. *J. Org. Chem.* **73**, 9482-9485 (2008).
59. Jacobsen, E.N., Kakiuchi, F., Konsler, R.G., Larrow, J.F. & Tokunaga, M. Enantioselective catalytic ring opening of epoxides with carboxylic acids. *Tetrahedron Lett.* **38**, 773-776 (1997).
60. Fodran, P. & Minnaard, A.J. Catalytic synthesis of enantiopure mixed diacylglycerols - synthesis of a major *M. tuberculosis* phospholipid and platelet activating factor. *Org. Biomol. Chem.* **11**, 6919-6928 (2013).
61. Gilbertson, S. & Chang, C.-W. Synthesis of New Disugar Phosphine Ligands and Their Use in Asymmetric Hydrogenation. *The Journal of Organic Chemistry* **60**, 6226-6228 (1995).

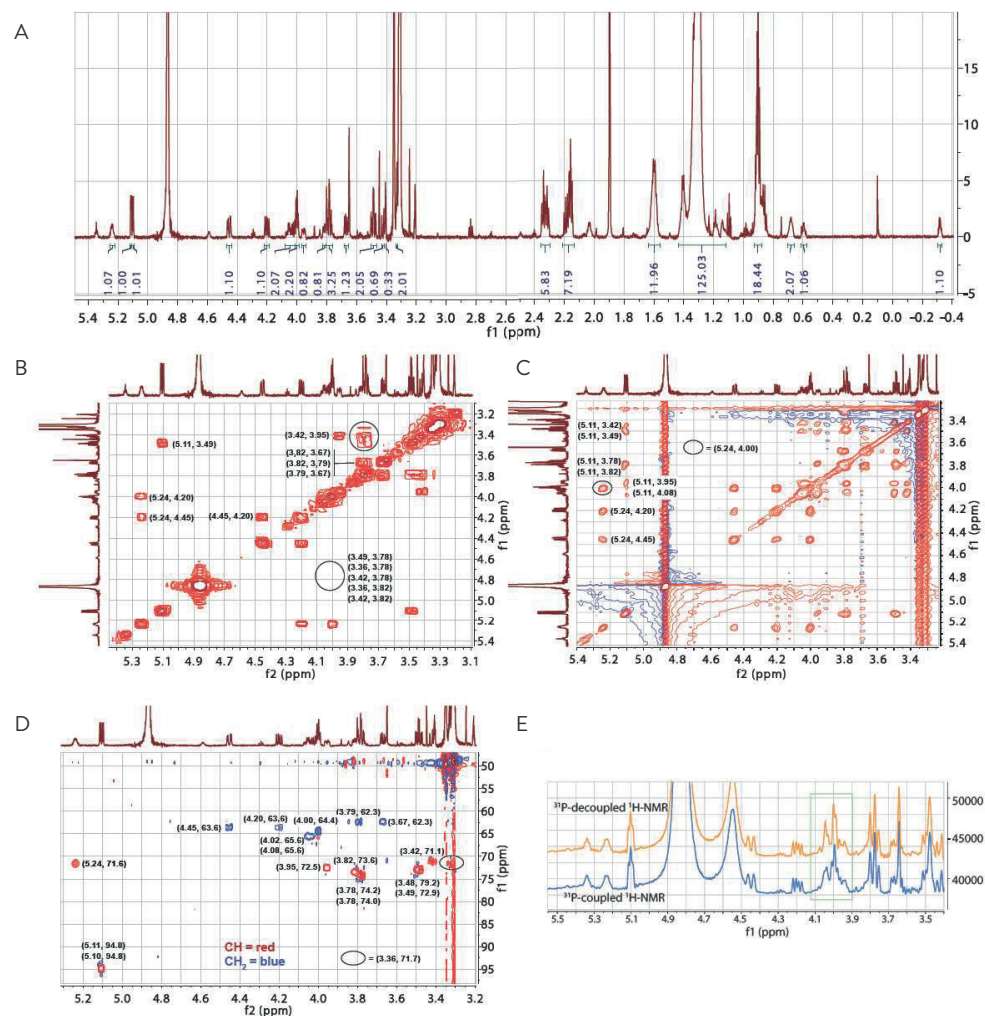
Supplemental material



Supplementary Figure 1. Identification of TLC-isolated lipids from *S. Typhi* as hexose phospholipids.

Material from the upper (R_f 0.26, A) and lower (R_f 0.22, B) band, isolated as shown in Fig. 1c, was subjected to multistage collisional mass spectrometry using nano-electrospray ionization mass spectrometry in the positive mode on an LXQ Linear Ion Trap Mass Spectrometer. The first stage collisional mass spectrum is shown. The interpretation is based on this spectrum and additional higher stage CID-MS experiments that are not shown.

Supplementary Figure 1: Identification of TLC-isolated lipids from *S. Typhi* as hexose phospholipids.



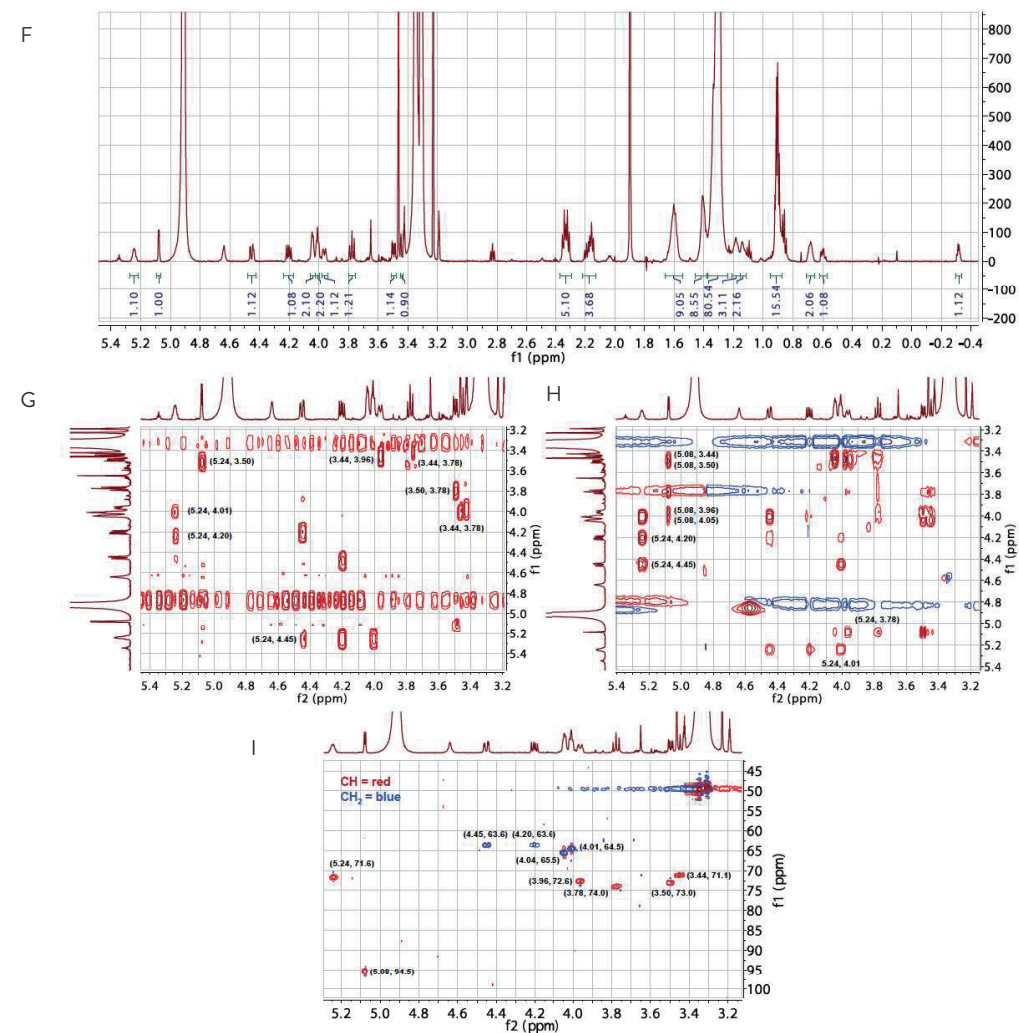
Identification of phosphatidyltrehalose from *S. Typhi* by NMR.

TLC-isolated lower (R_f 0.22) band lipid from *S. Typhi* was analyzed by (a) $^1\text{H-NMR}$, (b) COSY, (c) TOCSY with 100 ms spin-lock, (d) 135- Distortionless Enhancement by Polarization Transfer (DEPT)- heteronuclear single quantum coherence (HSQC), and (e) ^{31}P -coupled and ^{31}P -decoupled $^1\text{H-NMR}$.

$^1\text{H-NMR}$ (700 MHz, MeOH- d_4): δ 5.27 – 5.21 (m, 1H), 5.11 (d, $J = 3.7$ Hz, 1H), 5.10 (d, $J = 3.7$ Hz, 1H), 4.45 (dd, $J = 12.1$, 3.0 Hz, 1H), 4.20 (dd, $J = 12.0$, 6.9 Hz, 1H), 4.05 (ddd, $J = 11.8$, 5.6, 2.2 Hz, 2H), 4.00 (t, $J = 5.6$ Hz, 2H), 3.98 – 3.94 (m, 1H), 3.82 (ddd, $J = 10.3$, 5.3, 2.3 Hz, 1H), 3.80 – 3.76 (m, 3H), 3.67 (dd, $J = 11.8$, 5.2 Hz, 1H), 3.49 (td, $J = 9.8$, 3.8 Hz, 2H), 3.41 (dd, $J = 9.4$ Hz, 1H), 3.34 – 3.33 (m, 1H), 2.37 – 2.29 (m, 6H), 2.21 – 2.13 (m, 7H), 1.66 – 1.55 (m, 12H), 1.44 – 1.12 (m, 55H), 0.92 – 0.88 (m, 13H), 0.70 – 0.66 (m, 2H), 0.60 (ddd, $J = 4.3$ Hz, 1H), -0.32 (q, $J = 5.0$ Hz, 1H).

^{13}C (176 MHz, MeOH- d_4) - signals observed in the multiplicity-edited HSQC - : δ 94.8, 74.2, 74.0, 73.6, 72.9, 72.5, 71.6, 71.1, 65.6, 64.4, 63.6, 62.3, 38.7, 36.3, 34.7, 32.6 – 23.6 (multiple resonances), 27.3, 25.8, 22.6, 19.9, 16.6, 14.5.

Supplementary Figure 2: Identification of phosphatidyl trehalose and diphosphatidyl trehalose from *S. Typhi* by NMR

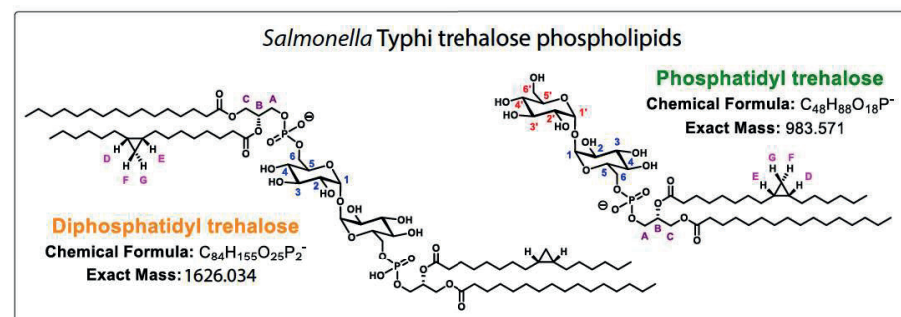


Identification of diphosphatidyltrehalose from *S. Typhi* by NMR.

TLC-isolated upper (R_f 0.26) band lipid from *S. Typhi* was analyzed by (f) $^1\text{H-NMR}$, (g) COSY, (h) TOCSY with 100 ms spin-lock, and (i) 135- Distortionless Enhancement by Polarization Transfer (DEPT)- heteronuclear single quantum coherence (HSQC).

$^1\text{H-NMR}$ (600 MHz, MeOH- d_4): δ 5.27 – 5.22 (m, 1H), 5.08 (d, $J = 3.7$ Hz, 1H), 4.45 (dd, $J = 12.0$, 3.0 Hz, 1H), 4.20 (dd, $J = 12.0$, 6.9 Hz, 1H), 4.06 – 4.02 (m, 2H), 4.01 (t, $J = 5.7$ Hz, 2H), 3.96 (dt, $J = 9.2$, 2.0 Hz, 1H), 3.78 (t, $J = 9.4$ Hz, 1H), 3.50 (dd, $J = 9.7$, 3.6 Hz, 1H), 3.44 (d, $J = 9.9$ Hz, 1H), 2.37 – 2.29 (m, 5H), 2.22 – 2.12 (m, 4H), 1.66 – 1.55 (m, 9H), 1.40 (d, $J = 7.6$ Hz, 8H), 1.37 – 1.24 (m, 80H), 1.19 (dd, $J = 8.5$, 4.6 Hz, 3H), 1.14 (s, 2H), 0.91 (td, $J = 7.0$, 4.5 Hz, 16H), 0.71 – 0.65 (m, 2H), 0.60 (ddd, $J = 4.5$, 3.7 Hz, 1H), -0.32 (q, $J = 5.0$ Hz, 1H).

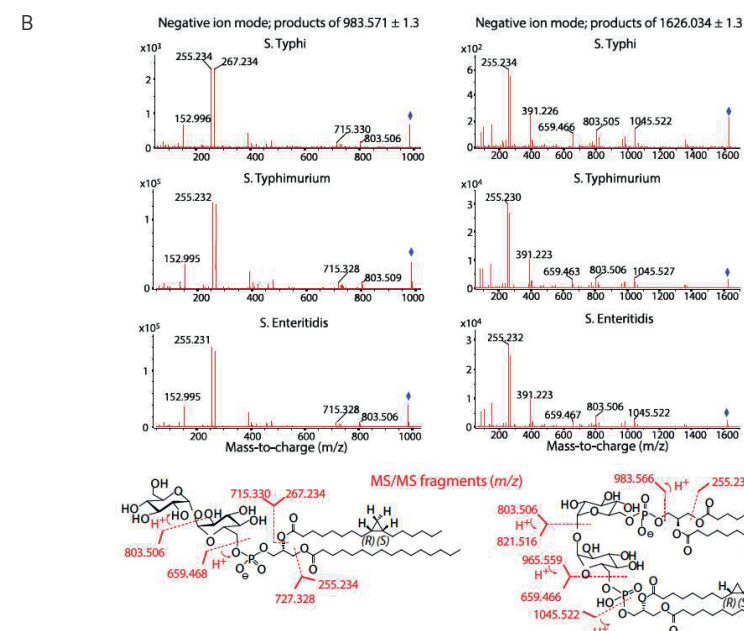
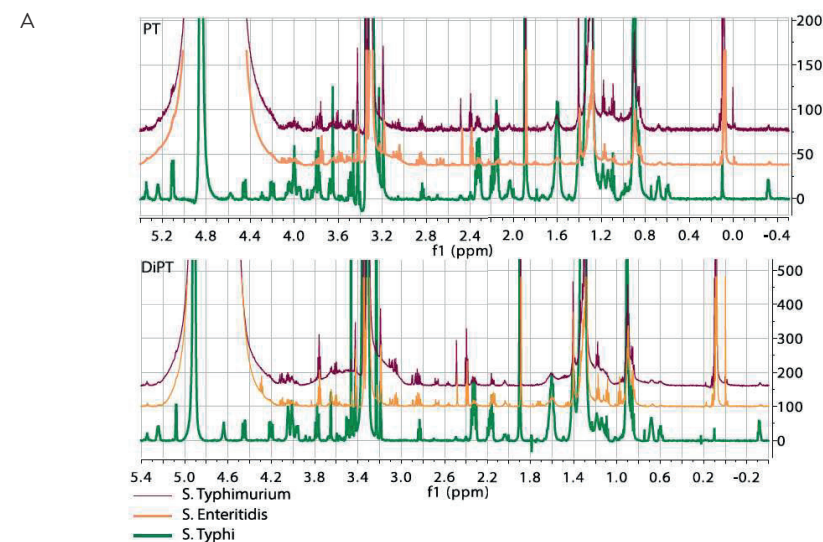
^{13}C signals (151 MHz, MeOH- d_4) - observed in the multiplicity-edited HSQC - : δ 95.4, 74.0, 73.0, 72.6, 71.6, 71.1, 65.5, 64.5, 63.6, 38.8, 36.4, 34.8, 32.9-23.5 (multiple resonances), 27.5, 25.9, 22.9, 19.8, 16.6, 14.3, 11.6.



Position	Phosphatidyl trehalose				Diphosphatidyl trehalose			
	multiplicity	1H (ppm)	$J_{H,H}$ (Hz)	^{13}C (ppm) ^a	multiplicity	1H (ppm)	$J_{H,H}$ (Hz)	^{13}C (ppm) ^a
1	d	5.11	3.7	94.8	d	5.08	3.7	94.5
2	dd ^b	3.49	3.8, 9.8	72.9	dd	3.50	3.6, 9.7	73.0
3	dd ^c	3.78	9.7	74.0	t	3.78	9.4	74.0
4	dd ^d	3.42	9.4	71.1	d	3.44	9.9	71.1
5	m	3.94-3.96		72.5	dt	3.96	2.0, 9.2	72.6
6a	m	4.02		65.6	m	4.02-4.06		65.5
6b	ddd ^c	4.08	2.2, 5.6*, 11.6	see 6a	m	see 6a		see 6a
1'	d	5.10	3.7	94.8				
2'	dd ^b	3.48	3.8, 9.8	72.9				
3'	dd ^c	3.78	9.7	74.2				
4'	m	3.36		71.7				
5'	ddd	3.82	2.3, 5.3, 10.3	73.6				
6'a	m	3.76-3.81		62.3				
6'b	dd	3.67	5.2, 11.8	see 6'a				
A	t	4.00	5.6	64.4	t ^e	4.01	5.7	64.5
B	m	5.21-5.27		71.6	m	5.22-5.27		71.6
C1	dd	4.45	3.0, 12.1	63.6	dd	4.45	3.0, 12.0	63.6
C2	dd	4.20	6.9, 12.0	see C1	dd	4.20	6.9, 12.0	see C1
D	m	0.66-0.70		16.6	m	0.65-0.68		16.5
E	m	0.66-0.70		16.6	m	0.65-0.68		16.5
Fcis	ddd	0.60	4.3		ddd	0.60	3.7, 4.5	
Gtrans	q	-0.32	5.0		q	-0.32	5.0	
C=O ^e				173.2				

Summary table: correlation of NMR data with structure.

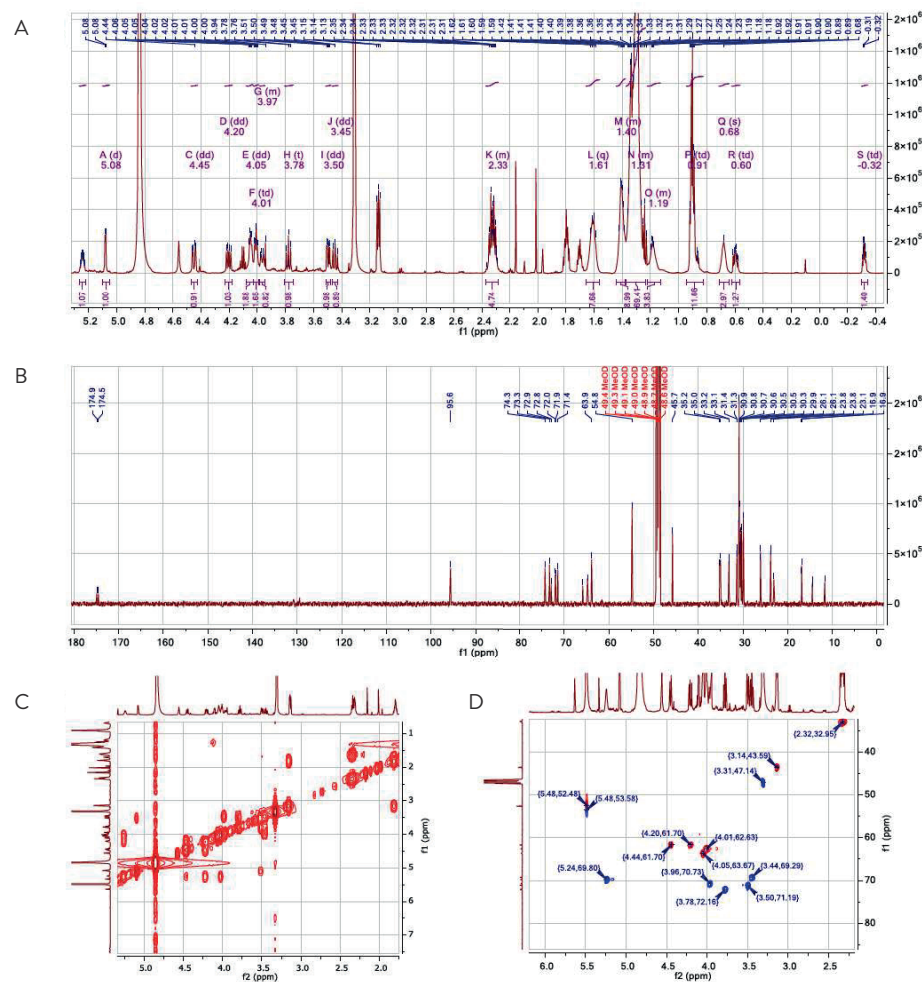
(j) NMR assignments of the lower and upper band isolated from *S. Typhi*. ^aassigned using the multiplicity-edited HSQC spectrum, ^bapparent doublet, ^capparent multiplet, ^dapparent triplet, ^eassigned using a Heteronuclear Multiple Bond Correlation (HMBC) spectrum, ^{*} = $^3J_{H,P}$ coupling, grey boxes = not applicable or not obtained.



Supplementary Figure 3. Comparison of PT and diPT from *S. Typhi*, *S. Enteritidis*, and *S. Typhimurium*.

(A) 1H -NMR spectra of TLC-isolated PT (upper panel) and diPT (lower panel) from *S. Typhi*, *S. Enteritidis*, and *S. Typhimurium* were merged and aligned. (B) Negative ion mode high accuracy collisional mass spectra of TLC-isolated PT (left panels) and diPT (right panels) from *S. Typhi*, *S. Enteritidis*, and *S. Typhimurium*.

Supplementary Figure 3: Comparison of PT and diPT from *S. Typhi*, *S. Enteritidis*, and *S. Typhimurium*



Supplementary Figure 4. NMR of synthetic diPT.

Synthetic diPT was analyzed with (A) ^1H NMR (600 MHz), (B) ^{13}C NMR (150 MHz), (C) gCOSY, (D) Multiplicity-Edited HSQC

^1H NMR (600 MHz, CD_3OD): = 5.24 (dtd, $J = 6.7, 5.1, 3.1$ Hz, 1H), 5.08 (d, $J = 3.8$ Hz, 1H), 4.45 (dd, $J = 12.0, 3.1$ Hz, 1H), 4.20 (dd, $J = 12.0, 6.9$ Hz, 1H), 4.05 (dd, $J = 6.2, 3.5$ Hz, 2H), 4.01 (td, $J = 5.5, 1.6$ Hz, 2H), 3.98 – 3.94 (m, 1H), 3.78 (t, $J = 9.4$ Hz, 1H), 3.50 (dd, $J = 9.7, 3.8$ Hz, 1H), 3.45 (dd, $J = 10.0, 9.0$ Hz, 1H), 2.37 – 2.29 (m, 4H), 1.61 (q, $J = 6.9, 6.2$ Hz, 7H), 1.44 – 1.38 (m, 8H), 1.37 – 1.24 (m, 65H), 1.22 – 1.13 (m, 4H), 0.91 (td, $J = 7.0, 4.2$ Hz, 11H), 0.68 (s, 3H), 0.60 (td, $J = 8.3, 4.1$ Hz, 1H), -0.32 (td, $J = 5.4, 4.1$ Hz, 1H).

^{13}C NMR (150 MHz, MeOD): = 174.9, 174.539, 95.7, 74.3, 73.3, 72.9, 72.9, 71.9, 71.4, 65.8, 65.8, 64.9, 64.75, 63.9, 54.8, 45.7, 35.2, 35.0, 33.2, 33.1, 31.4, 31.3, 30.9, 30.8, 30.7, 30.6, 30.5, 30.5, 30.3, 29.9, 26.1, 26.1, 23.8, 23.8, 23.1, 16.9, 16.9, 14.5, 14.5, 11.7.

Supplementary Figure 4: NMR of synthetic diPT

Gene mutated	Set 1: Kanamycin-selected mutants peak area ($\times 10^3$)			
	m/z 983 (PT)	m/z 1626 (DiPT)	m/z 674 (PE)	m/z 1389 (Cardiolipin)
Aas	1202	100	1289	1533
ClsA	1147	121	1632	1366
ClsB	0	0	1827	973
EntF	723	64	1919	1296
OtsA	214	23	1731	1651
OtsB	278	56	1957	1652
PagP	567	47	1831	1120
PgpA	1274	104	1677	1453
PgpB	932	103	1733	1586
PgpC	661	48	1919	1581
TreY	349	28	2072	1219
TreZ	833	87	2215	1417
WT	467	48	1740	1360

Gene mutated	Set 2: Chloramphenicol-selected mutants peak area ($\times 10^3$)			
	m/z 983 (PT)	m/z 1626 (DiPT)	m/z 674 (PE)	m/z 1389 (Cardiolipin)
Aas	392	33	718	636
ClsA	522	46	818	331
ClsB	0	0	728	581
EntF	560	51	919	724
OtsA	631	57	906	660
OtsB	42	7.6	933	623
PagP	278	25	791	701
PgpA	119	11	904	880
PgpB	316	31	891	668
PgpC	593	50	716	379
TreY	482	31	1052	794
TreZ	566	53	789	680
WT	36	6.9	972	657

Supplementary Figure 5. Abundance of PT, diPT, phosphatidylethanolamine (PE), and cardiolipin in *S. Typhimurium* single gene knockouts.

Total lipid extracts from the indicated *S. Typhimurium* single gene knockouts were analyzed by reverse phase HPLC-MS. Areas under the indicated extracted ion chromatograms are shown for PT (m/z 983.5714 \pm 10 ppm), diPT (1626.0338 \pm 10 ppm), phosphatidylethanolamine (PE) (674.4770 \pm 10 ppm), and cardiolipin (1389.9806 \pm 10 ppm).

Supplementary Figure 5: Abundance of PT, diPT, phosphatidylethanolamine (PE), and cardiolipin in *S. Typhimurium* single gene knockouts



Chapter 7

Discussion

In this chapter we discuss some recurring themes as described in this thesis. Topics that will be discussed are T cell receptors (TCR) and CD1 conservation (**chapters 2 and 4**), the use of tetramers (**chapters 3, 4, and 5**), the repeated discovery of CD1 restricted autoreactivity (**chapters 3 and 5**), and lipid adjuvants (**chapter 6**).

TCR and CD1 conservation

For a long time, researchers have been interested in conservation of TCRs and T cell restriction by polymorphic MHC molecules, two seemingly conflicting phenomena. In 1974 it was reported that donor A's T cells, specific for a viral antigen, cannot recognize the same viral antigen in donor B¹. This donor restriction of the T cells was explained by the polymorphic nature of the major histocompatibility (MHC) molecules, and implied TCR repertoire variability among donors. Non-polymorphic antigen presenting molecules and T cells that recognize these molecules were discovered about 15 years later, marking the discovery of "donor unrestricted T cells". These T cells do not bind classical MHC molecules but monomorphic or nearly monomorphic molecules like CD1, MR1 or HLA-E. T cells recognizing a complex of one of these monomorphic molecules with an antigen will recognize the same complex in another donor or in vitro using another donor's antigen presenting cells. So, by definition, all T cells that recognize CD1, MR1, or HLA-E are donor unrestricted. Initially, when only a few T cell clones that recognize these nonpolymorphic molecules were known, no special observations were made regarding their TCR repertoire. And while many donor unrestricted T cells with unique or diverse TCRs exist², a special category of donor unrestricted T cells is formed by "invariant T cells". TCR sequences shared by many donors were discovered without knowing their specificity³, and when these shared and often simple TCRs were demonstrated to interact with non-polymorphic antigen-presenting molecules⁴⁻⁶, invariant T cells quickly became a field of intense scientific investigation.

So far, three invariant TCRs have been described. Invariant T cells have two key properties. First, intradonor conservation; expansion of T cells with many similar TCR sequences in one person and second, interdonor conservation; similar TCR sequence patterns among nearly all humans⁷. The three invariant T cells known so far are the MR1 binding MAIT cell, the CD1d binding NKT cell, and the CD1b binding GEM. They all have stringent conservation of the α chain ("invariance") and a biased TRBV use of the β chain. Invariant α chains use no or only a few N nucleotides, which explains their existence in every individual⁸.

We hypothesized that the conservation of CD1 molecules (**chapter 2**) might have led to coevolution between TCR segments and CD1 molecules. Such a coevolution is suggested for the TRAV1 segment of the invariant MAIT TCR and MR1⁹. While for CD1c binding T cells it had been shown that TCRs using TRBV4-1 are preferentially used¹⁰, our data

shows a V segment bias for CD1b recognition (**chapter 4**). Reanalyzing the data of DeWitt *et. al.* also shows an increased population of TRBV4-1⁺ cells in 4 out of 4 donors in the CD1b-GMM tetramer binding population compared to the non-binding population. Interestingly, in mice no CD1b and CD1c homologs are present¹¹ and they also lack a TRBV4-1 homolog¹². To confirm if there is coevolution between TRBV4-1 and CD1b or CD1c, genomes of many other species should be studied. However, such a study is not yet possible because the CD1 and TCR loci in genomes other than human and mouse have not been sufficiently assembled because of the high number of paralogous genes.

The use of tetramers

Many studies, including my own, on antigen presentation and T cells that recognize antigen presenting complexes are using tetramers. The first tetramers, HLA-A2 molecule loaded with the HIV derived peptides Gag(77-85) or reverse transcriptase(309-317)¹³ were used in 1996. Four years later the first CD1 tetramers were described^{14,15}.

Currently, tetramers of all CD1 isoforms and of all HLA types can be ordered from commercial suppliers or requested from non-profit organizations like the NIH tetramer facility. CD1a, CD1b, and CD1c tetramers are always provided unloaded and must be loaded by the end user with the ligand of interest. Of note, unloaded CD1 is by no means "empty". CD1 molecules are unstable without a bound lipid. During synthesis in the endoplasmic reticulum and transport via the Golgi complex, resident cellular lipids from the cell line in which the CD1 molecules are produced are loaded. When the end user tries to load a new putative lipid antigen, it is difficult to know if the new molecule is loaded if no positive control cell line with known reactivity pattern is available. Often the tetramer with uncertain loading percentage will be used to isolate T cells, and those T cells will be independently evaluated for antigen specificity in for example a cytokine release assay using CD1-expressing antigen presenting cells. Only if the T cells turn out to have the desired antigen specificity, the tetramer can be formally validated.

Although tetramers are easily accessible and frequently used, this tool still has some specific drawbacks. T cells that functionally react to an antigen when it is presented by an antigen-presenting cells can be missed by tetramer staining. This is caused by the fact that the affinity needed for a T cell to bind to a tetramer is higher than the affinity needed to be activated by the antigen presenting molecule¹⁶. Furthermore, reactive T cells can be missed because of TCR downregulation during active disease¹⁷. These problems may explain the fact that frequencies of antigen specific T cells are systematically underestimated in some clinical studies¹⁸. To overcome these problems, many strategies are used to enhance the tetramer staining. Cells are treated with protein kinase inhibitors to inhibit TCR down regulation¹⁹. Furthermore, antibodies against co-receptors²⁰ and antibodies against the tetramers²¹

have shown to improve tetramer staining. From our own experience we know that adding a specific α CD3 antibody (clone: OKT3, specific for the ϵ subunit) to a FACS panel can also enhance tetramer staining. Another way of increasing sensitivity is making higher level multimers like dextramers²² which have up to 13 streptavidin molecules to bind antigen presenting molecules. Dextramers labeled with both fluorochromes and DNA barcodes were used to study >1,000 T cell specificities in one sample²³. To generate tetramers for mass cytometry (CYTOF), dodecamers were developed. With these dodecamers Huang *et al* identified a fraction of T cells which was tetramer⁻ but dodecamer⁺²⁴, supporting the notion that a higher level of multimerization promotes the binding of antigen presenting molecules to T cells.

With these considerations in mind, the fact that we did not find T cells specific for CD1b-BbGL-II in Lyme disease patients (**chapter 3**) does not imply that these cells do not exist.

Autoreactivity

Many autoreactive T cells against CD1 molecules have been discovered by many independent research groups. Whereas previously described CD1-autoreactive T cells were mostly specific for CD1a, CD1c and CD1d²⁵⁻²⁸, in this thesis we describe autoreactivity for CD1b recognizing T cells. The main hypothesis that has been put forward to explain this ubiquitous autoreactivity in the CD1 system is that CD1 restricted T cells follow a very different selection course in the thymus²⁹. Also, their positively selecting antigens are likely self- lipids.

With so many autoreactive, self-lipid-specific T cells present, one would expect that some autoimmune diseases are lipid related. Although the mechanism is unknown, the autoimmune disease psoriasis is linked to hyperlipidemia³⁰. In mice that are hyperlipidemic due to the lack of the ApoE protein and that express human CD1 molecules, psoriasis-like skin inflammation was found. The inflammation was CD1b-dependent. Likewise, a higher number of CD1b autoreactive T cells was found in psoriasis patients compared to healthy controls³¹. This could indicate that CD1 restricted T cells are involved in psoriasis. In systemic lupus erythematosus 30% of the patients have signs of hyperlipidemia when diagnosed, which increases to 60% three years after diagnoses³². In systemic lupus erythematosus, protective and pathogenic rolls have been described for NKT cells. In mice, a reduction of NKT cells is related to increased disease progression³³. However, a study in humans shows that NKT can induce anti-dsDNA antibodies in a CD40/CD40L- and CD1d-dependent manner³⁴. Around 40% of the systemic lupus erythematosus patients also suffer from antiphospholipid syndrome. These patients have antibodies against common cell membrane phospholipids³⁵. It is known that CD1c restricted CD4⁺ CD8⁻ T cells can provide help to CD1c⁺ B cells to produce IgG antibodies³⁶ but it is less well understood whether and

how CD1a- and CD1b-restricted T cells contribute to systemic lupus erythematosus. For rheumatoid arthritis (RA) it is known that there is a lower frequency of NKT cells in the peripheral blood of patients, and these NKT are less well able to produce Th2 cytokines^{37,38}. The role of T cells restricted to the other CD1 molecules in RA is unknown.

These observations of ubiquitous autoreactive cells raise the question how tolerance for self-lipids is maintained. Currently, there is no data that suggest that negative selection for CD1 binding autoreactive T cells takes place, so there may be no central tolerance-generating mechanism. Probably, peripheral tolerance is somehow protecting against autoimmune diseases. Four mechanisms of peripheral regulation of CD1-specific autoreactive T cells have been proposed³⁹. 1) Regulation of CD1 levels on the cell surface of the antigen presenting cell (APC). *In vitro* differentiation of monocytes to dendritic cells with GM-CSF and interleukin (IL)-4 caused an upregulation of CD1a, CD1b, and CD1c on the cell surface^{40,41}, but the levels on resting monocytes are low. On monocytes CD1a, CD1b, and CD1c are upregulated after stimulation via Toll Like Receptor (TLR) 2 and IL-2 β ^{42,43}. 2) Enhanced activation of CD1-specific T cells by cytokines. Both *in vivo* and *in vitro*, IL-12 production by the APC is a main driver of interferon (IFN)- γ production by the NKT cell⁴⁴. 3) Regulation of the amount of lipid antigen. Increased binding of fluorescently labeled TCR to APCs was detected after TLR activation by lipopolysaccharide (LPS)⁴⁵. LPS treatment of APCs induced upregulation of glycosphingolipid synthesis, which caused increased T cell activation⁴⁶. 4) Spatial separation of antigen and presenting molecule. CD1a autoreactive T cells, CD1a expressing Langerhans cells, and CD1a-presented lipids are present in the human skin. However, under normal conditions there is no activation of these T cells, presumably because the APC, antigen, and the T cells are physically separated in different layers of the skin⁴⁷.

Despite the minimal insights into negative selection of CD1-specific T cells, positive selection and development of CD1d and CD1b restricted T cells in the thymus has been described in mice⁴⁸. Several studies aimed to determine which lipids are involved in the positive selection of NKT cells. It is suggested that ether lipids derived from the peroxisome are necessary for NKT cell selection and activation. Mice lacking an enzyme important for synthesis of these lipids had lower numbers of NKT cells⁴⁹. Also isoglobotrihexosylceramide (iGb3) was believed to be involved in thymic selection of NKT cells in mice⁵⁰. However, that was convincingly disputed in another study⁵¹. Also, humans lack the enzyme that generates iGb3, and iGb3 has not been found in humans⁵². A mouse model was used to study positive selection of the CD1b restricted T cell clone DN1. DN1 TCR-transgenic mice that express CD1b on their hematopoietic cells and not on their thymic epithelial cells have more DN1 cells than mice that express CD1b on their thymic epithelial cells and not on their hematopoietic cells. This suggests that DN1 cells are positively selected

by hematopoietic cells which is different from MHC binding T cells which are positively selected by thymic epithelial cells⁵³. These findings are consistent with what is known about NKT cell positive selection by hematopoietic cells.

Molecular mechanisms of autoreactivity

Trimolecular structures have demonstrated that there are various way autoreactive T cells can bind to the presenting molecule. The CD1a autoreactive TCR BK6 binds on the A' roof without interacting with the lipid. Nonpermissive ligands were able to alter the TCR CD1a interaction by disrupting residues in the A' roof⁵⁴. However, CD1b autoreactive T cells make direct contacts with the self-lipids they recognize. The PG90 TCR has a small cavity that fits small phospholipid headgroups, like PA and PG^{55,56}. Bigger headgroups are not recognized by PG90, presumably because they don't fit in the small cavity. We described in **chapter 5** the escape channel which allows the BC8B TCR to recognize multiple phospholipids, including ones with big headgroups. PG90 does not have the escape channel and the TCR footprint is surrounding the entrance of the binding cleft^{55,56}. The autoreactive TCR 3C8 that binds to CD1c, binds to the area surrounding the F' portal, covering the F' portal itself. This can only happen when the ligand is buried in the binding cleft. Therefore, 3C8 TCR can bind to CD1c loaded with any lipid which has a small headgroup so it does not interfere with the TCR binding surface²⁸. Since autoreactive T cells that recognize CD1a loaded with headless lipids are well known²⁶, this could be a more general binding method of CD1a and CD1c autoreactive T cells but further studies are necessary to confirm this. The binding of the autoreactive clones BK6 and 3C8 are reviewed by Cotton *et al.*⁵⁷.

Lipids as adjuvants

Besides the potential use of lipids as antigen in a subunit vaccine, lipids are currently used as adjuvants. In **chapter 6** we describe the discovery of diphosphatidyl trehalose (diPT). Despite being a chemical compound that has never been described in any organism, based on superficial structural analogies with trehalose dimycolate, we discovered that diPT that can active cells via the macrophage inducible C-type lectin (Mincle) receptor. Complete Freud's adjuvant, is water-in-oil emulsion mixed with heat killed *Mycobacteria*⁵⁸. These heat killed *Mycobacteria* contain many Mincle ligands, like TDM and GMM⁵⁹. Although Complete Freud's adjuvant is very effective, the side effects make it undesirable for use in human vaccines. Side effects caused by Complete Freud's adjuvant are granulomas, necrosis in surrounding tissue, and abscesses⁶⁰. The CAF01 adjuvant is an adjuvant based on trehalose 6,6'-dibehenate (TDB), which is a synthetic analog of TDM⁶¹. CAF01 is a save adjuvant for humans that elicits a Th₁ response⁶². Because two effective adjuvants are based on lipids that bind to the Mincle receptor, diPT might be also a good candidate as lipid component of an adjuvant. We predict that subtle differences in physical and biological properties between TDM, TDB, and diPT justify further development in this direction. For

example, even though these three lipids activate Mincle with comparable potency, other receptors may be differentially affected by the three lipids and generate different responses *in vivo*. Thus, the effects of diPT as a prospective candidate for being incorporated in an adjuvant should be determined in future *in vivo* and *in vitro* studies.

References

1. Zinkernagel, R. M. & Doherty, P. C. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702 (1974).
2. DeWitt, W. S. et al. A Diverse Lipid Antigen-Specific TCR Repertoire Is Clonally Expanded during Active Tuberculosis. *The Journal of Immunology* **180**, 1800186 (2018). doi:10.4049/jimmunol.1800186
3. Porcelli, S. et al. Recognition of cluster of differentiation 1 antigens by human CD4-CD8- cytolytic T lymphocyte. *Nature* **341**, 447–450 (1989).
4. Bendelac, A. et al. CD1 recognition by mouse NK1+ T lymphocytes. *Science* **268**, 863–865 (1995).
5. Tilloy, F. et al. An Invariant T Cell Receptor α Chain Defines a Novel TAP-independent Major Histocompatibility Complex Class Ib-restricted α/β T Cell Subpopulation in Mammals. *Journal of Experimental Medicine* **189**, 1907–1921 (1999).
6. Van Rhijn, I. et al. TCR Bias and Affinity Define Two Compartments of the CD1b-Glycolipid-Specific T Cell Repertoire. *J Immunol* **192**, 4054–4060 (2014).
7. Van Rhijn, I. & Moody, D. B. Donor Unrestricted T cells: a shared human T cell response. *J Immunol* **195**, 1927–1932 (2015).
8. Greenaway, H. Y. et al. NKT and MAIT invariant TCR α sequences can be produced efficiently by VJ gene recombination. *Immunobiology* **218**, 213–224 (2013).
9. Boudinot, P. et al. Restricting nonclassical MHC genes coevolve with TRAV genes used by innate-like T cells in mammals. *PNAS* **113**, E2983–E2992 (2016).
10. Guo, T. et al. A Subset of Human Autoreactive CD1c-Restricted T Cells Preferentially Expresses TRBV4-1+ TCRs. *The Journal of Immunology* **170**, 0677 (2017). doi:10.4049/jimmunol.1700677
11. Bradbury, A., Belt, K. T., Neri, T. M., Milstein, C. & Calabi, F. Mouse CD1 is distinct from and co-exists with TL in the same thymus. *EMBO J* **7**, 3081–3086 (1988).
12. Bosc, N. & Lefranc, M.-P. The Mouse (*Mus musculus*) T Cell Receptor Beta Variable (TRBV), Diversity (TRBD) and Joining (TRBJ) Genes. *ECI* **17**, 216–228 (2000).
13. Altman, J. D. et al. Phenotypic Analysis of Antigen-Specific T Lymphocytes. *Science* **274**, 94–96 (1996).
14. Benlagha, K., Weiss, A., Beavis, A., Teyton, L. & Bendelac, A. In Vivo Identification of Glycolipid Antigen-Specific T Cells Using Fluorescent Cd1d Tetramers. *Journal of Experimental Medicine* **191**, 1895–1904 (2000).
15. Matsuda, J. L. et al. Tracking the Response of Natural Killer T Cells to a Glycolipid Antigen Using Cd1d Tetramers. *Journal of Experimental Medicine* **192**, 741–754 (2000).
16. Laugel, B. et al. Different T Cell Receptor Affinity Thresholds and CD8 Coreceptor Dependence Govern Cytotoxic T Lymphocyte Activation and Tetramer Binding Properties. *J. Biol. Chem.* **282**, 23799–23810 (2007).
17. Gallegos, A. M. et al. Control of T cell antigen reactivity via programmed TCR downregulation. *Nature Immunology* **17**, 379–386 (2016).
18. Rius, C. et al. Peptide-MHC Class I Tetramers Can Fail To Detect Relevant Functional T Cell Clonotypes and Underestimate Antigen-Reactive T Cell Populations. *The Journal of Immunology* **200**, 2263–2279 (2018).
19. Lissina, A. et al. Protein kinase inhibitors substantially improve the physical detection of T-cells with peptide-MHC tetramers. *J Immunol Methods* **340**, 11–24 (2009).
20. Wooldridge, L. et al. Enhanced immunogenicity of CTL antigens through mutation of the CD8 binding MHC class I invariant region. *Eur J Immunol* **37**, 1323–1333 (2007).
21. Tungatt, K. et al. Antibody Stabilization of Peptide-MHC Multimers Reveals Functional T Cells Bearing Extremely Low-Affinity TCRs. *The Journal of Immunology* **194**, 463–474 (2015).
22. Dolton, G. et al. Comparison of peptide-major histocompatibility complex tetramers and dextramers for the identification of antigen-specific T cells. *Clin Exp Immunol* **177**, 47–63 (2014).
23. Bentzen, A. K. et al. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nature Biotechnology* **34**, 1037–1045 (2016).
24. Huang, J. et al. Detection, phenotyping, and quantification of antigen-specific T cells using a peptide-MHC dodecamer. *PNAS* **113**, E1890–E1897 (2016).
25. De Lalla, C. et al. High-frequency and adaptive-like dynamics of human CD1 self-reactive T cells. *European Journal of Immunology* **41**, 602–610 (2011).
26. de Jong, A. et al. CD1a-autoreactive T cells recognize natural skin oils that function as headless antigens. *Nature Immunology* **15**, 177–185 (2014).
27. Mallevaey, T. et al. A Molecular Basis for NKT Cell Recognition of CD1d-Self-Antigen. *Immunity* **34**, 315–326 (2011).
28. Wun, K. S. et al. T cell autoreactivity directed toward CD1c itself rather than toward carried self lipids. *Nature Immunology* **19**, 397–406 (2018).
29. Bendelac, A., Savage, P. B. & Teyton, L. The Biology of NKT Cells. *Annu. Rev. Immunol.* **25**, 297–336 (2007).
30. Han, C. et al. Cardiovascular disease and risk factors in patients with rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. *The Journal of Rheumatology* **33**, 2167–2172 (2006).
31. Bagchi, S. et al. CD1b-autoreactive T cells contribute to hyperlipidemia-induced skin inflammation in mice. *J Clin Invest* **127**, 2339–2352 (2017).
32. Tselios, K., Koumaras, C., Gladman, D. D. & Urowitz, M. B. Dyslipidemia in systemic lupus erythematosus: just another comorbidity? *Seminars in Arthritis and Rheumatism* **45**, 604–610 (2016).
33. Oishi, Y. et al. Selective reduction and recovery of invariant Valpha24AlphaQ T cell receptor T cells in correlation with disease activity in patients with systemic lupus erythematosus. *The Journal of Rheumatology* **28**, 275–283 (2001).
34. Shen, L. et al. Invariant natural killer T cells in lupus patients promote IgG and IgG autoantibody production. *Eur. J. Immunol.* **45**, 612–623 (2015).
35. Pons-Estel, G. J., Andreoli, L., Scanzi, F., Cervera, R. & Tincani, A. The antiphospholipid syndrome in patients with systemic lupus erythematosus. *Journal of Autoimmunity* **76**, 10–20 (2017).
36. Sieling, P. A. et al. Human double-negative T cells in systemic lupus erythematosus provide help for IgG and are restricted by CD1c. *J. Immunol.* **165**, 5338–5344 (2000).
37. Linsen, L. et al. Peripheral blood but not synovial fluid natural killer T cells are biased towards a Th1-like phenotype in rheumatoid arthritis. *Arthritis Res Ther* **7**, R493–R502 (2005).
38. Yanagihara, Y., Shiozawa, K., Takai, M., Kyogoku, M. & Shiozawa, S. Natural killer (NK) T cells are significantly decreased in the peripheral blood of patients with rheumatoid arthritis (RA). *Clin Exp Immunol* **118**, 131–136 (1999).
39. De Jong, A. Activation of human T cells by CD1 and self-lipids. *Immunological Reviews* **267**, 16–29 (2015).
40. Pickl, W. F. et al. Molecular and functional characteristics of dendritic cells generated from highly purified CD14+ peripheral blood monocytes. *The Journal of Immunology* **157**, 3850–3859 (1996).
41. Zhou, L. J. & Tedder, T. F. CD14+ blood monocytes can differentiate into functionally mature CD83+ dendritic cells. *Proc Natl Acad Sci U S A* **93**, 2588–2592 (1996).

42. Roura-Mir, C. *et al.* Mycobacterium tuberculosis Regulates CD1 Antigen Presentation Pathways through TLR-2. *The Journal of Immunology* **175**, 1758–1766 (2005).
43. Yakimchuk, K. *et al.* Borrelia burgdorferi infection regulates CD1 expression in human cells and tissues via IL1- β . *Eur. J. Immunol.* **41**, 694–705 (2011).
44. Brigl, M. *et al.* Innate and cytokine-driven signals, rather than microbial antigens, dominate in natural killer T cell activation during microbial infection. *Journal of Experimental Medicine* **208**, 1163–1177 (2011).
45. Salio, M. *et al.* Modulation of human natural killer T cell ligands on TLR-mediated antigen-presenting cell activation. *Proc Natl Acad Sci U S A* **104**, 20490–20495 (2007).
46. De Libero, G. *et al.* Bacterial Infections Promote T Cell Recognition of Self-Glycolipids. *Immunity* **22**, 763–772 (2005).
47. De Jong, A. *et al.* CD1a-autoreactive T cells are a normal component of the human $\alpha\beta$ T cell repertoire. *Nature Immunology* **11**, 1102–1109 (2010).
48. Benlagha, K., Wei, D. G., Veiga, J., Teyton, L. & Bendelac, A. Characterization of the early stages of thymic NKT cell development. *Journal of Experimental Medicine* **202**, 485–492 (2005).
49. Facciotti, F. *et al.* Peroxisome-derived lipids are self antigens that stimulate invariant natural killer T cells in the thymus. *Nature Immunology* **13**, 474–480 (2012).
50. Zhou, D. *et al.* Lysosomal Glycosphingolipid Recognition by NKT Cells. *Science* **306**, 1786–1789 (2004).
51. Porubsky, S. *et al.* Normal development and function of invariant natural killer T cells in mice with isoglobotrihexosylceramide (iGb3) deficiency. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5977–5982 (2007).
52. Christiansen, D. *et al.* Humans Lack iGb3 Due to the Absence of Functional iGb3-Synthase: Implications for NKT Cell Development and Transplantation. *PLOS Biology* **6**, e172 (2008).
53. Zhao, J. *et al.* Mycolic acid-specific T cells protect against Mycobacterium tuberculosis infection in a humanized transgenic mouse model. *eLife Sciences* **4**, e08525 (2015).
54. Birkinshaw, R. W. *et al.* $\alpha\beta$ T cell antigen receptor recognition of CD1a presenting self lipid ligands. *Nature Immunology* **16**, 258–266 (2015).
55. Shahine, A. *et al.* A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Science Immunology* **2**, eaao1384 (2017).
56. Van Rhijn, I. *et al.* Human autoreactive T cells recognize CD1b and phospholipids. *PNAS* **113**, 380–385 (2016).
57. Cotton, R. N., Shahine, A., Rossjohn, J. & Moody, D. B. Lipids hide or step aside for CD1-autoreactive T cell receptors. *Current Opinion in Immunology* **52**, 93–99 (2018).
58. Opie, E. L. & Freund, J. An experimental study of protective inoculation with heat killed tubercle bacilli. *J Exp Med* **66**, 761–788 (1937).
59. Decout, A. *et al.* Rational design of adjuvants targeting the C-type lectin MinCLE. *PNAS* **114**, 2675–2680 (2017).
60. Apostólico, J. de S., Lunardelli, V. A. S., Coirada, F. C., Boscardin, S. B. & Rosa, D. S. Adjuvants: Classification, Modus Operandi, and Licensing. *Journal of Immunology Research* (2016). doi:10.1155/2016/1459394
61. Davidsen, J. *et al.* Characterization of cationic liposomes based on dimethyldioctadecylammonium and synthetic cord factor from M. tuberculosis (trehalose 6,6'-dibehenate)—A novel adjuvant inducing both strong CMI and antibody responses. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1718**, 22–31 (2005).
62. van Dissel, J. T. *et al.* A novel liposomal adjuvant system, CAF01, promotes long-lived Mycobacterium tuberculosis-specific T-cell responses in human. *Vaccine* **32**, 7098–7107 (2014).



Appendix

Nederlandse samenvatting

Dankwoord

Curriculum vitae

List of publications

.....

Nederlandse samenvatting

Doel van deze thesis

Is het mogelijk om lipiden (vetachtige stoffen) van ziekteverwekkende bacteriën te gebruiken in vaccins?

Achtergrond

Ons lichaam wordt elke dag blootgesteld aan ziektekiemen zoals virussen en bacteriën. Gelukkig is ons lichaam uitgerust met een geavanceerd immuunsysteem. Dit systeem bestaat uit verschillende soorten cellen die samenwerken om het lichaam te beschermen tegen infecties en daarmee ziekte. Ziektekiemen kunnen zich op twee manieren in ons lichaam schuilhouden. Sommige ziektekiemen bevinden zich buiten de cellen van ons lichaam, zoals een bacterie die in het bloed of op de huid leeft, en andere verschuilen zich in de lichaamscellen. Voor beide heeft het lichaam een tactiek.

Elke cel laat namelijk continu aan het immuunsysteem weten hoe het met hem gaat. Dit gebeurt door het presenteren van stukjes van eiwitten, zogeheten peptiden. Gespecialiseerde immuuncellen met een signaalfunctie (T-cellen) herkennen de gepresenteerde peptiden.

Eiwitten zijn de “werkers” van onze cellen. Zowel onze eigen cellen als ook bacteriën en virussen hebben hun eigen eiwitten. Mocht een cel in contact komen of geïnfecteerd raken met een bacterie of een virus dan zullen er ook peptiden van deze indringer aan de T-cellen gepresenteerd worden. De presentatie wordt gedaan door een speciaal eiwit, het MHC-molecuul. Dit molecuul kan een peptide binnen in de cel binden en dan naar de buitenkant van de cel gaan om het te presenteren aan voorbijkomende T-cellen. De T-cellen proberen vervolgens om dit MHC-molecuul mét peptide te herkennen. Als dit gebeurt, dan wordt het MHC-molecuul gebonden aan de T-cel door middel van een receptor op de T-cel (de T-cel receptor, ook een eiwit). De T-cellen die het peptide van de indringer op het celoppervlak vinden kunnen op twee manieren reageren, de cel die het peptide presenteert kan met indringer en al opgeruimd worden of hij kan andere cellen oproepen om te helpen met het opruimen van de indringer.

Naast het systeem om peptiden te presenteren aan T-cellen, hebben cellen ook een manier om vetachtige stoffen, zogenoemde lipiden, te presenteren aan T-cellen. Dit kunnen cel-eigen lipiden zijn maar ook lipiden van ziektekiemen. Het presenteren gebeurt in dit geval door CD1-moleculen. De mens heeft vier verschillende CD1-moleculen (CD1a, CD1b,

CD1c, CD1d) die verschillende soorten lipiden kunnen presenteren. Het aantal CD1-moleculen verschilt per diersoort zo heeft de muis twee verschillende CD1d-moleculen waar de kleine vleermuis (*Microchiroptera*) in totaal 26 CD1-moleculen heeft (CD1a: 17, CD1b: 2, CD1c: 0, CD1d: 5, CD1e: 2). Dit verschil in aantallen CD1-moleculen zou kunnen komen door het verschil in ziektekiemen die diersoorten gedurende de evolutie zijn tegen gekomen. Dit proefschrift richt zich voornamelijk op het CD1b-molecuul.

Onderzoek

Als we kijken naar welke lipiden er gebonden kunnen worden door het menselijk CD1b-molecuul, zien we dat dat naast lichaamsvreemde ook lichaamseigen lipiden zijn. Het herkennen van lichaamseigen lipiden en peptiden kan gevaarlijk zijn omdat dit ervoor kan zorgen dat het immuunsysteem zich keert tegen het eigen lichaam (auto-immuunziekte). In mijn onderzoek hebben we ontdekt dat lichaamseigen lipiden die door CD1b worden gebonden, en herkend worden door de T-cellen verrassend veel voorkomende lipiden zijn. Deze lipiden maken bijvoorbeeld onderdeel uit van het celmembraan, de wand van elke cel. Maar ook lichaamseigen lipiden die zich in de cel begeven kunnen door CD1b gepresenteerd worden aan het immuunsysteem. Dit kunnen bijvoorbeeld lipiden zijn die signalen binnen in de cel doorgeven.

De T-cel receptor bevindt zich op het oppervlak van een T-cel en is uniek is voor elke T-cel. Van al het onderzoek dat al is gedaan naar T-cellen die peptiden herkennen die gebonden zijn aan MHC-moleculen, weten we dat het niet mogelijk is om te voorspellen wat een T-cel receptor herkent. Er zijn gewoonweg zoveel T-cel receptoren dat het nog niet in kaart is gebracht en we nu dus niet van tevoren kunnen zien wat hij herkent. Verder is duidelijk dat een T-cel receptor maar één peptide in combinatie met een MHC-molecuul kan herkennen.

Uit ons onderzoek is gebleken dat dit anders ligt voor T-cel receptoren die CD1-moleculen binden. We hebben een mechanisme gevonden dat één T-cel receptor in staat stelt om CD1b te herkennen als er verschillende lichaamseigen lipiden aan gebonden zijn. Tevens hebben we gevonden dat één en dezelfde T-cel receptor zowel een lichaamseigen als een lichaamsvreemd lipide herkent als het gebonden is aan CD1b. Verder laten we zien dat T-cel receptoren die een bepaald eiwitsegment hebben, het zogenoemde TRBV4-1 segment, vaker aan CD1b binden dan T-cel receptoren met andere segmenten. Daardoor lijkt het erop dat bepaalde T-cellen toch een voorkeur hebben om bepaalde moleculen te herkennen.

Naast dat we hebben gekeken naar lipiden die aan het CD1b molecuul binden, hebben we ook gekeken naar andere manieren waarop lipiden het immuunsysteem beïnvloeden.

Twee tot nu toe onbekende lipiden, genaamd PT en diPT, zijn door ons geïdentificeerd. Deze lipiden zijn eerst gevonden in de ziekteverwekkende *Salmonella* bacteriën. Verder zijn deze lipiden ook aanwezig in bepaalde stammen van de *E. coli* bacterie. We hebben niet kunnen aantonen dat PT of diPT aan CD1b bindt, maar wel dat het aan een andere receptor bindt: de Mincle-receptor. Deze Mincle-receptor is aanwezig op professionele antigeen-presenteerde cellen. Als er een lipide bindt aan deze receptor dan beginnen deze cellen hulpstoffen uit te scheiden die het immuunsysteem extra activeren. Zo'n lipide dat het immuunsysteem extra stimuleert kan ook gebruikt worden in vaccins, waarbij het toevoegen van zo'n molecuul kan leiden tot een verbeterde bescherming.

Conclusie

Al het werk beschreven in deze thesis is gedaan als onderzoek naar de mogelijkheid om lipiden van ziekteverwekkende bacteriën te gebruiken in vaccins. Maar dit onderzoek laat ook zien dat lipiden afkomstig van een ziekteverwekker herkend kunnen worden door T-cellen die ook een lichaamseigen lipide herkennen. Als deze geactiveerd worden door vaccinatie, zou ook het onschuldige lichaamseigen lipide herkend kunnen worden wat dan eventueel zou kunnen leiden tot een autoimmuunziekte. Deze kruisherkenning is door ons aangetoond. Dit is natuurlijk niet wenselijk als het gaat om vaccins. Verder hebben we een nieuw lipide geïdentificeerd dat mogelijk als hulpstof kan dienen in een vaccin. Dit lipide is in relatief grote hoeveelheden aanwezig in een bekende ziekteverwekkende bacterie en was nog nooit eerder beschreven, wat suggereert dat het waarschijnlijk is dat er nog meer immuunstimulerende lipiden gevonden zouden kunnen worden in andere ziekteverwekkers.

Dankwoord

Aan het eind van dit project wil ik graag van de gelegenheid gebruik maken voor een aantal mensen te bedanken.

Allereerst **Ildiko**, bedankt dat je mij de mogelijk hebt gegeven om dit onderzoek onder jouw begeleiding uit te voeren. Ik heb erg genoten van jouw kennis en kunde. Ik bewonder hoe gemakkelijk jij tussen verschillende projecten kan schuiven. Van bindinginteracties tussen CD1 en de T-cel receptor naar het voor ons beide nieuwe bacteriologische werk. Verder heb ik mooie herinneringen aan de diverse wandelingen in de White Mountains met jou en Scott.

Willem, bedankt dat ik weer terug kon komen op de afdeling. Leuk dat je tijdens je vakantie in Boston tijd had voor een lunch.

Branch, thank you for allowing me to work in your lab. It was a great pleasure to work in such an inspiring environment. Not only did I learn from you how to think about research and conduct science, also my presentation skills improved a lot during my stay, thanks to your clear feedback.

Christine en **Can** bedankt dat jullie mijn AIO-begeleidingscommissie wilde vormen. Gelukkig heb ik jullie niet nodig gehad maar het is fijn om te weten dat er mensen zijn om je te helpen indien nodig.

Het merendeel van mijn tijd zat ik bij de immuno's van de UU, waar ik ook veel mensen wil bedanken. Als eerst mijn paranympen; **Steffi**, 2 jaar heb je mij geholpen bij de zoektocht naar T-cellen die CD1 konden binden. Soms leek het wel of ik in een lab werkte waar alles vanzelf werd aangevuld. Was de voorraad half op, dan kwam jij alweer met nieuwe voorraad aanlopen. Ik vind het indrukwekkend hoe makkelijk jij de Nederlandse taal leerde en het viel mij af en toe op dat ik (uit gewoonte) de enige was die nog Engels met je sprak. **Manon**, we begonnen tegelijk en we trokken tegelijk de deur achter ons dicht. Hoewel onze projecten erg verschillend waren kon je altijd wel meedenken als ik even vastliep.

Verder de ex-koffieclub **Anouk, Marit en Suus**, het was altijd fijn om even frisse lucht te happen en over wat anders te praten dan wetenschap. **Irene**, nu is het aan jou om de afdeling te vertegenwoordigen bij The Village. **Charlotte**, hoe vaak ik mij hoofd wel niet om het hoekje heb gestoken voor een vraagje... Bedankt voor alle antwoorden. The “new” AIOs **Jurriaan, Robin, Andreja, Nathalie, and Qingkang** good luck with finishing your thesis's. **Peter, Aad, en Daphne**, wat een fijne vraagbaken zijn jullie toch.

My “home away from home” was **the Moody lab**. Many thanks to **all the lab members** for their help. In particular I would like to thank two people; **Jeffrey**, wat was het fijn om samen met jou 3 maanden te knallen op het salmonella project. Als je ooit nog onderzoek gaat doen naar waterbeertjes, laat het me weten! **Tan-Yun**, thanks for all your help with the Mass spec, I am still impressed by your ability to see a mass and immediately see what fatty acid it possible could be.

All **the members of the Brenner lab**, thanks for answering all my questions. **Gerald**, thanks for all the nice sailing trips on the Charles river and watching movies on Sunday morning.

The work described in this thesis could not have been performed with a lot of help from **many collaborators**. In special but not limited to, “The Australians”; the **Rossjohn lab** and the **Godfried lab**. Our Lyme collaborators at the **RIVM, Diakonessen hospital, and Mass General Hospital**. The chemists from the **Minnaard lab** and **Peter Willemsen**, thanks for the endless supply of bacteria.

Het is altijd goed om te ontspannen naast het werk, hoe kan dat nou beter dan zeilend met de mannen van de **Houten Posse** of met een lunch met de extendend posse. De **I&X** in PAIX; onze vriendschap begon ongeveer tegelijk met mijn AIO project. In deze tijd is een hechte vriendschap ontstaan, met als een van de hoogtepunten toch wel jullie visite in Boston. **Jorrit en Koen**, onze “biertjes doen” die vaak eindigen in avonden vol sterke verhalen hielpen zeker ook mee aan de ontspanning. Stoom afblazen kon ik altijd op het foorballveld, zowel bij de grootveldteams van de afgelopen jaren als bij mijn mixedteammies van **Trukes Tijgers / Sanders Salamanders**.

Mijn schoonfamilie; **Erna**, bedankt voor alle steun en de leuke roadtrip (in die koude tipi). **Harm, Sylvie, Klaasje, Hillie**, niets werkt zo goed om tot rust te komen als het Drentse platteland. **Bart, Laurent en Charlotte** bedankt voor jullie steun en interesse.

Margreet, Maarten, Inge, en Marc-Jan, ook al begrepen jullie in het begin volgens mij niet veel van wat ik met vetdeeltjes van bacteriën moest, jullie waren altijd vol interesse. Bedankt daarvoor.

Papa en Mama, bedankt voor jullie onvoorwaardelijke steun bij alles wat ik doe. Ik weet dat het even schrikken was toen ik vertelde dat ik een jaar naar de VS wilde gaan. Maar jullie steunden me omdat jullie zagen dat dit voor mij een enorme kans was, en maakten er gelijk gebruik van om zelf een keer de VS te bezoeken.

Lieve lieve **Aniek**, je wist waar je aan begon toen wij wat kregen. Ik had net ja gezegd tegen dit project en het jaar naar Boston hing boven ons hoofd. Sommige mensen zien een IKEA-kast als relatietest. Wij deden de overtreffende trap; jij besloot alles hier achter te laten en met mij mee te gaan naar de VS. Ik heb in jou mijn grootste fan én criticus gevonden. Jij houdt me scherp. Bedankt voor alles. Nu op naar nieuwe avonturen met **Meike**.

Curriculum Vitae

Peter Reinink was born on September 22nd, 1985 in Nieuwegein. In 2002 he started at the University of Applied Sciences in Utrecht, studying Informatics. After a year he switched to studying molecular biology at the University of Applied Sciences in Utrecht and he obtained his bachelor degree in 2007. Hereafter, he worked as a technician in the group of Dr. Ildiko van Rhijn at the division of Immunology at the Faculty of Veterinary Medicine at Utrecht University. In 2009 Peter started with his master's in bioinformatics at the VU University in Amsterdam. During this master he did two internships, the first one was under supervision of Jorg Calis at the Theoretical Biology group at the Faculty of Science, Utrecht University. For his second internship he went to Germany, where he worked at Jacobs University in Bremen under supervision of Prof. Dr. Sebastian Springer. Peter graduated *cum laude* for his bioinformatics master in 2011. That same year he started working as a bioinformatician/technician in the group of Dr. Menno Creyghton at the Hubrecht Institute in Utrecht. In 2014 he returned to the division of Immunology of the Faculty of Veterinary Medicine at Utrecht University as a PhD candidate. The candidacy was under supervision of Dr. Ildiko van Rhijn. During his PhD project he visited the lab of Prof. Dr. D. Branch Moody in the department of Rheumatology, Immunology and Allergy at the Brigham and Women's Hospital in Boston, in the United States of America twice. In 2016 Peter was part of the organizing committee of the Veterinary Science Day, a congress for employees of the Faculty of Veterinary Medicine of Utrecht University. Since February 2019 Peter works in the discovery department at Gadeta.

List of publications

Peer reviewed

Part of this thesis

* Equally contributed

#CD1b presents self and *Borrelia burgdorferi* diacylglycerols to human T cells

Reinink, P., M. N. T. Souter, T.-Y. Cheng, T. Van Gorkom, S. Lenz, J. Kubler-Kielb, K. Strle, K. Kremer, S. F. T. Thijsen, A. C. Steere, D. I. Godfrey, D. G. Pellicci, D. B. Moody, and I. Van Rhijn. 2019. *European Journal of Immunology* (In Press).

#Discovery of Salmonella trehalose phospholipids reveals functional convergence with mycobacteria

Reinink, P.*, J. Buter *, V. K. Mishra, E. Ishikawa, T.-Y. Cheng, P. T. J. Willemsen, S. Porwollik, P. J. Brennan, E. Heinz, J. A. Mayfield, G. Dougan, C. A. van Els, V. Cerundolo, G. Napolitani, S. Yamasaki, A. J. Minnaard, M. McClelland, D. B. Moody, and I. Van Rhijn. 2019. *Journal of Experimental Medicine* jem.20181812.

#A T-cell receptor escape channel allows broad T-cell response to CD1b and membrane phospholipids

Shahine, A., **P. Reinink**, J. F. Reijneveld, S. Gras, M. Holzheimer, T.-Y. Cheng, A. J. Minnaard, J. D. Altman, S. Lenz, J. Prandi, J. Kubler-Kielb, D. B. Moody, J. Rossjohn, and I. Van Rhijn. 2019. *Nature Communications* 10: 56.

CD1b Tetramers Identify T Cells that Recognize Natural and Synthetic Diacylated Sulfoglycolipids from *Mycobacterium tuberculosis*

James, C. A., K. K. Q. Yu, M. Gilleron, J. Prandi, V. R. Yedulla, Z. Z. Moleda, E. Diamanti, M. Khan, V. K. Aggarwal, J. F. Reijneveld, **P. Reinink**, S. Lenz, R. O. Emerson, T. J. Scriba, M. N. T. Souter, D. I. Godfrey, D. G. Pellicci, D. B. Moody, A. J. Minnaard, C. Seshadri, and I. Van Rhijn. 2018. *Cell Chemical Biology* 25: 392-402.e14.

Epigenomic annotation of gene regulatory alterations during evolution of the primate brain

Vermunt, M. W., S. C. Tan, B. Castelijn, G. Geeven, **P. Reinink**, E. de Bruijn, I. Kondova, S. Persengiev, N. B. Bank, R. Bontrop, E. Cuppen, W. de Laat, and M. P. Creighton. 2016. *Nature Neuroscience* 19: 494-503.

#Mammalian CD1 and MR1 genes

Reinink, P., and I. Van Rhijn. 2016. *Immunogenetics* 68: 515-523.

Expression Patterns of Bovine CD1 In Vivo and Assessment of the Specificities of the Anti-Bovine CD1 Antibodies

Nguyen, T. K. A., **P. Reinink**, C. E. Messlaki, J. S. Im, A. Ercan, S. A. Porcelli, and I. Van Rhijn. 2015. *PLOS ONE* 10: e0121923.

F pocket flexibility influences the tapasin dependence of two differentially disease-associated MHC Class I proteins.

Abualrous, E. T., S. Fritzsche, Z. Hein, M. S. Al-Balushi, **P. Reinink**, L. H. Boyle, U. Wellbrock, A. N. Antoniou, and S. Springer. *European Journal of Immunology* 45: 1248-1257.

Role of peptide processing predictions in T cell epitope identification: contribution of different prediction programs

Calis, J. J. A., **P. Reinink**, C. Keller, P. M. Kloetzel, and C. Keşmir. 2014. *Immunogenetics* 1-9.

Large-Scale Identification of Coregulated Enhancer Networks in the Adult Human Brain

Vermunt, M. W. *, **P. Reinink** *, J. Korving, E. de Bruijn, P. M. Creighton, O. Basak, G. Geeven, P. W. Toonen, N. Lansu, C. Meunier, S. van Heesch, H. Clevers, W. de Laat, E. Cuppen, and M. P. Creighton. 2014. *Cell Reports* 9: 767-779.

Severe combined immunodeficiency in Frisian Water Dogs caused by a RAG1 mutation

Verfuurden, B., F. Wempe, **P. Reinink**, P. J. S. van Kooten, E. Martens, R. Gerritsen, J. H. Vos, V. P. M. G. Rutten, and P. A. Leegwater. 2011. *Genes Immun* 12: 310-313.

Conservation of mucosal associated invariant T (MAIT) cells and the MR1 restriction element in ruminants, and abundance of MAIT cells in spleen

Goldfinch, N., **P. Reinink**, T. Connelley, A. Koets, I. Morrison, and I. Van Rhijn. 2010. *Veterinary Research* 41: 62.

Functional CD1d and/or NKT cell invariant chain transcript in horse, pig, African elephant and guinea pig, but not in ruminants

Loringh van Beeck, F. A. *, **P. Reinink** *, R. Hermsen *, D. M. Zajonc, M. J. Laven, A. Fun, M. Troskie, N. J. Schoemaker, D. Morar, J. A. Lenstra, L. Vervelde, V. P. M. G. Rutten, W. van Eden, and I. Van Rhijn. 2009. *Mol Immunol* 46: 1424-1431.

The bovine T cell receptor alpha/delta locus contains over 400 V genes and encodes V genes without CDR2

Reinink, P., and I. Van Rhijn. 2009. *Immunogenetics* 61: 541-549.

Not Peer reviewed

In Silico method for identification of MHC class I-like molecules in whole genomes

Reinink, P., and I. Van Rhijn. 2016. *bioRxiv*.