THE UNIVERSITY OF CHICAGO

MOLECULAR AND CELLULAR DIVERSITY OF THE MR1-MR1T CELL AXIS IN ANTIMICROBIAL IMMUNITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

NICOLE ANNA LADD

CHICAGO, ILLINOIS

JUNE 2023

To my parents, for everything, but especially for allowing me to be curious.

Copyright© 2023 by Nicole Anna Ladd

All Rights Reserved

Freely available under a CC-BY 4.0 License

List of figures vii

List of tables xi

List of important abbreviations and definitions xii

Acknowledgements xiii

Abstract xvi

Chapter 1. Introduction 1

1.1 Molecular mechanisms of the T cell response to microbial infection 1

1.1.1 MR1-restricted T cells lie at the interface of the adaptive and innate immune systems 1

1.1.2 Conventional T cells recognize HLA molecules in complex with pathogen-derived peptides 2

1.1.3 T cells use sequence-variable T cell receptors to communicate with host cells 5

1.1.4 The various immune programs of αβ T cells 8

1.2 MR1 presents microbial small molecules to MR1-restricted T cells in a manner analogous to conventional MHC presentation of peptides 9

1.2.1 MR1 is a monomorphic MHC-related molecule that restricts MAIT cells 9

1.2.2 The riboflavin biosynthetic pathway is the origin of microbial MR1 ligands 11

1.2.3 Microbial MR1 ligands independent of the RBP may exist but are yet to be identified 17

1.3 MAIT cells are a subset of MR1-restricted T cells 19

1.3.1 MAIT cells use a semi-invariant TCR to recognize agMR1 19

1.3.2 Phenotypic properties of MAIT cells 22

1.3.3 MR1 tetramers have revealed heterogeneity in the MR1T cell compartment and allude to MR1T antigen selectivity 24

1.4 The potential for and progress toward MR1/MR1T-mediated therapies or vaccines against Mycobacterium tuberculosis 26

1.4.1 Eradication of tuberculosis will require more robust, accessible, cost-effective vaccines 26

1.4.2 Determinants of MR1-MR1T cell-mediated protective immunity toward M. tuberculosis infection remain unclear 27

1.5 Significance 28

Chapter 2. Metabolomic analysis of MR1 ligands reveals a class of indolyl-ribityllumazines and RBP-independent ligands 30

2.1 Introduction 30

2.2 Recombinant MR1 made in eukaryotic cells captures heterogeneous bacterial ligands 35

2.3 Metabolomic analysis recombinant MR1 loaded with microbial antigens reveals several ions of interest 40

2.4 The ribityllumazine profile of bacterially-loaded hpMR1 is distinct to the bacterium 45

2.5 Knockouts of the enzymes of the riboflavin biosynthetic pathway reveal the absolute dependence of MAIT cell agonism on Rib2ab but not on Rib1ab 48

2.6 PLIII, PLV, PLVI, and PLIX are likely 7-oxo-6-indolyl-8-D-ribityllumazines 56

2.7 PLX and PLVII are likely indolyl-ribityllumazines with unique chemical properties 60

2.8 Validation of PLV as an MR1 ligand and an agonist of the D481C7 MAIT clone 63

2.9 Isolation of select photolumazines from bacterial supernatants for structural and antigenic characterization 65

2.10 Discussion 70

2.11 Supplemental information 76

Chapter 3: MAIT and MR1T cell selectivity: A case study of DMRL and 5-deaza-DMRL 91

3.1 Introduction 91

3.2 Synthesis of deazalumazine, 2'-deoxy-deazalumazine, monomethyl deazalumazine, and deazaphotolumazine I 94

3.3 DZ, dDZ, mmDZ, and PLI stabilize MR1 on the cell surface 97

3.4 MR1T cell clones variably discriminate between DZ and DMRL 98

3.5 Impact of chemical modifications to DZ and PLI on activation of the D481C7 MAIT cell clone 100

3.6 MAIT cell recognition of DZ is MR1-dependent 102

3.7 DMRL, DZ, and 2'-deoxy-DZ exhibit differential dynamics in the MR1 binding pocket 103

3.8 Discussion 109

3.9 Supplemental information 114

Chapter 4. Pilot studies of single cell multi-omics of diverse MR1T cells from the peripheral blood of healthy human donors 121

4.1 Introduction 121

4.2 Single cell RNA sequencing and multi-omics of MR1T cells from the peripheral blood of two healthy human donors 122

4.3 Barcoded tetramers enable classification of antigen/tetramer specificity 127

4.4 A hyperexpanded cross-reactive MAIT clone exhibits hallmarks of degranulation and contrasts with the relative homogeneity of the remaining cross-reactive clonotypes 132

4.5 Tetramer reactivity reveals a more TCRα-diverse population of cross-reactive MR1T cells 135

4.6 The transcriptional and cell surface phenotype of non-MAIT, cross-reactive MR1T cells 138

4.7 The phenotype and of a small population of putative 5-OP-RU-agnostic MR1Ts 141

4.8 Discussion and ongoing work 143

4.9 Supplemental information 145

Chapter 5. Conclusions and future directions 154

5.1 Perspectives on the MR1 ligand repertoire and the dependence of MAIT cells on the riboflavin biosynthesis pathway (Relating Chapter 2) 154

5.2 MR1T and MAIT cell ligand selectivity suggests that there is more plasticity underlying ligand recognition than once appreciated (Relating Chapter 3) 156

5.3 Single cell multi-omics investigations of MR1-restricted T cells reveal the potential for greater TCR diversity and 5-OP-RU-agnostic clonotypes (Relating Chapter 4) 158

Chapter 6. Materials and methods 161

6.1 Materials and methods used in Chapter 2 161

6.2 Materials and methods used in Chapter 3 177

6.3 Materials and methods used in Chapter 4 187

Bibliography 190

List of figures

Figure 1.1 Properties and ligands of conventional αβ T cells 3

Figure 1.2 A structural perspective of the pMHC-Ia:TCR complex 5

Figure 1.3 V(D)J recombination is responsible for the diversity of TCRs and the CDR3 loops 7

Figure 1.4 Properties and ligands of some unconventional αβ T cells 10

Figure 1.5 The riboflavin biosynthetic pathway 12

Figure 1.6 Known MR1 ligands result from the condensation of 5-A-RU with small diketones, some of which can form a covalent bond with MR1 15

Figure 1.7 Additional MR1 ligands 16

Figure 1.8 MR1T cells recognize agMR1 using a "conserved" tyrosine in its CDR3α loop 20

Figure 2.1 The "co-infection" protocol yields MR1 loaded with microbial antigens which can be used for ligand and T cell characterization. 33

Figure 2.2 Validation of endogenous loading of MR1 in a mammalian expression system 37

Figure 2.3 Characterization of *M. smegmatis*-loaded MR1 derived from the mammalian expression system 39

Figure 2.4 Two parallel methods for the prioritization of ions eluted from the MR1 binding groove 44

Figure 2.5 The relative and absolute abundance of the ribityllumazines found in the MR1 binding groove is dependent on the bacterium added into the co-infection system 46

Figure 2.6 The activation of MAIT clones by *M. smegmatis* riboflavin biosynthesis pathway knockout strains 50

Figure 2.7 Functional properties and ribityllumazine profile of MR1 when loaded with *M. smegmatis* riboflavin biosynthesis pathway knockout strains 53

Figure 2.8 Annotation of the PLV MS2 spectrum and putative structures of PLV, PLVI, and PLIX 57

Figure 2.9 Properties of PLX 61

Figure 2.10 Validation of PLV as an MR1 ligand and MAIT antigen 65

Figure 2.11 Isolation of PLIII, PLV, PLVI, and PLIX from the supernatant of *E. coli* overexpressing *Rib1ab* and *Rib2ab* from *B. subtilis* 69

Figure 2.12 Summary of the structural and functional properties of the photolumazines 72

Supplemental Figure S2.1 Overview of the methods for loading MR1 with ligand 76

Supplemental Figure S2.2 MR1 construct and purification scheme 77

Supplemental Figure S2.3 Binding kinetics of *tn*-hpMR1 with MAIT and MR1T TCRs 78

Supplemental Figure S2.4 Activation of MAIT clones by *tn*-hpMR1 and *hs*-hpMR1 79

Supplemental Figure S2.5 The photolumazines identified in *hs*-hpMR1^{+Msmeg} are also apparent in *tn*-hpMR1^{+Msmeg} and *tn*-hpMR1^{+Ecoli} 80

Supplemental Figure S2.6 Molecular networks of ions eluted from the MR1 groove 81

Supplemental Figure S2.7 Variability of AUCs across injections of hpMR1 onto the mass spectrometer 82

Supplemental Figure S2.8 PLIII, PLV, PLVI, and PLIX can be found in the crude PLIII synthesis 83

Supplemental Figure S2.9 Comparison of the MS2 spectra of PLV, PLVI, and PLIX with PLIII 84

Supplemental Figure S2.10 Unique properties of PLX 85

Supplemental Figure S2.11 PLV synthesis route 86

Supplemental Figure S2.12 Isolation of photolumazines from bacterial supernatants 87

Figure 3.1 Proposed deaza-molecules and their aza- parent molecules 93

Figure 3.2 Synthesis of DZ, dDZ, DZPLI, and mmDZ 96

Figure 3.3 DZ stabilizes MR1 at the cell surface 98

Figure 3.4 MR1T and MAIT clones respond variably toward DZ and DMRL 100

Figure 3.5 MAIT cell recognition of DZ is MR1-dependent 103

Figure 3.6 Docking mode of ligands into the MR1 binding groove 105

Figure 3.7 DMRL, DZ, and dDZ vary in the dynamics of their ribityl moiety 108

Supplemental Figure S3.1 MAIT clones that did not respond to PLI also do not respond to DZPLI 114

Supplemental Figure S3.2 Validation of docking ribityllumazines into MR1 115

Supplemental Figure S3.3 Docking mode of ligands into the MR1 structure derived from a ternary complex 116

Supplemental Figure S3.4 Comparison of docking modes between MR1 donor structures derived from ternary and non-ternary complexes 117

Supplemental Figure S3.5 RMSD of ligands across the length of each independent simulation 118

Supplemental Figure S3.6 DZ is stable when stored at -80 ºC as assessed by MAIT cell potency 119

Figure 4.1 Graphical abstract of the approach taken for the characterization of MR1Ts using single cell RNA sequencing 123

Figure 4.2 Properties of the integrated data sets after cell filtering 126

Figure 4.3 Tetramer reads for binned clonotypes, split by donor 128

Figure 4.4 Reactivity and donor origin projected onto the UMAP dimensional reduction 130

Figure 4.5 A hyperexpanded MAIT clone exhibits properties of a T_{EMRA} phenotype 133

Figure 4.6 V(D)J gene usage by TRAV1-2$^-$ cross-reactive MR1T TCRs 137

Figure 4.7 Hierarchical clustering of the TRAV1-2$^-$ cross-reactive MR1T TCRs 139

Figure 4.8 TRAV1-2$^+$ and TRAV1-2$^-$ cross-reactive MR1Ts do not differ strongly with respect to their transcriptome and cell surface phenotype 140

Figure 4.9 Unique properties of the gene expression and cell surface phenotype of 5-OP-RU-agnostic MR1Ts 142

Supplemental Figure S4.1 Purification and quality control of oligo-conjugated streptavidin-PE 145

Supplemental Figure S4.2 Quality control of the DNA libraries and filtering of low-quality and non-T cells from the data sets 146

Supplemental Figure S4.3 Clonotype collapse and reactivity cutoffs 147

Supplemental Figure S4.4 Variability in tetramer reads across the two donors requires separate analysis of clonotype reactivity 148

Supplemental Figure S4.5 Collapsed clonotypes binned into each reactivity category 149

Supplemental Figure S4.6 Assigned reactivities of presumed type I NKT cells 150

Supplemental Figure S4.7 V(D)J gene usage by TRAV1-2^{+} cross-reactive MR1Ts 151

Supplemental Figure S4.8 V(D)J gene usage by TRAV1-2^{-} cross-reactive MR1Ts 152

Supplemental Figure S4.9 V(D)J gene usage by clonotypes with unclassified reactivity 153

List of tables

Supplemental Table S2.1 Exact masses for each MS1 parent ion for MR1 ligands 89

Supplemental Table S2.2 Exact masses for each fragment ion of MR1 ligands used for the MRM assay 90

Table 3.1 TCR rearrangements and sequences for MR1T cell clones. 99

Supplemental Table 3.1 AutoDock Vina scores for the top docking mode of each ligand modeled into the MR1 A′ pocket 120

Table 4.1 Summary of the reactivity binning of recovered αβ T cells 128

Table 4.2 Examples of gene usage and CDR sequences of expanded TRAV1-2^{-} cross-reactive MR1T clonotypes 136

List of important abbreviations and definitions

MR1	MHC-related protein I
MAIT	Mucosal-associated invariant T cell
MR1T	MR1-restricted T cell
RBP	Riboflavin biosynthesis pathway
Mtb	*Mycobacterium tuberculosis*
Msmeg / *M. smeg*	*Mycobacterium smegmatis*
Ecoli / *E. coli*	*Escherichia coli*
TCR	T cell receptor
Clonotype	A family of T cells bearing identical T cell receptors, as defined either by their shared amino acid sequence or shared nucleotide sequence
5-A-RU	An intermediate of the riboflavin biosynthesis pathway and the presumed precursor to the ribitylpyrimidine and ribityllumazine ligands
5-OP-RU	The ribitylpyrimidine condensation product of 5-A-RU and methylglyoxal; the most potent MAIT cell agonist known to date
PL	Prefix meaning “photolumazine”
RL	Prefix meaning “ribityllumazine"
MS1	The first step of tandem mass spectrometry, which detects intact parent ions
MS2	The second step of tandem mass spectrometry, wherein the signal corresponds to the abundance of ions produced from the fragmentation of parent ions
EIC	Extracted ion chromatogram; plot of the retention time against intensity of the signal for a provided m/*z*
Tetramer	Cell staining reagent made of four ligand molecules connected by a tetrameric streptavidin or streptavidin-variant scaffold; overcomes low-affinity ligand-receptor interactions by increasing the avidity
AUC	Area under the curve; integration of the EIC over a defined period of time

Acknowledgements

It takes a village to complete a Ph.D., and I am so lucky to have had a community that went above and beyond to support me throughout this journey. Of course, I first need to thank Erin Adams, who has helped form me into the scientist I am today and who has advocated for me since the first day I walked into her office. Thank you also to my committee members, Robert (Bob) Keenan, Engin Özkan, Anne Sperling, and (formerly) Michaela Gack, for helping guide me through my thesis and holding me to a high standard. I also want to extend a huge thank you to the BMB community, who have been welcoming, scientifically challenging, and always willing to lend a hand. And thank you to the people who hold us together, especially Jack Larkins, Lisa Anderson, Shani Charles, Amy Murphy, and Rebeca Marmolejo. There is no department without their hard work. My hope for graduate school was to join a community of truly collaborative and supportive individuals and the department has exceeded my expectations.

To my lab mates, current and former, who have molded me into a great biochemist and a passable immunologist: thank you for making our lab environment so wonderful to work in. Thank you for your patience with my high standards and with my silly mistakes, and for teaching me both how to listen and how to be a mentor. Special thanks go to Andrew Peev, who is on his way to becoming an amazing physician. You accomplished so much as an undergraduate and took everything in stride as the pandemic reshaped your college experience. I'm so glad we were able to work together, and I am so proud of you. Also to Drs. Christopher Boughter, Kristof Nolan, Sobhan Roy, Charlotte James, and Caitlin Castro, who taught me pretty much everything

I know, and for being patient even after I've asked them every question under the sun, multiple times over. And finally, a massive thank you to Sean Ryan, Idongesit Ekpo, and Hortencia Graves, who have listened to me ramble about mass spectrometry, MR1, and T cells (among other things) for hours and have entertained every one of my strange theories and have helped me experimentally more times than I can count. I am so proud to watch all of you grow into extremely talented and curious scientists, but I'm also thankful to call each of you my friends.

I also want to thank my mental health care team and my university support, without whom I would not be here to see the sun rise on this accomplishment. First, to Lisa Anderson and Joe Piccirilli, who supported me during my recovery from severe depression and ensured that I was able to continue pursuing my Ph.D. Without you, I would have flunked out of graduate school; I will never be able to repay you for going above and beyond in providing me support. Moreover, I need to extend my unwavering gratitude toward the mental health professionals I have worked with over the past seven years. I very nearly lost my battle with depression, but their kindness, patience, and encouragement is what has helped me finally achieve some peace. I owe my life to all of you.

To my friends, Jane, Brittany, Soo Ji, Kristof, Charlotte, Eric, Mel, Alex, Laura, Amy, Chris, Caitlin, Jordan, Jess, and everyone else who has supported my journey: thank you for keeping me sane. Jane, thank you for being my cynical partner-in-crime through this crazy period of our lives; you'll always be the Katya to my Trixie. And Brittany, thank you for being my sister through thick and thin. I love you and am so glad you're in my life.

Innumerable people have contributed to my scientific journey, but my long-term friend and mentor, Brent Krueger, has been a constant source of support in my life since we started working together nearly a decade ago. Thank you - for everything. You helped me become more confident in who I am as a scientist and who I am as a person, and you have been there for me - unwavering - through the highs and the lows. I hope that I have made you proud.

No one shares in this accomplishment as much as my family, especially my parents. Mom and Dad, there are no words for the immense gratitude I have for your support and for the sacrifices that you have made so that the boys and I can pursue our passions. And to my brothers - I watched each of you grow up since the day you came into the world. It makes me so joyful to continue seeing you develop into the young men that you are. I am so proud to be your daughter and sister. I love you all very much.

Finally, to my partner and best friend, Philipp: I am so lucky to have met you and to have had your companionship and strength through this journey. You have challenged me every day to be a better person and a better scientist. Thank you for making me laugh no matter how tough things became, and for being an adoring pet parent to Rosalind, Lilith, and Hemera. I love you and I am so excited for the next step in our lives.

Abstract

In lieu of peptides, the monomorphic MHC-I-like molecule MR1 is thought to present small molecule antigens to stimulate a semi-invariant, phenotypically distinct subset of αβ T cells called MR1-restricted T (MR1T) cells. The MR1 ligands identified to date derive from the riboflavin biosynthesis pathway and their presentation is therefore thought to be an indicator of bacterial infection, including that of *Mycobacterium tuberculosis*. The MR1-MR1T cell axis was once thought to be a highly conserved, innate-like immune mechanism due to the monomorphic nature of MR1, the small pool of validated ligands, and the bias toward usage of a single TRAV gene by a subset of MR1T cells. However, the data presented in this thesis suggest that the MR1-MR1T cell axis is not a binary sensor of infection, but rather a complex indicator of the cellular metabolome; specifically, this work expands our understanding of the molecular mechanisms underpinning donor-unrestricted T cell biology. In this thesis, I first describe a mass spectrometry-based platform for the investigation of the bacterially-derived ligand repertoire of MR1 which revealed a novel class of indolyl-ribityllumazine MR1 ligands. I then present a case study on MR1T cell selectivity by contrasting MR1T cell clone responses toward a natural ribityllumazine MR1 ligand and its deaza-analogue. Finally, I explore heterogeneity in the TCR diversity, antigen specificity, and phenotype of MR1T cells from the peripheral blood of healthy donors using single-cell multi-omics techniques. Altogether, this work addresses the diversity of MR1-restricted T cell antigen recognition and in doing so, challenges the paradigms of MR1T cell biology.

Chapter 1. Introduction

1.1 Molecular mechanisms of the T cell response to microbial infection

1.1.1 MR1-restricted T cells lie at the interface of the adaptive and innate immune systems

The primary responsibility of the immune system of any organism is to identify and eliminate threats to host homeostasis. In the context of pathogenic infection, immune cells must distinguish between the host ("self") and infectious agent ("non-self") and mount an appropriate immune response thereto [1, 2]. In vertebrates, these responses are orchestrated by innate and adaptive immune cells. Innate immune cells recognize broad molecular patterns and therefore can mount quick, nonspecific responses to infectious agents. In contrast, adaptive immune cells - such as conventional αβ T cells - are slow to respond but are remarkably pathogen-specific, a property which is imbued by their sequence-variable cell surface receptor. They are also capable of developing immunological memory, meaning that the maintenance of these cells allows for a quicker response to the same signal upon later challenge [3, 4]. However, the line segregating the adaptive and innate immune compartments is fuzzy, with some cells sharing properties of both compartments.

Indeed, over the last 30 years, several specialized subsets of "innate-like" T cells have been identified and characterized, including a class called MR1-restricted T cells (MR1Ts) and a major subset of MR1Ts called mucosal-associated invariant T (MAIT) cells. MAITs are poised to respond quickly to novel microbial threats, but do so with lower pathogen specificity than

conventional αβ T cells due to their restriction to a monomorphic protein, MR1, bound to a small class of non-peptidic antigens [5-7]. Unlike what we would expect for conventional T cells, the molecular mechanisms governing MAIT cell antigen recognition are relatively constant both within and between donors [8]. Thus, antigen engagement by MAIT cells is reminiscent of the pattern recognition of innate immune cells but is achieved through adaptive immune receptors. These qualities make these "donor-unrestricted" T cells tempting immunotherapeutic and vaccine targets, but harnessing their properties both requires a greater understanding of how MAITs and MR1Ts differ from conventional T cells and warrants further investigation into the greater diversity in the MR1T cell compartment.

1.1.2 Conventional T cells recognize HLA molecules in complex with pathogen-derived peptides

In lieu of direct recognition of foreign molecules by cell surface proteins, T cells are dependent on other host cells to present "danger" signals to them [9-11]. In the case of conventional αβ T cells, these danger signals come in the form of cell surface-expressed host proteins called Major Histocompatibility Complex molecules (**Figure 1.1**), which are named after the vertebrate-specific locus in which they are contained [12]. When referring specifically to human immunity, these are interchangeably referred to as "Human Leukocyte Antigen" (HLA) molecules. Conventionally, MHC molecules bind short peptides in a long, linear, surface-exposed groove and "present" them on the surface of antigen-presenting cells (APCs) to be probed by various immune cells, including αβ T cells (**Figure 1.1**) [13-15]. The bound antigens are sampled from a peptide pool either generated by the degradation of cytosolic proteins or taken in from the

extracellular space [16-18] and can derive from either a pathogen or the host proteome [19]. For these reasons, the repertoire of peptides presented on the cell surface by MHC molecules reflect the health status or local environment of the host cell.

MHC molecules fall into two primary classes, MHC-I molecules, which engage with "killer" (e.g. cytotoxic) T cells, and MHC-II molecules, which engage with "helper" T cells (**Figure 1.1**) [20, 21]. In humans, HLA-I molecules can be further categorized into three HLA-Ia genes (HLA-A, HLA-B, and HLA-C) and three HLA-Ib genes (HLA-E, HLA-F, and HLA-G), which

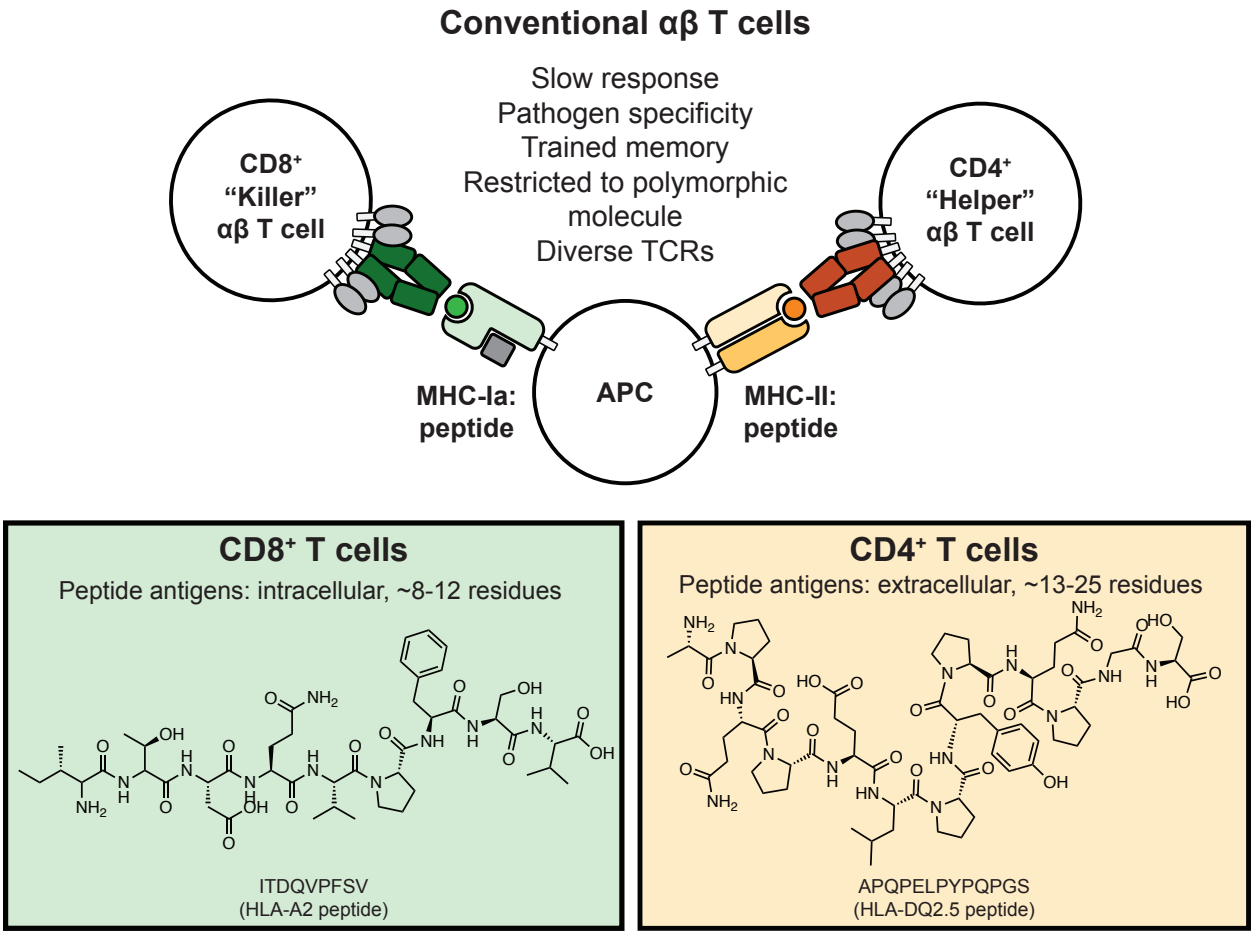


Figure 1.1 CD4+ and CD8+ T cells recognize HLA-II and HLA-I, respectively, bound to unique classes of peptides.

are distinguished in part by their degree of polymorphism and their cell surface expression patterns. HLA-Ib molecules are monomorphic or oligomorphic and show tissue-specific or context-specific expression patterns, alluding to specialized roles for these molecules [22]. In contrast, HLA-Ia molecules are expressed on almost all nucleated cells - which reflects their organism-wide importance in antigen surveillance - and display extreme levels of polymorphism [23]; indeed, with 4,000 to 5,000 uniquely encoded proteins per HLA-Ia gene currently identified, they are the most polymorphic genes in the human genome [24]. This polymorphism is concentrated in and around the peptide-binding groove, thereby influencing the identity of the peptides that are bound and presented on the cell surface and, in turn, the T cells that react thereto [25, 26].

Across species, MHC-I and MHC-II molecules have been shown to generally adopt the same fold consisting of two membrane-proximal immunoglobulin (Ig) domains (i.e. the scaffold) which support the characteristic peptide-binding groove (i.e. the platform) (**Figure 1.2a**) [5]. While MHC-I and MHC-II molecules share a global structure, they differ in their domain arrangement: MHC-II molecules are formed by two membrane-linked polypeptide chains, with each chain contributing one domain to each the scaffold and the platform (**Figure 1.1**) [27]. In contrast, MHC-I molecules are constructed of a three-domain, membrane-bound "heavy chain" (consisting of the entire platform and one scaffold domain) and a soluble, invariant "light chain" known as β-2-microglobulin or β_2m, which serves as the other scaffold domain[1] (**Figure 1.1,**

[1]When discussing MHC-I and MHC-Ib-like molecules in this thesis, it is implied that the molecule is complexed with β_2m unless otherwise noted.

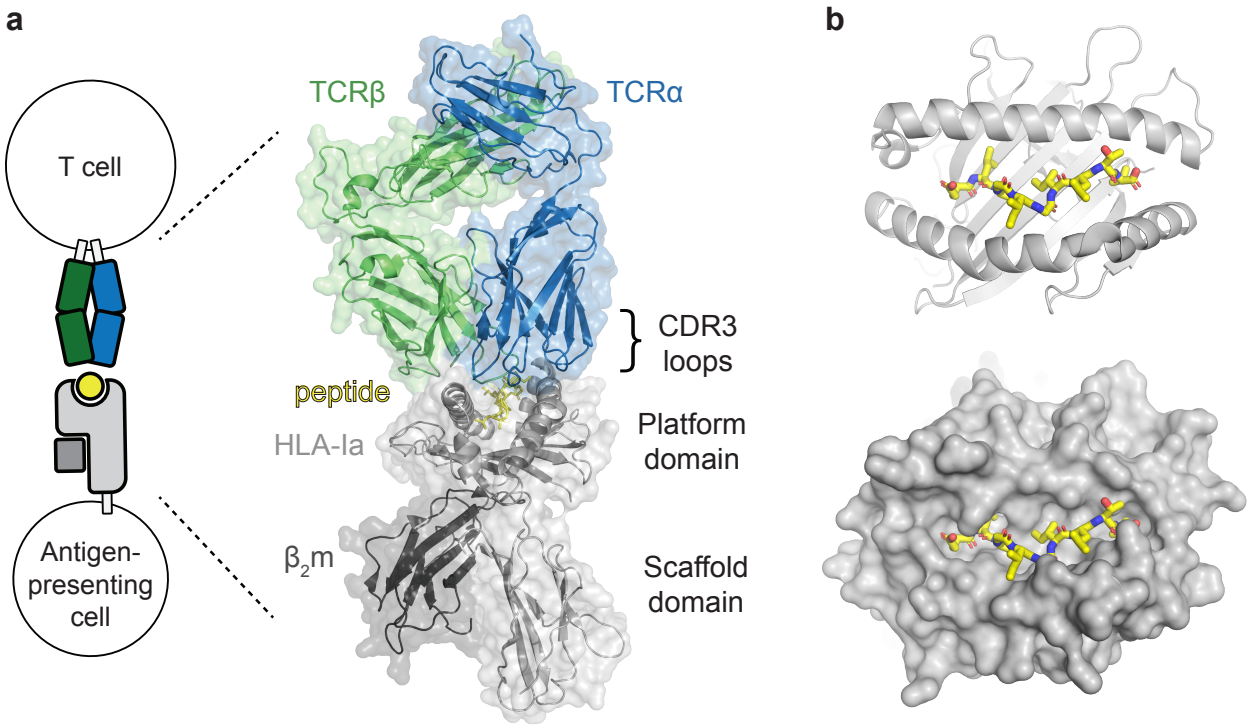


Figure 1.2 Orientation of the pHLA:TCR complex. a) Illustration of the engagement of pMHC with TCR (left) and a crystallographic structure of one such complex (PDB 3HG1). b-c) Structural features of the antigen bound within the peptide-binding groove of the platform domain shown as a cartoon (b) or a space-filling model (c).

Figure 1.2a) [13]. The peptide-binding groove of the MHC molecule is formed by a "floor" made of a β-sheet which supports two antiparallel alpha helices in which the peptide is cradled (**Figure 1.2b**) [5], and the complex is positioned on the cell surface such that an αβ TCR can make contact with both the peptide and the MHC molecule itself [28]. Thus, the MHC gene and allele are important not only for determining the peptide repertoire, but also for direct interactions with the antigen recognition machinery of the T cell.

1.1.3 T cells use sequence-variable T cell receptors to communicate with host cells

T cells which recognize peptide:MHC complexes (pMHC) displayed on other host cells do so via their T cell receptor (TCR) [29-31]. Each TCR belongs to a family of proteins with very similar

tertiary structures but highly variable amino acid sequences - especially at the positions which determine pMHC binding. There are estimated to be about 10^{11} total T cells in the human body carrying approximately 10^{10} unique TCRs ("clonotypes"), each at variable abundances [32]. Encoding for all of these receptors in the germline, however, would drastically increase the size of the human genome. To circumvent this, the adaptive immune system has evolved a specialized mechanism of recombination called V(D)J recombination to stochastically generate these receptors through combinatorial diversity [33, 34]. This mechanism is estimated to be able to encode for up to 10^{14} to 10^{18} possible unique human αβ T cell receptors [32] using only about 200 genes [35].

Conventional αβ T cells, the major class of pMHC-restricted T cells, bear the αβ TCR, which is a heterodimer of the TCR alpha chain (TCRα) and TCR beta chain (TCRβ). Each chain is individually generated by V(D)J recombination, wherein one variable (V) gene, one diversity (D) gene (only for the β chain), one junctional (J) gene, and one constant (C) gene are selected to form a full-length TCR chain transcript [36] (**Figure 1.3a**). This process thus encodes the combinatorial diversity of the TCR repertoire. In humans, there are 58 functional TCRα V (TRAV) genes, 60 TCRα J (TRAJ) genes, and one TCRα C (TRAC) gene; there are also 65 TCRβ V (TRBV) genes, 2 TCRβ D (TRBD) genes, 13 TCRβ J (TRBJ) genes, and 2 TCRβ C genes [35]. Selecting one of each of these segments results in up to 10^6 TCR gene combinations, though not all combinations are structurally compatible and this is therefore an overestimation. Nevertheless, even further diversity is introduced during the recombination event by stochastic editing of the region spanning the V-(D)-J joint via the excision of germline-encoded bases and/

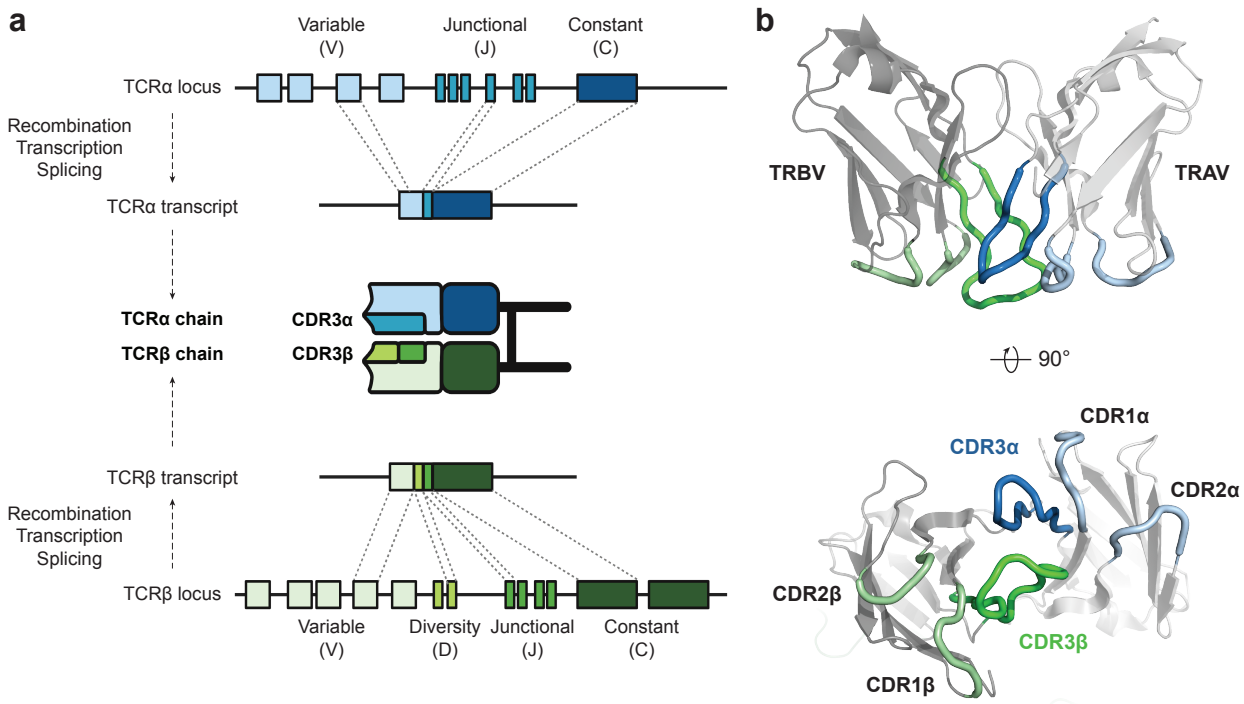


Figure 1.3 TCRs are generated by a V(D)J recombination. a) Generalized illustration of the combinatorial generation of the T cell receptor at a genetic, transcriptional, and protein level. b) Structure of the V(D)J component of an αβ T cell receptor (the constant domain is omitted for space), referred to in this text as the "platform domain." The CDR loops are visualized as ribbons; the germline-encoded (V gene-dependent) loops are shown in light colors and the CDR3s are shown in a darker color, consistent with the coloring of (a).

or the introduction of non-templated bases [37, 38]. This phenomenon is what allows the number of possible TCR rearrangements to reach the range of 10^{15} to 10^{18}. This editing of the V-(D)-J segments is also what encodes one of the three "complementarity-determining regions" (CDRs) of each TCR chain. The CDRs are loops of amino acids which are arranged such that they are the most distal from the T cell membrane and are poised to interact with pMHC complexes *in trans*. CDR1 and CDR2 are encoded by the germline of the V gene, but CDR3 spans the V-(D)-J intersection and thus is where the most variability is encoded [39-41] (**Figure 1.3b**). Accordingly, the CDR3 loops are also typically positioned such that they mediate contacts with both the MHC and the variable peptide and thus bear much of the responsibility for peptide selectivity [42, 43]. Collectively, the combinatorial response of sequence-variable TCRs toward

polymorphic MHC molecules loaded with diverse pathogen-derived peptides allows for great flexibility and modularity to respond to many different pathogenic threats.

1.1.4 The various immune programs of αβ T cells

When the TCR engages with an MHC in complex with a "non-self" peptide, an intracellular signaling cascade is initiated that results in - among other cellular responses - the production of secreted proteins such as cytotoxins and cytokines [44]; T cells in this acute state are referred to as being "activated." These molecules act directly by inducing apoptosis or lysing infected target cells or indirectly by contributing to and regulating the signaling network between immune cells [45, 46]. For conventional αβ T cells, the particular set of proteins that are upregulated is dictated by the developmental stage of the cell and the functional subset to which it belongs, which can often be distinguished by cell-surface markers [47, 48]. Analysis of these markers can therefore be used to subset T cells on any number of phenotypic properties, including antigen experience (memory phenotype), tissue residency status, and regulation by co-receptors [48].

Mature T cells can be broadly divided by their surface expression of CD4 and CD8, which act as co-receptors for the T cell receptor. $CD8^+$ killer (i.e. cytotoxic) T cells specifically recognize pMHC-I (**Figure 1.1**) and are responsible for inducing apoptosis or lysis of infected cells via the release of molecules including perforin, granzymes, and granulysin [45]. On the other hand, engagement of $CD4^+$ helper T cells with pMHC-II molecules leads to the activation of B cells and cytotoxic T cells to amplify the response to a pathogen [46]. The expression of CD4 and CD8 are thought to be mutually exclusive except for during some specialized developmental

stages [49, 50] and therefore $CD4^+CD8^+$ "double positive" T cells are rare outside of the thymus. However, a smaller population of $CD4^-$ $CD8^-$ "double negative" (DN) T cells, which bear neither co-receptor, do exist in the periphery. These cells can be restricted to MHC-I, MHC-II, or nonclassical MHC-Ib-related molecules such as CD1 molecules and MR1. In agreement with their the diverse antigens they can recognize, these cells variably exhibit an inflammatory or regulatory phenotype [51, 52]. Due to their heterogeneity in function and antigen recognition, however, their ontogeny and function in various immune diseases and infection responses are poorly understood. In all, the variability among T cell functional programs complements the intrinsic diversity of their antigen recognition machinery; it is this plasticity that makes them a central player in the immune response.

1.2 MR1 presents microbial small molecules to MR1-restricted T cells in a manner analogous to conventional MHC presentation of peptides

1.2.1 MR1 is a monomorphic MHC-related molecule that restricts MAIT cells

In addition to the peptide-presenting MHC molecules, monomorphic MHC-I-related molecules have evolved in vertebrates to present other classes of biomolecules to "innate-like" immune cells [53-55]. For example, members of the CD1 family present lipids to various classes of T cells including NKT-I cells [56], and the MHC-Ib-related molecule MR1 binds metabolites for presentation to MR1-restricted T cells, including a major subset known as mucosal-associated invariant T (MAIT) cells [57, 58] (**Figure 1.4**). Unlike the CD1 family, which is thought to have evolved very early, MR1 is only found in mammals and shares more sequence similarity with MHC-I than CD1 molecules [59-61], though both share the characteristic MHC-I fold [5]. The

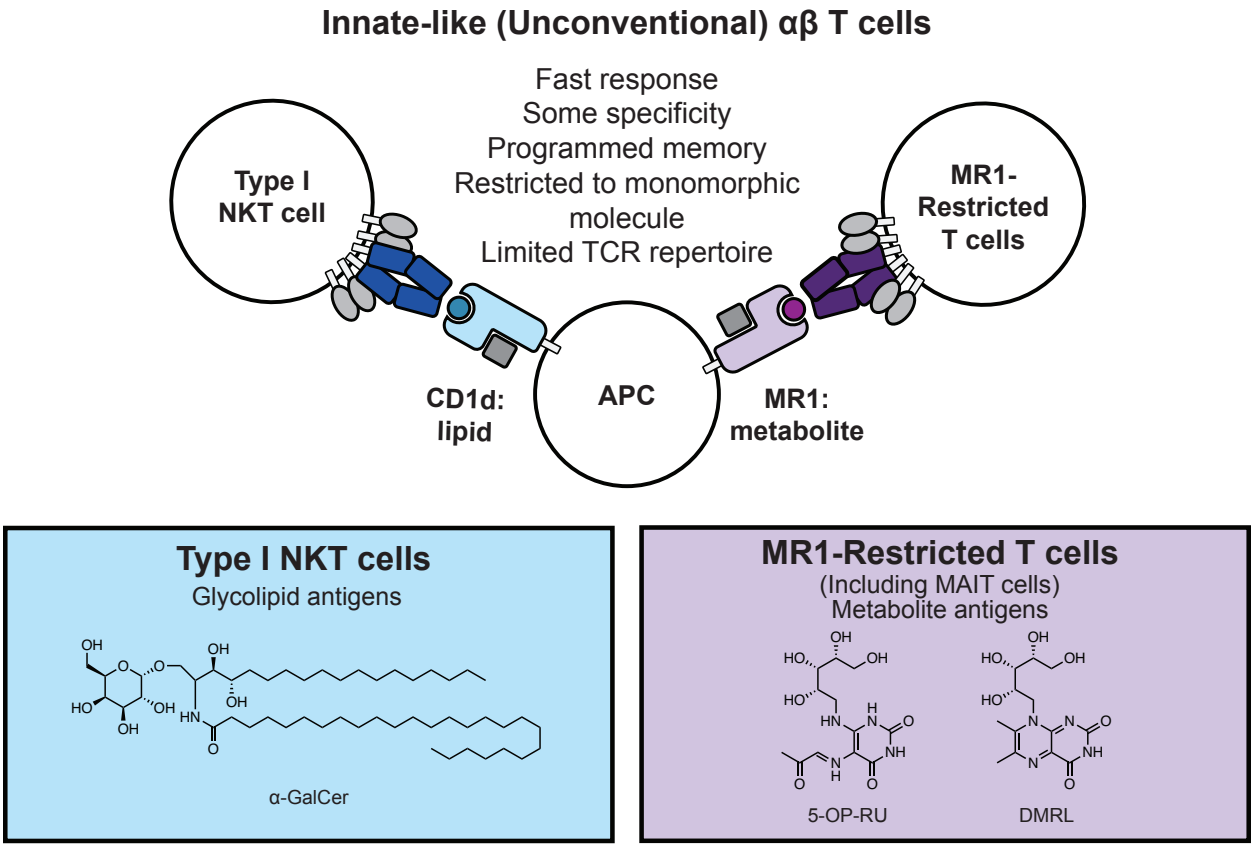


Figure 1.4 Innate-like T cells recognize non-peptidic antigens and have unique functional and phenotypic characteristics.

gene encoding for MR1 displays a striking degree of sequence conservation, illustrated by the fact that the human and mouse homologs share 90% amino acid sequence identity [59, 60]. This is particularly true of the residues lining the ligand binding groove, indicating strong selection pressure against diversification in this region [61-63]. The high degree of sequence conservation and monomorphic nature of MR1 together suggest that a small class of structurally conserved ligands bind in the groove and are responsible for an invariant mechanism of antigen presentation and - therefore - of antigen recognition.

1.2.2 The riboflavin biosynthetic pathway is the origin of microbial MR1 ligands

Indeed, in the early 2010s, MR1 was found to present a small family metabolites deriving from the riboflavin biosynthesis pathway (RBP) (**Figure 1.4**) [57, 58]. The RBP can be found in bacteria and yeast but is absent from the *Animalia* kingdom [64], indicating that MR1 is a messenger of microbial infection to MR1T cells. Riboflavin (RFN) - also known as Vitamin B_2 - is essential for both microbial and human metabolism because it is the precursor to flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), two cofactors/prosthetic groups for a diverse family of metabolically-imperative flavoproteins [65, 66]. It's therefore unsurprising that the deletion of non-redundant RBP genes from obligate autotrophs without supplementation of exogenous riboflavin or gene complementation is lethal [67]. However, riboflavin auxotrophs do exist; the distribution of riboflavin auxotrophy across a 16S rRNA phylogenetic tree shows many independent losses of the ability to make riboflavin, suggesting that the pathway is dispensable [68]. The essentiality of riboflavin, however, requires that the auxotrophs evolved alternative compensation mechanisms, such as membrane transporters, to uptake riboflavin from the environment in the absence of *de novo* biosynthesis [69]. Importantly, although riboflavin auxotrophs maintain a pool of riboflavin for their own metabolism, it does not seem that it can be biotransformed into MAIT cell agonists, as indicated by the inability of MAIT cells to respond toward riboflavin auxotrophs [70]. Thus, it is the presence of the biosynthetic pathway and its intermediates, and not the end product, that is required for generation of MR1 ligands.

The riboflavin biosynthetic pathway originates with GTP[2], which is hydrolyzed by GTP cyclohydrolase II (*Rib1a*) to form DARPP[3] [64, 71] (**Figure 1.5**). (Since RBP genes are annotated differently across bacterial species, they will be referred to here by the enzyme name/ function or the provided numerical annotation.) DARPP is then converted to ArPP[4] via *Rib2ab*, a bifunctional deaminase (*Rib2a*) and reductase (*Rib2b*). DARPP is dephosphorylated to 5-A-RU[5] by a haloacid dehydrogenase (*Rib3*) that was recently characterized in *Arabidopsis thaliana* [72], but is yet to be definitively characterized in prokaryotes. Although candidate HAD enzymes

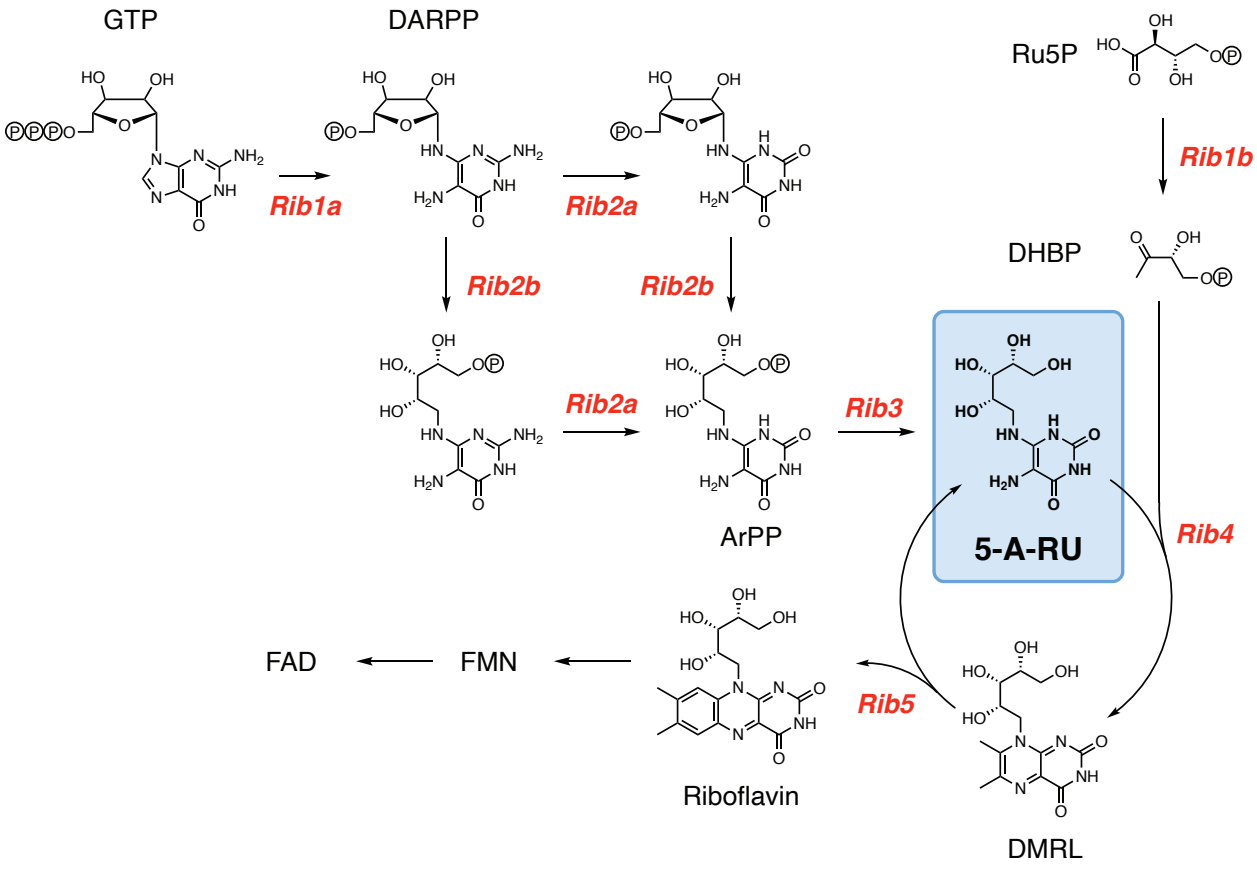


Figure 1.5 The prokaryotic riboflavin biosynthesis pathway. *Rib1ab* and *Rib2ab* are bifunctional enzymes.

[2] Guanosine triphosphate
[3] DARPP, 2,5-diamino-6-ribosyl-amino-4(3H)pyrimidinedione 5′-phosphate
[4] ArPP, 5-amino-6-ribityl-amino-2,4(1H,3H)pyrimidinedione 5′-phosphate
[5] 5-A-RU, 5-amino-6-ribityl-amino-2,4(1H,3H)pyrimidinedione

from *Escherichia coli* have been shown to dephosphorylate DARPP *in vitro*, it's not yet clear that this is the case *in vivo* since knockouts of these enzymes have no apparent phenotype [67, 73]. In any case, the production of 5-A-RU initiates a biosynthetic cycle wherein its condensation with DHBP[6] in a spontaneous manner or via lumazine synthase (*Rib4*) results in the production of a small ribityllumazine called DMRL[7]. DHBP is generated from Ru5P by DHBP synthase (*Rib1b*), which forms a bifunctional enzyme with *Rib1a*. Finally, two molecules of DMRL are converted into one riboflavin molecule and one 5-A-RU molecule by riboflavin synthase (*Rib5*), the latter of which re-enters the biosynthetic cycle.

The first paper that revealed the identity of bacterial MR1 ligands did so by mass spectrometric characterization of the ligands bound to MR1 when it was refolded in vitro with the supernatant of *Salmonella typhimurium* [57]. This study revealed a ligand called rRL-6-CH_2OH, a reduced ribityllumazine, which was inferred to be related to the RBP by structural similarity to DMRL (the RBP intermediate) and another small ribityllumazine, HMRL[8]; indeed, functional analysis determined all three of these molecules were able to stimulate Jurkat cells transduced with a MAIT TCR in an MR1-dependent fashion, though to variable extents. Two years later, Corbett et al. [58] cemented the centrality of the RBP in the generation of MAIT cell agonists using RBP enzyme knockout strains of *Lactococcus lactis*. When they evaluated the ability of the culture supernatant of these strains to activate Jurkat cells transduced with MAIT TCRs (Jurkat-MAIT cells), they found that knocking out genes upstream of 5-A-RU ablated reactivity toward the

[6] DHBP, 3,4-dihydroxy-2-butanone-4-phosphate
[7] 6,7-dimethyl-8-D-ribityllumazine
[8] 6-methyl-7-hydroxy-8-D-ribityllumazine

supernatant; meanwhile, the Jurkat-MAIT cells had equivalent reactivity toward the wild-type *L. lactis* supernatant and that of knockout strains downstream of 5-A-RU. Thus, 5-A-RU is the presumed requirement for the generation of microbial MAIT cell agonists.

Since the discovery of their origin, two major classes of ligands that have been shown to derive from 5-A-RU: the ribitylpyrimidines (monocyclic) and the ribityllumazines (bicyclic). 5-A-RU is a highly reactive molecule which can condense spontaneously with metabolites such as α,β-diketones which can presumably derive from either the bacterial or host environment, as is seen for DMRL biosynthesis [58] (**Figure 1.6a**). *In vitro*, this ketone/amine condensation occurs at the primary amine of 5-A-RU and results in the formation of the ribitylpyrimidine family of MR1 ligands [58, 74], which are typically the most potent MAIT cell agonists [8, 58]. The ribitylpyrimidines are intrinsically unstable in aqueous solution [58, 74, 75] but are captured in this form upon binding MR1 due to their unique ability to form a Schiff base with a lysine (K43) residue buried deep in the ligand-binding groove [57, 58, 76] (**Figure 1.6a-b**). In addition to stabilizing the ligand, this covalent bond also seems to promotes high ligand occupancy in the groove *in vitro*, though whether this is the case *in vivo* is unknown. Ribitylpyrimidine species include 5-OP-RU[9], which is formed from 5-A-RU and methylglyoxal, a ubiquitous metabolite, and 5-MOP-RU[10], which is the putative intermediate in the production of DMRL during the biosynthetic cycle of the RBP [58] (**Figure 1.6a**). Without MR1 to stabilize them, the ribitylpyrimidines undergo a spontaneous intramolecular cyclization which results in the

[9] 5-(2-oxopropylidenamino)-6-D-ribitylaminouracil
[10] 5-(1-methyl-2-oxopropylideneamino)-6-D-ribitylaminouracil

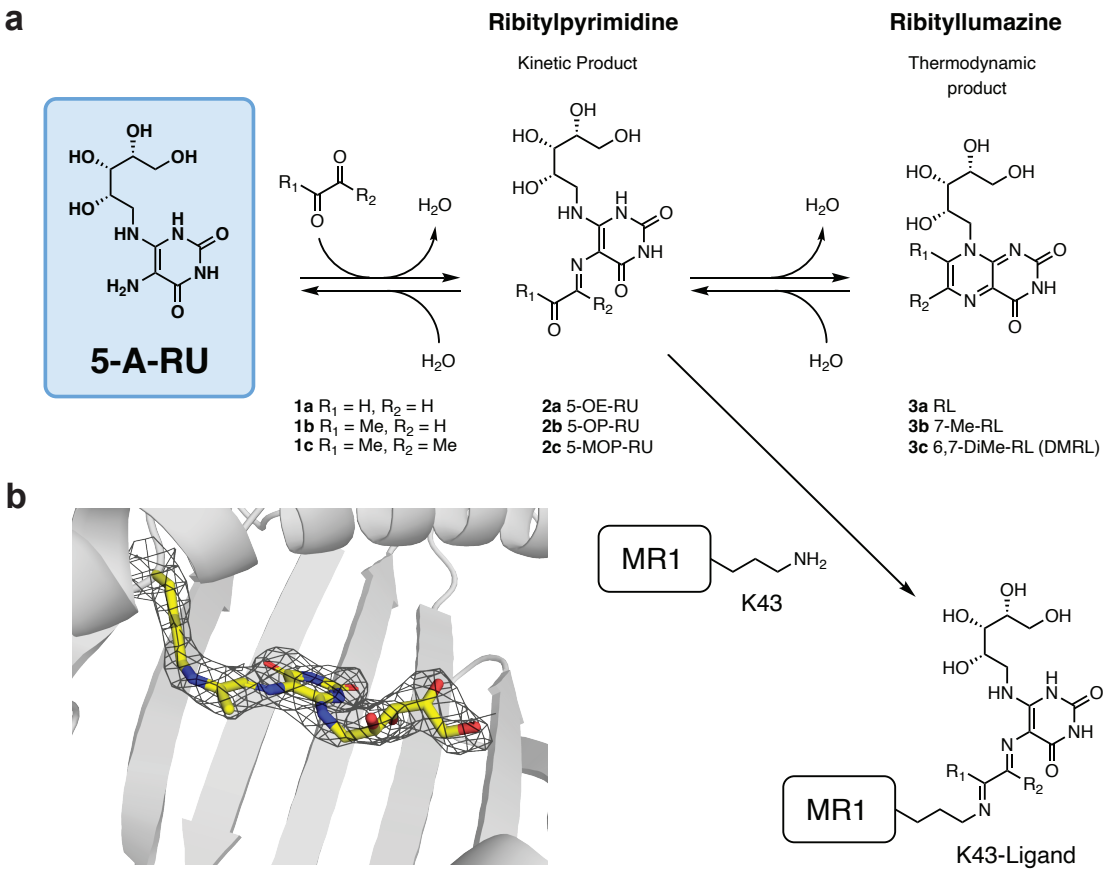


Figure 1.6 Generation of ribityllumazine and ribitylpyrimidine antigens. a) The 5-A-RU intermediate of the riboflavin biosynthetic pathway condenses with small α,β-diketones to give rise to the ribityllumazines, which then spontaneously undergo a second ketone/imine condensation to form the ribityllumazines. b) 2Fo-Fc map showing the covalent bond between K43 and 5-OP-RU (PDB 4NQC).

production of ribityllumazines [58, 74] (**Figure 1.6a**). These are bicyclic, aromatic structures with a conserved lumazine base but variable adducts at positions C6 and C7 of the ring system, depending on the metabolite with which 5-A-RU condenses. For example, 5-OP-RU cyclizes to form 7-Me-RL[11] [58]. Recently, ribityllumazines deemed "photolumazines" have been identified that have much larger chemical groups attached at C6 and C7 [77] (**Figure 1.7**). With the exception of one of these molecules, photolumazine I, the ribityllumazines characterized to date do not have a non-aromatic ketone or carboxylic acid and thus are not expected to form a Schiff

[11] 7-methyl-8-D-ribityllumazine

base or undergo amide coupling with the K43 residue of MR1; thus, their pocket occupancy *in vivo* in comparison with their covalently-bound ribitylpyrimidine counterparts is unclear. In addition to the ribitylpyrimidines and ribityllumazines, a small sampling of tricyclic flavins and flavin-like molecules deriving from the RBP (FO[12] and riboflavin itself) have also been shown to compete with other ligands for the MR1-binding groove, but have not been shown to stimulate MAIT clones (**Figure 1.7**) [77]. Further, how MR1 adapts structurally to these larger ligands is still unknown. Unfortunately, the contribution of different ligand classes to the T cell response toward microbes *in vivo* is also unknown because very few of these RBP-generated metabolites have been identified, much less characterized. Thus, it is imperative that the field defines additional MR1 ligands deriving from the RBP in order to truly understand the molecular mechanisms governing this axis and the role that it plays in the greater immune response toward microbes.

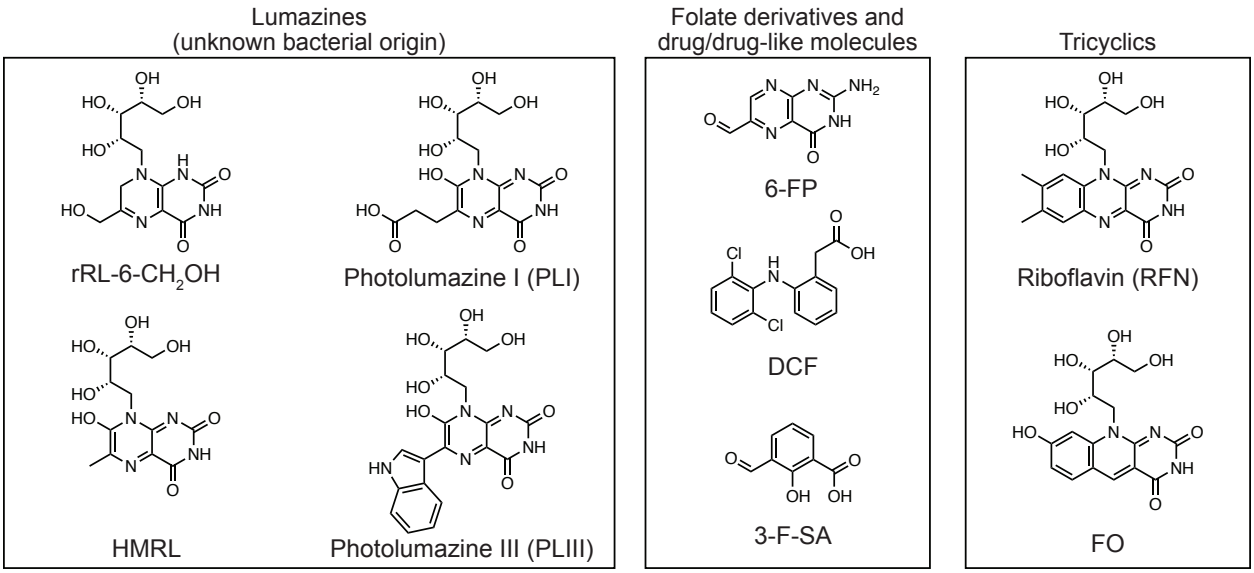


Figure 1.7 Additional MR1 ligands of natural and synthetic origin.

[12] Didemethyl-8-hydroxy-5-deazariboflavin

1.2.3 Microbial MR1 ligands independent of the RBP may exist but are yet to be identified

While the RBP-based generation of bacterially-derived MAIT cell agonists is widely accepted, there have been several recent observations that suggest that the ligand repertoire of MR1 may be more expansive. In fact, the first ligand to be crystallized with MR1 was a folate photodegradation product, specifically 6-FP[13], which is a bicyclic heterocycle with structural similarity to lumazines [57] (**Figure 1.7**); because of the absence of the ribityl chain, which is critical for MAIT cell antigen recognition, 6-FP is not known to stimulate MAIT clones [8, 57], but at least one non-autoreactive MR1T cell clone has been identified as an exception to this paradigm [78]. Further, *in silico* screening of drug and drug-like compounds containing carbonyl or carboxylate moieties capable of neutralizing the primary lysine of the K43 residue of MR1 identified a handful of small molecules that were able to both upregulate MR1 on the cell surface and stimulate rare MAIT cells [79] (**Figure 1.7**). However, the concentration of these compounds necessary to elicit an immune response is orders of magnitude higher than would be physiologically relevant. Nevertheless, these molecules have proven to be useful tools for the study of MR1T cells and shown that RBP-independent ligands are sometimes capable of stabilizing MR1 and stimulating MAIT cells [80].

Beyond explorations into the MR1 ligand repertoire itself, there have been several observations of MR1T cell clones bearing unique reactivity toward MR1. For example, several papers from the Sewall and de Libero labs have identified MR1-restricted T cell clones that are responsive

[13] 6-formylpterin

toward cancer cell lines [81, 82], implying the existence of cancer antigens capable of binding MR1 and stimulating a specialized subset of MR1T cells. Nucleobase adducts have been put forward as MR1 ligands generated under the conditions of metabolic/oxidative stress imposed by a cancerous cell, but the concentration of nucleosides necessary to elicit activation was also quite high, lending concern to the relevance of these ligands *in vivo* [83]. Additionally, in 2016, Meermeier et al. [70] discovered an MR1T cell clone that was responsive toward *Streptococcus pyogenes*, which is a riboflavin auxotroph (lacking all of the validated pathway enzymes). The ligand that elicited this response, however, is still unknown. Collectively, these studies imply that the MR1-MR1T cell axis is a complex sensor of the intracellular metabolome rather than a binary danger signal.

Thus, the ligandome of MR1 seems to be broader than previously appreciated, both in the context of immunity against bacteria and against other immunological threats such as cancer; these data therefore warrant the characterization of MR1 ligands and MR1-restricted T cell antigens. In **Chapter 2** of this thesis, I describe an untargeted mass spectrometric platform for MR1 ligand discovery and the characterization of indolyl-ribityllumazine compounds found in the MR1 groove. Further, I show the semi-quantitative characterization of the RBP-generated ligand repertoire of MR1 produced in the presence of divergent microbial species, and that some MR1 ligands may derive from a *Rib1ab*-independent but *Rib2ab*-dependent pathway.

1.3 MAIT cells are a subset of MR1-restricted T cells

1.3.1 MAIT cells use a semi-invariant TCR to recognize agMR1

In 1993, well before MR1 or its ligands had been discovered, a study by Porcelli et al. [84] described the isolation and characterization of the TCR repertoire of $CD4^{-}CD8^{-}$ T cells from the peripheral blood of several human donors. In this T cell subset, they discovered the enrichment of two semi-invariant gene rearrangements of the TCRα chain across donors, TRAV10/TRAJ18 and TRAV1-2/TRAJ33. Their abundance and limited TCR diversity indicated that these reservoirs of cells may share similar mechanisms of reactivity across individuals. Even then, the authors posited that the restricting ligand(s) may be non-polymorphic MHC-I-like molecules with limited antigen diversity. Indeed, the TRAV10/TRAJ18 recombination was soon found to be restricted to the nonclassical lipid-binding molecule CD1d; these cells are known as type I natural killer T (NKT) cells due to their expression of natural killer cell receptors [56, 85, 86] and fall under a greater umbrella of CD1d-restricted T cells (**Figure 1.4**). Likewise, cells bearing the TRAV1-2/TRAJ33 rearrangement were later identified as being restricted to MR1; further studies found that they are abundant (even outside of $CD4^{-}CD8^{-}$ compartment) in many different tissues but especially at mucosal sites and thus have been historically called "mucosal-associated invariant T" (MAIT) cells [87].

The delineation between MR1Ts and MAITs has been defined several ways in the literature, but for the purposes of this thesis, I will define human MAIT cells as a subset of MR1Ts which specifically utilize the TRAV1-2 gene (**Figure 1.8a**). The TRAV1-2 gene of MAIT cells is almost exclusively recombined with TRAJ33, TRAJ20, or TRAJ12 [88], and the resultant TCRα chain

forms a CDR3 loop with a "conserved" tyrosine at the eighth position of the CDR3α (Y95α), a residue which has been shown to be critical for recognition of and reactivity toward ribityl-containing antigens [63, 89, 90] (**Figure 1.8b**). Whereas the TCRα is nearly completely invariant, the TCRβ chain exhibits slightly more sequence variability, though the gene usage is skewed toward TRBV6 and TRBV20 [84].

Because the TCR repertoire of MAIT cells is restricted and the pool of known antigens is quite small, the ternary complex with antigen:MR1 (agMR1) has been dissected quite extensively [8, 75]. These crystallographic models show that MAIT TCRs bind MR1 on top of its platform

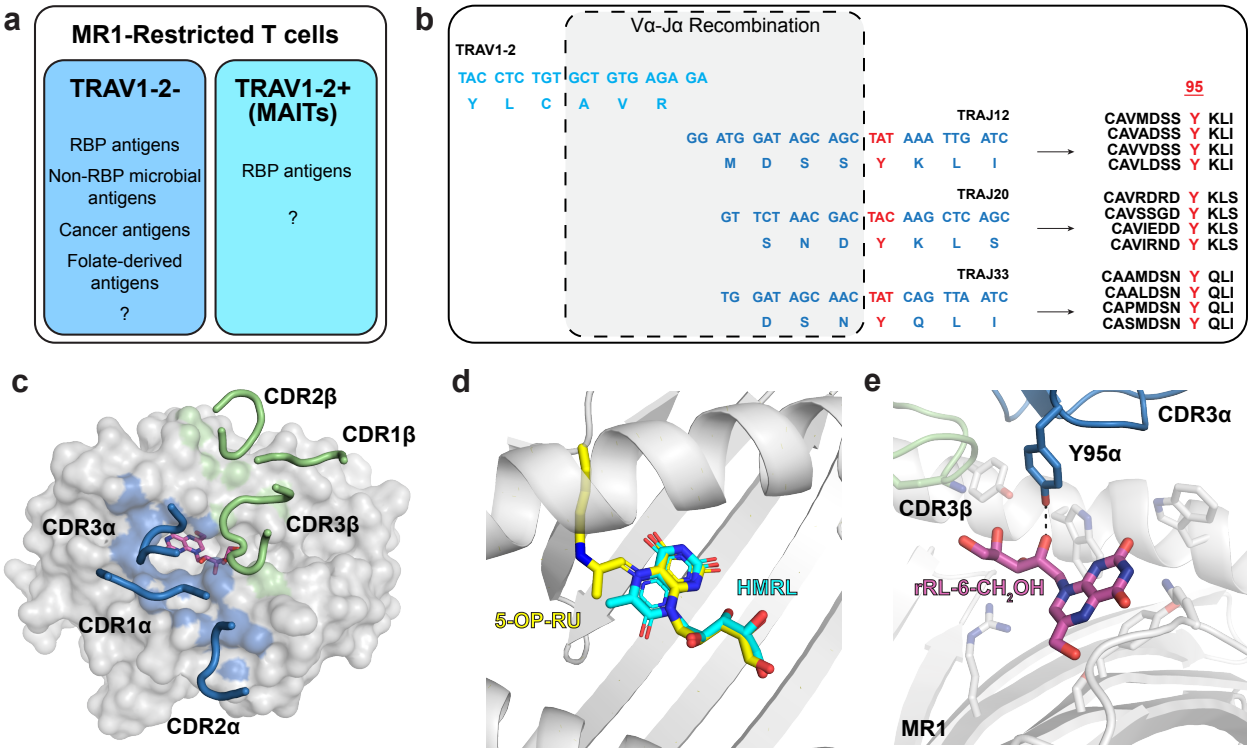


Figure 1.8 Molecular basis of MAIT cell reactivity. a) MAITs are a subset of all MR1Ts. b) MAIT TCR rearrangement between TRAV1-2 and a select number of TRAJ genes. c) Footprint of the CDR loops of a MAIT TCR on the MR1 binding groove (López-Sagaseta et al. 2013) d) Binding mode of HMRL and 5-OP-RU in the MR1 binding groove (PDBs 4NQC and 4L4V). e) Interaction of Y95 of the MAIT TCR with the 2' hydroxyl of the ribityl moiety of rRL-6-CH_2OH (López-Sagaseta et al. 2013).

domain, straddling the antiparallel α-helices; the TCRβ chain is positioned above the α_1 helix and the TCRα chain is poised above the α_2 helix of MR1, thus centering the CDR3s above the A' pocket of MR1 [63, 89, 90] (**Figure 1.8c**). Within the aromatic cradle of the A' pocket, the ribitylpyrimidines and ribityllumazines have been shown to bind in a similar manner, as indicated by the overlap of their shared uracil substructures, despite the fact that the ribitylpyrimdines are constricted by the Schiff base which they form with the primary amine of K43 [58] (**Figure 1.8d**). Interestingly, there is more space in the groove than is utilized by either the ribitylpyrimidines or the ribityllumazines, alluding to the fact that larger antigens may be able to bind in the same manner, though no crystallographic complexes with larger ligands have been elucidated. At least for these smaller ligands, however, these molecules are buried deep within the cleft, largely cordoned off from the TCR. Indeed, the only substructure for RBP-derived ligands that is known to be accessible to the TCR is the sugar-derived ribityl moiety, which protrudes from the protected MR1 binding groove through a gap in the aromatic cradle [63, 89, 90] (**Figure 1.8c**). It is the 2'-hydroxyl of the ribityl moiety which forms the critical hydrogen bond with Y95α in the CDR3α loop that is conserved across MAIT TCRs (**Figure 1.8e**). Modifying this Y95α residue to phenylalanine results in loss of MAIT cell reactivity and a decrease in the affinity of the TCR for MR1:ligand by an order of magnitude [8, 58, 90]. Likewise, modification of ligands to their 2'-deoxy form also decreases the stimulatory capacity of the ligand [75]. These findings are surprising considering that this bond is only responsible for 0.6% of the large buried surface area of the MR1:ligand-TCR interaction [75]. Thus, unlike pMHC-I and pMHC-II complexes, which form a composite surface of the peptide and MHC for probing by TCRs, the agMR1 surface is thought to be dominated by MR1 itself, and yet the loss

of one key contact can result in severe phenotypic consequences. The explanation for this effect may lie in the influence of the ligand on MR1 dynamics and, indirectly, engagement of the MAIT TCR with MR1:ligand complexes.

1.3.2 Phenotypic properties of MAIT cells

Although MAIT cells are only a subset of MR1Ts, they have been characterized much more extensively than the cells which fall outside of this definition. They are extremely abundant in humans, especially in the peripheral blood and liver where they have been shown to represent up to 10% and 45% of all T cells, respectively [91, 92]. Consistent with their designation as "mucosal-associated" cells, tissue-resident MAITs are also found in significant numbers at mucosal sites and are thus poised as a first line of defense against microbial pathogens [87]. They are able to initiate rapid effector functions upon encountering antigen, a property imbued to MAIT and other innate-like T cells by the PLZF transcription factor [93]. MAIT cells can gain some effector function in the thymus due to upregulation of PLZF (and therefore of RORγt and T-bet), but they undergo further maturation after egress to the periphery [94]. Here, they acquire an effector memory phenotype as implied (for human T cells) by the constitutive cell surface expression of CD45RO and the absence of CD62L [92]. In response to agMR1 (or MR1-independent activation by IL-12 and/or IL-18 [95]), IFN-γ- and TNF-secreting (Th1-like) MAITs dominate the immune response [96]. IL-17- and IL-22-secreting (Th17-like) MAITs do exist, but seem to be more tissue-restricted, such as to reproductive tract cells including the endometrium and cervix [97].

Approximately 80% of human MAITs are $CD8^+$ (mostly $CD8\alpha\alpha^+$), another 15% are $CD4^-$ $CD8^-$, and the final 5% are $CD4^+$ [7]. Although there have been few studies directly relating the influence of MAIT co-receptor usage on agMR1 recognition and effector function, a recent crystallographic study did show that CD8 binds MR1 in a manner analogous to its interaction with MHC-I [98]; this interaction enhanced the cytokine production of MAIT cells in response to RBP-derived antigens, but also was shown to be important for recognition of non-RBP-derived ligands. Further studies, however, will be required to determine the antigen recognition mechanisms and physiological roles of $CD4^-$ $CD8^-$ and $CD4^+$ MAITs. Aside from CD4/CD8 delineation, MAIT cells have also been characterized as those cells having high expression of CD161 (KLRB1), a property shared with natural killer (NK) cells, and, in some contexts, CD26 (DPP4) [96]. Indeed, the original studies on MAIT cells often defined them as $CD4^-$ $CD161^{hi}$ T cells expressing a TRAV1-2^+ TCR [99].

Unfortunately, this phenotypic definition does not encompass all MR1-restricted T cells, nor even all MAIT cells. Indeed, the expression of CD161 can depend on the tissue localization, maturity, and/or activation status of MAIT cells [7], and not all TRAV1-2^+ cells are MR1Ts. Further, considering the abundance of MAIT cells, the $CD4^+$ subset is not insignificant in number. Thus, unfortunately, studies using this definition of MAIT cells may have made conclusions about them based on populations which included non-MR1-restricted T cells and which excluded bona fide MR1-restricted T cells that did not fall into this phenotype, thus obscuring the biology of both MAIT and MR1T cells [7]. To overcome this limitation, recombinant agMR1 complexes have been produced and tetramerized using fluorescently-

labeled streptavidin to generate cell-staining reagents to unambiguously determine T cell antigen specificity in flow cytometric assays [76, 96]. Specifically, canonical MAITs have come to be classified as those T cells which bind MR1 loaded with 5-OP-RU ($MR1^{+5OPRU}$) and do not bind MR1 loaded with 6-FP ($MR1^{+6FP}$) [58, 78]. Thus, more recent studies of MAIT and broader MR1T cells which include MR1 tetramers are and will be more reliable regarding their phenotypic characterization.

1.3.3 MR1 tetramers have revealed heterogeneity in the MR1T cell compartment and allude to MR1T antigen selectivity

In addition to aiding in the unambiguous identification of MAIT cells, $MR1^{+5OPRU}$ and $MR1^{+6FP}$ tetramers have also revealed the existence of non-MAIT, MR1-restricted T cells with unique ligand selectivity. For example, Gherardin et. al. [78] identified an interesting $TRAV21^{+}$ $TRBV5\text{-}6^{+}$ MR1T clone from the peripheral blood of healthy human donors that was $MR1^{+5OPRU}$ tetramer-agnostic but was stained by the $MR1^{+6FP}$ tetramer, implying the existence of 5-OP-RU-agnostic MR1T cells. Importantly, this clone was not stained by an apo ("empty") MR1 tetramer, suggesting it was not autoreactive. 6-FP is not known to be a physiologically-relevant ligand [57] and therefore the role (and *in vivo* antigen reactivity) of this clonotype is unclear. The authors of this study also found a $TRAV36^{+}$ $TRBV28^{+}$ MR1T clone that was stained by the $MR1^{+5OPRU}$ tetramer but not by the $MR1^{+6FP}$ tetramer; due to its gene usage, this TCR did not have the CDR3α Y95 residue and was shown crystallographically to have a unique mechanism of engagement with $MR1^{+5OPRU}$. These studies suggest that MR1Ts exhibit ligand selectivity and therefore, 5-OP-RU may not be a perfect antigen proxy for all MR1-restricted T cells.

A new generation of MR1 reagents has more recently been developed with a heterogeneous mixture of ligands which are thought to be reflective of those which would naturally be loaded into MR1 in a particular biological context [77]. Interestingly, a tetramer of MR1 loaded with diverse *Escherichia coli* antigens ($MR1^{+Ecoli}$) was found to stain (in a donor-dependent manner) a greater proportion of $CD3^{+}CD4^{-}$ T cells than the $MR1^{+5OPRU}$ tetramer, which was attributed, at least in part, to the fact that the $MR1^{+Ecoli}$ tetramer stained a substantially greater population of TRAV1-2^{-} cells [77]. This, again, implies the existence of a distinct population of MR1T cells that are 5-OP-RU-agnostic, though the microbial antigen(s) toward which they are reactive is unknown. Interestingly, when $MR1^{+Ecoli}$ was compared with an analogous tetramer loaded with heterogeneous *Mycobacterium smegmatis* antigens ($MR1^{+Msmeg}$), these two reagents were able to differentially activate four canonical MAIT cell clones, suggesting that unique populations of antigens were loaded into each tetramer and that these MAIT clones were selectively activated by certain ligands [77]. Indeed, this observation agrees with those described by Gold et al. [100] that even within individuals, different bacterial species can elicit activation of unique TCR clonotypes. In addition to selectivity toward microbes, recent observations of MAIT and MR1T cell selectivity toward structurally similar ligands have revealed that this phenomenon does arise at the molecular level [75, 77, 78, 82]. However, the molecular mechanisms underlying this phenomenon are unknown.

In **Chapter 3** of this thesis, I put forward a case study of the differential MR1T cell clone responses toward DMRL, an RBP-generated MR1T agonist, and a synthetic deaza-form of

DMRL ("DZ"). Since the modification between the two ligands should not, based on crystallographic studies of other ligand:MR1-TCR complexes, be accessible to the TCR, we propose that the ligand bound to MR1 influences the pocket dynamics and therefore the molecular landscape that the TCR is able to probe. We suggest that it is this change that underlies the ability of MR1T TCRs to distinguish between MR1 ligands. Additionally, in **Chapter 4** of this thesis, I use single-cell RNA sequencing to explore the clonal diversity and phenotype of 5-OP-RU-reactive cells bearing TRAV1-2^{-} TCRs and of MR1T cells agnostic toward 5-OP-RU but reactive toward tetramers loaded with heterogeneous microbial antigens. Specific emphasis is placed on T cells reactive toward mycobacterial ligands due to our interest in the role of MR1T cells in the immune response toward *M. tuberculosis*.

1.4 The potential for and progress toward MR1/MR1T-mediated therapies or vaccines against Mycobacterium tuberculosis

1.4.1 Eradication of tuberculosis will require more robust, accessible, cost-effective vaccines

One of the most significant infectious diseases plaguing society is tuberculosis (TB), a disease of the lungs caused by *Mycobacterium tuberculosis* (*Mtb*). The number of annual deaths attributed to TB exceed 1.5 million globally and currently fall second only to COVID-19 deaths [101]. Though the global incidence of TB death has decreased over the past twenty years, it seems that the decrease in mortality is primarily driven by better health outcomes for those people co-infected with *Mtb* and human immunodeficiency virus (HIV), while those deaths caused by *Mtb* alone have stayed more or less constant [101]. The situation has been made worse by the rise of multidrug-resistant tuberculosis (MDR-TB), for which there are even fewer effective treatments

due to the strains' circumvention of classical antibiotic treatments [101, 102]. Moreover, the accessibility of TB vaccines and treatments have recently suffered due to redistribution of money and resources following the COVID-19 pandemic [103].

Currently, the only effective vaccine against *Mtb* is the Bacillus Calmette-Guérin (BCG) vaccine, which is a live attenuated strain of *Mycobacterium bovis*, the bovine tuberculosis bacterium. In the 100 years since its first use in humans, BCG has been shown to provide variable protection against a number of mycobacteria, including *M. leprae* (which causes leprosy) and *M. ulcerans* (which causes Buruli ulcer) in addition to *Mtb*, as well as viruses and yeast [104]. However, reports on the efficacy and protective duration of BCG against *Mtb* vary highly depending on the study cohort [104]; thus, there is a significant global effort to develop a new generation of vaccines against *Mtb* that perform more consistently [105].

1.4.2 Determinants of MR1-MR1T cell-mediated protective immunity toward *M. tuberculosis* infection remain unclear

Because MR1T cells recognize a monomorphic - and therefore publicly shared - HLA-related molecule, they - along with some γδ T cells and Type I NKT cells - are considered "donor unrestricted" T (DURT) cells. DURTs present tempting targets for immunotherapies and vaccine strategies because they are HLA allele-independent; thus, a single antigen could be used to target an entire semi-invariant class of T cells that is maintained by all individuals [106, 107]. Indeed, MR1T cells (or, at least, MAIT cells) have been studied extensively in the context of *Mtb* infection of mouse, human, and non-human primates (NHP), though their effect on health

outcomes is still inconclusive. In humans, it has been shown that an intronic single nucleotide polymorphism in the MR1 gene imparts susceptibility to *Mtb* in a cohort of Vietnamese adults [108], that MAIT cells are known to respond to *Mtb* or BCG-infected cells in an MR1-dependent manner [109], and that MAIT cells are less frequent in the blood of adults and children with active TB, implying tissue homing of these cells to the lungs [91, 109, 110]. Together, these observations suggest that MR1T cells may be important players in the adaptive immune response to *Mtb*. Likewise, in NHP models of BCG vaccination, one study has shown that peripheral MAIT cells of macaques display an activated phenotype after inoculation and have a stronger response after re-challenge [111]. However, this same study also showed little evidence of MAIT cell activation in granulomas (the tubercular lesion of the lungs), and a second study found no accumulation of MAIT cells in the lungs of children with active TB [110]. Thus, it is unclear if targeting MAIT cells alone with a vaccination strategy would be sufficient to initiate the immune response necessary for protection. A significant experimental limitation to these findings is that in many of these studies, MAIT cells were again defined on flow cytometric phenotype [112, 113]. Additional non-conventional MR1T cells may therefore be responding in an MR1-dependent manner but their contribution may not be captured. Therefore, the assessment of new *Mtb* vaccine candidates targeting the MR1-MR1T cell axis requires tetramer reagents to verify T cell antigen restriction and to avoid obscuring the contribution of non-MAIT MR1-restricted T cells.

1.5 Significance

As the leading cause of microbial infection-related deaths worldwide, tuberculosis has persisted

as a global health burden. Its eradication will require improvements over the current generation of BCG-based vaccines, yet key determinants of protective immunity toward its causative agent, *Mycobacterium tuberculosis*, remain unclear. Lung-resident lymphocytes likely play a critical role in the clinical outcome following exposure and as such, may provide a potential therapeutic avenue. One of the cornerstones of the antimicrobial adaptive immune response, particularly in mucosal tissues, lies in the interaction axis between MR1 and its cognate T cells. Thus, understanding the antigens presented by MR1 during the course of mycobacterial infection, the identity and phenotype of the T cells restricted thereto, and the molecular mechanisms governing these interactions are critical to harnessing this immune axis for a new generation of *Mtb* vaccines and immunotherapies.

Chapter 2. Metabolomic analysis of MR1 ligands reveals a class of indolyl-ribityllumazines and RBP-independent ligands

2.1 Introduction

As discussed in **Chapter 1**, in the classical T cell immune axis, the HLA molecule, complexed with β_2m, is responsible for presenting peptides to the T cell receptor. HLA molecules display high degrees of polymorphism, and variations between alleles are concentrated most in the peptide binding groove [114]. Therefore, the peptide repertoire of each allele varies depending on the biophysical properties ascribed by those polymorphisms, which then requires that peptide motifs must be determined empirically for each allele [115]. Fortunately, advances in liquid chromatography-tandem mass spectrometry (LC-MS/MS), and the data interpretation thereof, have made it easier to characterize these peptide pools [115, 116]. This is, in part, because peptides are composed of combinations of subunits (amino acids) with characteristic masses and isotope distribution patterns; thus, the mass difference between MS2 fragments of a given MS1 ion can generally be attributed to the addition or subtraction of an amino acid [117]. The substantial effort of defining binding motifs for HLA molecules is therefore due to the allelic diversity that need to be characterized rather than ligand identification [118].

In contrast to classical HLA molecules, MR1 is monomorphic and is highly conserved among mammals [59, 61]. These observations implied that MR1 did not coevolve with certain pathogens, but rather that it has evolved to bind and present a subset of conserved biomolecules

[61]. The biochemical nature of the ligands that MR1 presents remained elusive for nearly ten years after the discovery that MAIT cells are restricted to MR1. In part, determining their identity was challenging because MR1 ligands do not derive from known MHC ligand classes (*i.e.* peptides or lipids) and instead represent a novel category of antigens in the form of small organic molecules. Indeed, as discussed in **Chapter 1**, the first defined biological antigens for MR1 were derivatives of the microbial riboflavin biosynthesis pathway (RBP) [58]. Unlike peptides, for which the secondary structure can be readily inferred from their mass spectrometric data and their origin determined by referencing genome sequences, the characterization of the identity of small molecules is significantly more challenging to solve *de novo* due to the lack of relevant reference databases [119]. While some headway has been made in determining the ligandome of MR1, there is still much work to be done in determining the repertoire of ligands that MR1 presents during microbial infection.

The microbial MR1 ligands that activate MAIT cells that are known to date are found in two general categories, the ribitylpyrimidines and the ribityllumazines, which are formed by the condensation of the RBP intermediate 5-A-RU with small host- and/or microbe-derived metabolites (see **Chapter 1**). In a landmark paper by Corbett et al. in 2014 [58], the authors showed that supernatants from *Lactococcus lactis* RBP enzyme knockout strains cultures variably stimulated a MAIT TCR-expressing Jurkat line. In particular, strains depleted of the two enzymes upstream of 5-A-RU production (*Rib1ab* and *Rib2ab*) were unable to activate the MAIT-Jurkats (as measured by the upregulation of surface CD69). In contrast, the wild-type and *Rib4* knockout (*Rib4* is immediately downstream of 5-A-RU biosynthesis) potently induced

upregulation of CD69. In accordance with these data, they were able to show the depletion of an ion with m/*z* 329.1103 Da from only the *Rib1ab*, *Rib2ab*, and *Rib3* supernatants. Although this ion could represent one of two isomers, each of which has the ability to stimulate MAIT cells (5-OP-RU, a potent agonist and ribitylpyrimidine, or rRL-6-CH_2OH, an agonist and reduced ribityllumazine), the authors argued that 5-OP-RU was the biologically-relevant molecule. This, with other complementary data, was the first evidence for the importance of the RBP in MAIT cell antigen generation.

However, this was a very targeted exploration of MR1 ligands, one which has resulted in the bias of studies toward 5-OP-RU in lieu of other MR1 ligands. However, there is currently no direct evidence that 5-OP-RU can be found in the MR1 binding pocket during microbial infection. While 5-OP-RU can be loaded effectively in recombinant MR1 protein, there are concerns related to the biological relevance of this ligand *in* vivo due to its inherent chemical instability. In the case of extracellular infections, it is unlikely that 5-OP-RU reaches newly synthesized MR1 molecules in the endoplasmic reticulum of host cells before these molecules cyclize. Transport of the ribitylpyrimidine form to the ER seems particularly unlikely if uptake is by passive retrograde transport through acidic endosomal compartments, where the pH (~4.5-6.5 pH units) [120] would be low enough to shift the equilibrium of the reversible ketimine condensation away from 5-OP-RU and toward the cyclized 7-methyl-8-D-ribityllumazine (7-Me-RL) product. Given the potency of 5-OP-RU toward MAIT cells and the functional consequences thereof, the question of whether it can be found in the MR1 pocket during microbial infection is important to address.

In addition to the question of 5-OP-RU occupancy, the observations outlined in **Chapter 1** argue that a more global exploration of MR1 ligands, and the MR1T cells which they restrict, is warranted. In pursuit of this, attempts have been made to use MR1 as a "molecular fishing rod" wherein the protein is used to capture molecules from a complex mixture insomuch that the ligands can then be characterized for their structural and functional properties [121]. In the Adams lab, we have developed several systems to express recombinant MR1 such that it is loaded with a heterogeneous pool of microbial ligands (**Supplemental Figure S2.1**). In one approach, we express a chimeric bovine-human construct wherein the α_1 and α_2 domains - *i.e.* the ligand-binding platform - are human and the α_3 and β_2m domains are bovine, was deemed "human platform" MR1 or "hpMR1" via baculoviral transduction of eukaryotic cells [122]. By co-culturing ("co-infecting") these cells with bacteria, a heterogeneous pool of antigens is

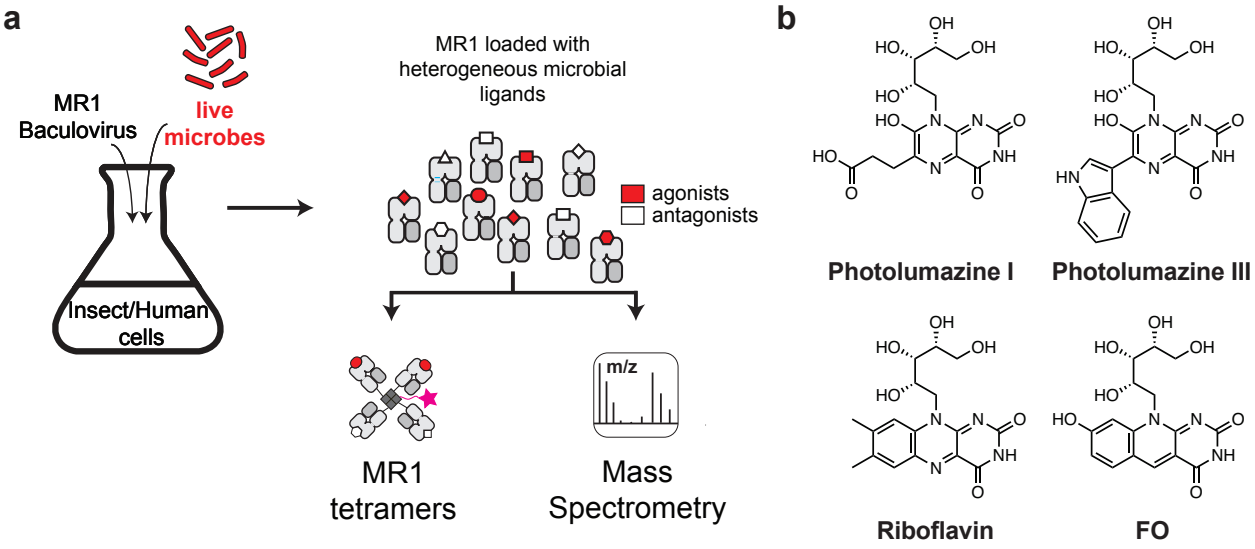


Figure 2.1. Capture of heterogeneous ligands by MR1 from live microbes. a) MR1 is expressed in a professional protein-producing cell line by baculovirus. By omitting antibiotics and adding live microbes, MR1 captures a mixture of ligands, some of which have the capability to stimulate MR1T cells. Protein can be used for mass spectrometric analysis and generation of cell-staining reagents. b) Some of the MR1 ligands characterized through this approach in Harriff *et al.* 2018.

sampled from the bacterial species in question [77] (**Figure 2.1a**). While this system has the flexibility to study MR1 antigens from many bacterial species, we are focused primarily on ligands that arise from *Mycobacteria* due to the implications of the MR1-MR1T cell axis in immunity toward *Mycobacterium tuberculosis* (*Mtb*) (**Chapter 1**). We have primarily used *Mycobacterium smegmatis* (*M. smeg* or *Msmeg*) as a model for *Mtb* due to its genetic similarity to *Mtb* and enhanced safety profile (BSL1 instead of BSL3).

Previously, a protein expression system was established in the lab by Dr. Cara Froyd whereby baculovirus-infected insect cells expressing MR1 were co-cultured ("co-infected") intentionally with *M. smegmatis*. This system generated MR1 protein that captured microbial antigens such that MAIT cells were productively activated from this material [77]. In order to compare the MR1 ligands derived from *M. smegmatis* to those from a different bacterium, she also optimized the co-infection conditions for *Escherichia coli* (e.g. *Ecoli*) due to its evolutionary divergence from *Mycobacteria* and ease of use. Once purified, the ligands bound in the MR1 pocket were then characterized by mass spectrometry by Drs. Curtis McMurtrey and William Hildebrand at the University of Oklahoma Health Sciences Center. After searching for ions structurally related to riboflavin, Drs. Froyd, McMurtrey, Mehaffey (of Dr. Karen Dobos' lab) and Harriff (of Dr. David Lewinsohn's lab) used this pipeline to identify - and later validate - five novel MR1 ligands: riboflavin, hesperidin, 7,8-didemethyl-8-hydroxy-5-deazariboflavin (cofactor FO), photolumazine I (PLI), and photolumazine III (PLIII), which were loaded to different extents in the MR1 binding groove (**Figure 2.1b**) [77]; whereas riboflavin and hesperidin (not shown) were shown to be antagonistic toward the four MAIT clones tested, the photolumazines selectively

activated two of the five MAIT clones tested. Altogether, this study resulted in a novel MR1 reagent that is loaded with molecules that are more likely to be reflective of what MR1 experiences *in vivo*, and developed a mass spectrometric pipeline for ligand identification.

Following this work, one of the outstanding questions was whether the host-microbe interactions and metabolic environment of the insect cell protein expression system was generating ligands that are relevant in mammals. In this chapter, I describe the parallel development of mammalian-produced recombinant MR1 reagents and the mass spectrometric analysis of the ligands captured within these reagents, as well as a reassessment of our previous data sets to mine for additional MR1 ligands. I show that a class of indolyl-ribityllumazines is abundant in only the bacterially-loaded protein and is dependent on several enzymes of the RBP for production and capture by MR1. Further, I describe the validation of molecular isomers of one of the synthetically-achievable ligands, PLV, for its ability to stabilize MR1 on the cell surface and to activate a MAIT clone. For the ligands which could not be synthesized, I describe my attempts to isolate and characterize these compounds from fractionated bacterial supernatants. Beyond insights into the ligand repertoire presented by MR1 in the context of unique bacterial infections, this study also contributes to the growing body of evidence toward MR1-restricted T cell antigen selectivity (discussed further in **Chapter 3**) and suggests that a class of ligands exist that are independent of 5-A-RU, the key metabolite that is thought to give rise to all MAIT ligands.

2.2 Recombinant MR1 made in eukaryotic cells captures heterogeneous bacterial ligands

To follow up on the discovery of the photolumazine compounds identified by Harriff et al., I

translated this work into a mammalian expression system to compare the ligand repertoire to that of the insect cells. To transition the MR1 expression system to a mammalian host, I employed a modified baculovirus strategy (BacMam) to express transgenes in transduced mammalian cells [123]. Specifically, I used this system to transduce *HEK293T* or *Expi293F* cells to produce a single-chain, secreted construct of human platform MR1 (*hs*-hpMR1, where the *hs* prefix indicates that the protein was made in a *Homo sapiens* cell line) and purified the protein in a manner similar to the insect cell protein (*tn*-hpMR1, where the *tn* prefix indicates that the protein was made in the *Trichoplusia ni* Hi-5 cell line) (**Supplemental Figure S2.2**) [77, 122].

I first determined whether this system could capture MR1 antigens when they were added to the culture medium during the course of protein expression. To do so, I doped DMRL (isolated from bacterial supernatants in-house) [124] into the baculovirus-induced cultures, then purified the protein and assessed it for its ability to stimulate MAIT cell clones in an IFN-γ Spot assay. (A description of the various "Spot" assays, which are variations of the EliSpot assay, used in this thesis can be found in the **Materials and methods**.) For these experiments and all other activation assays in this thesis, I used a panel of MAIT clones generated in the lab of Drs. Deborah and David Lewinsohn (Oregon Health and Sciences University) that were originally expanded on *Mtb* and are MR1-restricted. I found that *hs*-hpMR1 purified from cultures doped with increasing concentrations of DMRL were capable of stimulating MAIT clones and that the activation was dose-dependent, indicating that the *Expi293F* cells are capable of ligand uptake and transfer to MR1 (**Figure 2.2a**). In contrast, when riboflavin, an MR1 ligand that does not stimulate any of the clones in our MAIT cell panel, was added to culture, there was no activation

induced at any concentration of riboflavin (**Figure 2.2a**). Hence, hpMR1 produced by baculoviral transduction in mammalian cells was able to capture ligands added to the culture medium and specifically activated MAIT clones only when endogenously loaded with antigens which were stimulatory toward these MAIT clones.

Since endogenous loading of known antigens was successful, I then proceeded to express *hs*-hpMR1 in *Expi293F* cells co-incubated with *M. smegmatis* to see if I could capture microbial ligands in this manner, similar to what we had done with *T. ni* cells [77, 122]. I found that adding

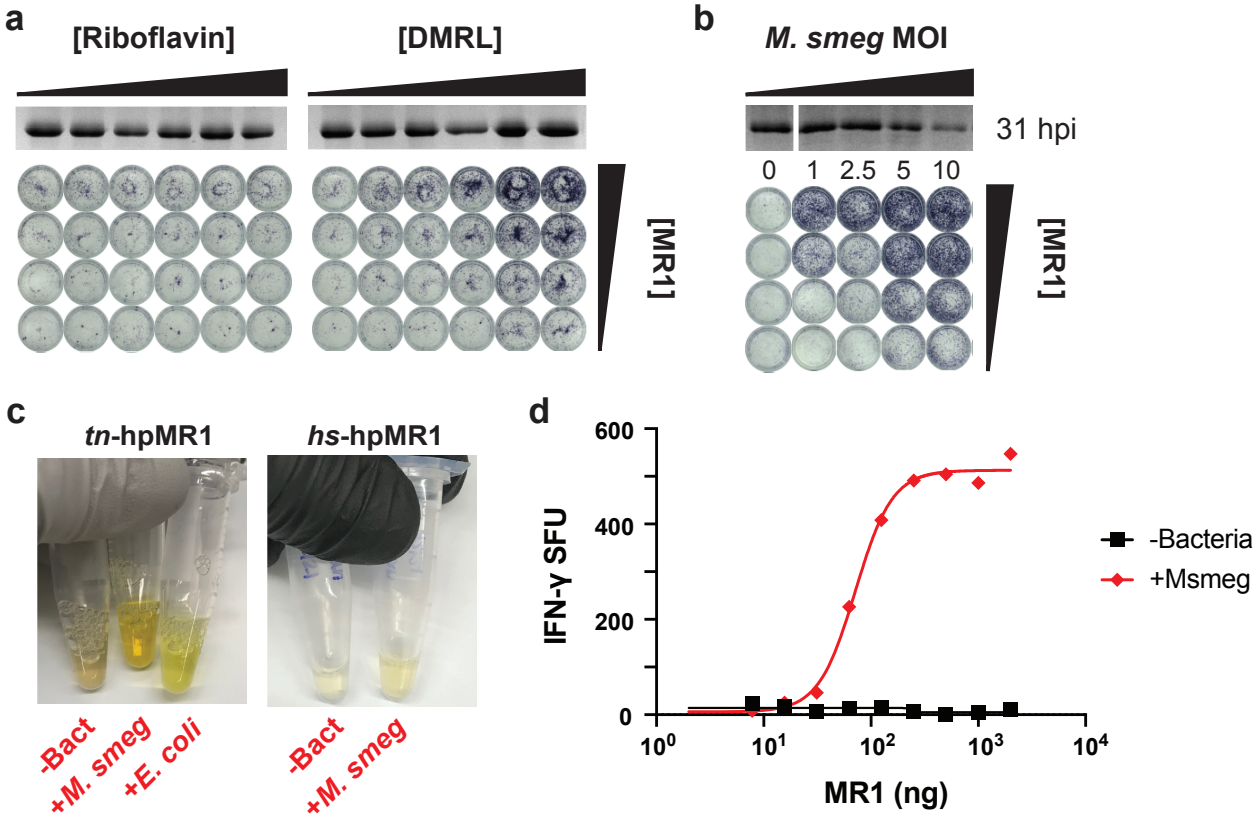


Figure 2.2. MR1 produced in *Expi293F* cells is capable of capturing antigens from its environment. a) ExSpot measuring the IFN-γ production of MAIT cell clones in response to small-scale preparations of MR1 loaded with increasing concentrations of riboflavin or DMRL. b) Optimization of bacterial multiplicity of infection (MOI) used for the co-infection, as determined by ExSpot-mediated activation of MAIT cell clones. c) Purified hpMR1 made recombinantly in the presence of bacterial antigens shows a distinct yellow coloration likely indicative of the ligands that are captured in the pocket. d) Activation of MAIT clones by *hs*-hpMR1^{-Bact} and *hs*-hpMR1^{+Msmeg}) tetramers by ExSpot, using IFN-γ production as the readout.

M. smegmatis at an MOI of 50 (bacteria:*Expi293F*) at 31 hours past baculoviral induction, I was able to capture bacterial ligands that activated MAIT clones without significantly sacrificing protein yield (**Figure 2.2b**). After optimizing the purification scheme, I was able to yield approximately 2-3 mg/L *Expi293F* culture of the mammalian *M. smegmatis*-loaded protein (*hs*-hpMR1^{+Msmeg}) or its paired negative control, (*hs*-hpMR1^{-Bact}) (**Supplemental Figure S2.2**). The purified protein showed a slight yellow coloration, likely due to the presence of *M. smegmatis*-derived ligands; we also see this effect for purified *tn*-hpMR1 produced in the presence of bacteria, but the effect is much more pronounced (**Figure 2.2c**) Using plate-bound tetramers of MR1, I found that *hs*-hpMR1^{+Msmeg} was able to stimulate MAIT clones, as determined by their production of IFN-γ, whereas the paired control was unable to do so (**Figure 2.2d**). Thus, MR1 expressed in mammalian cells is capable of capturing stimulatory antigens from *Mycobacterium smegmatis*.

I then tried to further characterize the molecular interactions between MR1 and MAIT TCRs. To do so, I first used biolayer interferometry (BLI) to determine the dissociation constant (K_D) of the interaction between *hs*-hpMR1^{+Msmeg} and several MAIT TCRs that we produce recombinantly in the lab. *hs*-hpMR1 control protein was unable to bind any of the TCRs at any concentration (**Figure 2.3a**). While the signal was weak, the *M. smegmatis*-loaded protein did bind to the each of the TCRs tested (**Figure 2.3b**), and the fact that the unloaded protein did not bind the TCRs indicated that the small binding that we did see for the loaded protein was from a specific interaction conferred by the ligand in the MR1 pocket and not autoreactivity. I couldn't determine explicit affinities despite titrating the analyte to 32-64 μM, which is well above the

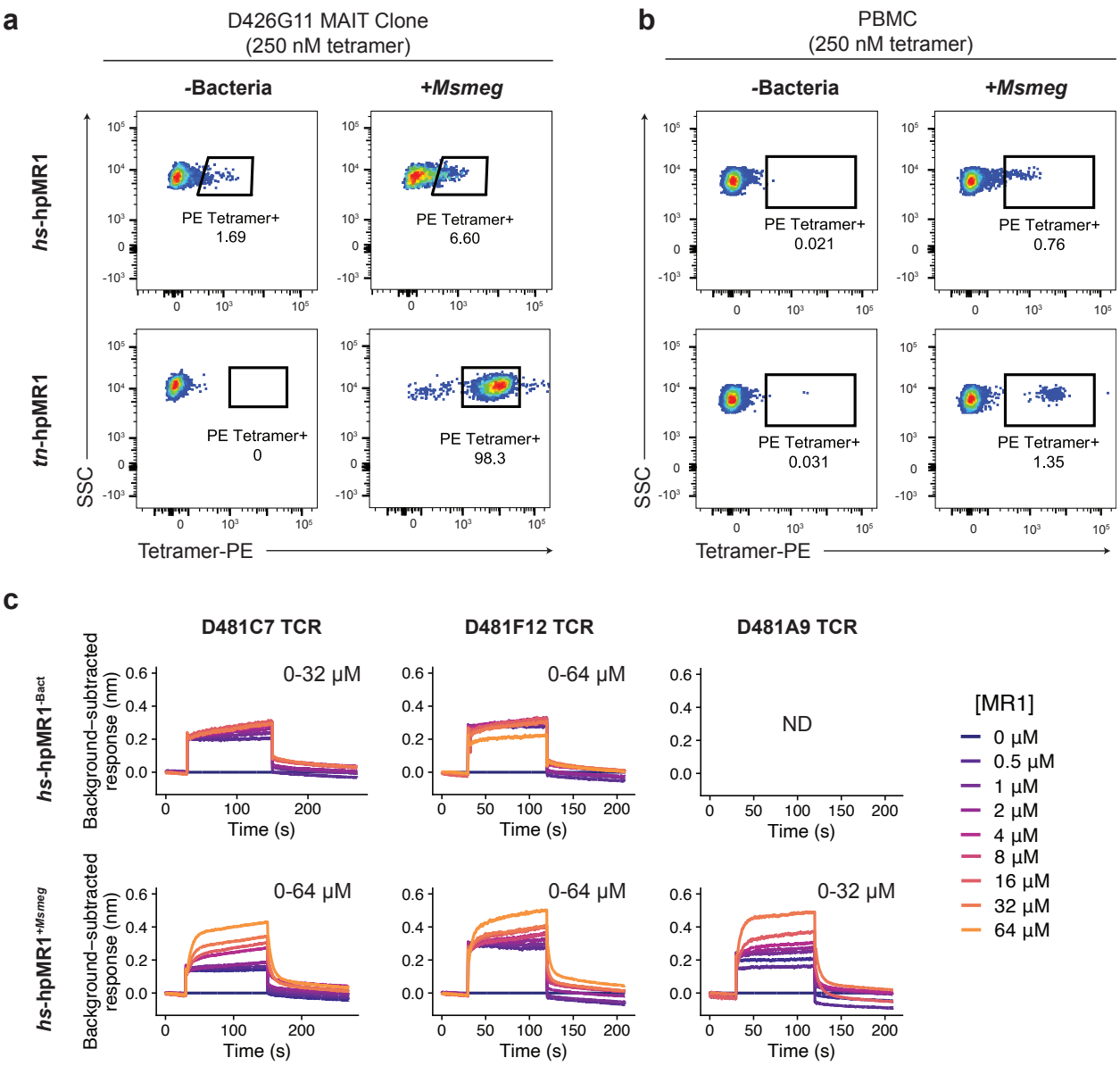


Figure 2.3. Comparison of the biophysical and functional characterization of *hs*-hpMR1 and *tn*-hpMR1 proteins loaded with *M. smegmatis* antigens. a-b) Comparison of staining of a human MAIT clone (a) or human PBMC (b) with *hs*-hpMR1 and *tn*-hpMR1, both in their loaded and unloaded formats. c) Comparison of *hs*-hpMR1$^{\text{-Bact}}$ and *hs*-hpMR1^{+Msmeg} binding to recombinant MAIT TCRs.

affinity that I had expected based on the 1-10 μM K_D interaction between the same TCRs and *tn*-hpMR1 loaded with *M. smegmatis* antigens (**Supplemental Figure S2.3a-b**). This result was somewhat surprising because the protein was clearly able to robustly activate T cells on the same order of magnitude as the insect cell-derived protein (**Supplemental Figure S2.3a-b**). This

suggested to me that at least some of the mammalian reagent is loaded with stimulatory ligand but that the process of ligand loading may be less efficient in the mammalian cells than in the insect cells. While it's possible that stimulatory ligands do not confer a strong interaction with TCR, this model would not align with observed trends of other MR1-MAIT TCR interactions [8, 63, 78, 89]. Because of the limitations to the *in vitro* binding experiment, I went on to validate the loading of the microbial antigens by using this reagent to stain clonal and primary T cells from the peripheral blood of healthy human donors. Consistent with the observation of a low affinity of *hs*-hpMR1^{+Msmeg} for the MAIT TCR, I found that the staining of a MAIT cell clone by fluorophore-labeled *hs*-hpMR1^{+Msmeg} tetramers required a significantly higher concentration of reagent (at minimum, 500 nM - 1 μM) in comparison with *tn*-hpMR1^{+Msmeg} (approximately 50-100 nM), and even at this concentration, the *hs*-hpMR1^{+Msmeg} tetramer never saturated the MAIT clone (**Figure 2.3c**). In contrast, PBMC-derived T cells were stained at the same concentration of the mammalian and insect cell tetramers (**Figure 2.3d**), though this is likely due to their relative infrequency (on average, around 1% of all T cells) in PBMC. In all, these data suggest that *hs*-hpMR1^{+Msmeg} is loaded with ligands derived from *M. smegmatis* and can robustly stimulate MAIT cell clones despite apparent low ligand occupancy. Thus, we began to investigate the identity of these ligands using mass spectrometric and metabolomics approaches.

2.3 Metabolomic analysis recombinant MR1 loaded with microbial antigens reveals several ions of interest

Since our data suggested that *hs*-hpMR1^{+Msmeg} was loaded with microbial ligands capable of stimulating MAIT clones, I generated three independent biological replicates of *hs*-hpMR1^{+Msmeg}

(and *hs*-hpMR1^{-Bact}) and verified that they were able to stimulate MAIT clones (**Supplemental Figure S2.4**). Following this validation, each sample was sent to my collaborator Jason Krawic, who is a graduate student in the lab of Dr. William Hildebrand at the University of Oklahoma. Jason performed LC-MS/MS as Dr. McMurtrey had previously done for loaded *tn*-hpMR1 samples [77]. Briefly, he injected 10 μg of intact *hs*-hpMR1 for low pH reverse-phase nano-scale LC-MS/MS. The chromatography was achieved using a C18 column developed with an acetonitrile gradient, and ion spectra were then collected in data-dependent acquisition mode in negative ion polarity. The reason for injecting intact *hs*-hpMR1 (in lieu of extracting the ligands) is to prevent biasing the detectable ions away from labile compounds which are unstable outside of the MR1 groove. Each biological replicate of the matched *hs*-hpMR1^{+Msmeg} and *hs*-hpMR1^{-Bact} proteins were injected in triplicate. Collecting data on several biological replicates provided us with increased statistical power to compare the ions from the loaded and unloaded protein, power that we don't have for the insect cell-derived protein, since only one biological replicate was used. Given that mass spectrometry is so sensitive, removing background ions is an absolute necessity, else there are too many ions to sift through manually. Thus, we mined the data for ions that were highly enriched in the *M. smegmatis*-loaded protein, manually validated the data, and then used metabolomics approaches to find other ions with similar MS2 spectra to these molecules.

First, we generated a volcano plot to determine enrichment of ions in *hs*-hpMR1^{+Msmeg} in comparison with its unloaded (*hs*-hpMR1^{-Bact}) counterpart. For each ion found to be highly and significantly enriched (in the upper-right quadrant of the volcano plot), we manually validated

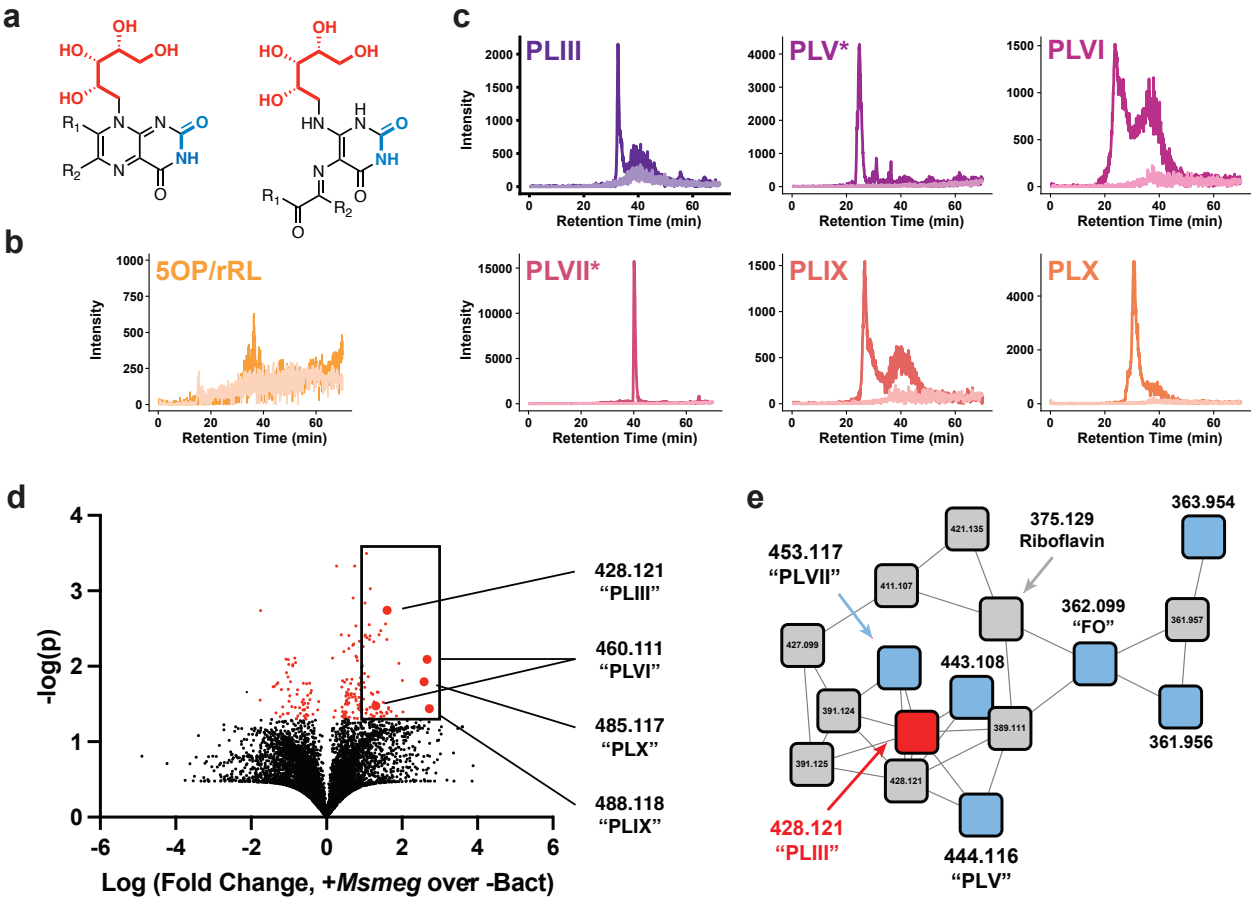
the parameters listed below in order to justify continuing study on them. We also analyzed the ions with a large fold-change over the unloaded protein but which did not reach a significant *p* value, but none of these ions ended up meeting our criteria. The requirements for inclusion on our shortlist included:

1. Must be singly charged (e.g. [M-H]-).
2. Must have a high-quality MS2 spectrum.
3. Must have the fragmentation hallmarks indicative of the ribityllumazines and ribitylpyrimidines (**Figure 2.4a**):
    a. A fragment of 41.9987 Da, which represents the amide in the uracil backbone of the pyrimidine/lumazine/flavins
    b. A neutral loss of 134.0549 Da, which represents the loss of the ribityl moiety from the molecule.
4. Must be absent from all nine *hs*-hpMR1^{-Bact} technical replicates.
5. Must be present in all three technical replicates of a biological replicate, and in at least two of the biological replicates.
    a. Must have a distinct peak in the extracted ion chromatogram (EIC) to be considered "present" (**Figure 2.4b-c**)

Only four ions deriving from the volcano plot fulfilled our criteria, m/*z* = 428.1212, m/*z* = 460.1110, m/*z* = 488.1048, and *m/z* = 485.1175 (**Figure 2.4d**). The first, photolumazine III (PLIII) (**Figure 2.1c**), was described and validated as an MR1 ligand by Harriff et al. from their

mass spectrometric analysis of the insect cell-generated reagent (more highly enriched in *tn*-hpMR1^{+Msmeg} but was also present in *tn*-hpMR1^{+Ecoli}) [77]. The identification of this ion was relatively straightforward because PLIII was isolated and characterized over 50 years ago from *Achromobacter petrophilum* supernatants [125]. The latter three ions were not described as MR1 ligands in the previous study, although in retrospect, each of these ions were found in the *tn*-hpMR1 sets as well, which lent us more confidence that these were relevant ligands to pursue (**Supplemental Figure S2.5**). We deemed these molecules photolumazine VI (PLVI), photolumazine IX (PLIX), and photolumazine X (PLX), respectively. Interestingly, the ion corresponding to 5-OP-RU/rRL-6-CH_2OH (m/*z* 329.1103 Da) was not found in this upper quadrant; indeed, the MS1 intensity was so low that it was not selected for fragmentation (**Figure 2.4c**). We therefore focused on PLVI, PLIX, and PLX for further characterization.

We then searched for other ions that were structurally related to PLVI, PLIX, and PLX using Global Natural Product Social (GNPS) Molecular Networking [126, 127], which creates networks of ions based on the similarity of their MS2 fragmentation patterns. Of note, this analysis therefore requires that a given ion was abundant enough to have been selected for fragmentation in data-dependent acquisition mode. In the network generated from the *hs*-hpMR1^{+Msmeg} (and matched controls) alone, PLVI, PLIX, and PLX did not cluster with any other ions, and PLIII is associated only with an ion of m/*z* 391.1250, which we know from previous work to be 8-hydroxymethyl-riboflavin (**Supplemental Figure S2.6a**) [77]. Since these ions all had hallmarks of ribityllumazines, we were surprised to see that they weren't clustered together. Therefore, we attempted to better populate the networks by integrating the *hs*-hpMR1 and *tn*-



f

		Present	Clustered	Enriched	Present	Clustered	Enriched	Present	Clustered	Enriched
Ion	m/z	*tn*-hpMR1^{+Ecoli}			*tn*-hpMR1^{+Msmeg}			*hs*-hpMR1^{+Msmeg}		
PLIII	428.1212	+	+	Insufficient number of biological replicates	+	+		+	+	+
PLV	444.1161	+	+		+	+	+			
PLVI	460.1110	+			+		+	+		+
PLVII	453.1153	+	+		+	+	+			
PLIX	488.1062/442.1001	+			+		+	+		+
PLX	485.1175	+			+			+		+

Figure 2.4 Prioritization of ions eluted from *hs*-hpMR1 for structural characterization. a) Ribityllumazine (left) and ribitylpyrimidine (right) showing the 41.9987 Da fragment corresponding to the uracil moiety (blue) and the neutral loss of 134.0549 Da, indicative of the ribityl moiety (red). b-c) EICs of relevant ions from *hs*-hpMR1^{+Msmeg} or *tn*-hpMR1^{+Msmeg}. Dark lines indicate the *M. smegmatis*-loaded protein, while the lighter lines indicate the unloaded matched control. Asterisks indicate the EIC is from *tn*-hpMR1 preps. d) Volcano plot showing log fold change of ions in the *M. smegmatis*-loaded protein relative to the unloaded negative control. Points in red are statistically significant, and the box indicates ions with greater than 10-fold enrichment in *hs*-hpMR1^{+Msmeg} which are also statistically significant. The two points for PLVI indicate two different retention times. e) GNPS network containing PLIII. Gray, found in unloaded MR1 or only one replicate of *hs*-hpMR1^{+Msmeg}; blue, found in loaded *tn*-hpMR1 protein but not in unloaded MR1; red, found in loaded *tn*-hpMR1 and *hs*-hpMR1 but not unloaded MR1. f) Summary figure describing where each ion was identified in the mass spectrometric data analysis.

hpMR1 data sets. Still, PLVI, PLIX, and PLX clustered on their own (**Supplemental Figure S2.6b**). PLIII, however, clustered with several ions that were specifically found in the bacterially-loaded protein, as well as with riboflavin (**Figure 2.4e**), as was seen in the networks produced by Harriff et al. These ions showed an m/*z* of 444.1160 Da, which we named photolumazine V (PLV) and an m/z of 453.1153, which we named photolumazine VII (PLVII). These ions were found only in the *tn*-hpMR1^{+Msmeg} and *tn*-hpMR1^{+Ecoli} data sets (**Figure 2.4e**), not in MR1 produced in the mammalian system. However, since their MS2 spectra showed hallmarks of the ribityllumazine/pyrimidine/flavin compounds, we also pursued these ions. Altogether, these data demonstrated that PLIII, PLV, PLVI, PLVII, PLIX, and PLX were specifically found to be highly enriched in the microbially-loaded recombinant MR1 (**Figure 2.4f**) and seemed to be ribityllumazine-like compounds and therefore we chose to pursue these ions for functional and structural characterization.

2.4 The ribityllumazine profile of bacterially-loaded hpMR1 is distinct to the bacterium

We then went on to quantify the relative abundance of each ion and to ask whether the protein expression system and the bacterial species affected the profile of ions captured by MR1. To make these comparisons, we determined the AUC of the extracted ion chromatogram (EIC) for known and novel bacterial MR1 ligands from both the *hs*-hpMR1^{+Msmeg} and *tn*-hpMR1^{+Msmeg} data sets (**Supplemental Figure S2.7**). While these are not absolute quantifications, we are able obtain relative abundances because we used the same mass of protein for each injection. To determine the ligand profile of each protein, we normalized each ligand's average AUC to the sum of the average AUCs. Interestingly, the profile of the ribityllumazines in the *M. smegmatis*-

loaded protein was nearly identical in the *tn*-hpMR1 and *hs*-hpMR1 preps (**Figure 2.5a**), indicating that the expression system did not bias the loading of the known ligands relative to one another. However, if we consider the raw AUCs rather than the normalized profile, it is very clear that the overall ion intensity is significantly lower in the *hs*-hpMR1 preps despite loading the same amount of protein onto the column (**Figure 2.5b**), which supports my hypothesis that

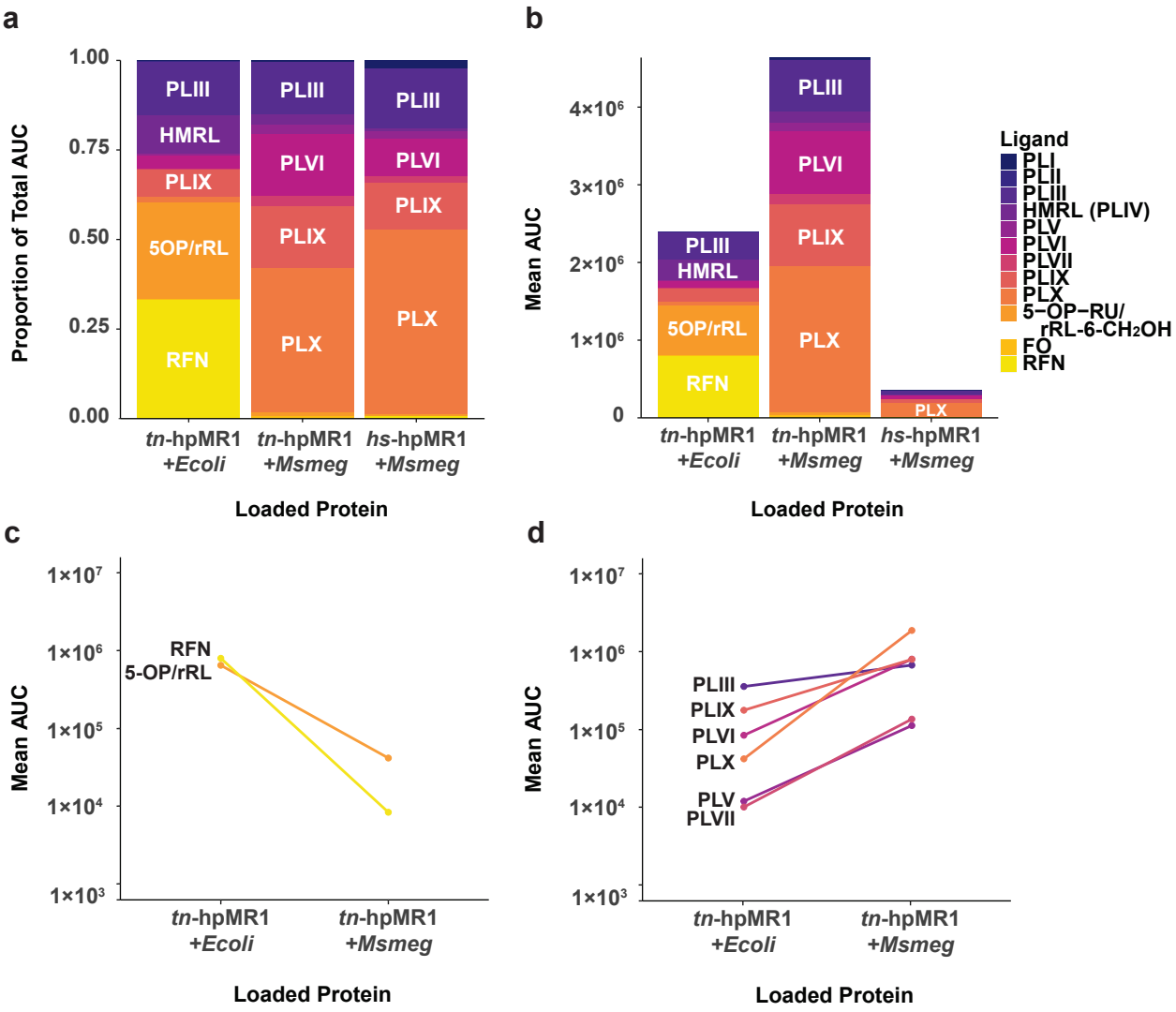


Figure 2.5. Ribityllumazine abundance across preparations and between microbes. a) Ribityllumazine profile (average AUC for each ion as a percent of the sum of the average AUCs) for each of the *tn*-hpMR1$^{+E.\ coli}$, *tn*-hpMR1^{+Msmeg}, and *hs*-hpMR1^{+Msmeg}. b) Average ribityllumazine abundance (not normalized) for each of the three proteins. c, d) Average AUC for 5-OP-RU/rRL-6-CH_2OH and riboflavin (c) and photolumazines (d) compared between in insect cell-generated proteins.

the mammalian cell-generated protein has lower bacterial ligand occupancy.

We then analyzed the data for *tn*-hpMR1^{+Ecoli} collected by Harriff et al. and compared it to that of *tn*-hpMR1^{+Msmeg} to look for differences in the MR1 profile of known ligands when different bacteria were used in the co-culture system. Strikingly, we found that the profiles of the *tn*-hpMR1^{+Msmeg} and *tn*-hpMR1^{+Ecoli} proteins varied significantly from one another, with the *E. coli*-loaded protein being dominated by the 5-OP-RU/rRL-6-CH_2OH and riboflavin ions and the *M. smegmatis*-loaded protein being skewed more toward the photolumazines (**Figure 2.5a**). When considering the non-normalized average AUCs, the photolumazines are significantly higher in the *M. smegmatis*-loaded protein; this effect is clearest for PLVI ($p = 0.0348$), PLIX ($p = 0.0127$), and PLX ($p = 0.0154$) (**Figure 2.5d**). In contrast, the 5-OP-RU/rRL-6-CH_2OH ion is 1-1.5 orders of magnitude less abundant in *tn*-hpMR1^{+Msmeg} than it is in the *E. coli*-loaded protein, but this effect is not significant ($p = 0.1159$) (**Figure 2.5c**). Additionally, riboflavin is significantly enriched in the *E. coli*-loaded protein ($p = 0.0196$); this observation is emboldened by the fluorescent profile of ligands in the MR1 pocket on a TLC. Indeed, only the *tn*-hpMR1^{+Ecoli} protein (or material eluted from this protein by boiling and pelleting the aggregate) shows the yellow fluorescent spot indicative of riboflavin (**Supplemental Figure S2.7c**). This experiment also showed a large blue fluorescent species specifically in the ligands eluted from *tn*-hpMR1^{+Msmeg}, though the identity of this species is unknown. Collectively, these data suggest that the MR1 ligand repertoire is tuned to the bacterium present in a more adaptive manner than previously appreciated. These observations may explain the MAIT cell discrimination between *E. coli*- and *M. smegmatis*-loaded MR1 observed by Harriff et al.: MR1T cells with the ability to

discriminate between ligands would react differently toward MR1 loaded with distinct profiles of metabolites. Further discussion of MR1T cell selectivity can be found in **Chapter 3**. Collectively, these data suggest that the MR1 ligand repertoire is tuned to the bacterium present in a more adaptive manner than previously appreciated.

2.5 Knockouts of the enzymes of the riboflavin biosynthetic pathway reveal the absolute dependence of MAIT cell agonism on *Rib2ab* but not on *Rib1ab*

We were then interested in whether these photolumazines arose from the riboflavin biosynthetic pathway, and how manipulation of this pathway in *Mycobacteria* would modulate the response by MAIT and MR1T cell clones. In *Mycobacteria*, *Rib1ab* (RibAB, MSMEG_3072), *Rib2ab* (RibD, MSMEG_3067), and an uncharacterized phosphatase (*Rib3*) are the enzymes upstream of 5-A-RU, *Rib4* (RibE, lumazine synthase, MSMEG_3073, MSMEG_6598) and *Rib5* (RibE, riboflavin synthase, MSMEG_3071) are downstream of the key intermediate [128] (**Figure 2.6a**). Dr. Melissa Chengalroyen, in the labs of our collaborators Drs. Valerie Mizrahi and Digby Warner at the University of Cape Town, generated *M. smegmatis* and *M. tuberculosis* strains from which she had knocked out and subsequently complemented the riboflavin biosynthetic enzymes. Melissa, along with Nurudeen Oketade, Dr. Carolina Mehaffy, and Megan Lucas of Dr. Karen Dobos' lab at Colorado State University, validated the knockouts by proteomics and whole genome sequencing (Chengalroyen et al., *in preparation*). These strains were also characterized for their ability to activate MAIT clones in an EliSpot assay; consistent with their location downstream of 5-A-RU, the Δ*Rib4/4O*[14], and Δ*Rib5 M. smegmatis* strains were able to stimulate

[14] Δ*Rib4/4O* refers to the knockout of *Rib4* and a putative *Rib4* ortholog (*Rib4O*)

MAIT clones at the same levels as wild-type (**Figure 2.6b**). In contrast, knockout of *Rib1ab* or *Rib2ab* genes from *M. smegmatis* strains completely ablated the ability to activate MAIT clones (**Figure 2.6b**). As expected, the complement strains, wherein the knockout strains were complemented with a plasmid carrying the knocked-out gene, were able to rescue the activity of MAIT clones toward each of the Δ*Rib1ab* or Δ*Rib2ab* strains. In contrast to the activation results for the canonical MAIT clones, the TRAV12-2^{+} D462E4 clone appeared to be affected by the knockout of *Rib5,* producing less IFN-γ in response to this knockout strain, and this effect was rescued by the corresponding complement strain. This was surprising as *Rib5* acts at the final step of the riboflavin biosynthetic process (well after 5-A-RU), converting 6,7-dimethyl-ribityllumazine (DMRL) to riboflavin. While riboflavin has been shown to bind MR1 *in vitro* - as determined by its ability to compete with and therefore deplete the activity induced by a MAIT cell-stimulating ligand (Harriff et al. 2018) - it is not known to stimulate MAIT or MR1T clones, including the clones used in this panel. Thus, the ligand(s) which stimulate the D462E4 clone which are dependent on *Rib5* for their biosynthesis are unclear. Similar to this clone, the D520E10 clone also showed a decreased response to *Rib4* and *Rib5* knockout. In these cases, the corresponding complement strains were unable to rescue the activity to wild-type levels (**Figure 2.6b**); the reason for this effect is unclear. Together, these data are consistent with the model wherein 5-A-RU is the key intermediate responsible for generation of canonical MAIT activating ligands, however suggests that non canonical MR1T clones are activated by other RBP-derived ligands that are dependent on *Rib4* and/or *Rib5* for their biosynthesis.

In order to validate these observations, the Dobos lab also generated lysates from these strains

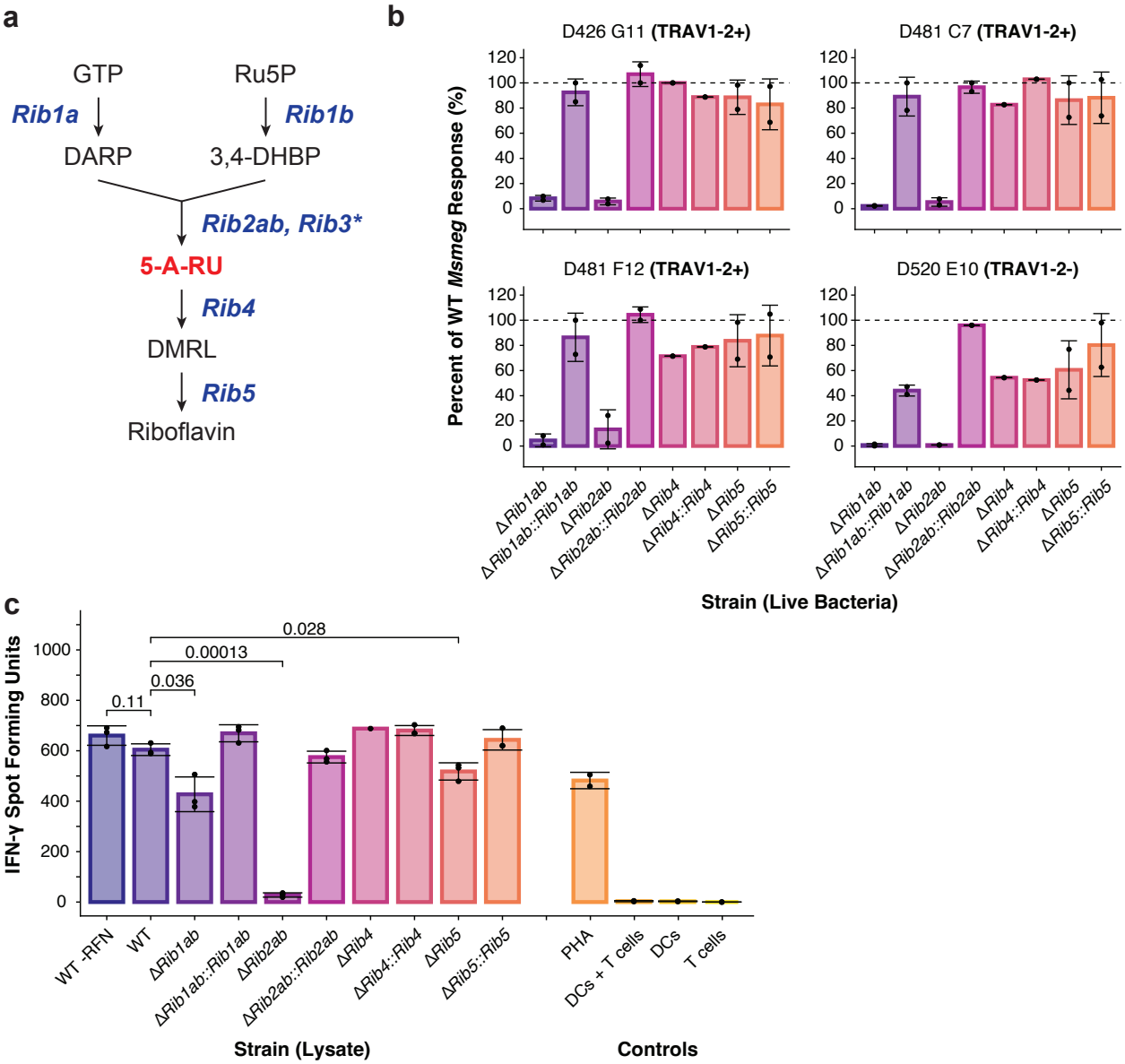


Figure 2.6. Activation of MAIT clones by *M. smegmatis* RBP enzyme knockouts. a) Riboflavin biosynthetic pathway. Rib3* indicates the unidentified phosphatase. b) Activation of MAIT and MR1T clones by the WT and RBP knockout strains (and relevant genetic complements) of *M. smegmatis*. The TRAV gene usage is provided alongside the clone name. c) Activation of the D426G11 MAIT clone the lysates of WT and RBP knockout strains (and relevant genetic complements) of *M. smegmatis*. The *p*-value (as determined by an unpaired *t*-test, n=3 observations per strain) is displayed above the bars indicating relevant statistical comparisons.

for use in an EliSpot with one of the MAIT clones, D426G11. When they were tested, we found that Δ*Rib5* and ΔRib*4/Rib4O M. smegmatis* strains were able to activate the clone at wild-type levels, and the Δ*Rib2ab* strain ablated activation, all consistent with what was seen in the EliSpot

with live bacteria (**Figure 2.6c**). However, the lysate from the Δ*Rib1ab* strain was still able to activate the D426G11 clone in this assay, though not to the same extent as wild type strains (**Figure 2.6c**). This was surprising, and it suggested that some MR1 ligands could arise through a *Rib1ab*-independent, *Rib2ab*-dependent manner. These lysates are much more concentrated and less complex than adding in live bacteria, so the higher specific activity seen for the lysate may be explained by ligand enrichment. Based on these data, I then hypothesized that *hs*-hpMR1 loaded with these strains via the co-infection system would bear similar activation patterns to the lysate, as *hs*-hpMR1 is similarly enriched for ligands in comparison with live bacteria.

I therefore pursued the loading of these strains into *Expi293F* cells as I had done for the wild-type (WT) *M. smegmatis* strain. Ablation of riboflavin biosynthetic genes (except lumazine synthase) is lethal in *M. smegmatis* since riboflavin is an essential vitamin and precursor to enzymatic cofactors FMN and FAD. Thus, in order to co-culture the knockout strains with *Expi293F* cells, I added exogenous riboflavin to the culture media in order for the bacteria to survive (Chengalroyen et al., *in preparation*). In order to control for this additional variable, I first prepared mid-scale cultures of *hs*-hpMR1^{-Bact} and *hs*-hpMR1^{+Msmeg} under both high riboflavin and control (no excess riboflavin) conditions. The addition of riboflavin decreased the potency of the *hs*-hpMR1^{+Msmeg} slightly (as determined by its maximum response and the EC_{50} (**Figure 2.7a**), which reflects what we saw of the bacterial lysates from WT *M. smegmatis* grown high riboflavin (**Figure 2.6c**). As expected, the addition of excess riboflavin had no consequence on the response of MAIT cell clones toward *hs*-hpMR1^{-Bact}. I then prepared *hs*-hpMR1$^{+MsmegΔRib1ab}$, *hs*-hpMR1$^{+MsmegΔRib2ab}$, and *hs*-hpMR1$^{+MsmegΔRib5}$, all under high-riboflavin

conditions. *hs*-hpMR1$^{+MsmegΔRib2ab}$ was unable to activate MAIT clones in any capacity and the *hs*-hpMR1$^{+MsmegΔRib5}$ protein activated at - if not above - wild-type levels (with high-riboflavin conditions) (**Figure 2.7b**). Surprisingly, when measured by tetraSPOT, we observed a small amount of signal for *hs*-hpMR1$^{+MsmegΔRib1ab}$, which we didn't expect because this enzyme precedes the 5-A-RU intermediate (**Figure 2.7b**) (at 400 ng, the significance between *hs*-hpMR1$^{+MsmegΔRib1ab}$ and *hs*-hpMR1^{-Bact} was 0.0155 by an unpaired *t*-test; n=3 observations per protein). Because this effect was small, I confirmed this by ExSPOT, which is more sensitive, though it also increases the background activation seen by *hs*-hpMR1^{-Bact}. Consistent with the results from the activation of MAIT clones with strain lysates, I saw residual activation of MAIT cell clones by this protein (**Figure 2.7c**). We believe that this higher specific activity is due to a higher concentration of relevant MR1 agonists/ligands in the lysate and MR1 protein relative to an entire bacterium. Thus, the activation of MAIT clones by lysates of and MR1 loaded with RBP knockout strains of *M. smegmatis* reveal that at least some MR1 ligands with MAIT cell-activating potential are *Rib1ab*-independent, but that they all seem dependent on *Rib2ab*. The effect on non-canonical TRAV1-2^{-} MR1T cells has yet to be explored.

To complement the activation results of the RBP knockout strains, I also sent the same recombinant MR1 protein produced from co-culture with the wild type and KO *M. smegmatis* strains (with exogenous riboflavin added) to Jason Krawic for characterization by LC-MS/MS. Unsurprisingly, the introduction of riboflavin into the culture media during co-infection with wild-type *M. smegmatis* led to an increased abundance of riboflavin in the MR1 pocket. However, the addition of riboflavin ablated the - (albeit already low) - intensity of the 5-OP-RU/

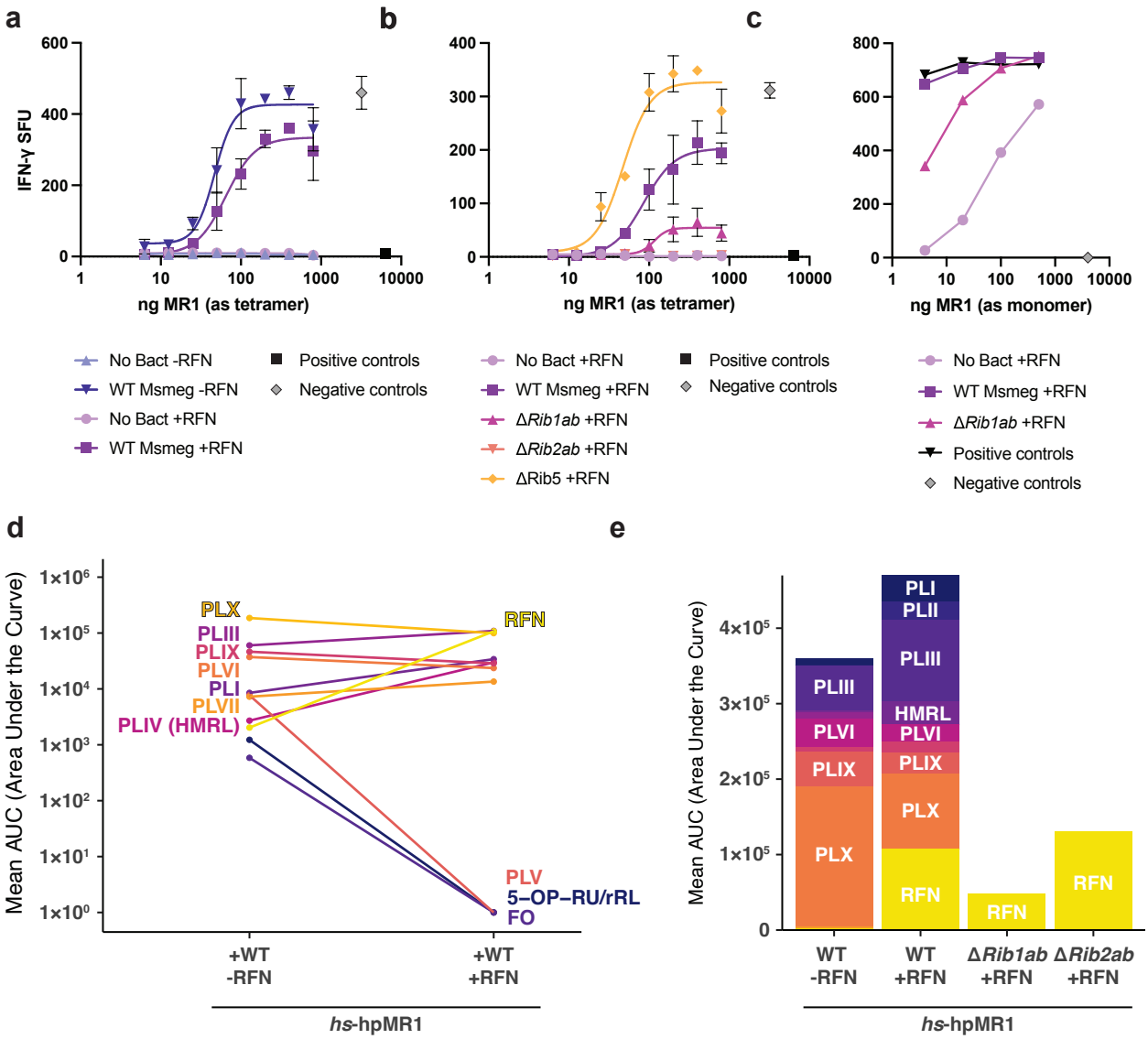


Figure 2.7 Functional properties and antigen repertoire of MR1 loaded with *M. smegmatis* RBP enzyme knockout strains. a) TetraSpot evaluating the response of a mixture of MAIT clones toward MR1 loaded with WT *M. smegmatis* or no bacteria in the presence or absence of a exogenous riboflavin. b) TetraSpot evaluating the response of a mixture of MAIT clones toward MR1 loaded with *M. smegmatis* variants (or controls). c) TetraSpot evaluating the response of a mixture of MAIT clones toward MR1 loaded with *M. smegmatis* variants (or controls). d) Mean AUC of each of the ribityllumazines in the absence (left) or presence (right) of exogenous riboflavin. e) Non-normalized mean AUCs for each of the relevant ribityllumazine/pyrimidine/flavins.

rRL-6-CH_2OH ion, the FO (a known RBP-derived metabolite and MR1 ligand) ion, and the PLV ion (**Figure 2.7d**), but the intensity of the remainder of the known ions were more or less the same. Although the slight decrease in activation of MAIT clones toward this protein (as compared with *hs*-hpMR1^{+Msmeg} made in low-riboflavin conditions) could be attributed to the

loss of these ligands from recombinant MR1, the contribution of alternative ligands that occupy the pocket due to their displacement makes this difficult to conclude. Nevertheless, it is important to note that *hs*-hpMR1^{+Msmeg} made in high-riboflavin conditions is still able to robustly activate MAIT cells (**Figure 2.7a**) despite the absence of the 5-OP-RU/rRL-6-CH_2OH ion. This critical finding establishes that 5-OP-RU/rRL-6-CH_2OH is not the central MAIT/MR1T-activating ligand in *M. smegmatis* and instead, other ligands that derive from 5-A-RU or another *Rib2ab* precursor are responsible for this strong activation. The reason that 5-OP-RU/rRL-6-CH_2OH, FO, and PLV are depleted in conditions with high concentrations of riboflavin is unclear; I hypothesize that riboflavin's role in biological redox reactions promotes scavenging of reactive oxygen species, thereby modifying the cellular metabolome [129-131]. Although these ligands were lost, the remaining photolumazines we identified earlier were still present in the protein prepared in high-riboflavin conditions, suggesting their generation is not dependent on the same environment as 5-OP-RU/rRL-6-CH_2OH, FO, and PLV.

Analysis of the knockout strains of the RBP enzymes upstream of 5-A-RU resulted in depletion of the photolumazines as determined by LC-MS/MS, suggesting that these ligands are, in fact, dependent on the RBP for their production (**Figure 2.7d-e**). This also suggests that the residual activation that was observed for *hs*-hpMR1$^{+MsmegΔRib1ab}$ (**Figure 2.7b-c**) is not attributable to the photolumazines or other known ribityllumazines/pyrimidines/flavins. The unknown ligand(s) are responsible for the activation of *hs*-hpMR1$^{+MsmegΔRib1ab}$ yet are dependent on *Rib2ab*, suggesting that they derive from alternative substrates for *Rib2ab.* Indeed, these substrates may arise from the putative (unvalidated) GTP cyclohydrolase II enzyme (MSMEG_3471) that has been

annotated in the *M. smegmatis* genome [132]. This enzyme does not seem to be compensatory for *Rib1ab*, however, since known ligands were not detected in the *hs*-hpMR1$^{+Msmeg\Delta Rib1ab}$ protein by LC-MS/MS. These alternative substrates may result in ligands with unique structural properties to the typical ribitylpyrimidines/lumazines/flavins which are therefore weeded out from our analysis by virtue of their MS2 spectra being unique. These data should be mined further going forward and alternative approaches to the LC-MS/MS chromatography and ionization should be attempted to detect unique molecular species.

Altogether, these data indicate that the photolumazines that are enriched in the bacterially-loaded MR1 samples are dependent on the riboflavin biosynthetic pathway. It was somewhat surprising that the photolumazines were present in the *hs*-hpMR1^{+Msmeg} protein made in high riboflavin conditions, because there is a well-described FMN-binding riboswitch that is thought to control the transcription of this pathway [133, 134]. Thus, in such a riboflavin-saturated environment, it would make sense for the pathway to be expressed at low or negligible levels. Nonetheless, almost all of the photolumazines were present in the LC-MS/MS data and this protein was able to stimulate MAIT cells very robustly. These data also indicate that while *Rib2ab* is necessary for production of MR1 ligands with the potential to activate MAIT cell clones, some ligands may be independent of *Rib1ab* for their production. Determining the structure and function of these alternative ligands will be extremely interesting from a microbiological and molecular perspective of this immunological axis.

2.6 PLIII, PLV, PLVI, and PLIX are likely 7-oxo-6-indolyl-8-D-ribityllumazines

The dependence of the photolumazines on the *Rib1ab* and *Rib2ab* RBP enzymes for their production suggested these molecules are ribityllumazine-based compounds and therefore they may be MAIT cell-activating ligands. To further investigate their role in MR1 presentation and activation of MAIT cells, we pursued determining their structures. In the untreated LC-MS/MS analysis of the crude synthesis material of PLIII (performed by Drs. Haihong Jin and Aaron Nielsen of the Oregon Health and Sciences University Chemistry Core) for the Harriff et al. study, we detected the signatures of the PLV, PLVI, and PLIX ions (**Supplemental Figure S2.8a**). When compared with the ion spectra of PLVI and PLIX from the *hs*-hpMR1^{+Msmeg} data (and that of PLV from the *tn*-hpMR1^{+Msmeg} data), we found that the fragmentation spectra aligned (**Supplemental Figure S2.8b**), indicating that they derived from the same molecules. Thus, given that PLIII is a ribityllumazine with an indole attached at the C6, the presence of these related molecules in the crude synthesis of PLIII suggests that these three MR1 ligands are, too, indolyl-ribityllumazine compounds. Importantly, this crude material robustly stimulated both classical (MAIT) and non-classical MR1T clones (not shown); thus, in order to determine whether PLV, PLVI, and/or PLIX were contributing to this potency, we were required to determine their structures.

The mass difference between PLV ([M-H]- m/*z*, 444.1161 Da) and PLIII ([M-H]- m/*z*, 428.1212 Da) is 15.9949 Da, suggesting that these two molecules differ only by a single oxygen, likely in the form of a hydroxyl group (in place of a hydrogen) (**Supplemental Figure S2.9a**). Likewise, the mass difference between PLVI ([M-H]- m/*z*, 460.1110 Da) and PLIII is 31.9898 Da,

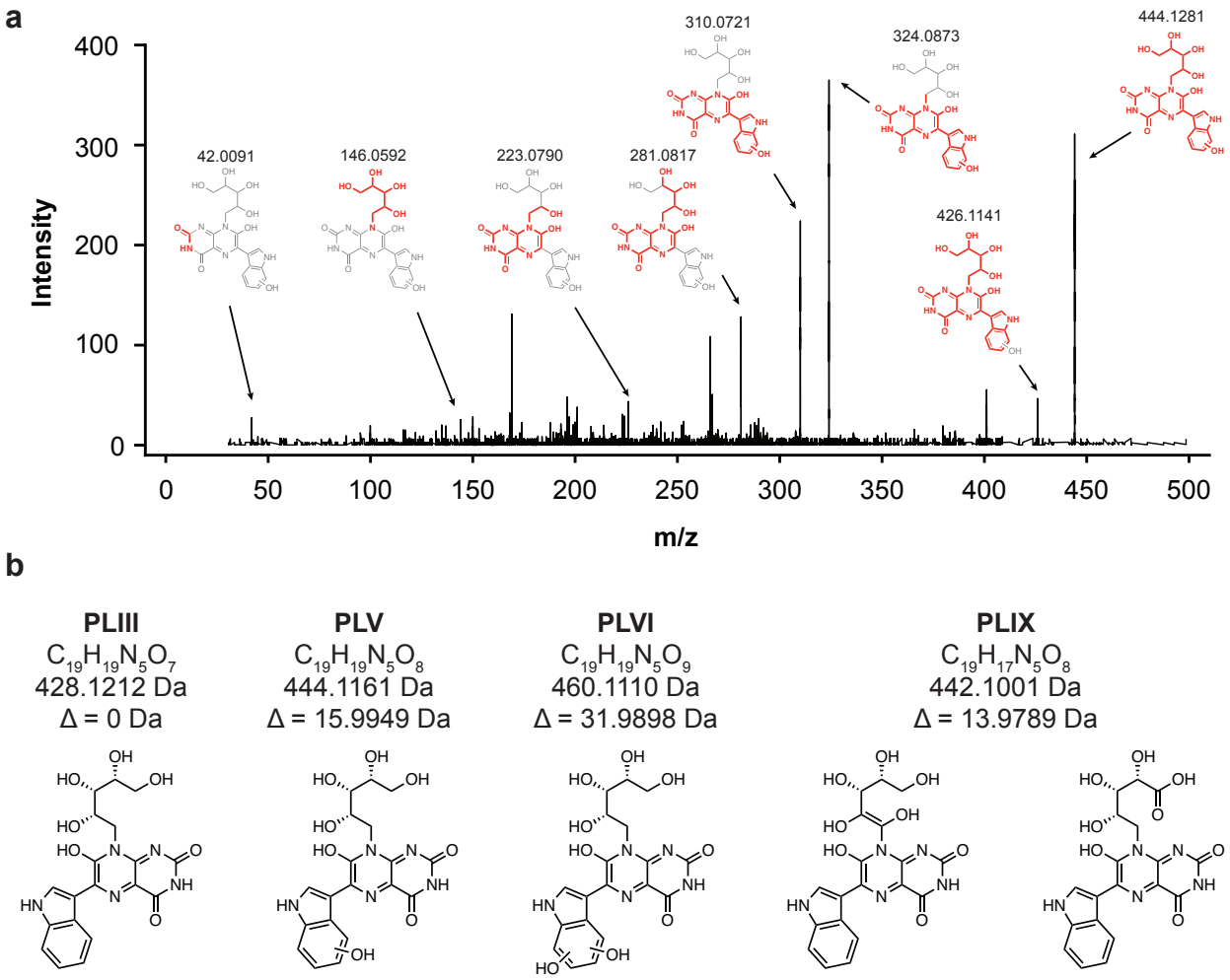


Figure 2.8 Putative molecular formulae and masses of each of the photolumazines. Δ values indicate the m/*z* difference between the molecular ion ([M-H]-) and PLIII.

suggesting that PLVI differs from PLIII by two oxygens, probably two hydroxyls. If we consider PLIII to be the "parent" structure, it becomes clear that there are only two segments upon which the additional hydroxyls could be installed: either upon the indole ring or upon the ribityl moiety. The PLV MS2 pattern suggests that its hydroxyl was most likely positioned at the C4, C5, C6, or C7 of the indole moiety (**Figure 2.8a**), but there was ambiguity as to which of these isomers could actually be found in the MR1 pocket. It may actually be several of these isomers together, as the EIC of the ion eluted from MR1 suggests there may be more than one distinct retention

time. In contrast, there does not seem to be more than one retention time PLVI, suggesting that it may be a single isomer loaded into MR1. sNevertheless, solving the structure of the biologically-relevant PLVI isomer faces the same challenges as for PLVI: it is unclear where the extra hydroxyls are installed. Interestingly, the MS2 fragmentation pattern of PLVI looks slightly different than for the other photolumazines, with low intensity for the ions resulting from the neutral losses of 120.0232 and 134.0549, which are typically two ions with the highest intensity (**Supplemental Figure S2.9b**). Nevertheless, these neutral losses are still apparent, which suggests that the hydroxyls may be installed upon the indole ring rather than the ribityl moiety. If we make this assumption, there are six combinations of positioning: PLVI-4,5, PLVI-4,6, PLVI-4,7, PLVI-5,6, PLVI-5,7, and PLVI-6,7. (If we assume C2 could be occupied by a hydroxyl, there would be ten possible combinations.) Thus, the intrinsic structural ambiguity limitation of mass spectrometry makes it harder for us to determine the biologically active isomer of PLV and PLVI.

Unlike PLV and PLVI, which overwhelmingly ionize into their [M-H]- ions, PLIX is primarily detected as a formic acid adduct. EICs of the 488.1048 Da ion and 442.1001 Da ion, which correspond with the [M+FA-H]- and [M-H]- ions for PLIX, respectively, show an identical retention time (from the PLIII crude synthesis data) (**Supplemental Figure S2.8a**). Further, the molecular ion is barely detectable in the MS2 spectrum of the m/*z* 488.1048 Da ion, which is indicative of the ablation of the formic acid upon collision-induced dissociation (fragmentation); rather, the parent mass here seems to be [M-H]- m/*z* 442.1001 Da (**Supplemental Figure S2.9c**). Thus, the mass difference between PLIX ([M-H]- m/*z*, 442.1001) and PLIII is 13.9789 Da ; the

most likely formula indicates that the difference between PLIX and PLIII is a single oxygen and a ring-double bond somewhere in the molecule. Assuming that PLIX is also a true ribityllumazine, it's unclear where the additional degree of unsaturation could be placed. While there is a possible isomer wherein the hydroxyl is placed on the indole ring and the conjugated electrons are rearranged to introduce an additional ring-double bond, this does not seem the most likely scenario. Rather, it seems that both the extra hydroxyl and ring-double bond may be localized on the ribityl moiety, as indicated by a unique neutral loss of 148.0408 Da in place of the typical neutral loss of 134.0549 Da, which corresponds with the native riibtyl moiety (**Supplemental Figure S2.9c**). This putative structure would most likely have either a carboxylic acid at C5' (in place of the hydroxyl usually at this position) or have a ketone installed at C1'. This latter isomer would most likely appear as the -enol form with a double bond between C1' and C2', changing the orbital hybridization of each of these carbons to sp^2 (**Figure 2.8b**). Since the 2' hydroxyl of the ribityl moiety forms one of the key interactions of the agMR1 complex with MAIT TCRs, the implications of a sterically-"locked" 2' hydroxyl moiety on T cell activation could be quite drastic [75, 135].

In summary, photolumazines III, V, VI, and IX are likely 7-oxo-6-indolyl-8-D-ribityllumazines that meet our inclusion criteria for structural elucidation and seem to be abundant in the MR1 ligand binding groove when bacteria are co-cultured in the expression media. Their relation to one another is cemented in the fact that they are all found in the crude PLIII synthetic material. In retrospect, the fact that PLV, PLVI, and PLIX are found in the PLIII crude synthetic material is not necessarily surprising, as 4-hydroxytryptophan (HT), 6-HT, and 7-HT are known oxidized

contaminants of commercial tryptophan [136]. It would be understandable if similar oxidation products were contaminants of the commercially-available indole-3-glyoxylic acid that was condensed with 5-A-RU for the PLIII synthesis. Nevertheless, due to the structural ambiguity of PLV, PLVI, and PLIX, we must therefore rely on other synthetic products and bacterial extracts to characterize the ligand both structurally and for their antigenicity.

2.7 PLX and PLVII are likely indolyl-ribityllumazines with unique chemical properties

The remaining ion that was significantly enriched in the mammalian protein, PLX ([M-H]- m/*z* 485.1175), was not found in the PLIII crude material. I observed that the MS2 fragmentation spectra of PLIX and PLX look similar to one another, suggesting that PLX, too, may be an indolyl-ribityllumazine or related molecule (**Figure 2.9a**). Like PLIX, PLX has very low, if not negligible, parent ion intensity. However, while it's possible that the precursor ion is also a formic acid adduct (resulting in m/*z* 439.1166), this ion is not detected as a molecular (MS1) ion such that a retention time can be correlated to that of the 485.1175 ion. A fragment of 439.1166 hardly appears in the MS2 spectrum, indicating that it is likely not the molecular ion (**Figure 2.9a**). Indeed, the molecular ion seems to be 443.1253 Da, which cannot be accounted for by known adduct species. However, there is a low-abundance fragment of 457.1337 Da which, alternatively, may be the precursor ion. Therefore, if we assume that the PLX precursor ion is not an adduct, the predicted molecular formula of PLX, $C_{19}H_{18}N_8O_8$, requires three additional nitrogens, an oxygen, and two ring-double bonds in comparison to PLIII (for which the molecular formula is $C_{19}H_{18}N_5O_7$). There are alternative predicted molecular formulas, *i.e.* $C_{20}H_{18}N_6O_9$, but the mass error is higher. In either case, the fact that PLX has an even-numbered

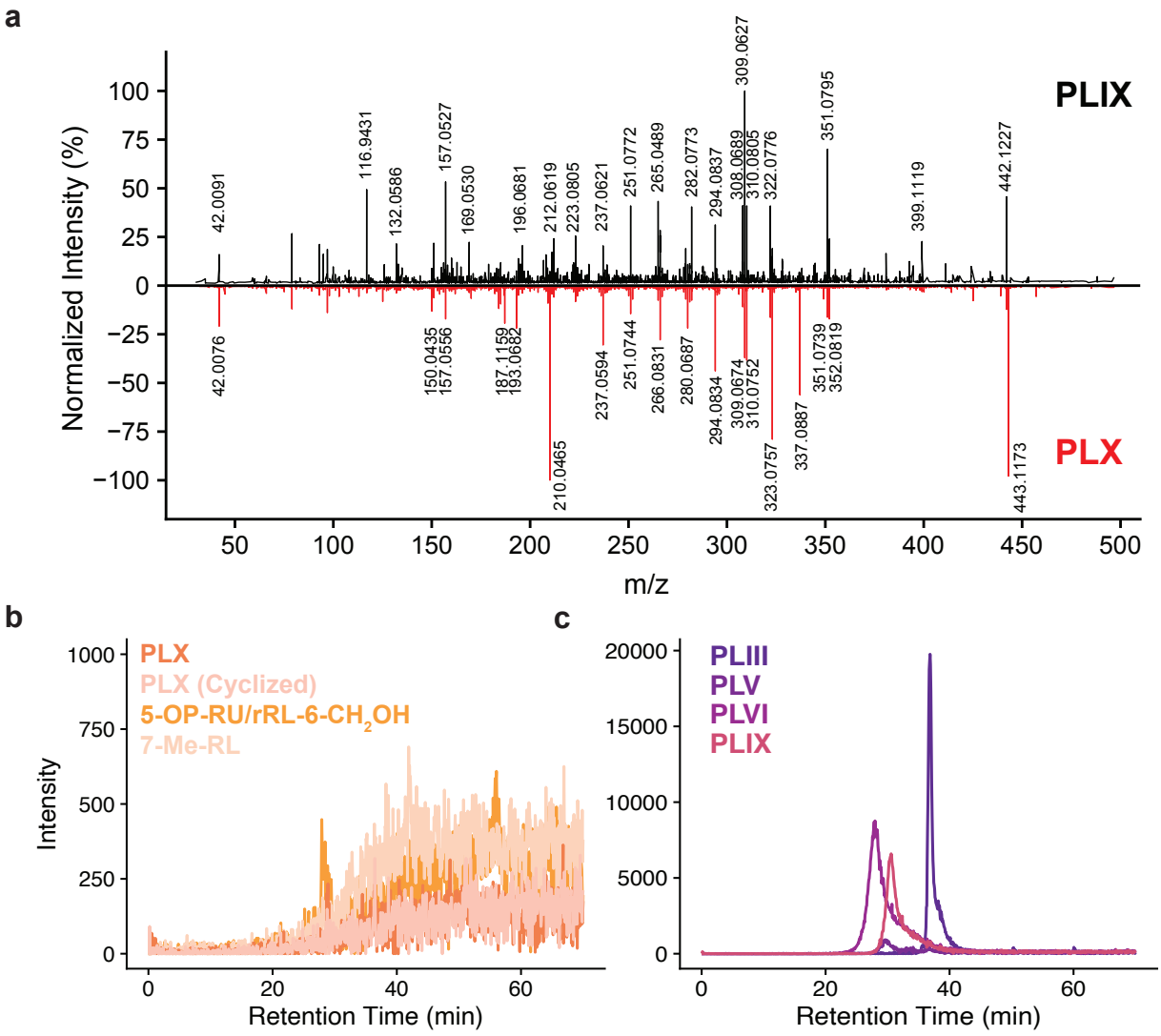


Figure 2.9 Properties of PLX. a) Fragmentation patterns of PLIX (top) and PLX (bottom). b) The EICs of the PLX (485.1175) and 5-OP-RU/rRL-6-CH_2OH (329.1103) ions for the dataset collected on the MR1 ligand extract using acetonitrile. c) The EICs of PLIII, PLV, PLVI, and PLIX ions for the dataset collected on the MR1 ligand extract.

nominal mass - *i.e.* an even number of nitrogens (per the nitrogen rule) - is interesting, as this sets it apart from PLIII, PLV, PLVI, and PLIX, which each have an odd number of nitrogens. One of my working hypotheses is that PLX has a carbamic acid or formamide group attached to C7 in place of the hydroxyl which is typically found at this position; there is literature precedent for this sort of molecule, as "russupteridine-yellow I," a 6-amino-7-formamide-8-D-

ribityllumazine, was found in *Russula* mushrooms [137-139] (**Supplemental Figure S2.10a**). I suspect that PLVII, the only other suspected indolyl-ribityllumazine with an even number of nitrogens ([M-H]- m/*z* 453.1153), shares structural properties with PLX (**Supplemental Figure S2.10b**), but I will focus here only on PLX because we did not find PLVII in *hs*-hpMR1^{+Msmeg}, only in *tn*-hpMR1^{+Ecoli} and *tn*-hpMR1^{+Msmeg} preps (**Figure 2.4c**, **Figure 2.4e**).

Several observations have led me to believe that PLX has unique chemical properties that distinguish it from PLIII, PLV, PLVI, and PLIX. For example, when I extracted the ligands from recombinant MR1 by chemical (organic solvents) or physical (boiling) methods and Jason injected these samples onto the LC-MS/MS, the ion intensity for PLX disappeared completely (**Figure 2.9b**). In contrast, photolumazines III, V, VI, and IX were all still present in the extract (**Figure 2.9c**). Thus, the process of ligand extraction affects the PLX molecule - by chemical modification, cyclization, degradation, or some other means - and results in its disappearance from the extracted material. Alternatively, it's also possible that the ligand was never released from the MR1 groove and remains in the pellet of aggregated protein. Interestingly, this same phenomenon is observed for the 5-OP-RU/rRL-6-CH_2OH ion (m/*z* 329.1103) (**Figure 2.9b**) without a synchronous rise in the 7-Me-RL ion, suggesting that the molecule responsible for this ion may still be associated covalently with MR1 in the pellet of the aggregated protein. Indeed, the pellet of aggregated protein was yellow, suggesting that some ligands were not eluted from the pocket like PLIII, PLV, PLVI, and PLIX were (**Supplemental Figure S2.10c**). However, when I solubilized and 3 kDa-filtered the pellet of aggregated protein and Jason injected these on the mass spectrometer, we still did not see either 5-OP-RU, PLX, or their cyclized forms

(-18.0105 Da) (**Supplemental Figure S2.10d**). While the exact reasons for this effect are unclear, this could be the result of a technical limitation or a biological effect of the ligand being covalently bonded to MR1 (assuming that bond is stable in the solvent that it was solubilized in), which should not have passed through the 3 kDa filter. Altogether, these results suggest that PLX may be unstable in the absence of the MR1 pocket and/or could possibly be a ribitylpyrimidine capable of forming a Schiff base with the K43 residue of MR1.

Further evidence toward the unique chemical nature of PLX (assuming it is, in fact, an indolyl-ribityllumazine/pyrimidine) lies in the fact that it is not found in any amount in the crude PLIII synthetic material (**Supplemental Figure S2.8b**), whereas PLIII, PLV, PLVI, and PLIX were abundant (**Supplemental Figure S2.8a**). I find it most likely that the starting material - and not necessarily the chemical synthesis methodology - was the cause for this effect since synthesis of PLX would require a nitrogen source that would not be necessary for the remaining PLs in the mixture. However, it is possible that PLX was generated as a byproduct of this reaction but was degraded or otherwise modified before the sample was analyzed by LC-MS/MS and thus the ion disappeared. Together, these observations support the hypothesis that PLX has chemically unique features from PLIII, PLV, PLVI, and PLIX, and, given its abundance by LC-MS/MS, makes it an intriguing - if difficult - target for structural elucidation.

2.8 Validation of PLV as an MR1 ligand and an agonist of the D481C7 MAIT clone

In order to determine the functional properties of PLV, Dr. Aaron Nielsen and Jordan Devereaux of the OHSU Chemistry Core synthesized and purified each of the PLV isomers. Specifically

PLV-4, PLV-5, PLV-6, and PLV-7 were synthesized, which had a hydroxyl installed at C4, C5, C6, and C7, respectively (**Figure 2.10a**, **Supplemental Figure S2.11**). PLV-2 was not synthesized because the MS2 ion fragmentation pattern was best fit when the hydroxyl was positioned at the other four positions. However, it is worth revisiting this isomer because the C2 position of the indole is one of the most reactive atoms of this functional group due to the free electron pair on the adjacent nitrogen. After purification from the remaining isomers Dr. Meghan Cansler (of Drs. David and Deborah Lewinsohn's labs at OHSU) tested their ability to stabilize MR1 on the cell surface. In the absence of ligand, MR1 is primarily sequestered in the endoplasmic reticulum [140, 141]. Upon ligand uptake, MR1 egresses to the cell surface, where it can be detected by flow cytometry using an α-MR1 antibody [99]. PLV-4 was able to induce only very minor stabilization of MR1 at the cell surface, and though PLV-5, PLV-6, and PLV-7 did show some ability to do so, none induced MR1 egress as efficiently as PLIII or the positive control 6-FP (**Figure 2.10b**). Nonetheless, this indicated that any of the ligands should have the potential to be responsible for MR1T cell agonism.

In order to determine which, if any, of these isomers had agonistic properties, we tested each for its ability to stimulate the D481C7 MAIT clone, which was activated by the crude PLIII synthetic material. The isomers displayed variable potency, which decreased as the hydroxyl group was shifted from C7 around to C4 of the indole ring (**Figure 2.10c**). Notably, though the low potency seems seems attributable its ability to stabilize MR1 on the cell surface, the differential reactivity seen toward PLV-5, PLV-6, and PLV-7 suggests that the D481C7 MAIT clone is displaying ligand selectivity based on the position of the hydroxyl. Thus, PLV is a bona

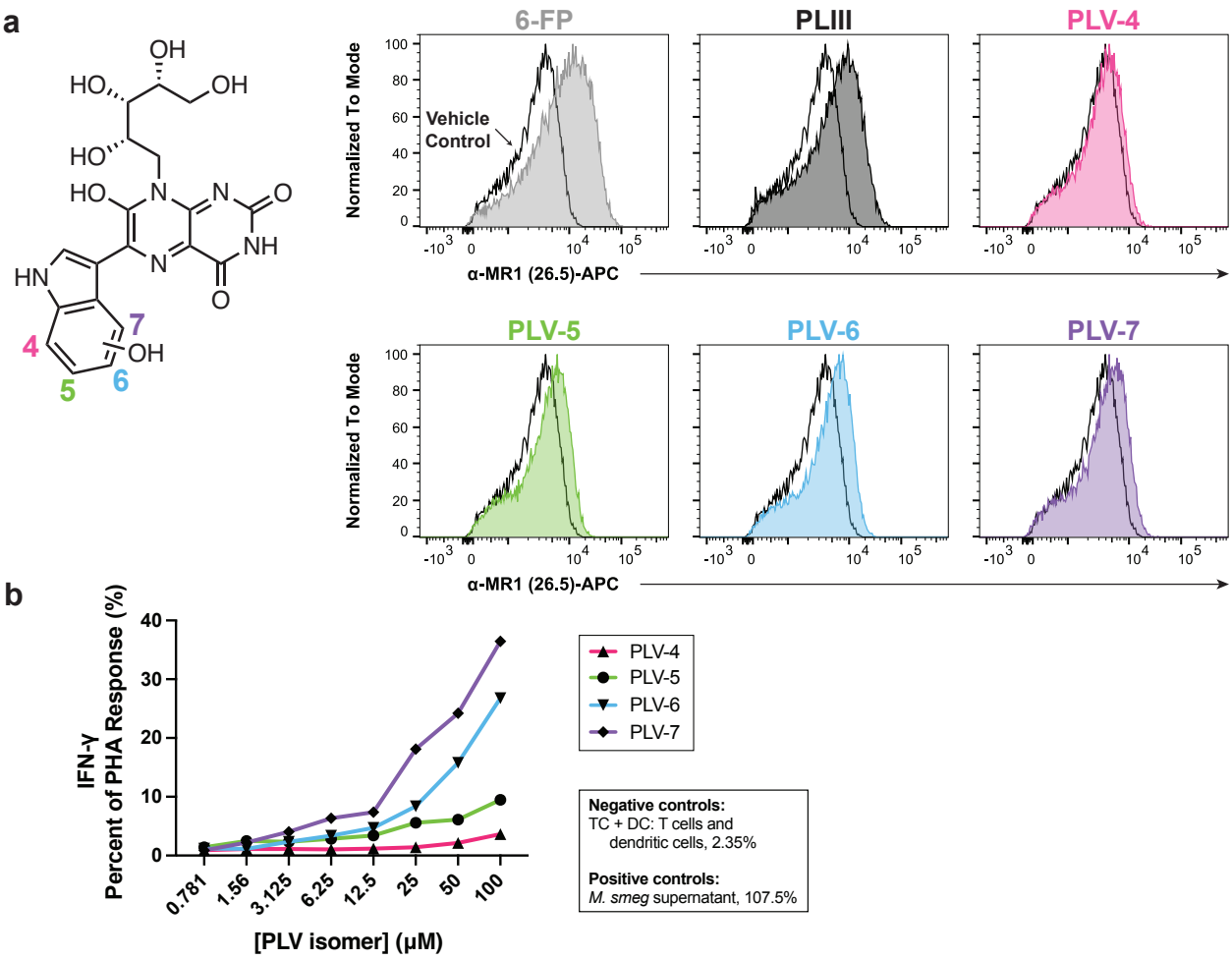


Figure 2.10 Validation of PLV as an MR1 antigen. a) Stabilization of MR1 on the cell surface by PLV isomers, with 6-FP, PLI, and PLIII as positive controls and NaOH as a negative control. b) Structure of PLV indicating where hydroxyls could be installed on the indole moiety. c) Activation of the D481C7 MAIT clone by PLV isomers.

fide MAIT cell agonist for at least one MAIT cell clone. However, more work needs to be done to determine which isomer(s) bacteria are capable of biosynthesizing, which are loaded into MR1 *in vivo*, and whether these two variables are dependent on the metabolism of unique bacteria.

2.9 Isolation of select photolumazines from bacterial supernatants for structural and antigenic characterization

Unfortunately, because of the isomer uncertainty and structural ambiguity of PLVI and PLIX,

respectively, these ligands are currently synthetically unachievable. Therefore, I asked whether we could isolate these ligands from bacterial supernatants such that we could structurally and functionally characterize them. We routinely produce 6,7-dimethyl-8-D-ribityllumazine (DMRL) in-house for *in vitro* studies by overexpressing recombinant *Rib1ab* and *Rib2ab* from *Bacillus subtilis* in BL21 *E. coli* [124]. This elicits an increase in ARP, the immediate precursor to 5-A-RU (**Figure 2.6a**). The addition of exogenous glucose then produces an excess of dihydroxy-butanone phosphate (DHBP), which enters into the biosynthetic cycle with 5-A-RU to form DMRL. Thus, DMRL (and other fluorescent molecules, including riboflavin) are produced in large quantities by the overexpression of the two enzymes in this system (**Figure 2.11a, Supplemental Figure S2.11a**). Since the photolumazines seemed to be dependent on the riboflavin biosynthetic pathway, I thought that they may be generated as a byproduct of this preparation. When Jason ran LC-MS/MS on the 3 kDa-filtered culture supernatant, we immediately found that PLIII, PLVI, and PLIX were quite abundant and that it was possible that PLV was there as well, though seemingly in a lower quantity (**Supplemental Figure S2.11b**). Interestingly, the four ions that we detected are the same four that were found in the PLIII crude synthetic material and in the MR1 protein extract. Unfortunately, PLX and PLVII were not found at any point in either crude or fractionated material, further indicating that these species may be transient or stable only when bound to MR1 (**Supplemental Figure S2.11c**). Nevertheless, this system seemed to be a viable source of antigen from native source, so I scaled up the preparation (and matched controls) and I went on to optimize the expression and determine a purification strategy of PLIII, PLV, PLVI, PLIX using thin-layer chromatography (TLC) and LC-MS/MS (performed by Jason) to track the photolumazines throughout purification.

First, it was clear from the fluorescent profile of the supernatants on a TLC that the overexpression of both enzymes was required to enrich these fluorescent molecules to the extent that they would be useful for purification purposes (**Supplemental Figure S2.12a-b**). This is consistent with previous reports in the field that the *Rib1ab*-mediated step of the RBP is the rate-limiting step [142]. I then attempted several modifications to the expression system including introducing additives including indole and tryptophan (as a potential indole source), varying the exogenous glucose concentration, varying the IPTG (expression inducer) concentration, shortening the length of expression, and overexpressing only *Rib1ab* or *Rib2ab*. The only clear difference that I found in the fluorescent profiles on these TLCs was that the exclusion of exogenous glucose resulted in the depletion of many of the spots that we consider to be contaminants (such as DMRL, $R_f = 0.21$ in 80:15:5 MeCN:water:EtOAc, teal color) (**Supplemental Figure S2.12b**). This modification did not change the intensity of the spots we think to be the indolyl-ribityllumazines, including the dense sea-foam colored spot running near the solvent front ($R_f = 0.58$ in 80:15:5 MeCN:water:EtOAc), which co-migrates with the PLIII standard (**Supplemental Figure S2.12d, S2.11e**). Indeed, when assessed by LC-MS/MS, we found that the unsupplemented supernatant had approximately the same AUC values for PLIII, PLV, PLVI, and PLIX, but that the DMRL ion was diminished in comparison with the glucose-supplemented supernatant (**Supplemental Figure S2.11f, S2.11g**; only DMRL and PLIII are shown for clarity). It is possible to speculate that the indolyl-ribityllumazines, therefore, do not depend on the DHBP synthase function of *Rib1ab* (specifically, *Rib1b*). In any case, overexpressing *B. subtilis Rib1ab* and *Rib2ab* recombinantly in *E. coli* results in *de novo*

biosynthesis of PLIII, PLV, PLVI, and PLIX.

Literature from early characterizations of DMRL and riboflavin led me to purify the photolumazines first from the crude, 10 kDa-filtered supernatant using a Florisil matrix developed with acetone (**Figure 2.11b**) [143-145]. Indeed, when assayed by LC-MS/MS, Jason found PLIII, PLV, PLVI, and PLIX to elute early from the column but again did not find PLVII and PLX (**Supplemental Figure S2.11e-f**). From here, I attempted a number of different types of chromatography for the fractionation of this material, including anion exchange, silica, and reverse-phase chromatography, but with the help of Dr. Karen Dobos at Colorado State University, we found that preparative-scale HILIC (hydrophobic interaction chromatography) was the next best step for separation of the photolumazines from other material and contaminants (**Figure 2.11c-d**). Dr. McMurtrey and Jason then developed a targeted metabolomics protocol for the detection of these compounds with which we analyzed each fraction from the HILIC column and found PLIII, PLV, PLVI, and PLIX in fractions 40 and 41, and another ribityllumazine, photolumazine II, which has a phenol group at C6 [146, 147] in fractions 47 and 48 (**Figure 2.11e**). Smaller, more hydrophilic ribityllumazines such as DMRL and 6-hydroxy-7-methyl-8-D-ribityllumazine (HMRL) didn't elute until Fraction 59-60 (**Figure 2.11e**). Thus, it seems that the ribityllumazines with aromatic groups at their C6 position were enriched in and around fractions 40-50.

I then tested each fraction from the HILIC column in an IFN-γ EliSpot assay with a mixture of four MAIT clones to determine whether the fractions containing our known photolumazines had

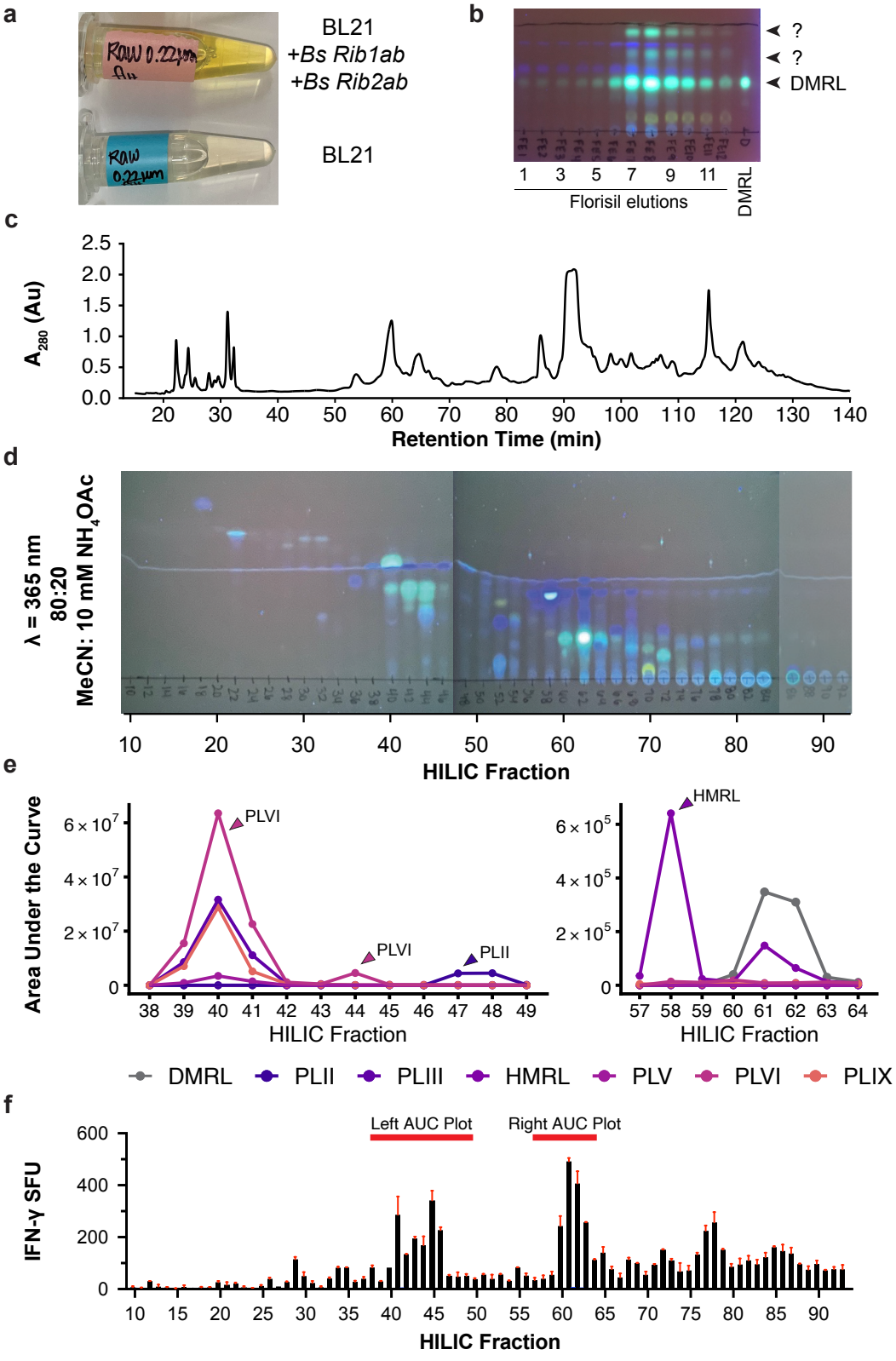


Figure 2.11 Purification of photolumazines III, V, VI, and IX from bacterial supernatants. *(continued on next page)*

Figure 2.11 ***(continued)*** a) Image of 0.22 μm-filtered supernatants from untransformed BL21 *E. coli* (left) or BL21 transformed with plasmids carrying *B. subtilis* RibA (*Rib1ab*) and RibD (*Rib2ab*) genes. b) TLC plate under 365 nm light showing the elution of fluorescent compounds from a Florisil column to which 3 kDa-filtered supernatant was applied. The spot for DMRL is a commercial standard. Question marks refer to compounds with similar emission wavelengths as DMRL and PLIII but which are unknown. c) Chromatogram (280 nm absorbance) of the preparative-scale HILIC column to which we applied Florisil-purified RibA/D supernatant. d) TLC plate under 365 nm light of every other fraction of the preparative HILIC column shown in (c). e) Activity of HILIC fractions resuspended at the same mass per volume in an EliSpot assay with a four-clone mix of MAIT cells. Fractions 40-48 are highlighted in red. Blue bars (very small) show activation by each fraction in the presence of the α-MR1 blocking antibody.

stimulatory potential. There was a local maximum of activation concentrated between Fractions 41 and 46, but the fact that Fraction 40 did not activate the clones indicated to us that the PLVI and PLIX - which were found in this fraction - did not have stimulatory potential or had weak agonism toward the MAIT clones (**Figure 2.11f**). The activation, however, was clearly MR1-dependent, as the activity of MAIT clones toward dendritic cells incubated with these fractions was abolished in the presence of the α-MR1 blocking antibody (the blocked signal is so minimal that it is difficult to see when plotted) (**Figure 2.11f**). It is possible that one of these two molecules is agonistic while the other is antagonistic and the competition between them is responsible for the depleted signal. Despite this, the purification of PLIII, PLVI, and PLIX from a native bacterial source is an ongoing pursuit such that we can use these ligands in structural and functional studies.

2.10 Discussion

In this chapter, I have summarized the work that I - with many, many collaborators - have done to pursue novel MR1 ligands originating from species of *Mycobacteria*. and to investigate the dependence of the MR1 ribityllumazine profile on bacterial species and on enzymes of the RBP. The LC-MS/MS pipeline that has been developed previously and utilized here has allowed us to

search for not only uncharacterized MR1 ligands but completely novel molecules. In collaboration with others, we were able to demonstrate the indolyl-ribityllumazines are bona fide, *Rib1ab*- and *Rib2ab*-dependent MR1 ligands and are highly enriched when MR1 is loaded with *M. smegmatis* ligands. We also showed that PLV is capable of stabilizing MR1 on the cell surface and can generate an MR1-dependent immune response by MR1T cell clones. Although some of the photolumazines are structurally ambiguous, I have speculated on the nature of their chemical structures. Further, I have generated a purification scheme for PLIII, PLV, PLVI, and PLIX from bacterial supernatants, which will aid us in future studies regarding their antigenic properties and their mechanism of presentation by MR1 to MR1T TCRs. A summary of the properties of the photolumazines can be found in **Figure 2.12**. In addition to the photolumazines, we have shown that MR1 may carry ligands that are dependent on *Rib2ab* for their biosynthesis but arise in the absence of *Rib1ab*, even (seemingly) in the absence of a compensatory *Rib1ab* ortholog. Altogether, these results a new shine light on the ligand repertoire of MR1 using an "antigen fishing" method that more closely recapitulates the interaction between bacterial and host cells than those which have been used previously.

One of the more surprising findings from this study was that the *M. smegmatis* Δ*Rib1ab* strain was able to induce a small amount of IFN-γ production by MAIT clones, but that the Δ*Rib2ab* strain was unable to activate clones in any capacity. When MR1 loaded with ligands from these strains of bacteria were analyzed by mass spectrometry, the photolumazines were absent from both datasets, indicating that there was no compensatory enzyme for *Rib1ab*, at least for known and detectable MR1 ligands. Instead, these two results suggest that there may be alternative

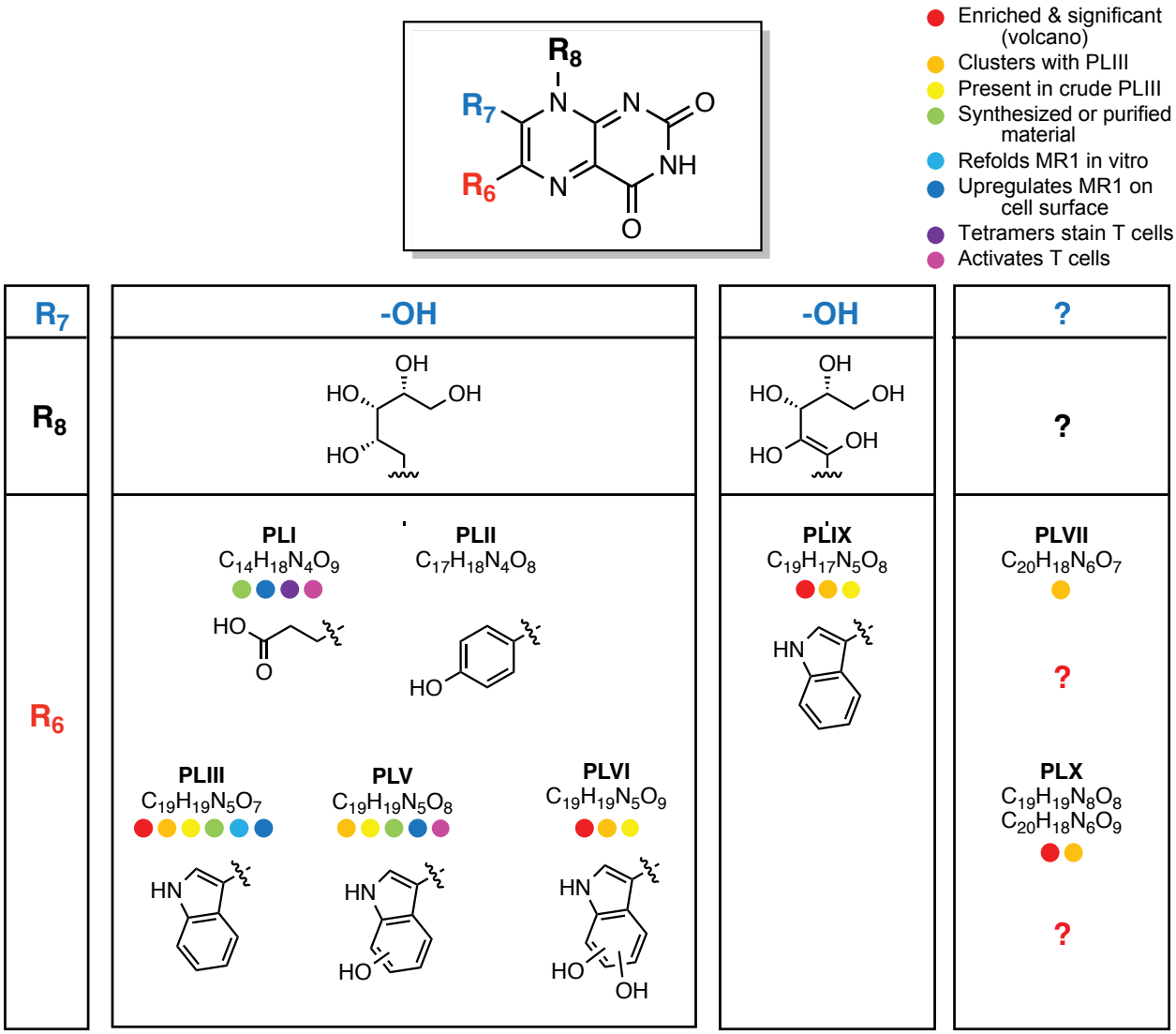


Figure 2.12 Summary of the structural and functional properties of the photolumazines.

substrates for *Rib2ab*, and that it is *Rib2ab*, not 5-A-RU, that is the central dependency for the production of MAIT cell-activating ligands. This said, since the potency of *hs*-hpMR1$^{+Rib1ab}$ was significantly lower compared with that of *hs*-hpMR1$^{+\text{WT }Msmeg}$, it's clear that *Rib1ab*-dependent ligands still account for the majority of observable MAIT cell reactivity. Although these results contradict the findings of Corbett et al. [8], it is not unreasonable that MR1 ligands in the supernatant of the *L. lactis* knockout strains used in their study were too dilute to activate a

MAIT TCR-transduced Jurkat line. Indeed, we were unable to detect MAIT clone activation by the Δ*Rib1ab* strain when live bacteria were used as the stimulus.

Although this work will need to be validated further, the potential for a novel class of microbial MR1 ligands is quite exciting. If my hypothesis is true, novel *Rib2ab* substrates would result in an entirely new subset of MR1 ligands, which could have large implications for the field as it continues to identify ligands which impact MAIT and MR1T cell responses in various contexts. Nevertheless, these results still do not explain the existence of MR1T cell clones reactive toward riboflavin auxotrophs such as *Streptococcus pyogenes* [70], which currently has no annotated *Rib2ab* gene (KEGG Pathway). Thus, there is work still yet to be done in investigating alternative pathways and enzymes from which MR1 ligands can arise.

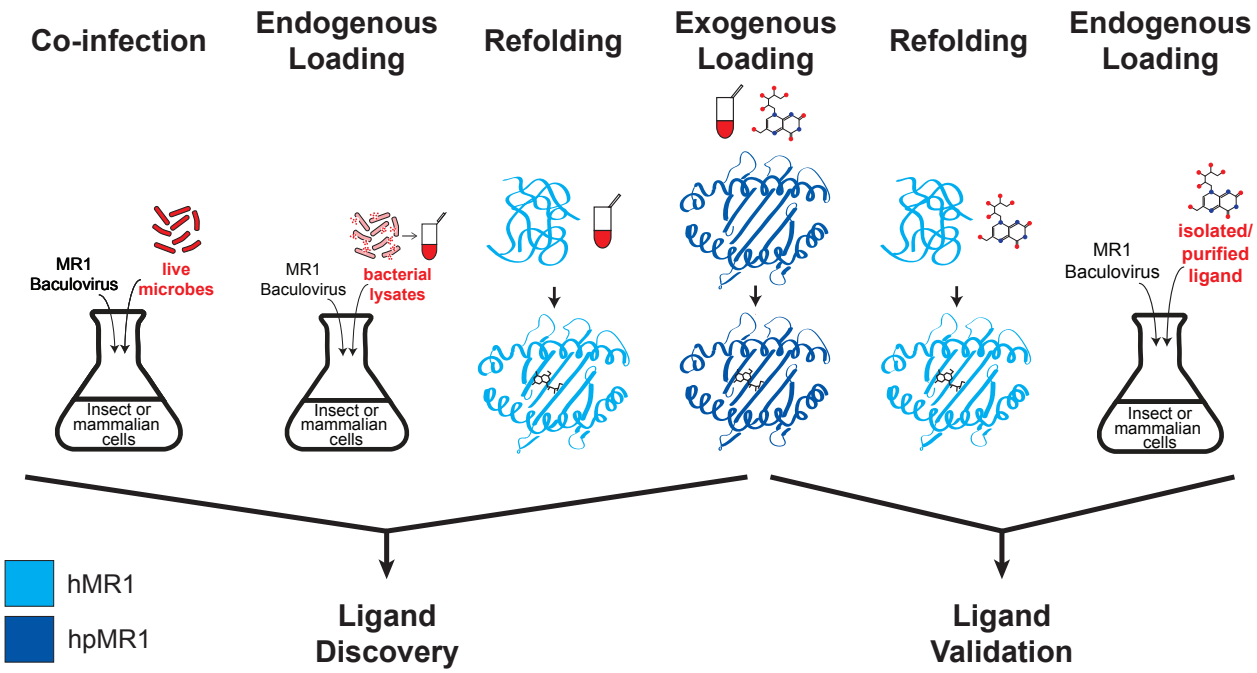
In all of this, it is important to remember that the photolumazines do not account for all activating ligands in the MR1 ligand-binding groove. One of the primary caveats of our work is that abundance does not necessarily correlate with the ability of a ligand to stimulate T cells. Thus, there may be a variety of low-abundance ligands that we are not capturing because of our abundance inclusion criteria. Likewise, we are likely ignoring structurally unique ligands due to our focus on ions with familiar hallmarks of ribityllumazines/pyrimidines/flavins. Further, our chromatographic methods and the ion polarity we use here may be skewing our detected ions away from relevant agonists. Finally, we are limited by our ability to screen whether a ligand is agonistic or antagonistic because our panel of MR1T cell clones are primarily TRAV1-2^{+} and thus likely use the same conserved recognition mechanism as has been elucidated dozens of

times for ribitylpyrimidines/lumazines. Thus, what we determine to be an “antagonist” may stimulate a unique subset of MR1Ts that have not been previously characterized. This is discussed further in a case study on differential activation of MR1T clones in **Chapter 3**. While these concessions are all valid, there is no method to assess the entire ligandome of MR1 at once; thus, determining the ligand repertoire under these conditions still provides novel insights for the field.

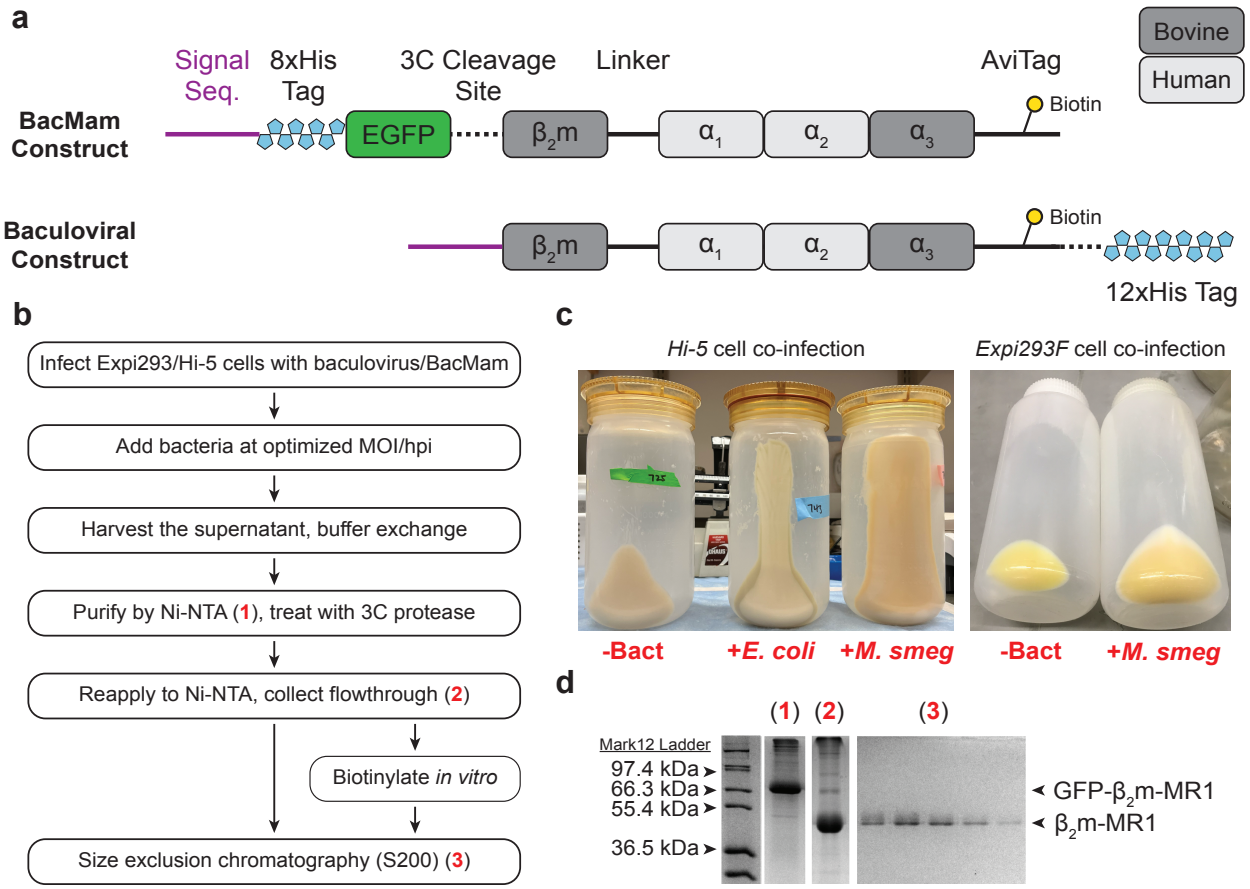
Indeed, we were able to identify eight novel MR1 ligands through our mass spectrometric pipeline. With the exception of PLI, PLII, and PLIII, these compounds are completely novel. PLIII was characterized first in the 1950s from a purine-requiring mutant of *Achromobacter petrophilum* [125]. While the authors of this study did not understand the origin of the fluorescent molecule they had characterized, they did speculate that the indole arises from tryptophan metabolism due to its connection to the lumazine core at its 3-position rather than its highly reactive 2-position. In the absence of known enzymatic machinery, the simplest explanation for the origin of PLIII is that 5-A-RU condenses spontaneously with indole-3-glyoxal (I3G) or indole-3-glyoxylic acid (I3GA). I3G and I3GA, however, are not known to arise directly from any enzymatic processes and therefore are relatively uncharacterized in comparison with other indole metabolites which have been studied extensively due to their importance as microbiome-generated molecules in the gastrointestinal tract [148]. Hydroxylated versions of I3G - which would be necessary to give rise to PLV and PLVI - have also never been characterized; however, it is also possible that PLIII is a “parent molecule” which is further derivatized to form PLV, PLVI, and PLIX. Indeed, there exist both bacterial and mammalian

enzymes capable of hydroxylating indoles such as [149, 150], which could explain the installation of extra hydroxyls, at least for PLV and PLVI. It is also possible that PLIII is modified to form PLVII and PLX, though perhaps this modification is made by host cells, which is why we can only detect it from injections of intact MR1 protein. Though the mechanism is unclear, our work clearly shows that bacteria (at least *E. coli*) can generate large amounts of PLIII, PLV, PLVI, and PLIX in a host cell-independent manner from overexpression of *Rib1ab* and *Rib2ab*. We hope that, going forward, we can use this system and the purification scheme I developed to produce large enough amounts of PLIII, PLV, PLVI, and PLIX for their structural validation and functional assessment.

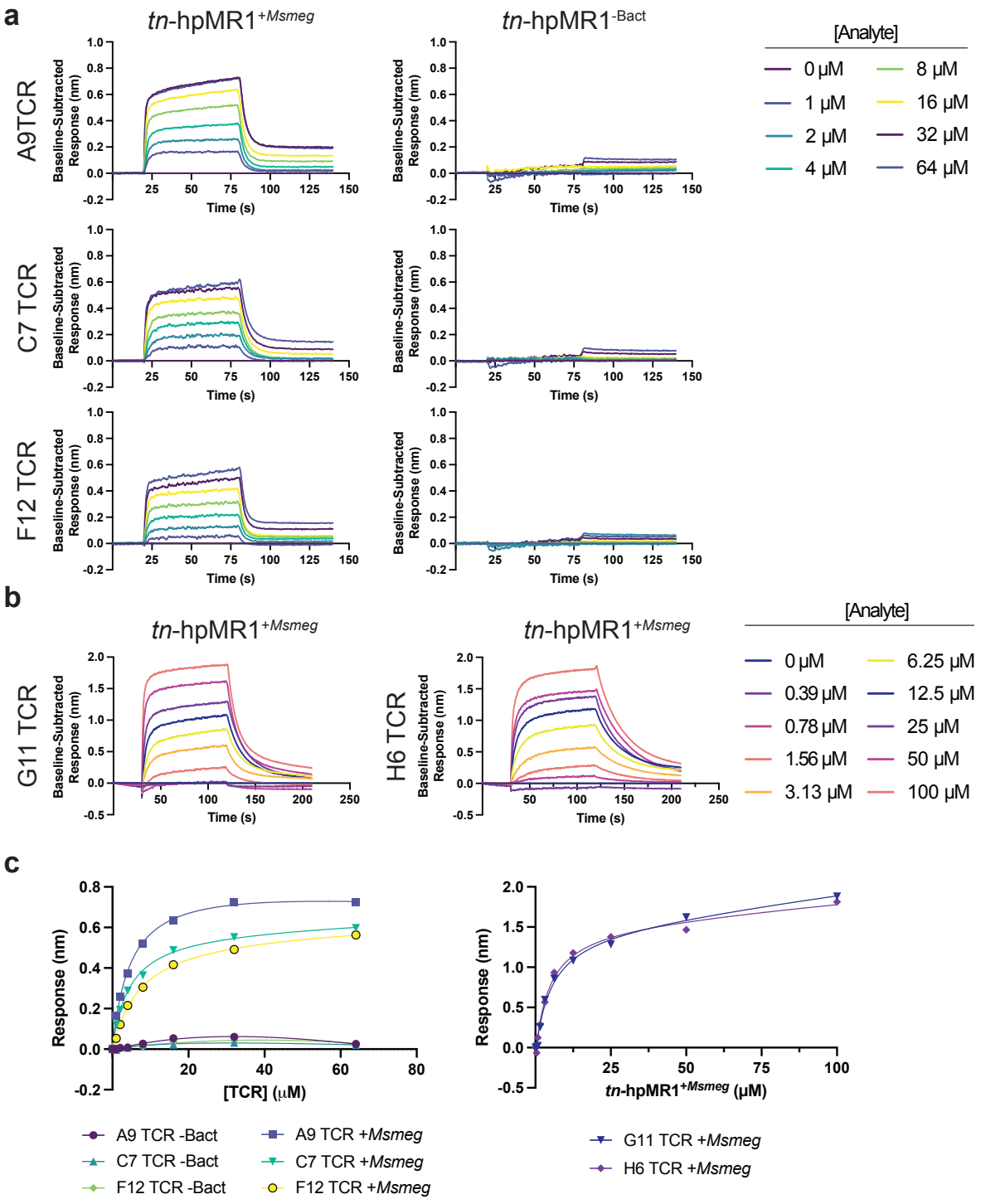
2.11 Supplemental information



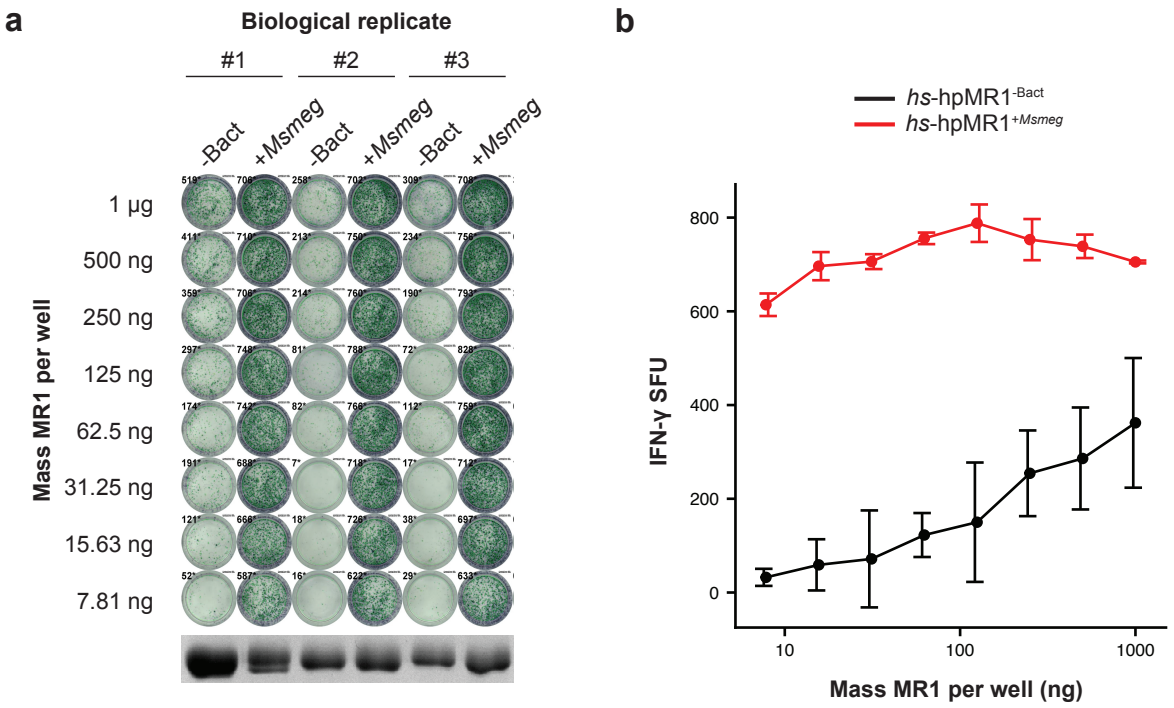
Supplemental Figure S2.1 Approaches to loading MR1 with antigen(s) of interest. a) Co-infection refers to the addition of live microbes to the culture of a cell line recombinantly expressing MR1, typically in a soluble form. b) Endogenous loading involves adding bacterial lysates (or other ligand sources) or purified ligand to the culture of a cell line recombinantly expressing MR1, typically in a soluble form. c) Human β_2m and the ectodomain of human MR1 can be expressed as inclusion bodies in *E. coli*; the complex can be refolded *in vitro* but this only occurs in the presence of a ligand. Thus, it can capture antigens from mixtures or pure ligands. d) Bovine β_2m and the chimeric human platform (bovine-human) MR1 ectodomain can also be expressed in *E. coli* and the complex refolded *in vitro*. However, it can do so in the absence of a ligand. Thus, the protein can be loaded with a mixture of ligands or a single ligand after purification.



Supplemental Figure S2.2. Construct and purification of hpMR1. a) BacMam and baculoviral human platform (hp) MR1 construct, consisting of human α_1 and α_2 domains and bovine α_3 and β_2m domains, as well as various tags for purification or other relevant protein modifications. b) General purification scheme for hpMR1 in either expression system. *In vitro* biotinylation was only performed for tetramer generation and was not performed for protein analyzed by mass spectrometry. c) Images of pellets from the expression of *tn*-hpMR1 and *hs*-hpMR1 after being co-cultured with various bacteria. d) Representative SDS-PAGE gels of *hs*-hpMR1 at each stage of protein purification highlighted in (b).

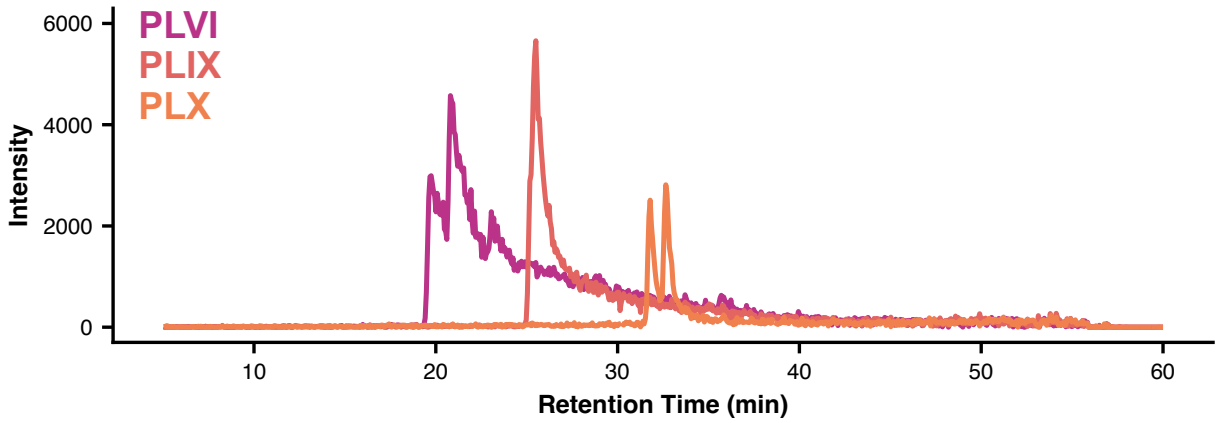


Supplemental Figure S2.3 Biophysical properties of *tn*-hpMR1^{+Msmeg} interactions with MAIT TCRs. a) BLI curves and equilibrium binding curve of *tn*-hpMR1^{+Msmeg} (or the *tn*-hpMR1^{-Bact} control) with TCRs deriving from three of the four MAIT clones we use in our activation assays; MR1 is immobilized and the TCR is the analyte. Performed by Dr. Cara Froyd. b) BLI curves and equilibrium binding curve of *tn*-hpMR1^{+Msmeg} with one TCR deriving from the MAIT clone activation panel; TCR is immobilized and MR1 is the analyte. c) Equilibrium binding curves of the experiments from (a) and (b), respectively; equilibrium response was determined as 2 seconds before the disassociation step began.

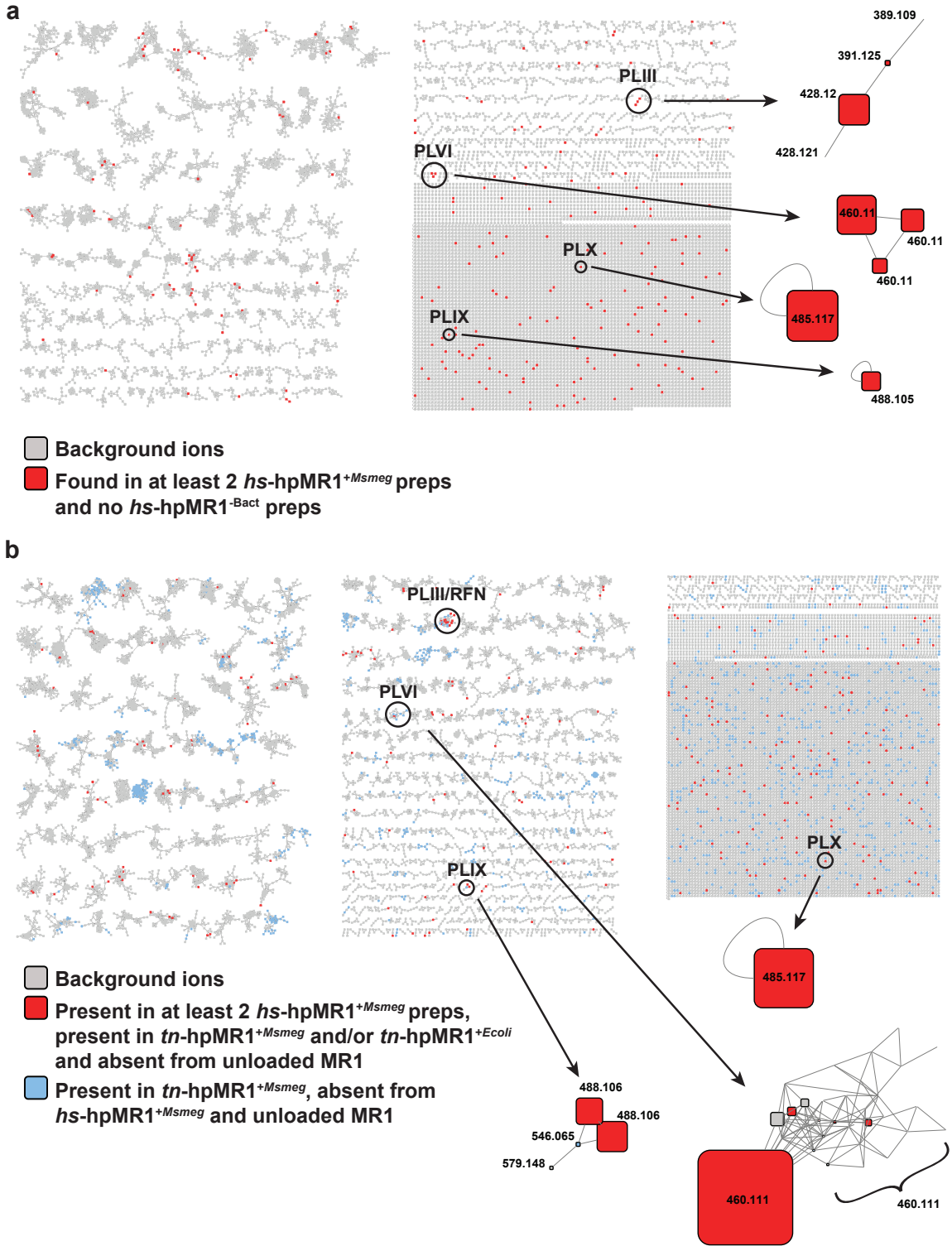


Supplemental Figure S2.4 Validation of the three *hs*-hpMR1 replicates. a) Titration series of the *M. smegmatis*-loaded and unloaded MR1 protein from each of the three biological replicates in an ExSPOT assay with a mixture of four MAIT clones. The SDS-PAGE gel at the bottom was run immediately after the assay with what was calculated to be 1 µg of protein. The -Bact replicate #1 shows higher levels of background, but it seems to be due to a miscalculation in the amount of protein added for the ExSpot assay since the corresponding band on SDS-PAGE is darker/heavier than the others. b) Quantification of the response seen in panel A. The response of T cell clones toward *hs*-hpMR1^{+Msmeg} is still near saturation, even at 7.8 ng of protein. While it is not shown here, the activation does titer out at picogram quantities. The best signal to noise ratio is achieved using approximately 50 to 150 ng of MR1 per well.

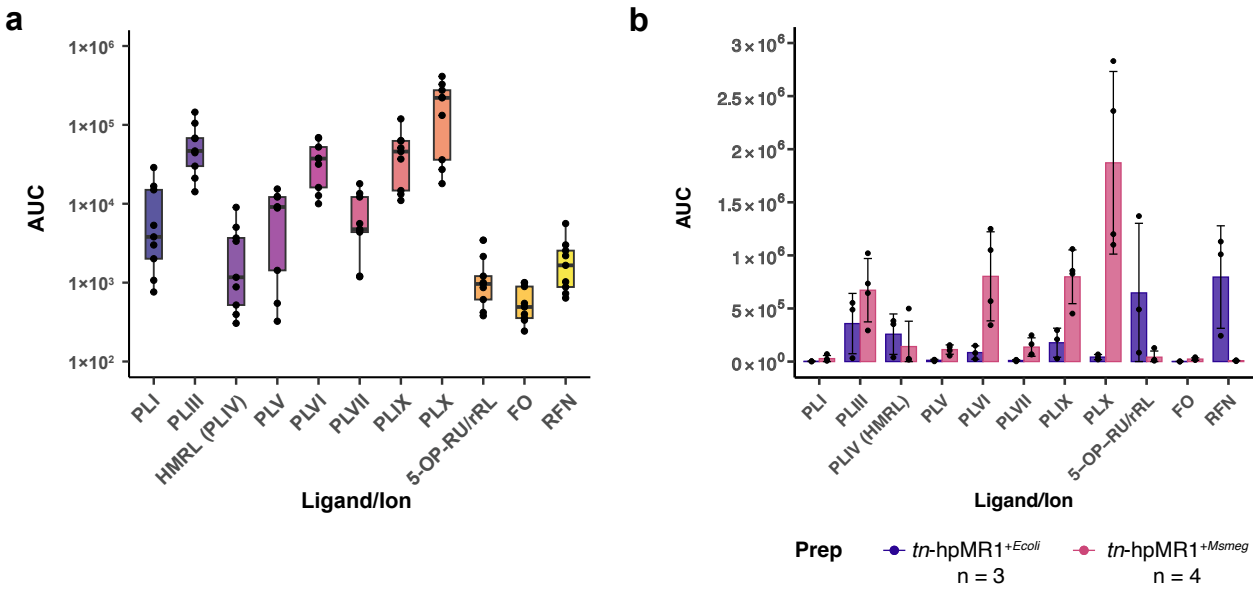
Supplmental Figure 2.5



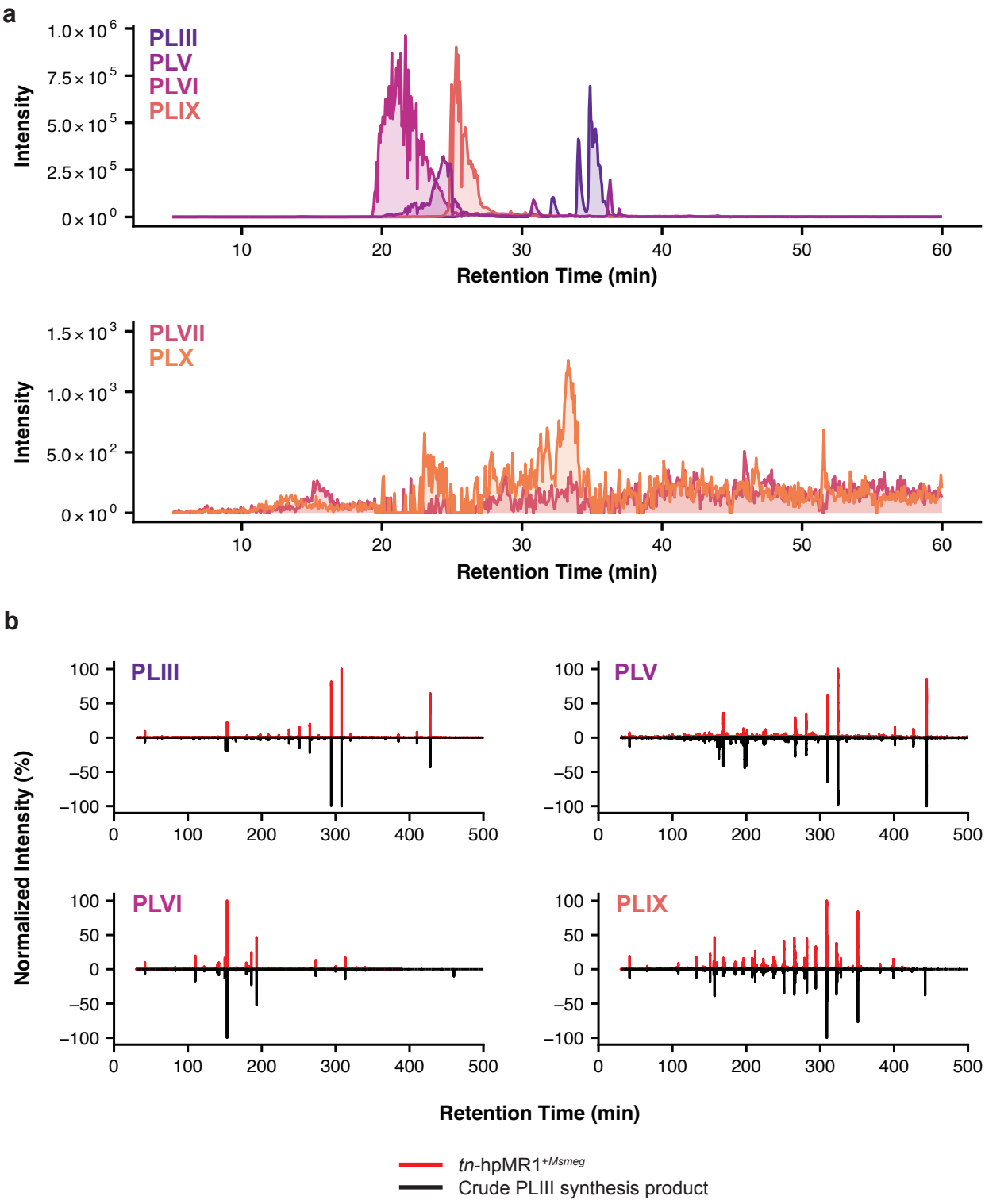
Supplemental Figure S2.5 EIC of PLVI, PLIX, and PLX from one of the *tn*-hpMR1^{+Msmeg} injections.



Supplemental Figure S2.6. GNPS clustering of ions eluted from *hs*-hpMR1 and/or *tn*-hpMR1. a) Global network for clustering of only *hs*-hpMR1 ions (with and without *M. smegmatis*). b) Global network for merged *hs*-hpMR1 and *tn*-hpMR1 GNPS analysis. The size of the node corresponds with the number of times the ion was selected for MS2, which correlates but does not perfectly represent their abundance, since the number of times an ion is selected for MS2 is dependent on the other abundant ions in the time window in which it is eluting.



Supplemental Figure S2.7 Range of AUCs observed for each loaded protein. a) Box-and-whisker plot showing the variability of abundance for each ion across the 9 technical replicates (3 biological replicates, each with 3 technical replicates) for *hs*-hpMR1^{+Msmeg}. b) The variability of abundance for each ion across the 4 technical replicates (1 biological replicates, each with 4 technical replicates) for *tn*-hpMR1^{+Msmeg} (pink) or the 3 technical replicates (1 biological replicate with 3 technical replicates) for *tn*-hpMR1^{+Ecoli} (purple). The small points represent the raw data points, whereas the larger point represents the mean AUC and the error bars indicate the standard deviation.



Supplemental Figure S2.8. Presence of PLV, PLVI, and PLIX in the PLIII crude synthesis data. a) EIC of PLIII, PLV, PLVI, and PLIX (both the [M-H]- species and the adduct) in the PLIII crude synthesis data. b) MS2 spectral matching between the MR1 protein eluate and the PLIII crude synthesis ions.

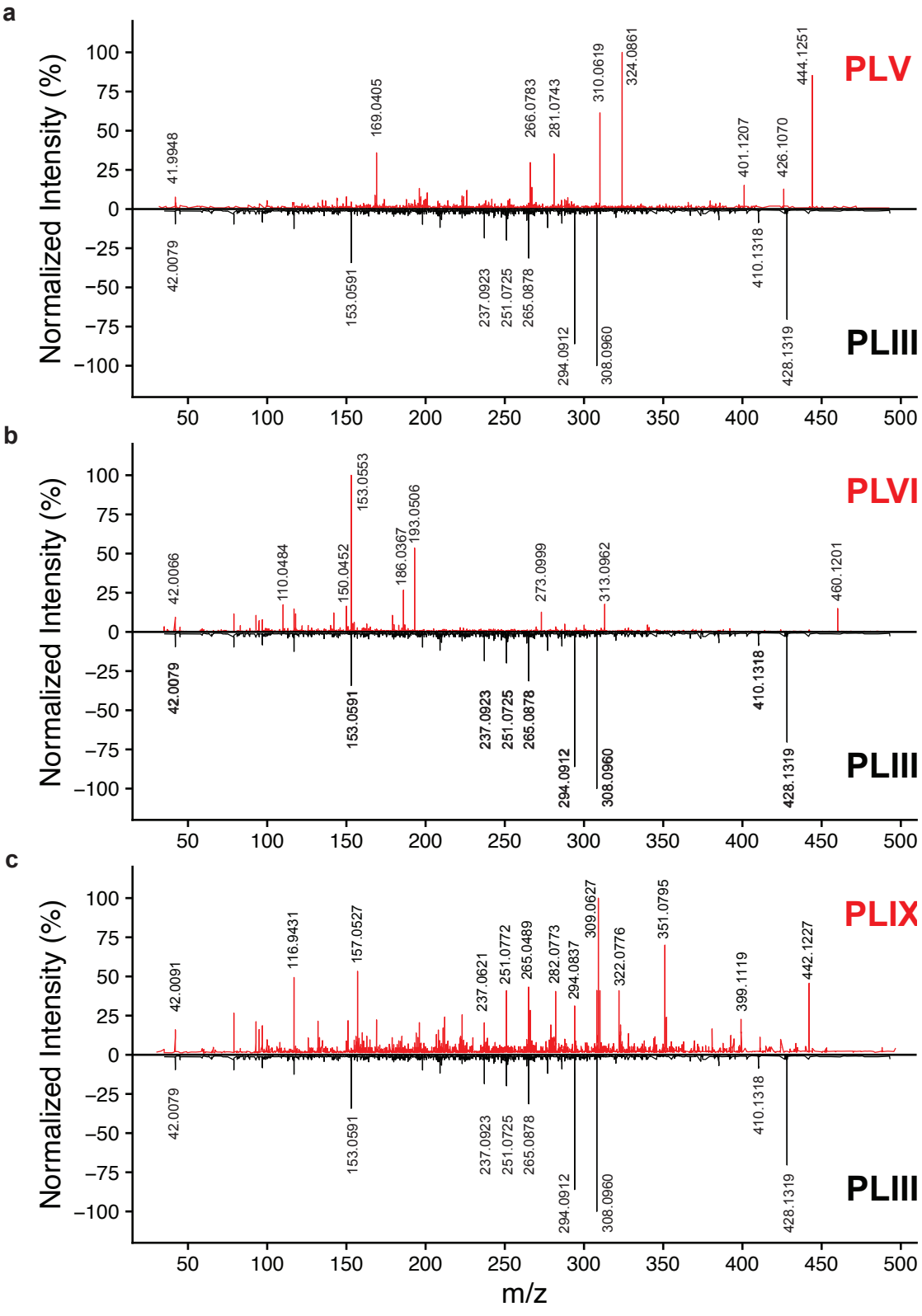
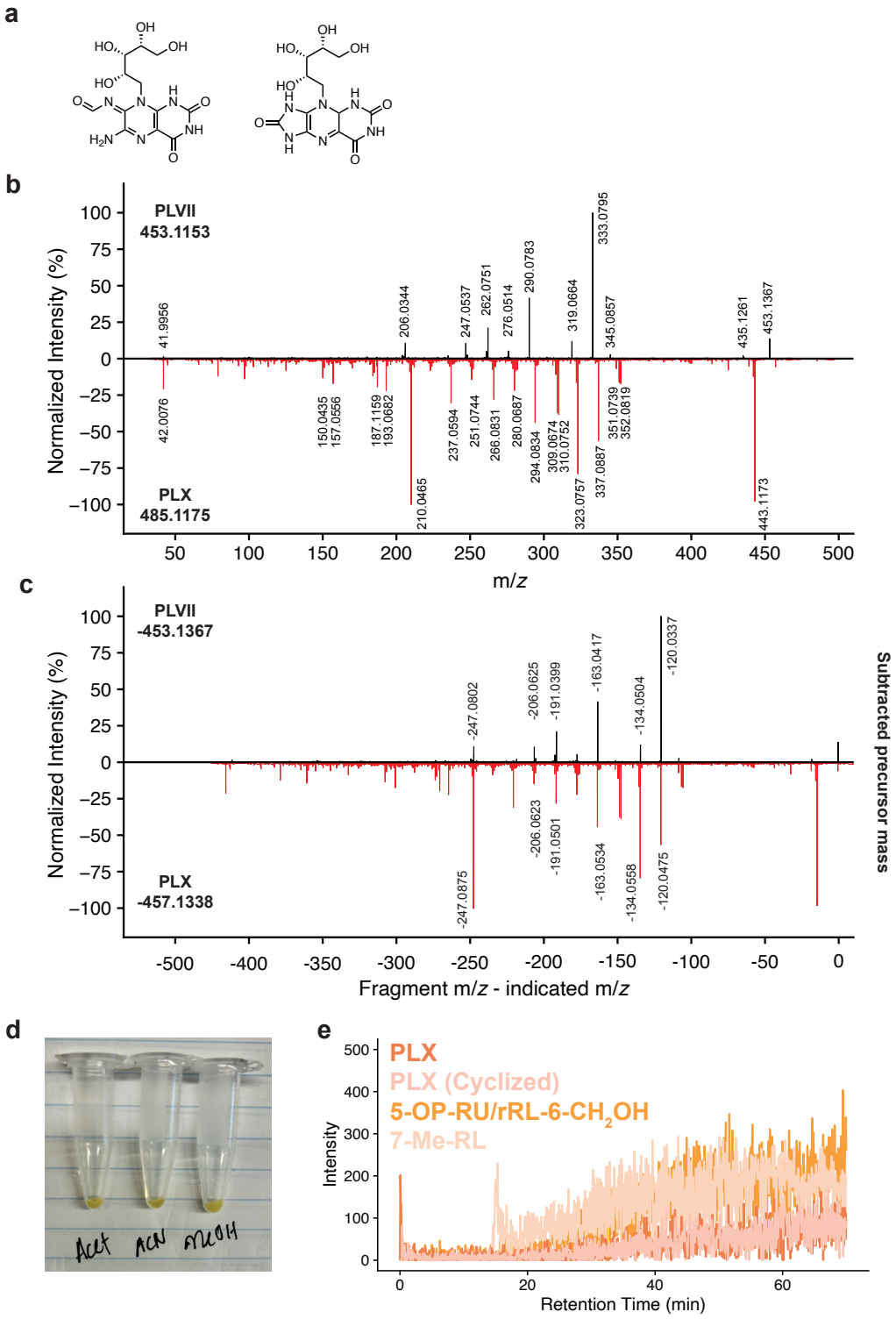
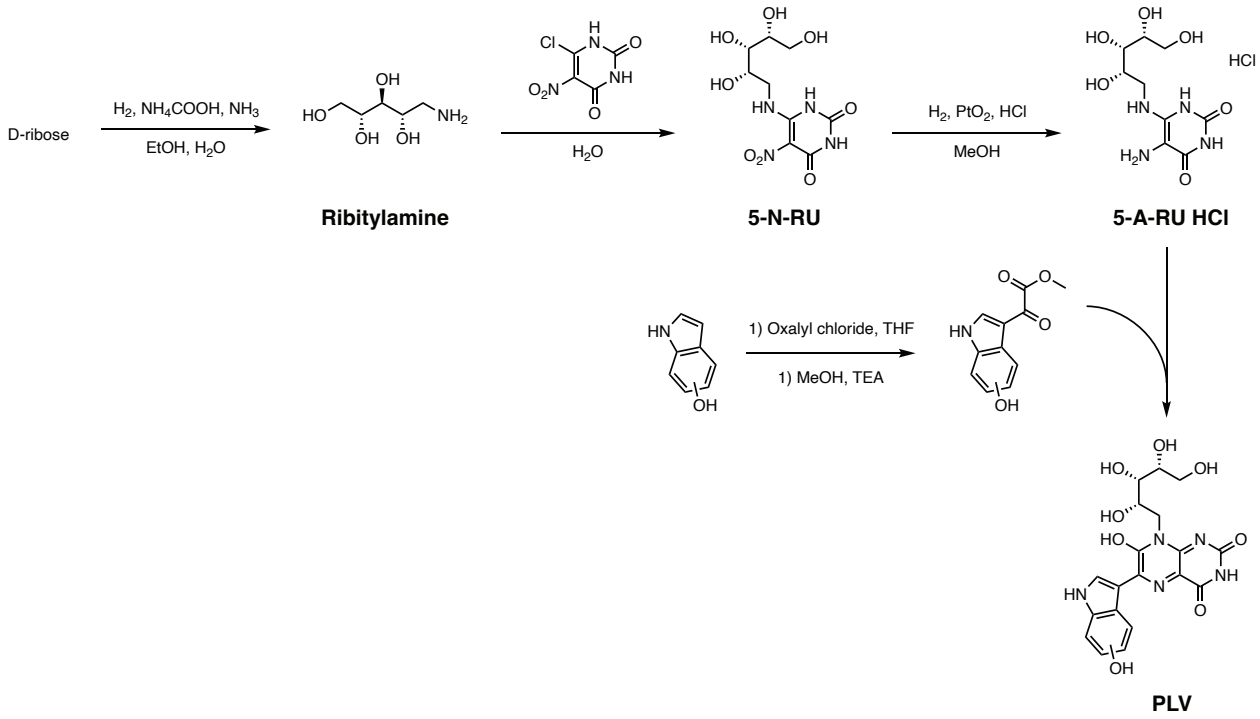


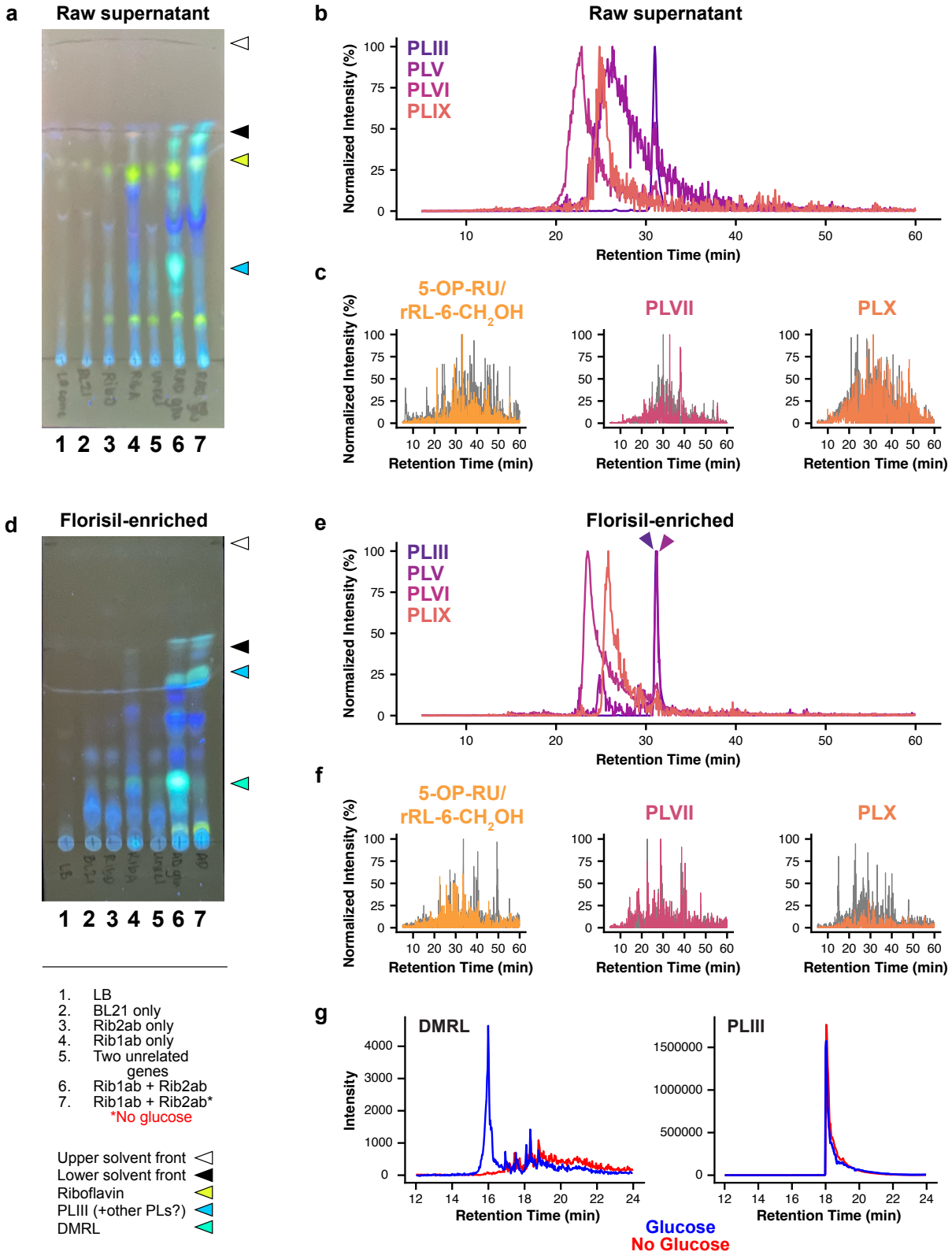
Figure S2.9 MS2 fragmentation pattern of PLV (a), PLVI (b), and PLIX (c) in comparison with PLIII.



Supplemental Figure S2.10 Unique properties of PLX. a) Structures of "russupteridine-yellow I" and "russupteridine-yellow IV." b) MS2 spectral alignment of the raw PLVII and PLX. c) MS2 spectral alignment of the raw PLVII and PLX after subtracting the precursor ion mass. d) Picture of pelleted protein after extraction of ligands by chemical methods. e) EIC of PLX and 5-OP-RU/rRL-6-CH_2OH and their putative cyclized products from the injection of the 3 kDa-filtered solubilized pellet.



Supplemental Figure S2.11 Chemical synthesis of PLV isomers.



Supplemental Figure S2.12 Isolation of photolumazines from the supernatant of *E. coli* overexpressing *B. subtilis Rib1ab* and *Rib2ab* enzymes. *(continued on next page)*

Supplemental Figure S2.12 ***(continued)*** a) TLC of the 3 kDa-filtered supernatants of *E. coli* overexpressing different combinations of *Rib* enzymes or unrelated genes, and either in the presence or absence of glucose. The plate was imaged under 365 nm light for compound detection. A key describing the sample order and the colored triangles can be found below panel (d). b) EIC of PLIII, PLV, PLVI, and PLIX from the 3 kDa-filtered supernatant. c) EIC of parent ion (colored) or the putative cyclized product (gray) for 5-OP-RU/rRL-6-CH_2OH, PLVII, and PLX from the 3 kDa-filtered supernatant. d) TLC of the samples from (a) after purification over a Florisil column. The plate was imaged under 365 nm light for compound detection. e) EIC of PLIII, PLV, PLVI, and PLIX from the Florisil-eluted material. The largest peak for PLV and PLIII overlap and are hard to distinguish; two colored arrows indicate the presence of both peak at 31-32 minutes. f) EIC of PLIII, PLV, PLVI, and PLIX from the 3 kDa-filtered supernatant. c) EIC of parent ion (colored) or the putative cyclized product (gray) for 5-OP-RU/rRL-6-CH_2OH, PLVII, and PLX from the Florisil-eluted material. g) EIC of DMRL and PLIII from the Florisil-enriched supernatants when expressed in the presence (blue) or absence (red) of glucose.

Compound	[M-H]-	[M+H]+	[M+FA-H]-	[M+FA+H]+
Lumichrome	241.0730	243.0877	-	-
DMRL	325.1154	327.1300	-	-
HMRL	327.0946	329.1092	-	-
5-OP-RU	329.1103	331.1249	-	-
FO	362.0994	364.1140	-	-
Riboflavin	375.1321	377.1467	-	-
PLI	385.1001	387.1147	-	-
PLII	405.1052	407.1198	-	-
PLIII	428.1212	420.1358	-	-
PLV	444.1158	446.1304	-	-
PLVI	460.1097	462.1243	-	-
PLVII	453.1153	455.1299	-	-
PLIX	442.1001	444.1147	488.1041	490.1187
PLX	485.1176	487.1322	-	-

Table S2.1. Exact masses for each relevant parent ion; 1 proton = 1.007282675 Da; 2 protons = 2.01456535 Da.

Compound	Fragment ions									
Lumichrome	199.0704	198.0647	170.0717	134.894	41.9999					
DMRL	209.0313	205.0673	190.8674	146.8774	144.8619	130.8824	118.8937	42.0004		
HMRL	309.0838	207.0519	193.0362	164.046	150.0301	42.0003				
5-OP-RU	191.0555	177.0383	109.0391	42.0001						
FO	344.0873	254.0564	242.0568	228.0413	199.0505	185.0349	184.0273	42.0004		
Riboflavin	255.0843	241.073	212.0833	198.065	184.0875	157.078	42.0003			
PLI	367.0968	341.1164	342.1019	265.0607	251.0444	221.0686	207.0525	164.045	150.0654	41.9965
PLII	285.0578	271.0412	258.0471	242.0515						
PLIII	308.0755	294.0599	265.0702	251.0549	153.044	41.9988				
PLV	426.1035	401.1088	324.0708	310.0575	281.0673	266.0629	169.0398			
PLVI	313.0771	273.082	193.0346	186.0176	153.0339	110.0338				
PLVII	333.0832	319.0665	290.0752	290.0752	262.077	247.0545	206.0367			
PLIX (Adduct)	442.0969	351.0568	322.053	310.0532	309.0483	308.0401	282.0597	265.0341	157.0389	
PLX	443.1215	337.0828	323.0754	310.0722	309.0654	294.0772	266.0812	210.0447	193.0628	42.0088

Table S2.1. Exact masses for each relevant fragment ion.

Chapter 3: MAIT and MR1T cell selectivity: A case study of DMRL and 5-deaza-DMRL

3.1 Introduction

In contrast to conventional αβ T cells, MAIT cells, a major subset of MR1-restricted T cells, use very limited diversity in their TCR repertoire (typically TRAV1-2 rearranged with TRAJ33, TRAJ20, or TRAJ12 and limited β chain diversity), resulting in a semi-conserved mechanism of antigen recognition across this unique T cell subset [84, 87]. This TCRα rearrangement encodes a "conserved" tyrosine residue (Y95α) in the CDR3α loop which is reportedly responsible for forming a critical hydrogen bond with the 2' hydroxyl of the ribityl moiety of riboflavin-related antigens [58, 63, 89]. Consistent with these observations, modifying Y95 to phenylalanine disrupts the ability of Jurkat cells transduced by MAIT TCRs to be activated by MR1 ligands and has been shown to decrease the affinity of agMR1-TCR interactions by as much as an order of magnitude [8, 58, 90]. Likewise, MR1 ligands such as folate metabolites which lack a ribityl moiety, or derivatives of ribityllumazines where the 2'-hydroxyl of the ribityl moiety has been ablated, induce diminished T cell activation and decrease the affinity of the ligand:MR1-MAIT TCR reaction [58, 75, 151]; thus, these ligands are typically defined as being antagonistic. Even just changing the stereochemistry of the ribityl group hydroxyls can have drastic effects on these variables [135]. Aside from Y95α, the CDR3β of some MAIT TCRs have also been shown to contact ribityllumazines/ribitylpyrimidines through the 4'-hydroxyl and/or 5'-hydroxyl of the ribityl moiety [8]; these contacts have been suggested to finely tune the interface but are

insufficient to stimulate MAIT clones in the absence of the 2'-hydroxyl, at least for two TCRs that were studied [75]. Otherwise, these ligands are sequestered deep in the ligand-binding pocket of MR1, rendering them inaccessible to the CDR3 loops [8, 58, 63, 89]. Thus, given the essentiality of Y95 for MAIT recognition and activation of riboflavin-related compounds, it would make sense that the MR1-restricted T cell repertoire is skewed toward TCR rearrangements encoding for this structural hallmark.

This paradigm of MAIT TCR reactivity has led to a trend of ligands being binned as "agonists" or "antagonists," which implies that MAIT cells respond in a binary fashion to MR1 ligands containing a ribityl moiety [152]. However, this model does not explain why the ribitylpyrimidines, such as the "model" MR1 ligand, 5-OP-RU, is (generally) so much more potent than the ribityllumazines. There have also been several reports of MAIT cells being able to discriminate between ligands (even very structurally similar ligands), leading to the hypothesis that MR1 ligands cannot be binned in such a fashion [75, 77, 78, 81]. For example, Harriff et al. [77] found that D481C7, a MAIT cell clone, was activated by two novel ribityllumazines, Photolumazine I (PLI) and Photolumazine III (PLIII), with larger adducts on the lumazine core than the other described ribitylpyrimidine and ribityllumazine MR1 ligands. Whereas the remaining MAIT clones in their panel were unresponsive toward these ligands, a TRAV12-2^{+} MR1T clone had moderate activity toward them. Other TRAV1-2^{-} MR1T cells have also been shown to adopt these unique reactivities, including a TRAV21^{+} MR1T cell clone was found to be activated by the folate derivatives 6-FP and acetyl-6-FP but agnostic toward 5-OP-RU in the absence of autoreactivity [78]; this reactivity pattern is surprising, as it is the opposite of the

MAIT cell activation trends typically seen toward these ligands [152]. It therefore seems that an MR1 ligand that is an agonist for one TCR may be an antagonist for another. Thus, MAIT and MR1T cell activation is likely not a binary response toward the presence or absence of a ligand bearing a ribityl moiety; it is more likely a probe of the dynamic molecular landscape generated by the combination of MR1 and the ligand. Because this more adaptive-like recognition may result in the expansion of select subsets of MR1T cells in a ligand dependent manner *in vivo* [153, 154], investigating the molecular basis of antigen discrimination by MAIT and MR1T TCRs is critical to determining the role of different classes of MR1 ligands.

The molecular mechanisms underpinning ligand discrimination are unknown but may result from a combination of local and allosteric conformational changes and/or changes in the dynamic

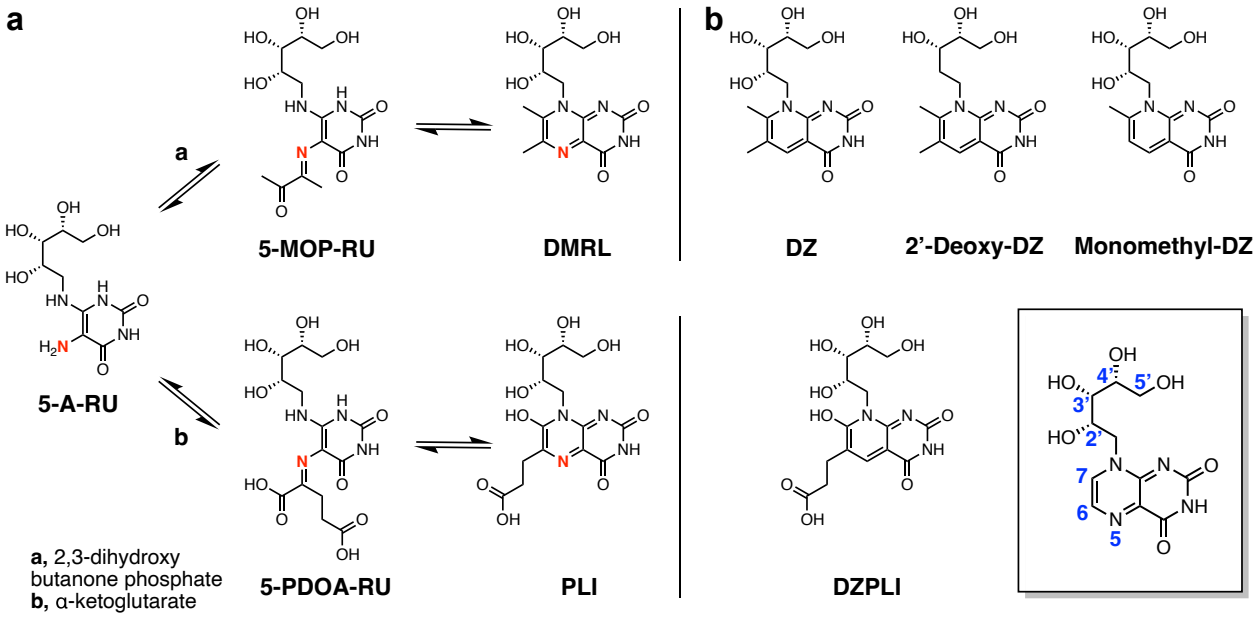


Figure 3.1. Proposed chemical structures for deaza-molecules. (a) Reaction scheme of the reversible 5-A-RU conversions to DMRL via 5-MOP-RU, a validated ribitylpyrimidine intermediate, and to PLI via a proposed 5-PDOA intermediate. The nitrogen (N5) that was modified for this study is emphasized in red. (b) Proposed chemical structure of DZ, monomethyl-DZ, 2′-deoxy-DZ, and DZ-PLI chemical structure. The boxed structure is the core ribityllumazine (RL) structure with atom numbering in blue.

nature of the ligand:MR1 complex [75]. Unfortunately, this question is difficult to address because few ligands capable of stimulating any MAIT TCR have been formally described [155]. Thus, in the absence of additional bacterial ligands, synthetic variants of known ligands provide unique tools to determine the molecular rules underlying MAIT and MR1T TCR engagement. Here, we sought to assess MR1T cell ligand selectivity using modified deaza-analogues of known ribityllumazine MR1 ligands. We tested the ability of these analogues to activate MR1T cells and stabilize MR1 on the cell surface, then used *in silico* techniques to model how minor ligand modifications may contribute to our cellular observations. We found that each MAIT and MR1T cell clone had different patterns of reactivity toward DMRL or the deaza-analogue (DZ). *In silico* molecular docking and molecular dynamic simulations demonstrate that ligand modifications that are buried deep in the ligand binding groove may result in altered dynamics of the ribityl moiety, thus possibly contributing to this differential recognition. Thus, the use of synthetic analogues to known bacterial MR1 ligands have the potential to provide insight into the molecular underpinnings of ligand:MR1 interactions with MR1T TCRs.

3.2 Synthesis of deazalumazine, 2'-deoxy-deazalumazine, monomethyl deazalumazine, and deazaphotolumazine I

The common precursor for the ribityllumazines and the ribitylpyrimidines is 5-A-RU, the key intermediate of the riboflavin biosynthetic pathway that is thought to give rise to a number of MR1 ligands. *In vivo*, 5-A-RU condenses with 2,3-dihydroxy-butanone phosphate (DHBP) in either an enzymatic or spontaneous manner to form DMRL through a ribitylpyrimidine intermediate known as 5-MOP-RU (**Figure 3.1a**, top) [58, 156]. Likewise, it is thought that other

ribityllumazines - such as Photolumazine I, a ligand described by Harriff et al. [77] - arise in a similar fashion. Indeed, retrosynthesis suggests that PLI arises from the condensation of 5-A-RU with α-ketoglutarate to generate a 5-(2-imino-pentanedioate)-8-D-ribityluracil (5-PDOA-RU) intermediate, which then spontaneously cyclizes to produce the corresponding ribityllumazine (**Figure 3.1a**, bottom) [77, 146].

In order to synthesize deaza-forms of lumazines, we took inspiration from preparations of lumazine synthase inhibitors where the 5-nitrogen of a non-ribityl lumazine core was replaced with a carbon (see **Figure 3.1** inset box for atom numbering) [157]. Specifically, we replaced the 5-nitrogen of the known ribityllumazine ligand 6,7-dimethyl-8-D-ribityllumazine (DMRL) or a monomethyl version of this ligand, 7-Me-RL, with a carbon to produce the known compounds 6,7-dimethyl-8-D-ribityldeazalumazine (deazalumazine or DZ) and 7-methyl-8-D-ribityldeazalumazine (monomethyl-DZ or mmDZ), respectively (**Figure 3.1b**). It is worth noting that the original design was to create a 5-deaza version of 5-OP-RU, but this compound spontaneously cyclized during synthesis to produce monomethyl-DZ. Additionally, we prepared 2′-deoxy-DZ (dDZ), lacking the 2′-hydroxyl of the ribityl moiety critical for ribitylpyrimidine and ribityllumazine contacts with MAIT TCRs. Finally, in order to study the effect of the 5-deaza-modification across ligands, we also synthesized a deaza-version of photolumazine I (DZPLI).

Syntheses of analogues DZ, 2′-deoxy-DZ, DZPLI and monomethyl-DZ are summarized in **Figure 3.2**. Substitution of commercially available 6-chlorouracil **1** with ribitylamine or 2′-

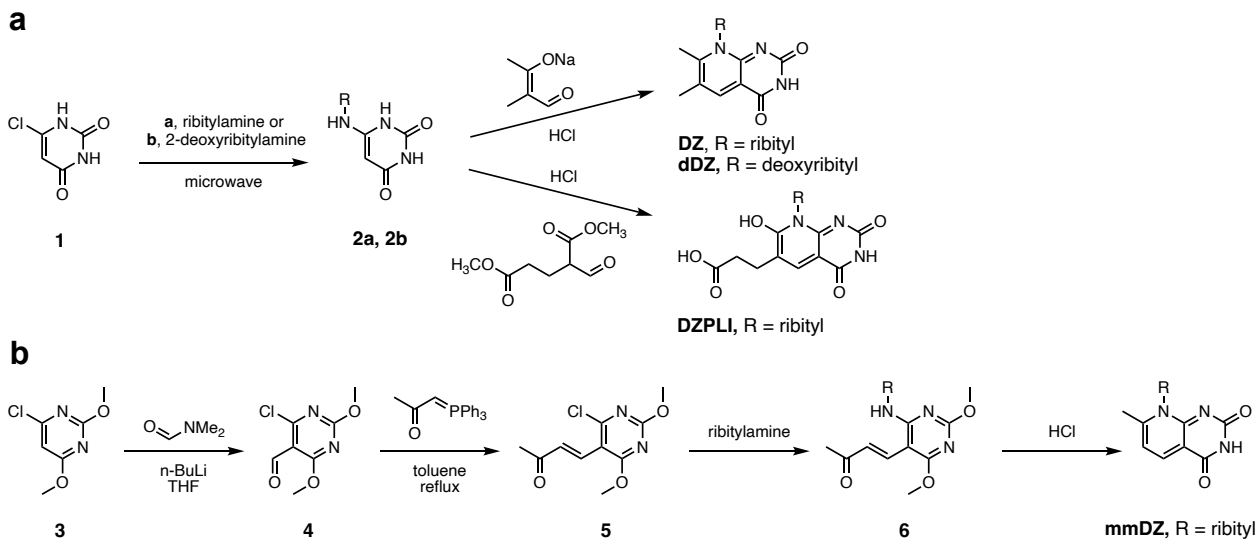


Figure 3.2. Synthesis of deazalumazine (DZ), 2′-deoxy-deazalumazine (2′-deoxy-DZ), and deazaphotolumazine I (DZPLI). (a) Chemical synthesis scheme for DZ, 2′-deoxy-DZ, and DZ-PLI. (b) Chemical synthesis scheme for monomethyl-DZ. Synthesis details for both schemes are described in the "Experimental procedures".

deoxyribitylamine in water at 150 ° C in a microwave reactor led to the corresponding ribityluracil **2a** or 2′-deoxyribityluracil **2b** (**Figure 3.2a**). The use of a microwave reactor rather than conventional heating boosted the reaction yield tenfold. Compounds **2a** and **2b** were then reacted with the sodium salt of 2-methyl-3-oxo-butanol under acidic conditions at 100 ° C to give DZ and 2′-deoxy-DZ, respectively. DZPLI was formed by reacting ribityluracil **2a** with 1,5-diethyl-2-formylpentanedioate under acidic conditions at 100 ° C. Monomethyl-DZ was prepared in four synthetic steps as illustrated in **Figure 3.2b**. The initial step, formation of aldehyde **4**, was accomplished via formylation of commercially available 4-chloro-2,6-dimethoxypyrimidine **3** by direct ortho-lithiation with n-butyllithium in tetrahydrofuran (THF) followed by formylation with N,N-dimethylformamide (DMF). Subsequently, the Wittig reaction was accomplished by reacting aldehyde **4** with commercially available (acetylmethylene) triphenylphosphorane; the product **5** was obtained in high yield. Compound **6** was obtained by chloro-displacement of compound **4** with ribitylamine. The crucial part of the synthesis was the acidic removal of the

compound **6** methoxy groups to give monomethyl-DZ. After screening a variety of conditions, *e.g.*, BBr_3 in CH_2Cl_2 and various concentrations of HCl, it was found that 6 M hydrochloric acid was able to convert the pyrimidine ring to the corresponding uracil with acceptable yield. Using HPLC analysis, the purities of DZ, 2′-deoxy-DZ, monomethyl-DZ and DZPLI were determined to be greater than 95%.

3.3 DZ, dDZ, mmDZ, and PLI stabilize MR1 on the cell surface

In the absence of ligand, MR1 is sequestered in the ER and endosomal compartments, but ligand availability triggers egress to the cell surface through an unknown mechanism [140, 141]. Therefore, we looked at whether or not these synthetic ligands were capable of stabilizing MR1 on the cell surface, similar to what has been observed previously for the MR1 ligand 6-formylpterin (6-FP) and for 5-OP-RU [8, 140]. Airway epithelial (BEAS-2B) cells overexpressing MR1 [140] were incubated with saturating concentrations of 6-FP, DZ, 2′-deoxy-DZ, monomethyl-DZ, PLI, or DZ-PLI then surface stained for MR1 using the α-MR1 26.5 antibody [99]. Here, we found that each ligand except DZPLI was able to stabilize MR1 on the cell surface (**Figure 3.3**). Although it has been suggested that Schiff base formation between the ligand and K43 of the MR1 pocket is necessary for MR1 egress from the ER [141], DZ, dDZ, and mmDZ - which do not have a compatible moiety for ketone/amine condensation with the primary amine of K43 - are clearly able to upregulate MR1 on the cell surface. Though it is not yet known whether or not PLI can form a Schiff base with MR1, it is also clearly able to upregulate MR1 on the cell surface. Therefore, we expect that, with the exception of DZPLI, these ligands are competent for presentation to T cells by MR1 on the cell surface.

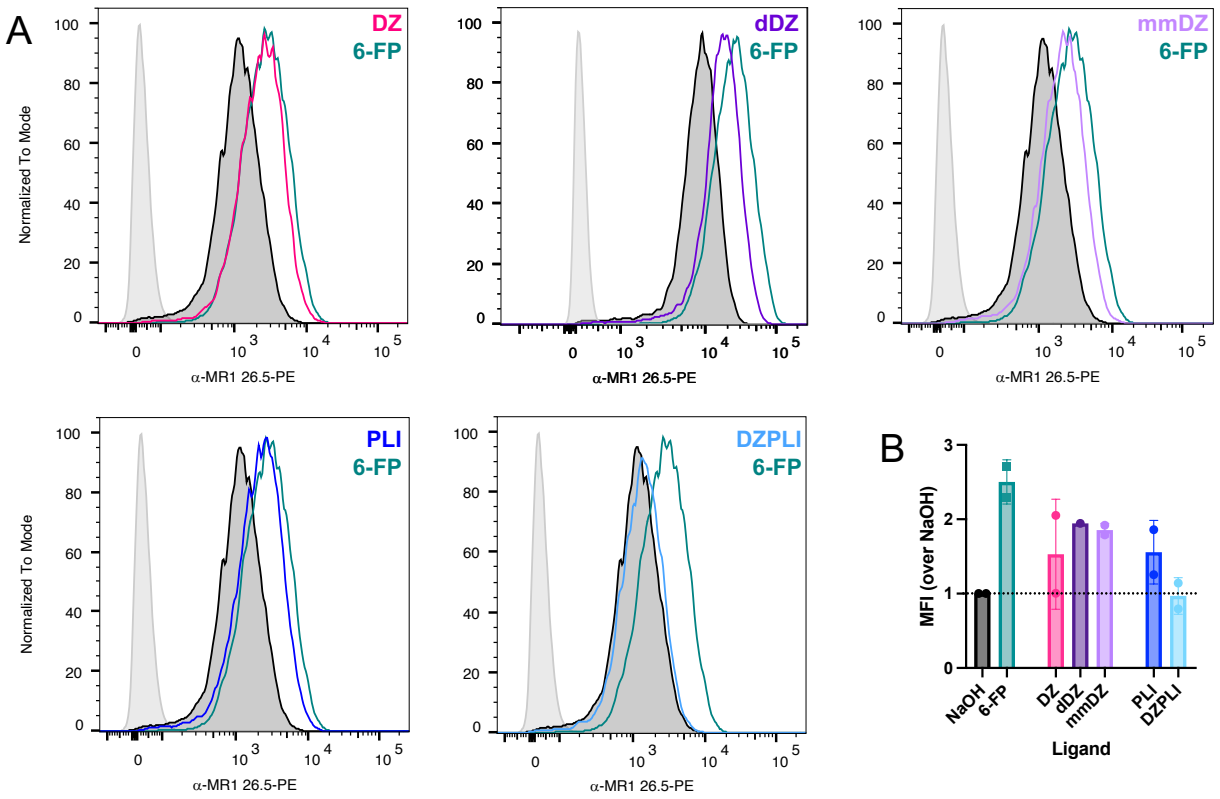


Figure 3.3. Cell surface stabilization of MR1 by DMRL, PLI, and their analogues. a) WT BEAS-2B cells overexpressing MR1-GFP were incubated with indicated ligands at 100 μM for 16 hours. MR1 on the cell surface was detected by surface staining with the α-MR1 26.5 antibody conjugated to APC. Dark grey shaded histograms represent the NaOH vehicle on MR1-overexpressing cells. Light gray shaded histograms represent the FMO. b) Median fold increase (MFI) of α-MR1 antibody staining for each ligand relative to an NaOH vehicle.

3.4 MR1T cell clones variably discriminate between DZ and DMRL

To determine whether or not DZ is a MR1T cell antigen, we then tested the synthetic DZ molecule and its parent structure DMRL for their ability to stimulate a panel of MAIT and MR1T cell clones in an IFN-γ ELISpot assay with human monocyte-derived dendritic cells (DC) as the antigen presenting cell. D426G11, D481A9, D481C7, and D481F12 are previously described MAIT cell clones where the TRAV1-2 α-chain is paired with diverse β-chains and CDR3 amino acid sequences (**Table 3.1**) [77, 100]. Despite the minor nitrogen to carbon substitution, the

Clone name	TRAV	CDR3α	TRAJ	TRBV	CDR3β	TRBJ
D426G11	1-2	CAVRDSNYQLIW	33	6-4	CASSDSGESGTEAFF	1-1
D481A9	1-2	CAAMDSNYQLIW	33	20-1	CSARQGAESREQY	2-7
D481C7	1-2	CAVSLQDYKLSF	20	6-4	CASSPSGGDYNEQF	2-1
D481F12	1-2	CAVRDSDYKLSF	20	4-2	CASSQIAGGQQETQY	2-5
D462E4	12-2	CAVRDAGNMLTF	39	29-1	CSVGGDSLIGNQPQHF	1-5

Table 3.1. TCR rearrangements and sequences for MR1T cell clones.

response of each MR1T cell clone to DC pulsed with DZ varied in comparison to the same clone's response toward DMRL (**Figure 3.4a**). There were no significant differences in the responses to DMRL or DZ for D481F12 ($p = 0.91$). D426G11 responded to DZ, but significantly less well than to DMRL ($p < 0.0001$). The response of D481A9 to DMRL was also significantly higher than to DZ ($p < 0.001$); in fact, this clone did not respond to DZ at any of the tested concentrations. In contrast, D481C7 had a significantly more robust response to DZ than to DMRL ($p = 0.0007$). To determine whether this phenomenon was also observed for TRAV1-2$^-$ MR1T cell clones, we also tested the D462E4 clone, which expresses TRAV12-2 paired with TRBV29-1 (**Table 3.1**) [70]. Interestingly, D462E4 was selective toward DZ, as it did not respond to DMRL at any tested concentration ($p = 0.0008$) (**Figure 3.4**). These data suggest that DZ is capable of acting as an MR1T cell agonist in a TCR-dependent manner, and that a minor chemical substitution that is not expected to be TCR-accessible is nonetheless able to modulate MR1T cell responses.

3.5 Impact of chemical modifications to DZ and PLI on activation of the D481C7 MAIT cell clone

We were then interested in how other deaza-molecules and/or further derivatization of DZ would

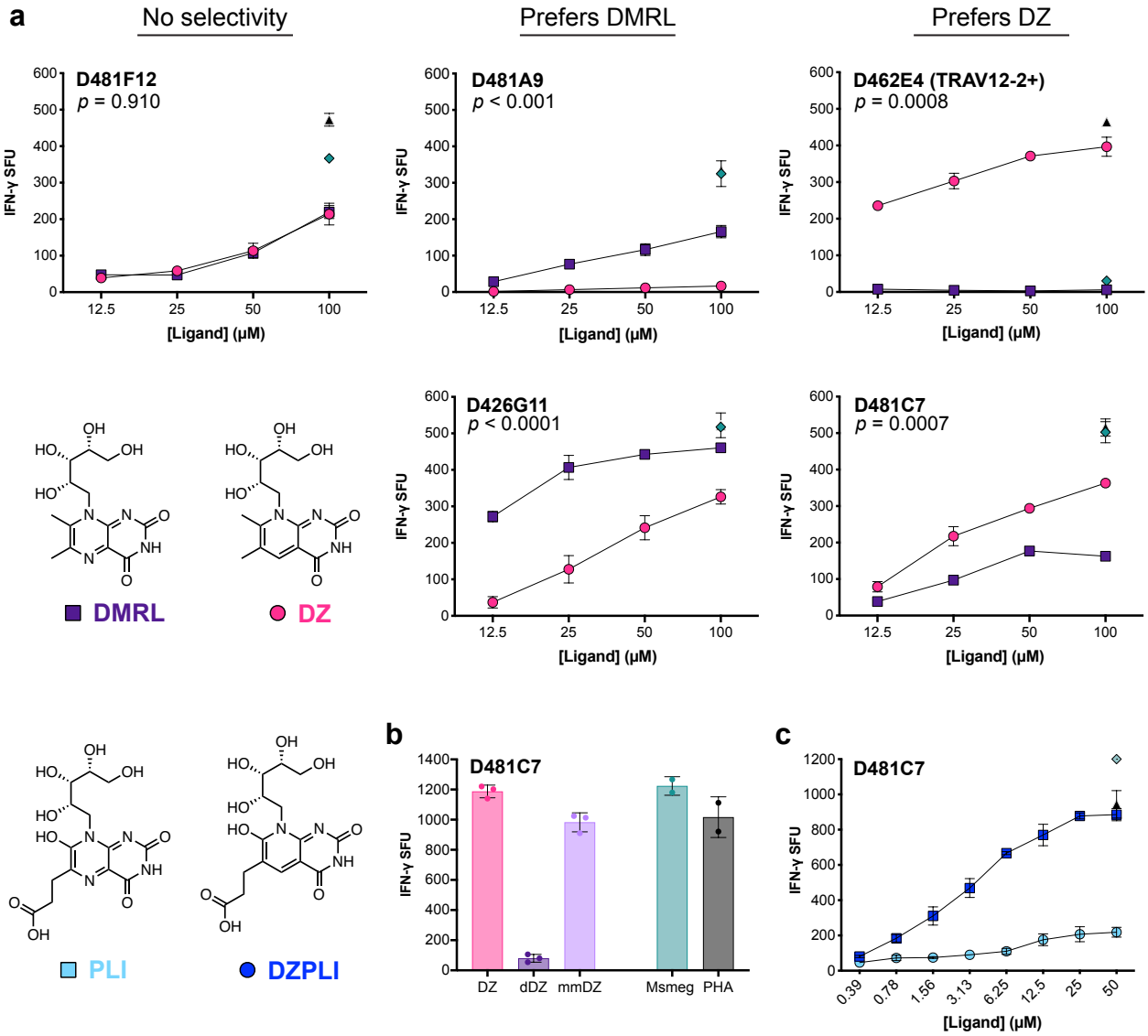


Figure 3.4. MR1T cell responses to DZ and derivatives thereof. (a) The IFN-γ response by 1× 10^4 MR1T cell clones was measured in an ELISpot assay with 1× 10^4 DC pulsed with the indicated concentrations of DMRL or DZ. *P* values were determined by simple linear regression of the slope and elevation of for DZ and DMRL. (b, c) The IFN-γ response by 2× 10^4 D481C7 MAIT cell clones was measured in an ELISpot assay with either (b) 1× 10^4 DC pulsed with DZ, 2′-deoxy-DZ, or monomethyl-DZ at 100 μM or (c) 2× 10^4 DC pulsed with PLI or DZ-PLI at the indicated concentrations. PHA and supernatant from *M. smegmatis* were used as described as positive controls for all assays. Error bars indicate the mean and standard deviation of technical replicates. Data shown are representative of 3 independent experiments.

influence MAIT cell activation. Since the D481C7 MAIT cell clone was robustly stimulated by DZ (**Figure 3.4a**) and PLI [77], we used this clone in an ELISpot assay measuring its IFN-γ production toward 2′-deoxy-DZ, monomethyl-DZ, PLI, and DZPLI. Removing the 2′-hydroxyl group from the ribityl chain of DZ to generate 2′-deoxy-DZ reduced the response of the D481C7 clone tenfold (**Figure 3.4b**). This is consistent with other reports that have determined the importance of the 2′-hydroxyl group for MAIT TCR recognition [58, 63, 75, 89, 90]. In contrast, modifying DZ to generate monomethyl-DZ had minimal impact on the response of the D481C7 clone (**Figure 3.4c**), but it is difficult to interpret this result in the absence of a direct comparison of this clone's response to the mmDZ (5-aza-) parent molecule, 7-Me-RL.

Previously, we demonstrated that only the D481C7 clone had robust responses to PLI, while the other clones had little to no response [77]. MR1T cell clones that did not recognize PLI also did not recognize DZPLI (**Supplemental Figure S3.1**). Whereas the D481C7 clone had a significantly higher response to DZ when compared to DMRL itself, the response of the D481C7 clone to DZPLI was reduced nearly fivefold when compared to PLI (**Figure 3.4c**). Since we did not observe MR1 cell surface stabilization by DZPLI, the reason for this residual activation is unclear, but it may be explained by an incompatibility between the kinetics of MR1 egress dynamics (which can vary across ligands, as shown previously [75]) and the time point at which we stained the cells for flow cytometric analysis. Together, these results show that, similar to what has been observed for other ribitylpyrimidine and ribityllumazine compounds, the 2'-hydroxyl of DZ is an important contact for its recognition by a the D481C7 MAIT clone TCR.

3.6 MAIT cell recognition of DZ is MR1-dependent

We next confirmed the requirement for MR1 in MAIT cell responses to DZ in two ways. First, we tested MAIT cell responses within the context of blocking with the α-MR1 26.5 antibody. Here, when human DCs were incubated with the α-MR1 26.5 antibody prior to being pulsed with DZ, the response by both the D481C7 and D426G11 MAIT cell clones was completely abrogated (**Figure 3.5a**). Second, we tested MAIT cell responses in an airway epithelial (BEAS-2B) cell line from which MR1 had been knocked out by CRISPR/Cas9 [158]. We first demonstrated that both MAIT cell clones could recognize DZ presented by BEAS-2B cells and human DCs (**Figure 3.5b**). As expected based on the results shown in **Figure 3.4a**, the response of the D426G11 clone was less than that of the D481C7 clone to both DC and BEAS-2B cells. Although the response to DZ-pulsed BEAS-2B cells, a non-professional antigen presenting cell, was lower than the response to DZ-pulsed DCs for both clones, MAIT cells are nonetheless clearly capable of recognizing DZ presented by these cells. When either MAIT cell clone was incubated with DZ-pulsed BEAS-2B:ΔMR1 cells, however, they were unable to produce IFN-γ (**Figure 3.5c**). Confirming the requirement for MR1, the IFN-γ response by both MAIT cell clones was restored when MR1 expression was reconstituted in the BEAS-2B:ΔMR1 cells (**Figure 3.5c**). Together, these results demonstrate that, as with other ligands including DMRL, professional antigen presenting cells and airway epithelial cells are capable of presenting DZ to MAIT cells in an MR1-dependent manner. Thus, DZ is a bona fide MR1T cell antigen and its continued study will contribute to the realization of ligand discrimination among distinct MAIT and MR1T TCRs.

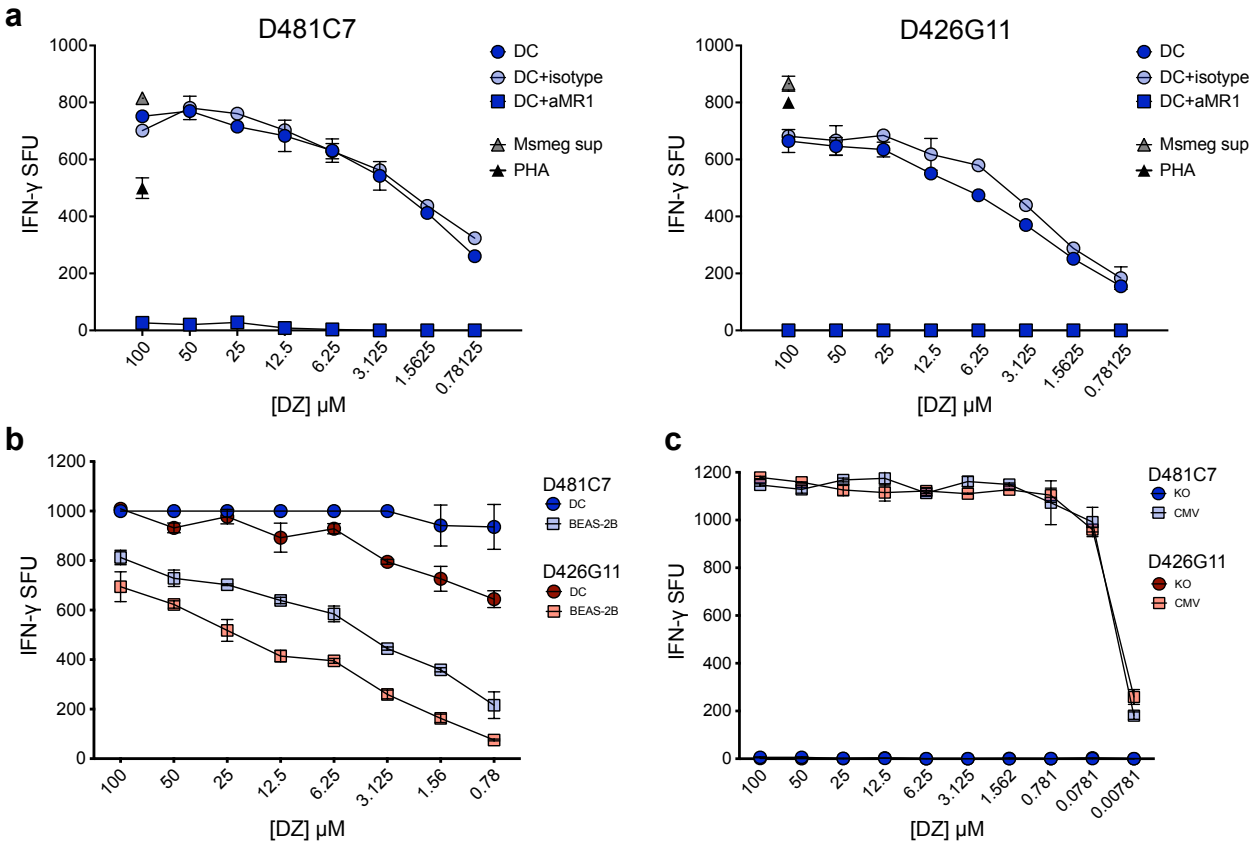


Figure 3.5. MAIT cell recognition of DZ is MR1-dependent. (a) The IFN-γ response by 2× 10^4 D481C7 or D426G11 T cells was measured in an ELISpot assay with 1× 10^4 DC pulsed with DZ at the indicated concentration. Where indicated, the α-MR1 26.5 antibody or isotype control was added one hour prior to adding DZ to the wells. PHA and supernatant from *M. smegmatis* were used as described as positive controls for the assay. (b) The IFN-γ response by D481C7 or D426G11 was measured in an ELISpot assay with 1× 10^4 DC or 1× 10^4 wild-type BEAS-2B cells pulsed with DZ at the indicated concentration. (c) The IFN-γ response by D481C7 or D426G11 was measured in an ELISpot assay with 1× 10^4 BEAS-2B:ΔMR1 or BEAS2B:ΔMR1+MR1 cells pulsed with DZ at the indicated concentrations. Error bars indicate the standard deviation of technical replicates. Data shown are representative of 3 independent experiments.

3.7 DMRL, DZ, and 2'-deoxy-DZ exhibit differential dynamics in the MR1 binding pocket

In order to gain insight into a possible mechanism behind the differential agonistic properties of these chemically-modified forms of DMRL, we employed computational approaches to provide atomistic insights into our cellular observations. Limited ligand availability and the difficulty of refolding MR1 with ribityllumazines [79, 90] made co-crystallization of the ligand with MR1 unfeasible, so AutoDock Vina [159] was used to model their docking. Vina's predictions are based on the calculated chemical potentials of all possible ligand conformations, and the output

provides an internal scoring system to rank binding modes (in "affinities," measured in kcal/mol). While these Vina scores are given in kcal/mol, the scoring function of AutoDock Vina does not necessarily correlate well with experimental affinities, and instead should be considered as an internal reference score [159]. For each dock, the highest-ranking conformation was chosen for each ligand; the corresponding affinities can be found in **Supplemental Table S3.1** and visualizations of these models are shown in **Figure 3.6** and **Supplemental Figures S3.2** and **S3.3**. Further details for this docking protocol can be found in the **Materials and methods**, and coordinates of the docks are provided in the supplemental material of Jin and Ladd et al. [160].

We began by testing the performance of AutoDock Vina by removing the ligands and TCRs from crystal structures containing agMR1 complexes, then docking the ligands back into these "donor structures" to see how well the dock agreed with the experimental data. We selected the A′ pocket of MR1 as the docking search space, as this is where the ribityllumazines and pyrimidines have been shown to bind the groove [57, 58, 63, 89, 90]. Docked ligands that have been previously crystallized show good agreement with experimentally determined structures, including docking HMRL back into human $MR1^{K43A}$ (PDB 4LCW, ternary complex) (**Figure S3.2a**); docking human MR1 (PDB 4L4V, ternary complex) (**Figure S3.2b**); and docking 5-OP-RU back into $MR1^{K43A}$ (PDB 4NQD, ternary complex) (**Figure S3.2c**). Notably, rRL-6-CH_2OH docked back into chimeric bovine-human MR1 (PDB 4LCC, ternary complex) (**Figure S3.2d**) is flipped about the plane of the ring that is observed in the experimental crystal structure [89]. However, the MR1 used to crystallize this complex was loaded with a heterogeneous mixture of ligands (including rRL-6-CH_2OH, per mass spectrometric analysis)

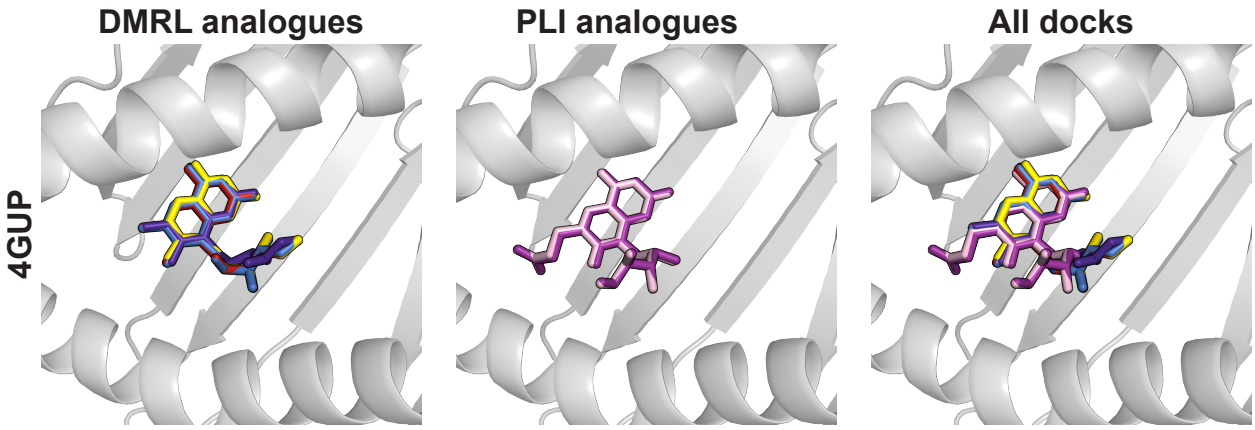


Figure 3.6. Docking mode of ligands into 4GUP donor structure. Left, DMRL (red), DZ (blue), 2'-deoxy-DZ (purple) and monomethyl-DZ (yellow) docked into 4GUP. Center, PLI (light pink) and DZPLI (purple) docked into 4GUP. Right, Overlay of all ligand docking modes showing overlap of the lumazine rings.

and therefore the electron density may represent the average of that of many ligands, possibly obscuring the positioning of the ligand in the pocket. Thus, we were confident that the AutoDock Vina was able to recapitulate experimental docking modes for structures with unambiguous, homogeneously-loaded ligands.

We then proceeded to dock the ligands synthesized and characterized in this paper into human MR1 (PDB 4GUP, which was crystallized in the absence of a TCR [57]). Comparing the DMRL, DZ, dDZ, and mmDZ docked in the 4GUP donor structure, the lumazine cores overlap generally, but DMRL is particularly aligned with DZ, presumably because the C6 and C7 substituents are identical for these two compounds (**Figure 3.6a**). Additionally, the ribityl chain of 2′-deoxy-DZ (purple) is posed differently from that of DMRL, DZ, and monomethyl-DZ, presumably owing to the absence of the 2′ hydroxyl group. Given the similarity of the DZ and 2′-deoxy-DZ binding modes, we reason that the inability of 2′-deoxy-DZ to stimulate MR1T cell clones (**Figure 3.4b**), while still being able to stabilize MR1 at the cell surface (**Figure 3.3**), is likely due to changes in

TCR contacts with the ligand and/or MR1, as is supported by crystallographic and functional evidence [58, 75, 90]. We also attempted docking of PLI and DZPLI and observed that in the 4GUP donor structure, they were essentially superimposable (**Figure 3.6b**) and the cyclic core of the molecule overlapped well with those of the DMRL analogues (**Figure 3.6c**).

We then wanted to test whether the structural adaptations of the MR1 ligand groove upon TCR binding - such as the rotation of MR1 residue Y152 and relaxation of the α_2 helix, which together "pry open" the groove [90] - affected the docking mode of the ligands. We therefore docked our ribityllumazine compounds into human MR1 that was crystallized with HMRL and a MAIT TCR (though we stripped the TCR from the PDB for the purposes of docking). Again, the DMRL analogues aligned well upon one another, with the dDZ ribityl moiety rearranged (**Figure S3.3a**). Only minor changes in ligand orientation were observed in comparison with the 4GUP docks: the DMRL analogues docked into the 4L4V donor structure experience a slight rotation about C7 (see **Figure 3.1** inset for atom numbering), and ~0.5 Å translation of each ligand, though they stayed in the same plane (**Figure S3.4**; only DMRL is shown for clarity). This resulted in a shift into the pocket (with greater movement of the uracil ring) and reorganization of the ribityl moiety. With respect to PLI and DZPLI, the cyclic cores were no longer superimposable as was seen for the 4GUP docks; rather, DZPLI was rotated about the uracil ring relative to PLI (**Figure S3.3**). When compared with the 4GUP dock, each ligand underwent conformational rearrangement of its propanoate acid group but only DZPLI experienced an additional rearrangement for its ribityl moiety (**Figure S3.4**). The fact that DZPLI was docked successfully into MR1 despite its inability to stabilize MR1 at the cell surface (**Figure 3.3**) is not necessarily

surprising, considering the biophysical and cellular parameters that cannot be recapitulated *in silico*. Nonetheless, these data suggest that this docking strategy is sensitive to minor changes in both protein and ligand structure, lending confidence toward these models.

To validate the docking mode of the DMRL analogues, we employed all-atom molecular dynamics of the agMR1 complexes to simulate the temporal evolution of each ligand in the binding pocket. We performed these simulations on the 4GUP docks to prevent bias of the starting model based toward protein alterations upon TCR engagement [90]. Simulation details can be found in the **Materials and methods** section. Starting from the structures of each ligand docked into the 4GUP donor structure, we simulated each system for 80 ns, running simulations in triplicate to minimize stochastic effects from standard velocity initialization [161, 162]. Ligand RMSD was stable over time, indicating that agMR1 complexes are stable *in silico* (**Figure S3.5**), adding further confidence to our docked models. Interestingly, over the course of these simulations, we found that DMRL, DZ, and 2′-deoxy-DZ adopt differential dynamics within the MR1 binding pocket. Using root mean square fluctuation (RMSF) of each atom within the ligand, we quantified this dynamic motion (**Figure 3.7b–d**, left) and found that in contrast to DMRL, the RMSF of atoms composing the bicyclic core of DZ are greater; conversely, the RMSF of atoms composing the ribityl moiety of DZ are smaller than those of DMRL. When comparing DZ and 2′-deoxy-DZ, ablation of the 2′-hydroxy enhances plasticity of the ribityl moiety without much change to the dynamics of the deazalumazine core. Through alignment of the dynamics of each ligand by its cyclic core, we visualized the extent of these fluctuations in the ribityl chains (**Figure 3.7e–g**) for a representative simulation of each ligand (simulations

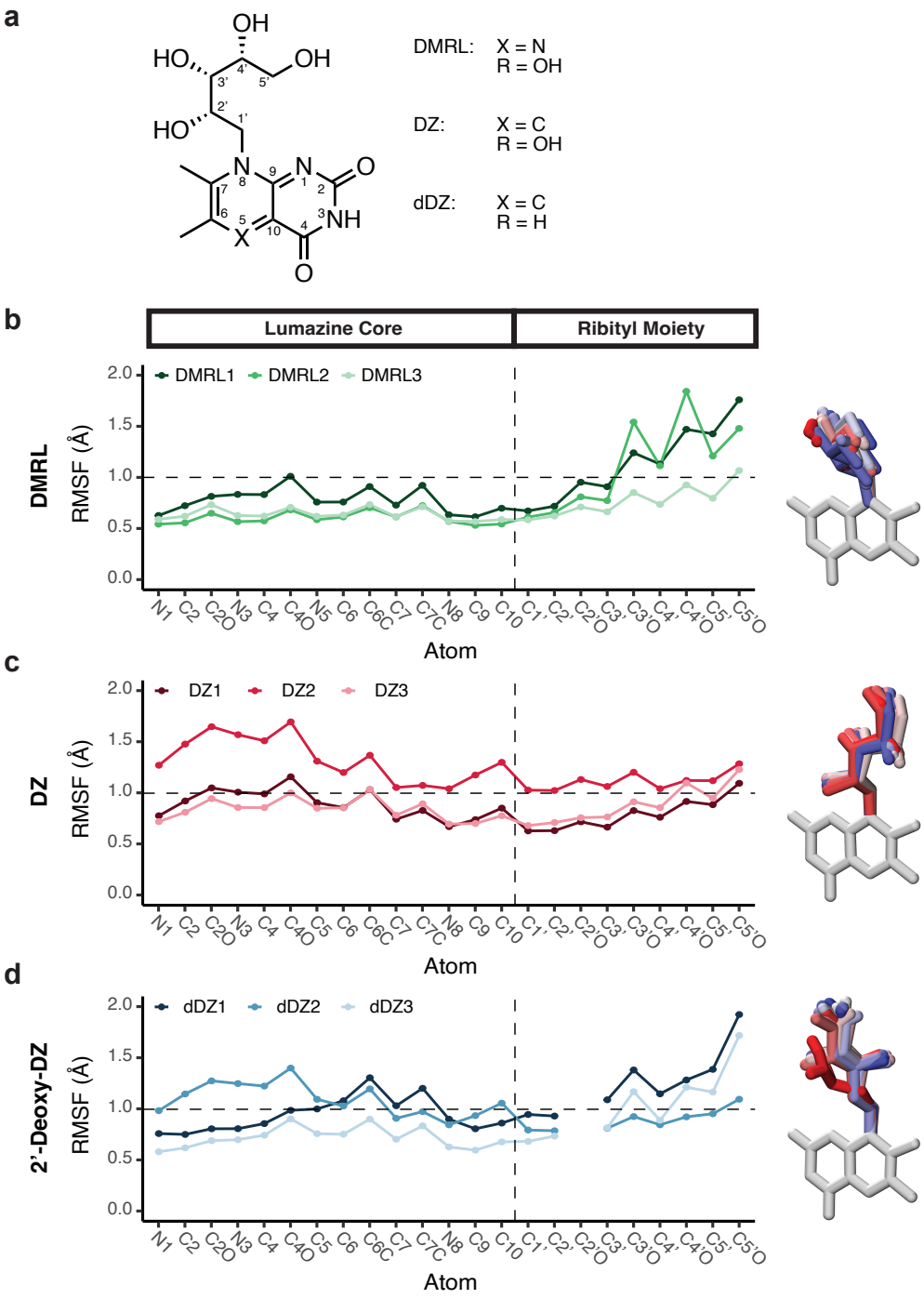


Figure 3.7. Molecular dynamics simulations of DZ, 2′-deoxy-DZ, and DMRL ligands docked to MR1 identify differential flexibility of each ligand within the binding pocket. (a) Chemical representation of DMRL and its derivatives; numbering of adduct atoms is relative to the core atoms (i.e. the methyl group on carbon 7 is named "C7C" and the 3′ hydroxyl is referred to as "C3′O"). (b–d) Left panels show RMSF calculations (indicative of atom dynamics) for each simulation replicate over the course of each 80 ns MD simulation for DMRL (b), DZ (c) or 2′-deoxy-DZ (dDZ) (d). The horizontal line at 1 Å does not represent any parameter; it only serves as a visual guide to compare across plots. Right panels show visualization of dynamic states adopted by the ribityl chain in a representative simulation for each ligand (simulations DMRL2, DZ1, and dDZ1), evolving from red to blue across time.

DMRL2, DZ1, and 2′-deoxy-DZ1, i.e. ddZ1). Taking snapshots every 10 ns of simulated time, we observed that the ribityl moiety of DMRL and 2′-deoxy-DZ were quite dynamic, though that of DMRL sampled space mostly through transposition of the entire ribityl moiety, while that of 2′-deoxy-DZ sampled space by rotation of the 5′ hydroxyl (reflected in the RMSF patterns). The ribityl chain of DZ, on the other hand, was strikingly rigid in all three of the simulation replicates. Our observation of enhanced fluctuations of the core-distal ribityl moiety atoms due to 2′-deoxy modification (in comparison with DZ) is consistent with those seen in previous crystallographic studies with of ternary complexes with synthetic deoxyribityl-versions of 5-OP-RU, as well as with molecular dynamics simulations performed on those ternary structures [75]. Just as hypotheses (based on molecular dynamics data and crystallographic evidence) have been presented with respect to peptide-dependent MHC dynamicity and TCR selectivity [163], these data suggest that minor modifications to MR1 ligands can alter their dynamics in the MR1 pocket, which may in turn affect TCR recognition and contribute to the phenomenon of TCR selectivity that we have observed experimentally.

3.8 Discussion

Microbial metabolites presented by MR1 and recognized by MR1T cells represent a novel and largely unexplored class of T cell-activating ligands, but targeting MR1T cells for therapeutic or vaccine-related applications requires a better understanding of structure and stability of ligands interacting with MR1 and the MR1T TCR to induce activation. While ribitylpyrimidines like 5-OE-RU and 5-OP-RU are potent antigens, their instability *in vitro* makes them challenging candidates for these applications [58, 74]. Furthermore, while there are now multiple recent

studies demonstrating that treatment with 5-OP-RU expands MR1T cells and protects against infection with organisms like *Francisella tularensis* and *Legionella longbeachii* in a mouse model (e.g. [164-166]), it is unclear whether these findings will translate to humans. For example, in a mouse model, treatment with 5-OP-RU does not result in protection against *Mycobacterium tuberculosis*, despite similar MR1T cell expansion [167]. Additionally, a recent study in macaques demonstrated that treatment with 5-OP-RU not only failed to provide a therapeutic benefit in the context of *Mycobacterium tuberculosis*, it also resulted in functional impairment and exhaustion of MR1T cells [168]. Together, these studies suggest that there is still much to understand with regard to the use of 5-OP-RU or other ligands for clinical purposes.

The majority of the functional and molecular work done in the MR1 field has focused on 5-OP-RU and high affinity MAIT TCRs, leaving the ribityllumazine and other unidentified classes of ligands under-explored. For example, 26 of the 49 existing agMR1-TCR structures contain 5-OP-RU or molecular variants of 5-OP-RU, and another 10 contain 6-FP or acetyl-6-FP, toward which MR1T reactivity is rare [78]. While it is a useful tool for contrasting with other MR1T cell ligands, the physiological relevance of 6-FP (and, some would argue, 5-OP-RU) remains unclear. Due to a combination of the emphasis on the potent ribitylpyrimidines and technical limitations surrounding loading ribityllumazines into recombinant MR1, there are unfortunately only three ternary crystal structures are in complex with ribityllumazines. Thus, more work needs to be done to identify bona fide, physiologically-relevant bacterial ligands and determine the molecular mechanisms underlying their recognition by MAIT and MR1T TCRs. In addition to limited ligand representation among crystallographic structures, 29 of the 48 structures of the

ternary complex currently available in the Protein Data Bank (PDB) are of the same TCR (known as “A-F7” or “F7"), leaving the diversity of MR1T cell TCR molecular recognition strategies largely unexplored. Thus, generation of synthetically modified ligands provides useful tools for both targeting of MR1T cells and for studying the complexity of the agMR1-TCR interactions. Here, we reveal that even minor chemical changes to TCR-inaccessible positions of MR1 ligands affect antigenicity of these ligands, highlighting the high level of antigen selectivity exhibited by MR1-restricted T cells. Interestingly, certain MR1- restricted T cell clones were agnostic to these chemical modifications, indicating the importance of studying the agMR1-TCR interaction closely.

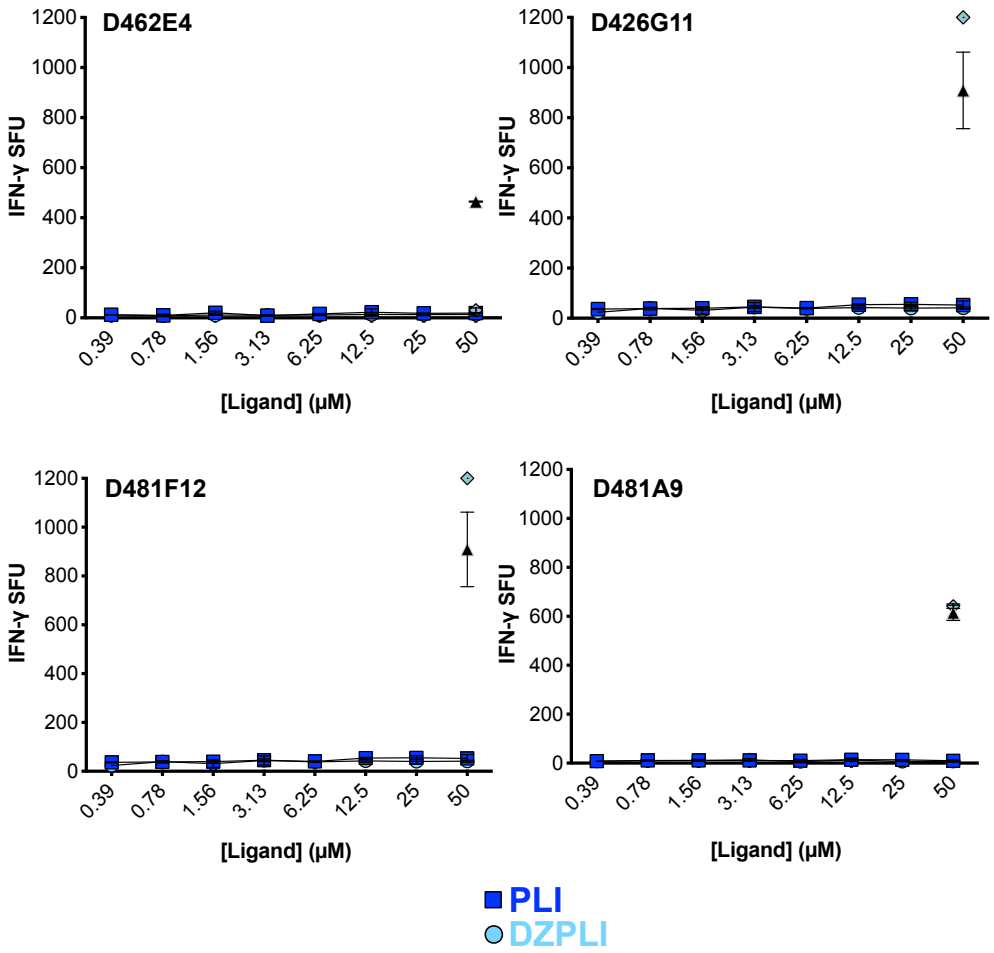
Our data support a model where minor ligand modifications affect antigen potency [74, 75, 151]. Previous work has shown that the potency of a deaza-5-OP-RU analogue was decreased considerably when assayed in an in vitro MR1T cell reporter assay with a Jurkat cell expressing a single MR1T TCR [74]. However, analysis of this molecule in the context of multiple MR1T cell TCRs could provide insight into whether the deaza-modification has different functional impacts for TCRs. When we made a similar modification to the 5-nitrogen of PLI to generate DZPLI, we also saw significantly decreased potency in the ability to activate MR1T cells, though this appears to be driven by failure to surface stabilize MR1 and not caused by TCR selectivity. The same was not true when we modified DMRL in the same way to generate DZ. While two of our MR1T TCRs exhibited a similar decrease in recognition of DZ, the deaza-version of DMRL was equally or more potent for three different MR1T TCRs. This was unexpected based on the results of Mak et al., who found that a 5,6-deaza version of 5-OP-RU was about 1,000 times less

potent toward Jurkat cells transduced with a MAIT TCR in comparison with the original 5-OP-RU compound, despite robust cell surface stabilization of the compound.

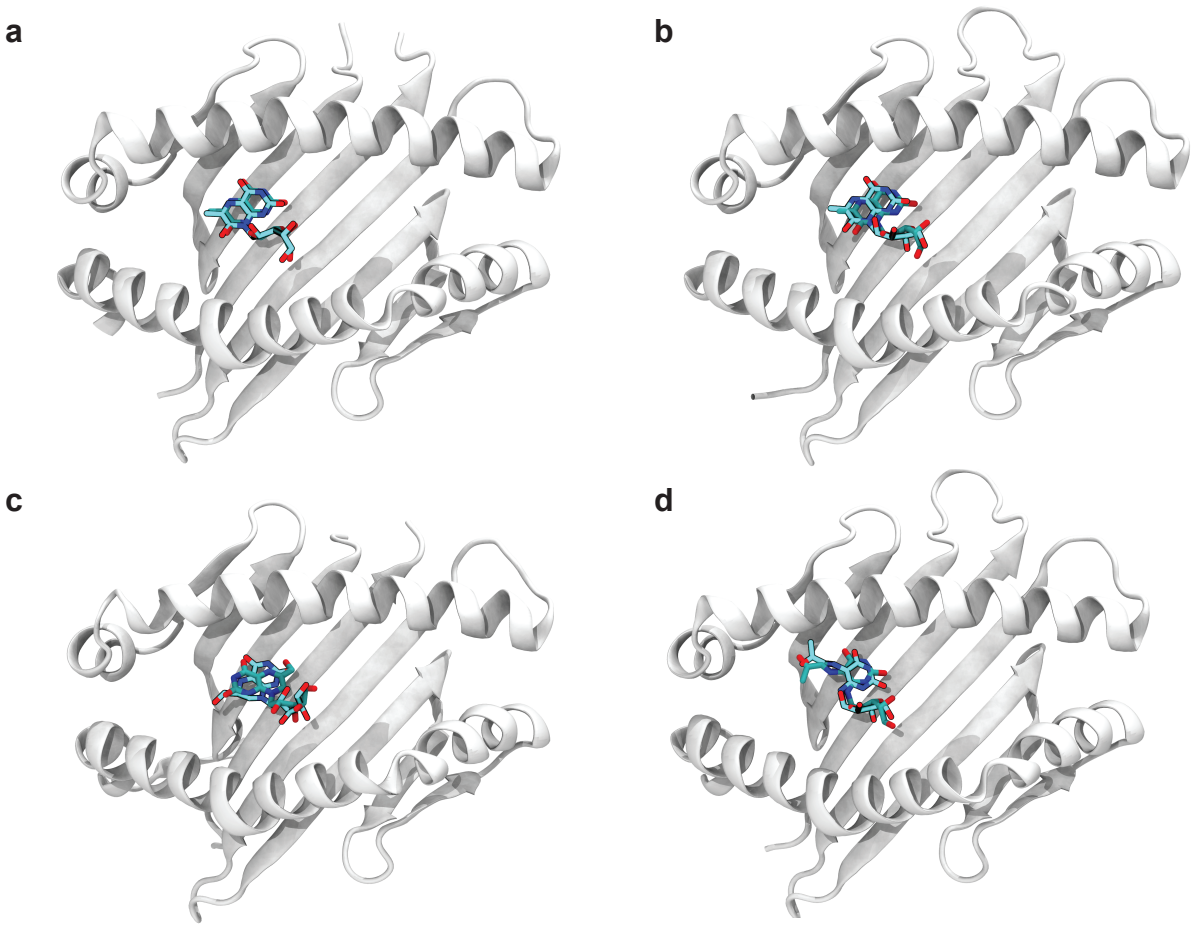
While all of the TCRs described in this paper are capable of recognizing the highly potent 5-OP-RU ligand at log-fold higher levels, these data clearly indicate that MR1T cells can discriminate between structurally similar ligands, though we do not yet understand the molecular determinants of this phenomenon. This is particularly striking because the derivatization of DMRL to DZ involves modification of the ligand at a location that is likely bound deep in the A′ pocket of MR1, and therefore inaccessible to the TCR. We find that one possible explanation of this observation is that modifications to the cyclic core may affect the dynamics of both MR1 and more distal atoms of the ribityl group, leading to differential CDR3 interactions with this group and therefore variability in MR1T cell reactivity. This is consistent with previous reports of simulations done on the MR1-TCR ternary complex with 2′-deoxy-5-OP-RU and 5′-deoxy-5-OP-RU, wherein the distal atoms of the ribityl moiety exhibited greater fluctuation in the simulations with the former [75]. However, simulation replicates were not performed in this study, making it hard to determine whether observation was due to stochastic effects. Nevertheless, these data, if underpowered, still align with our own, and suggest that similar trends for MR1 ligand dynamics exist despite structural impositions by the TCR. Collectively, our data contribute to growing evidence that MR1T cells can be ligand-permissive or exclusive, and suggest that structural differences between unique ribityllumazines may be communicated to the TCR through changes in the dynamic properties of the ribityl moiety.

Beyond unanswered questions regarding MR1T-TCR ligand discrimination events *in vitro*, the effect of these modifications on the ability of ligands to be adjuvanted or transported *in vivo* remain. Legoux et al. [169] demonstrated that 5-OP-RU could rapidly reach the thymus of mice after being painted on the skin of the thorax, but the mechanism by which this happens is not yet understood. Delivery of 5-OP-RU in this way resulted in an increase in the numbers and activation of early MR1T cell precursors from the thymus [169]. However, data demonstrating the selective clonal expansion of MR1T cells with distinct TCRs following infection with different microbes [100], and vaccine studies with 5-OP-RU [164-168] suggest that there are likely differences from what happens naturally in the context of natural infection with diverse microbes producing discrete ligands. Further, it is not clear whether exogenous delivery of a ligand with broad and potent MR1T cell activation (e.g. 5-OP-RU) will be the optimal way to target MR1T cells for therapeutic or vaccine purposes. Taken together, our data demonstrate that analysis of multiple MR1T cell antigen analogues in the context of diverse MR1T T cell receptors will be important to improving our understanding of the stability and relationship to activation of MR1T cells.

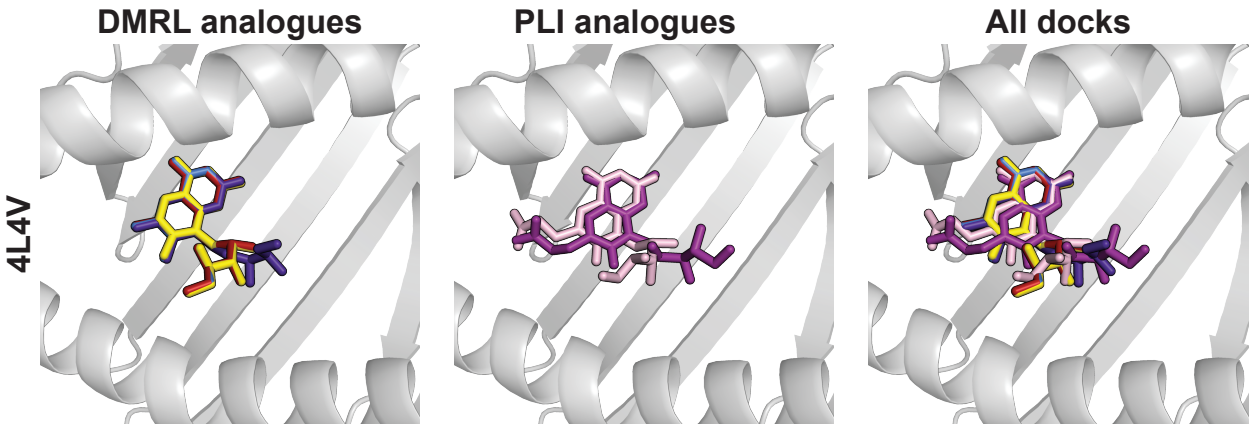
3.9 Supplemental information



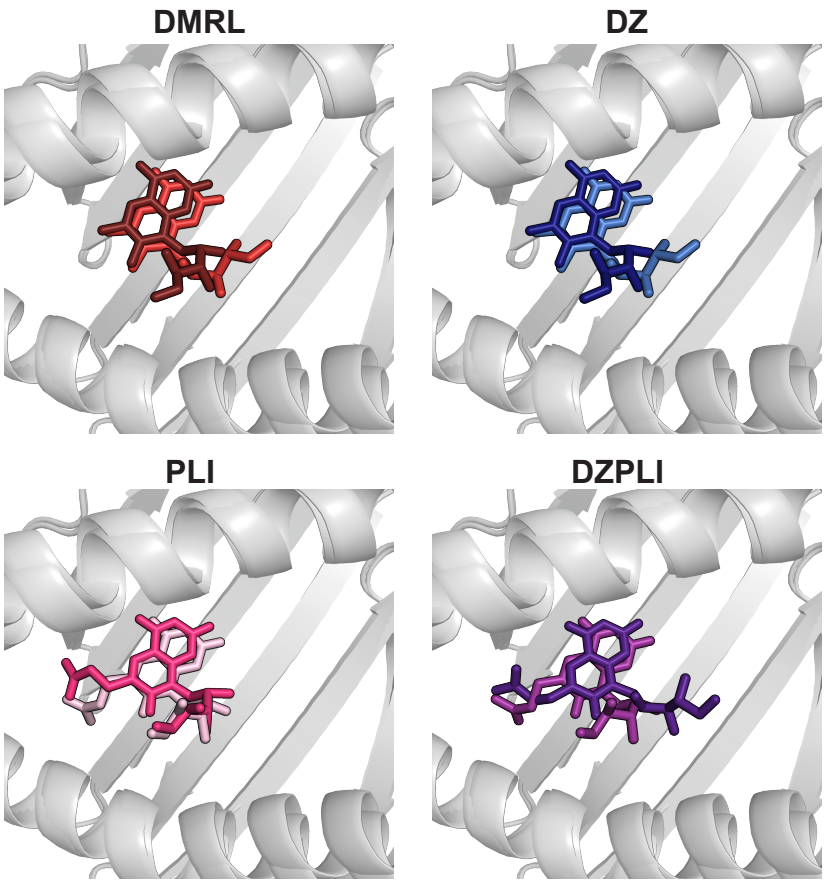
Supplemental Figure S3.1. DZPLI doesn't activate clones which do not respond to its parent molecule, PLI. The IFN-γ response by 2×10^4 of each MR1T cell clone was measured in an ELISPOT assay with PLI or DZ-PLI at varying concentrations. PHA and supernatant from *M. smegmatis* were used as described as positive controls for all assays. Error bars indicate the mean and standard deviation of technical replicates. Data shown are representative of 3 independent experiments.



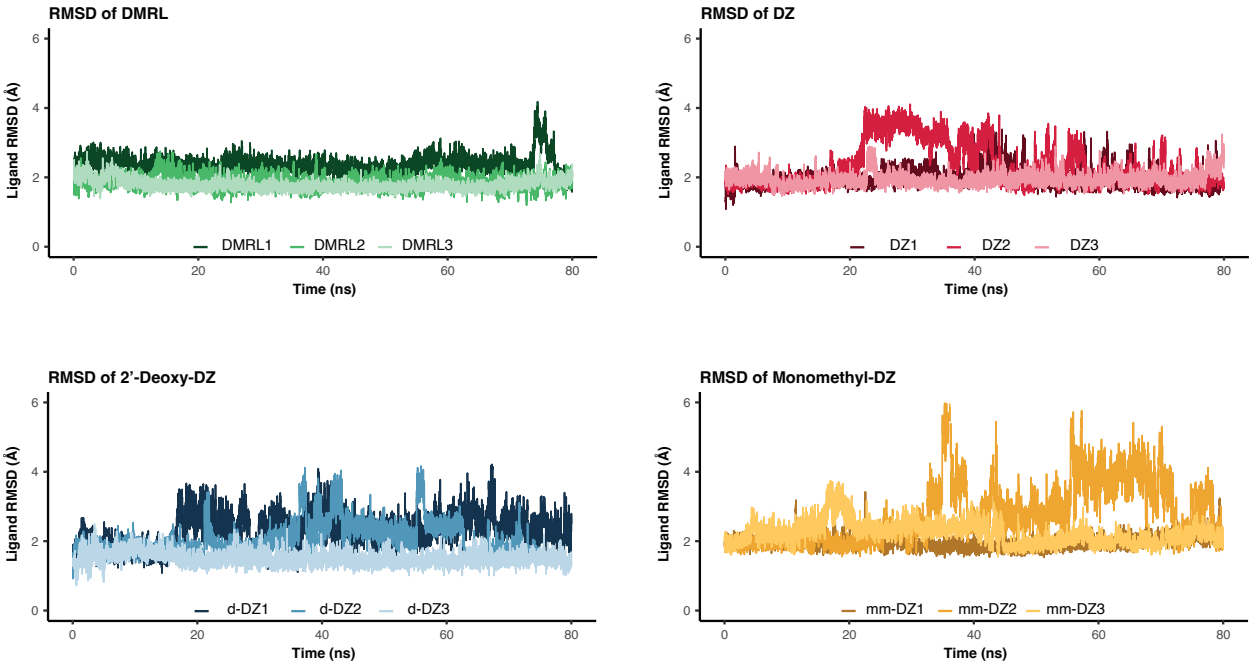
Supplemental Figure S3.2. Comparisons between docking results and experimentally determined structures of ligand-MR1 complexes. The crystal structure ligand is shaded (dark cyan) and the docked ligand structure is vibrant (cyan). a) HMRL docked into WT MR1 (PDB: 4L4V). b) HMRL docked into MR1^{K43A} (PDB: 4LCW). C) 5-OP-RU docked into MR1^{K43A} (PDB: 4NQD). d) rRL-6-CH_2OH docked into chimeric human-bovine MR1; 4LCC remains coplanar with the experimentally determined structure, but adopts a conformation ~180° flipped relative to that seen in the crystal structure.



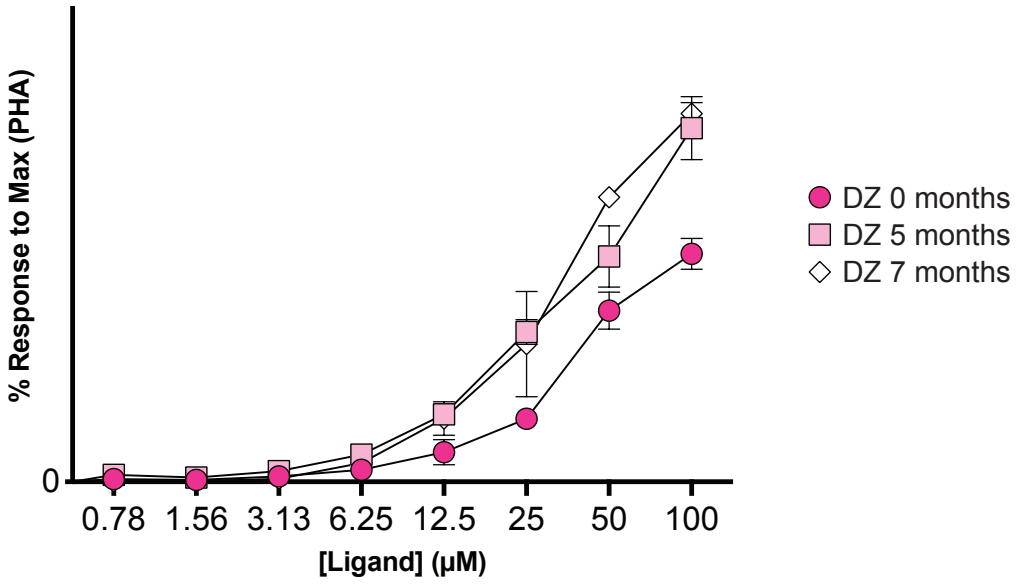
Supplemental Figure S3.3. Docking mode of ligands into 4L4V donor structure. Left, DMRL (red), DZ (blue), 2'-deoxy-DZ (purple) and monomethyl-DZ (yellow) docked into 4L4V. Center, PLI (light pink) and DZPLI (purple) docked into 4L4V. Right, Overlay of all ligand docking modes showing lack of overlap between DMRL analogue and PLI/DZPLI lumazine rings.



Supplemental Figure S3.4. Comparison of docking mode of DMRL, DZ, PLI, and DZPLI into each donor structure. The lighter color indicates the ligand has been docked into 4GUP; the darker color indicates it has been docked into 4L4V.



Supplemental Figure S3.5. Root mean square deviation (RMSD) calculations of each ligand within the binding pocket of MR1 highlight their stability within the pocket over the course of each simulation. RMSD for each ligand across the course of each simulation replicate (i.e. "DMRL1") are shown for DMRL, DZ, 2'-deoxy-DZ, and monomethyl-DZ.



Supplemental Figure S3.6. DZ is able to stimulate the D481C7 MAIT clone after storage at -80° C. The IFN-γ response by 1×10^4 D481C7 MAIT clones was measured in an ELISPOT assay with 1×10^4 DC pulsed with the indicated concentrations of ligand. Here, the same synthesis batch of DZ was tested either immediately after synthesis or after 5 or 7 months of storage in dimethylsulfoxide (DMSO) at − 80 ° C. See **Materials and methods** for details.

Complex (Donor structure)	5-OP-RU	rRL-6-CH_2OH	HMRL	DMRL	DZ	Deoxy DZ	mm DZ	PLI	DZ PLI
hβ2m/hMR1:6-FP (4GUP)	-	-	-	-8.3	-8.4	-8.7	-8.3	-8.9	-9.0
hβ2m/hMR1:HMRL-TCR (4L4V)	-	-	-7.2	-7.4	-8.1	-7.9	-7.5	-6.8	-7.3
hβ2m/hMR1K43A:HMRL-TCR (4LCW)	-	-	-7.7	-	-	-	-	-	-
hβ2m/hMR1K43A: 5-OP-RU-TCR (4NQD)	-6.8	-	-	-	-	-	-	-	-
bβ2m/bhMR1: rRL-6-CH2OH-TCR (4LCC)	-	-6.8	-	-	-	-	-	-	-

Supplemental Table S3.1. AutoDock Vina score (in kcal/mol) for the top docking mode of each ligand modeled into the MR1 A′ pocket. DDZ, 2′-deoxy-DZ; MMDZ, monomethyl-DZ.

Chapter 4. Pilot studies of single cell multi-omics of diverse MR1T cells from the peripheral blood of healthy human donors

4.1 Introduction

Recent improvements in accessibility and cost efficacy of single cell RNA sequencing (scRNA-seq) experiments have led to a boom in studies utilizing this method. scRNA-seq allows us to understand the diversity and abundance of many RNA transcripts with single cell resolution, and thus can assess heterogeneity even within a seemingly homogeneous cell population. In the field of immunology, and especially for T cell biologists, this has been an extremely powerful technique because it is able to pair transcriptomic data with the specific sequence of the TCR, the highly variable antigen receptor borne by these cells. Additionally, the development of oligo-conjugated antibodies compatible with the workflows for library generation from single cell suspensions can reveal the abundance of surface-localized proteins, therefore overcoming the limitations of extrapolating RNA transcript abundance to protein abundance and function. More recently, oligo-conjugated tetramers of MHC molecules have been used to provide additional data regarding antigen specificity, making scRNA-seq an incredibly powerful tool for linking antigen specificity with the TCR sequence and cellular properties with high throughput [170]. sc-RNAseq has therefore become an increasingly common complement (or replacement) to lower throughput, more resource-intensive techniques such as the cloning of antigen-specific T cells.

While this evolution has started to make its way into the field of MR1-restricted T cells, only one

so far (still only released as a pre-print) has focused on MR1 antigen specificity [171]. However, this study suffered from the limitations discussed in **Chapter 1.3** in that although they enriched for MR1T cells by sorting and sequencing $MR1^{+5\text{-}OP\text{-}RU}$ $tetramer^{+}$ T cells, they analyzed only the $TRAV1\text{-}2^{+}$ cells, opting to remove the remainder from their analysis because they could not verify whether these cells were actually MR1-restricted or were contaminants, thus obscuring the underlying diversity of MR1Ts. We therefore wanted to ask questions regarding the heterogeneity of MR1T cells with respect to their transcriptome, expression of cell surface markers, TCR repertoire, and antigen specificity in healthy donors. In this chapter, I describe two pilot experiments investigating the diversity of MR1Ts sorted from the peripheral blood of healthy human donors. This work was done in collaboration with Sean Ryan and Idongesit Ekpo, two graduate students in the lab, and Hortencia Graves, one of our research technicians.

4.2 Single cell RNA sequencing and multi-omics of MR1T cells from the peripheral blood of two healthy human donors

In order to achieve this multimodal data set, we generated single cell-barcoded DNA libraries for the transcriptional phenotype ("Gene Expression"/"GEX" library) and the paired αβ TCR sequence ("VDJ" library) (**Figure 4.1**). In addition, before entering the cell-partitioning microfluidics chamber, we stained the cells with oligo-conjugated antibodies; these commercial reagents make it possible to determine the expression of cell surface receptors of desired markers at single-cell resolution ("Antibody-Derived Tag"/"ADT" library). Finally, we used custom MR1 tetramers to determine the antigen specificity of each cell (**Figure 4.1**). Specifically, we generated oligo-conjugated streptavidin-phycoerythrin (SA-PE) variants (each with a unique

oligo) in order to make fluorescent, barcoded tetramers of MR1 reagents from our lab and the NIH Tetramer Core Facility (see **Materials and methods, Supplemental Figure S4.1**). With these tetramers, we were able to enrich our population of interest by sorting on fluorescent tetramer^{+} (PE^{+}) cells, enhancing our recovery of relevant T cells. Further, because the conjugated oligo is compatible with the same pipeline as that which is used for the ADT library, we were able to explicitly determine whether each cell or clonotype is restricted to MR1 or not,

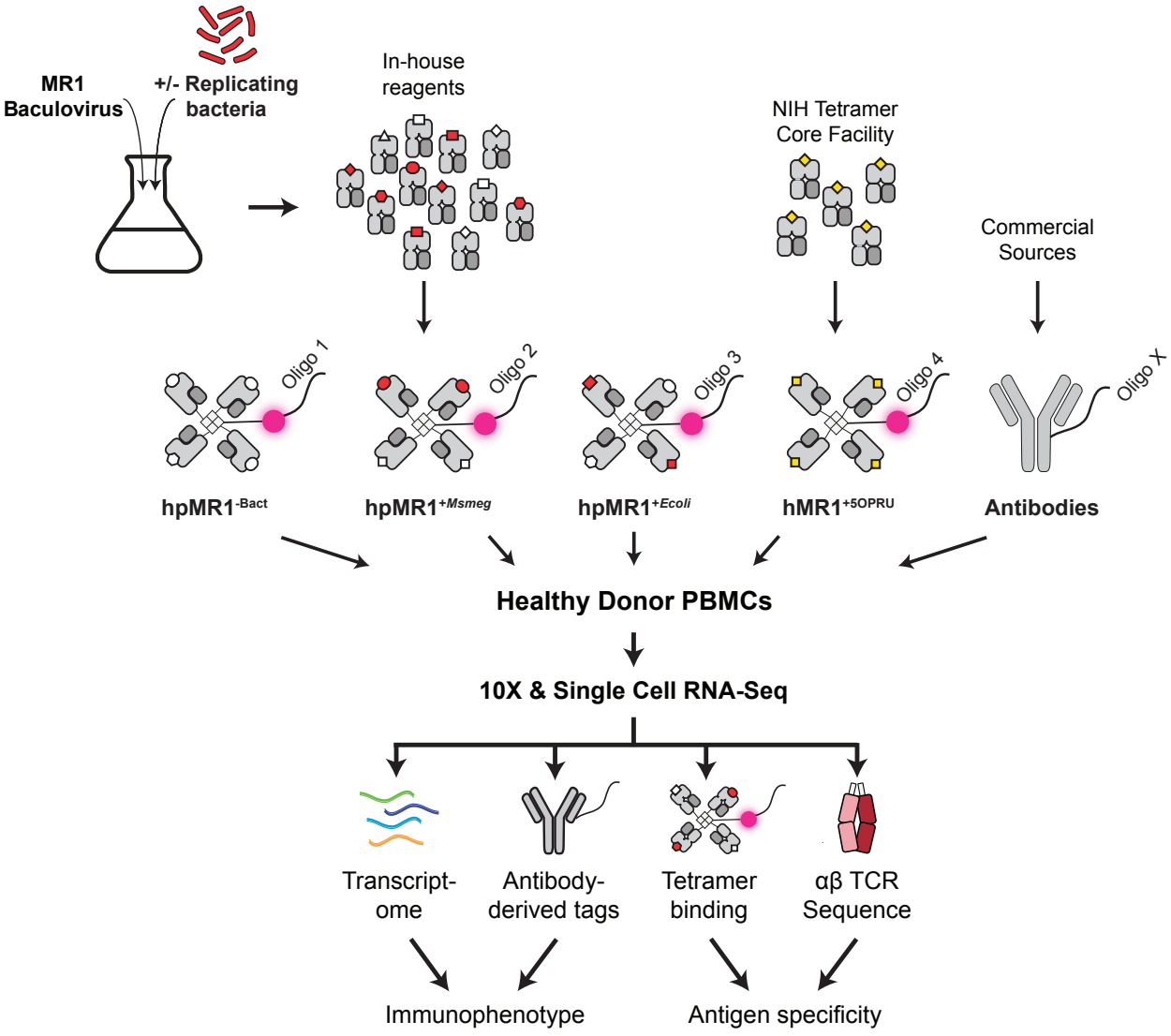


Figure 4.1 Experimental setup for the characterization of MR1T cells from the blood of healthy donors.

independent of phenotypic features or its TCR sequence (**Figure 4.1**). Technically, the ADT and tetramer specificity data both derive from a single library, but they will be referred to here separately to distinguish between antibody binding and tetramer binding. Altogether, these four libraries - GEX, VDJ, ADT, and tetramer specificity - and the sequencing derived from them (referred heretofore as "assays") provide views into the phenotype of MR1T cells at the single-cell level.

Although we hoped to use the mammalian MR1 reagents for these experiments, it was clear that they were not potent staining reagents despite their ability to robustly activate MAIT clones (**Chapter 2.2**); therefore, I made tetramers of the insect cell-produced MR1 reagents, either "unloaded" (*i.e.* loaded with unrelated insect cell ligands which cannot stimulate our panel of MAIT clones; *tn*-hpMR1^{-Bact}) or loaded with either *E. coli* or *M. smegmatis* (*tn*-hpMR1^{+Ecoli} and *tn*-hpMR1^{+Msmeg}, respectively). I also made a tetramer of the publicly-available NIH Tetramer Core Facility's human MR1 reagent loaded with 5-OP-RU (hMR1^{+5OPRU}) to serve as a positive control and to distinguish possible 5-OP-RU-agnostic cells from 5-OP-RU-reactive cells (**Figure 4.1**). In this chapter, these reagents will be referred to as MR1-Bact, MR1-Ecoli, MR1-Msmeg, and MR1-5OPRU for simplicity. Sean Ryan and I stained each PBMC sample with a cocktail of all of the barcoded tetramers and then a mixture of barcoded antibodies relevant for characterization of MR1T cells. We sorted live, CD3$^+$, tetramer-PE$^+$ cells, which were then loaded into the microfluidics system of the 10X Chromium Controller. We used the Chromium Single Cell 5' v2 workflow to generate high-quality DNA libraries (**Supplemental Figure S4.2a**), which were finally sequenced by next-generation sequencing (see **Materials and**

methods).

Notably, the two experiments were technically performed with three key differences: in the first experiment (Donor 1, *i.e.* D1), the oligo was not removed from the protein conjugation reaction using a high-resolution method and thus it is more likely that unconjugated oligo contaminated the cell staining mixture. This could result in non-specific binding of unconjugated oligo to cells and and therefore more non-specific reads corresponding therewith. In contrast, during the second experiment (Donor 2; D2), I used size-exclusion chromatography to remove excess oligo (**Supplemental Figure S4.1a**), allowing for complete purification of the streptavidin-PE away from the excess oligo. Second, the sample from D1 was freshly isolated (not cryopreserved), whereas the sample from D2 was frozen, which could influence the cell surface receptor and/or transcriptional phenotype. Finally, the samples were sorted on two different cytometers, which could differentially influence cell stress and/or the sorting fidelity. Thus, any of these three variables could be the cause of observable batch effects and further work will need to be performed to validate any findings.

After sequencing, processing, and cell filtering (including removal of cells expressing high levels of lysozyme and those binding significantly to the IgG1κ isotype antibody; **Supplemental Figure 4.2**), approximately 12,000 cells remained between the two experiments. The two datasets integrated well at the gene expression level, as indicated by the distribution of cells deriving from each experiment across the clusters of the dimensional reduction (**Figure 4.2a-b**). Further, the distribution of the recovered paired αβ TCR sequences suggested that the remaining

clusters and cells were predominantly T cells (**Figure 4.2c**). Indeed, they show high expression of CD3ε (**Figure 4.2d**) and generally display the mutually exclusive expression of CD4 (clusters 5 and 9) or CD8 (or neither, *e.g.* $CD4^{-}$ $CD8^{-}$ "double negative" (DN) T cells), as measured both by the ADT and GEX assays **(Figure 4.2e**). Thus, these two datasets - despite having significant technical variability - nonetheless integrate well at the gene expression level. However, in order to determine the biological relevance of patterns in the transcriptional and antibody binding data, it was first necessary to determine antigen restriction to delineate MAITs, MR1Ts, and any possible contaminating populations of T cells.

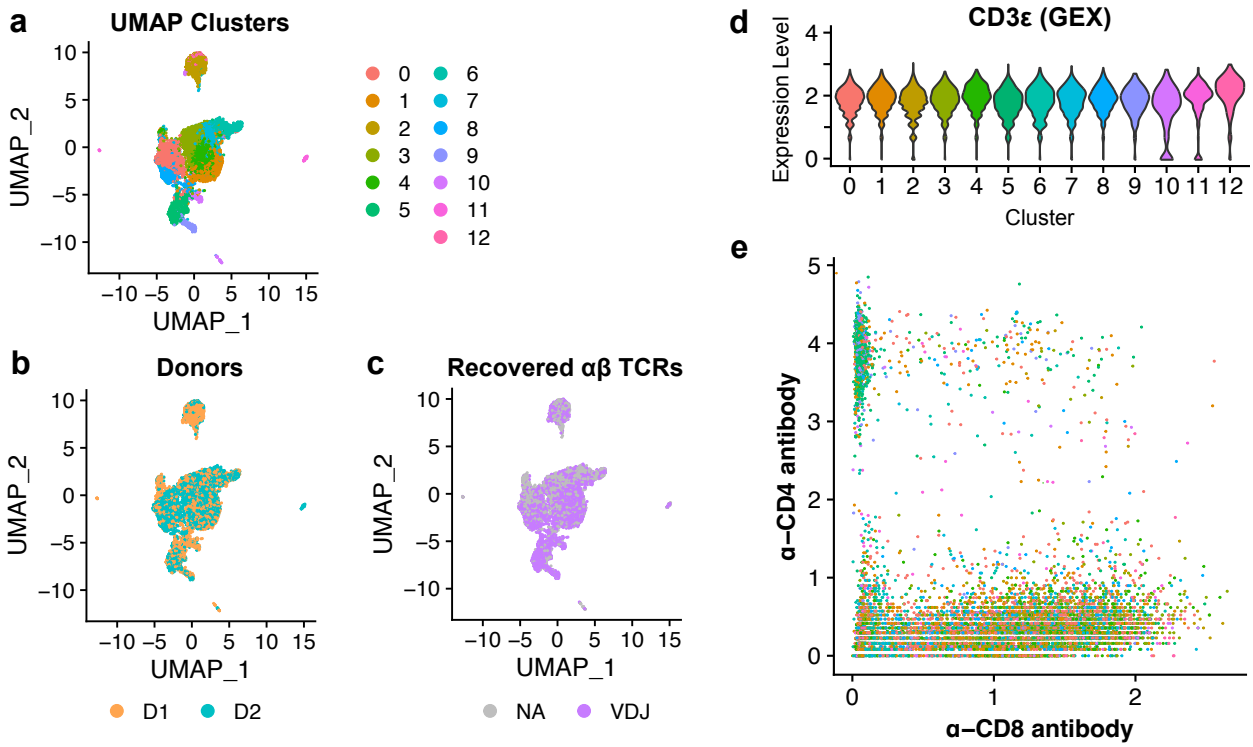


Figure 4.2 General properties of the integrated datasets. a) Dimensional reduction of the two data sets together. b) Origin of each cell mapped onto the UMAP dimensional reduction. c) V(D)J recovery status for each cell mapped onto the UMAP dimensional reduction. d) Cell surface gene expression reads (GEX) for CD3ε. e) Scatter plot of the α-CD4 antibody and α-CD8 antibody reads for filtered cells. Cells are colored by their cluster as in panel (a).

4.3 Barcoded tetramers enable classification of antigen/tetramer specificity

In order to determine tetramer specificity, I first had to account for variability among all cells of a particular clonotype. To do so, I removed cells that showed high staining of the unloaded tetramer, then I collapsed each clonotype each into a single point by averaging the normalized reads of each tetramer across all of the cells of a particular clonotype (as called by the amino acid sequence, not nucleotide sequence) (**Supplemental Figure S4.3**). I did this separately for each dataset because there was a significant batch effect for the tetramer-associated reads between the two datasets that could not be reconciled (**Supplemental Figure S4.4**), regardless of how the data were normalized (data not shown). After collapsing each clonotype, I empirically determined cutoffs above which the average (normalized) tetramer reads had to fall in order to be classified as binding each reagent (**Supplemental Figure S4.3**) based on the distribution of tetramer binding among the collapsed clonotypes. Generally, I found that the MR1-5OPRU tetramer gave very good separation, showing clear and distinct positive and negative populations. In contrast, it was harder to define appropriate cutoffs for the heterogeneously-loaded tetramers, as there was no distinct population. This is not necessarily surprising since these tetramers are made from MR1 that is loaded with variable ligands, which presumably sometimes results in lower avidity and therefore fewer reads per cell, but it did make the interpretation difficult.

After defining appropriate thresholds for each clonotype, I binned the clonotypes into five reactivity subsets as defined by their high expression of loaded tetramer and low expression of unloaded (-Bact) tetramer: *M. smeg* only (MS), *E. coli* only (EC), 5-OP-RU only (FOP), *M. smeg* and *E. coli* but not 5-OP-RU ("double positive" or DP), and "cross-reactive" cells binding 5-OP-

Donor	Reactivity	Classifier	# Clonotypes	Total # cells	Pct. clonotypes as singlets/doublets	Pct clonotypes that are TRAV1-2^{+}
D1	*M. smeg* only	MS	93	118	96.8%	6.5%
	E. coli only	EC	59	79	93.0%	8.5%
	M. smeg & *E. coli*	DP	38	43	97.4%	7.9%
	5-OP-RU only	FOP	4	4	100%	25%
	5-OP-RU and *M. smeg* and/or *E. coli*	XR	157	1936	78.9%	72.6%
	Unclassified	NA	353	915	94.3%	7.1%
	No V(D)J Info	No VDJ	-	2157	-	-
D2	*M. smeg* only	MS	20	20	100%	30%
	E. coli only	EC	15	15	100%	0%
	M. smeg & *E. coli*	DP	3	3	100%	33.3%
	5-OP-RU only	FOP	11	12	100%	100%
	5-OP-RU and *M. smeg* and/or *E. coli*	XR	660	2911	70.9%	87.0%
	Unclassified	NA	345	407	96.8%	14.8%
	No V(D)J Info	No V(D)J	-	2501	-	-

Table 4.1 Summary of the number of clonotypes (and cognate cells) binned into each reactivity, as well as the distribution of singlets/doublets and TRAV1-2 usage among each clonotype.

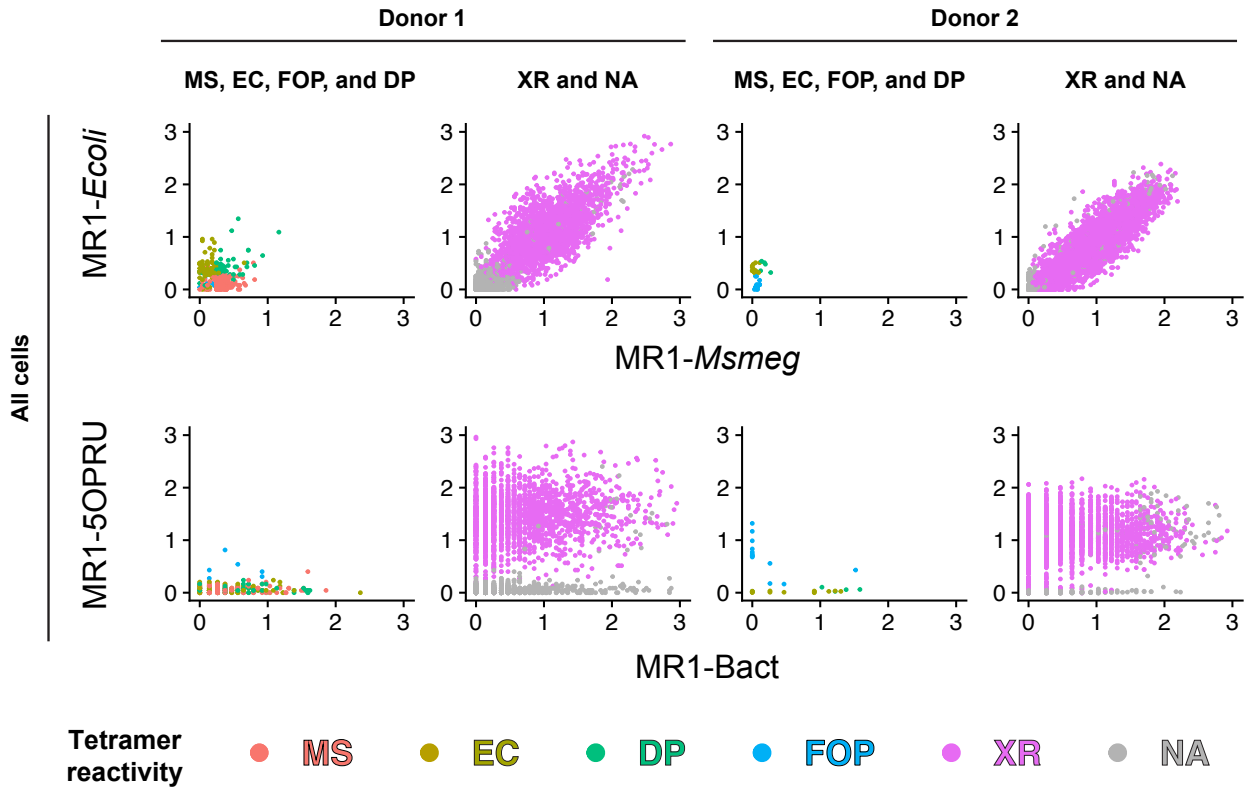


Figure 4.3 Cells binned into each reactivity category, shown as scaled counts (not raw counts) for clarity. The reactivities determined for each clonotype (**Supplemental Figure S4.5**) were appended onto each cell represented by the clonotype.

RU and *M. smeg* and/or *E. coli* (XR). Any cells that did not have tetramer reads sufficient to meet my criteria and/or had high reads corresponding to the control tetramer (**Table 4.1**) were considered "unclassified" (NA) clonotypes (**Supplemental Figure S4.5**). I then appended the determined clonotype reactivity onto each individual cell it represented (**Figure 4.3**) and mapped this metadata onto the dimensional reduction (**Figure 4.4**). Here, I saw that the cells annotated as having EC, MS, and DP reactivities fell into a pattern more similar to the NA cells than the XR cells (**Figure 4.4a**). indicating that the transcriptional phenotype of the cells in these bins are more similar to one another and to the NA cells than to the transcriptomes of the XR-binned clonotypes. This may indicate that my cutoffs were either too stringent (classifying more cells as NA than are really "unclassifiable") or too lenient (categorizing "unclassifiable" cells as EC, MS, or DP). Further, in looking at the donor distribution of each reactivity, I found that while the few 5-OP-RU-exclusive (FOP) cells almost exclusively derived from the Donor 2, the DP and EC cells arose primarily from Donor 1 (**Figure 4.4b**). Interestingly, there at least a handful of MS cells from both donors (**Table 4.1**). Given these biases with respect to transcriptional profile and donor origin, I therefore wanted to establish the fidelity of my tetramer reactivity thresholds to ensure that I wasn't generating a substantial number of false positives in these bins.

To verify the robustness of this method, I searched for T cells expressing the canonical NKT TCR gene recombination (TRAV10/TRAJ18/TRBV25) in the Donor 1 dataset and assessed the reactivity that I had appended to these clonotypes. I was confident that there would be at least a few NKT cells represented in the data because in this experiment (but not the Donor 2 dataset) we had included a CD1d tetramer loaded with PBS57, the synthetic analog to the canonical

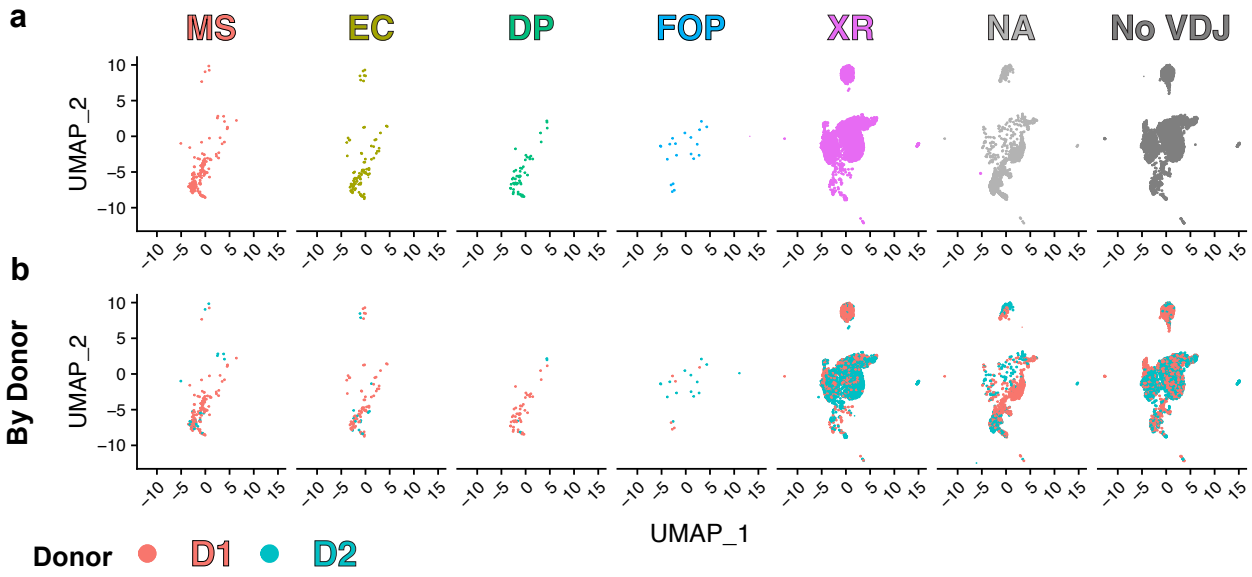


Figure 4.4 Distribution of reactivities across clusters and donors. The reactivity determined for each clonotype was re-appended onto the appropriate cells, then the dimensional reduction was split by this variable. a) Distribution of reactivity across the dimensional reduction, colored by the reactivity. b) Distribution of reactivity across the dimensional reduction, colored by donor.

ligand α-GalCer, to provide an "out-group" for our tetramer reactivity. Unfortunately, the barcode had degraded from this tetramer and thus we could not recover any reads from the sequencing data. However, the fluorophore was intact, as indicated by its ability to stain a Jurkat line carrying a CD1d-sulfatide-restricted TCR (performed by Augusta Broughton, Mayuri Viswanathan, and Caitlin Castro after this experiment was completed; data not shown). Thus, even though NKT cells are less abundant than MAITs/MR1Ts in the blood, I expected to see at least a few from the Donor 1 dataset. Indeed, of the thirteen clonotypes (two of which were represented by four cells each, with the remainder being singlets), four singlets were designated as either MS (1 cell), EC (2 cells) or DP (1 cell). The DP cell showed high reads for the MR1-Bact tetramer and the MS cell was only just above the MR1-Msmeg low cutoff, suggesting that these may be a non-specific interactions (**Supplemental Figure S4.6**). However, the two singlets being annotated as EC-reactive clearly show some above-baseline reads for the MR1-Ecoli

tetramer without high reads for the MR1-Bact, MR1-5OPRU, or MR1-Msmeg tetramers (**Supplemental Figure S4.6**). Whether this reflects a real interaction or not is unclear. Nevertheless, these observations are in line with my expectation that at lower clonotype frequencies, we will see more false positives. However, because we are interested in rare clonotypes, particularly those T cells which are agnostic toward 5-OP-RU, we expect to need to look at low-abundance clonotypes and therefore should also expect false positives in our populations of interest. Thus, an increased number of false positives was acceptable to me in order to seek out rare clonotypes, but because the proportion of singlet clonotypes in the EC, MS, DP, and FOP categories (**Table 4.1**) is quite high, we should take caution when making broad conclusions about these cells as a population, and should prioritize high-frequency clonotypes when possible.

While there were likely false positives identified for the clonotypes with defined reactivities, I also found that many of the recovered clonotypes - the vast majority of which were singlets or doublets - fell into the unclassified category, meaning that they had low reads for the unloaded tetramer but also had low reads for any of our other tetramers **(Figure 4.3, Figure 4.4**). These cells were also overwhelmingly TRAV1-2^{-} **(Table 4.1**), suggesting that at the very least, the oligo appended to the 5-OP-RU tetramer, which captures most or all of the MAIT cells, was mostly intact (if it were degraded or the cutoffs were too stringent, we would have expected to see more TRAV1-2^{+} cells in this category) **(Table 4.1**). Considering, however, that these cells were positively sorted as live CD3^{+} tetramer^{+} T cells and that all of the tetramer reagents show good separation from negative controls by flow cytometry (**Figure 2.3**), it is surprising that this

many clonotypes fall into this category. It may be that the dropout rate for the amplicons corresponding with these tetramers is so low due to the specific amplification and enrichment of these transcripts that the sensitivity is high enough that it's detecting interactions that we could not see by flow cytometry. It is also possible that my interpretation and classification of these data were too stringent. It seems most likely to me, however, that this can be explained by technical variability between the two experimental replicates, because it's clear that most of the NA cells arise from the first dataset. I hypothesize that this is the result of the FACS sorter we used for the first experiment based on its technical limitations (the MACSQuant Tyto cannot appropriately define populations of cells based on FSC/SSC - *e.g.* cell size - and it's therefore more likely to positively sort contaminating cell populations that would typically be removed by droplet-based sorters). Irrespective of the reason, this result highlights the importance of the barcoded tetramers for determining the true nature of MR1 restriction for the recovered cells.

4.4 A hyperexpanded cross-reactive MAIT clone exhibits hallmarks of degranulation and contrasts with the relative homogeneity of the remaining cross-reactive clonotypes

Although some cells were seemingly contaminants, the vast majority of the clonotypes fell into the "cross-reactive" (XR) category (**Table 4.1**), since nearly all cells that bound the 5-OP-RU tetramer also bound both the *E. coli* and the *M. smegmatis* tetramers in both donors **(Figure 4.3, Supplemental Figure S4.4**). This finding is consistent with dual-colored tetramer analysis of PBMC T cells by flow cytometry (data not shown). When the XR population was subset and re-clustered, most cells are distributed in an experiment-independent manner (**Figure 4.5a**), but there is a large pool of cells (cluster 1) specifically from Donor 1. Cluster 1 is composed

primarily of a hyperexpanded MAIT clone (Clonotype 4, "ct4") **(Figure 4.5a-b)**, though some of the cells from this clonotype were present in other clusters as well. In fact, Clonotype 4 represented nearly half of the recovered V(D)J sequences from D1. From here, I will refer to this as the "homogeneous" population, whereas the remaining clusters (0 and 2-10) I will refer to as

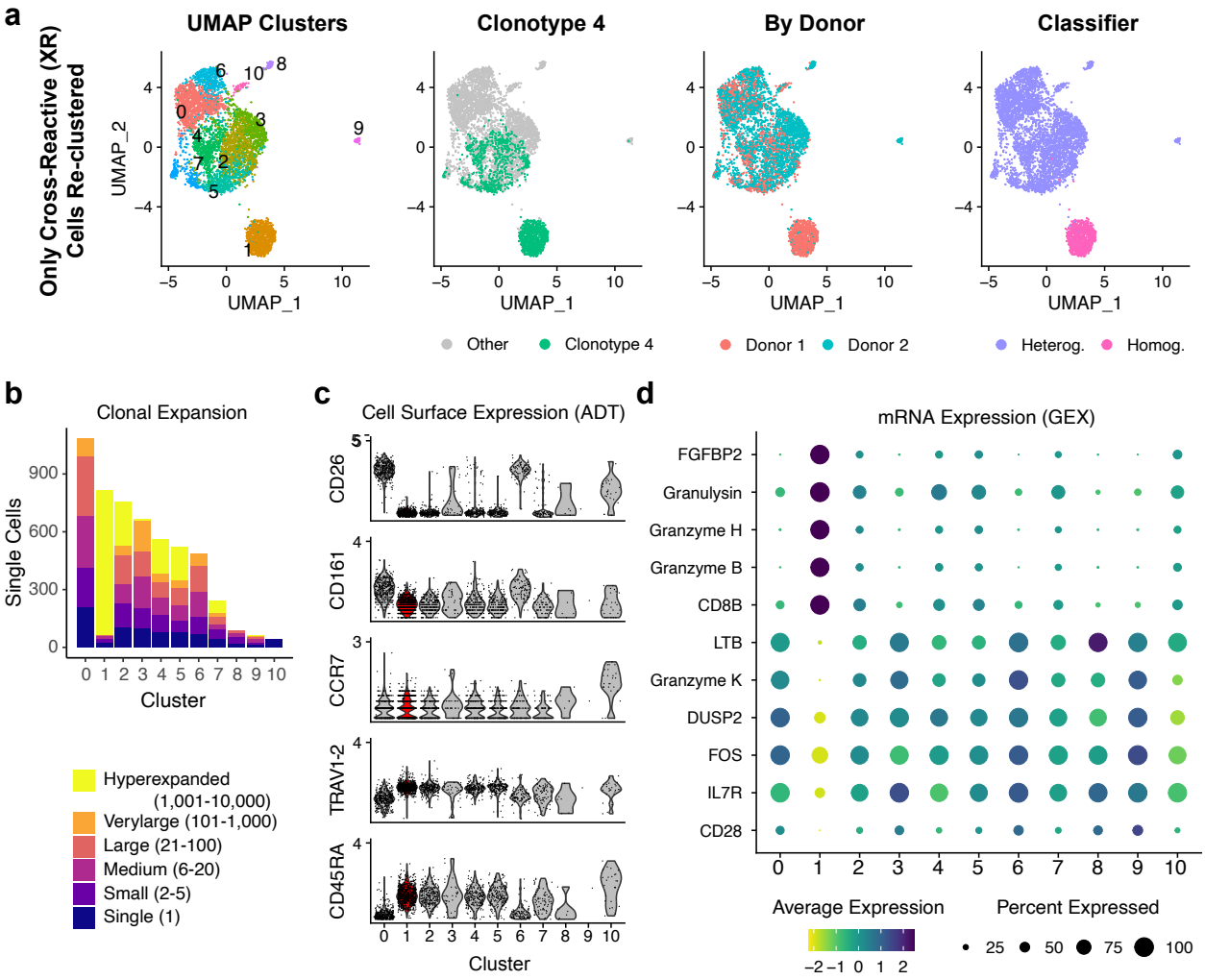


Figure 4.5 Properties of the hyperexpanded clonotype (ct4). a) Dimensional reduction of only the XR cells (left); localization of ct4 across the dimensional reduction (center left); distribution of XR cells arising from each donor across the dimensional reduction (center right); illustration of the binned "homogeneous" and "heterogeneous" cells (right). b) Representation of clonal expansion across each cluster the numbers next to each label in the legend represent the number of cells per clonotype represented by each category. c) Cell surface (ADT) reads for each cluster only for Donor 1 (from which ct4 originated; illustrated as such to not obscure the data, due to the batch effect of the ADT assay between the two donors). The violin plots for cluster 1 (containing majority ct4) are highlighted in red, though this is sometimes obscured by the data points. Cluster 9 was dropped because there was only one cell belonging to donor 1 in this cluster. f) Gene expression (GEX) reads for each cluster for the differentially-expressed genes (and CD28) segregating cluster 1 and the remaining clusters.

the “heterogeneous” population (**Figure 4.5a**, far right); the transcriptional and cell surface landscape of the heterogeneous population will be discussed further in **Chapter 4.6**.

This homogeneous subset is predominantly negative for the cell surface expression of CD26 and CD161 - two characteristic MAIT markers - and CCR7 **(Figure 4.5c**). Although CD161 has historically been utilized as a MAIT cell marker, it has been reported to be downregulated on the cell surface after MAIT cell activation, which here would be consistent with the hyperexpanded nature of the clone [172, 173]. In contrast, this population also shows significantly enhanced expression of CD45RA in comparison with the heterogeneous population **(Figure 4.5c**). Notably, this seems to be somewhat unique to Clonotype 4 since the other clusters containing cells of this clonotype also show higher CD45RA expression, and the signal for this marker is very low for all cells from Donor 2 (data not shown). At the transcriptional level, the homogeneous subset exhibits enriched transcript expression of FGFBP2, GNLY, GZMB, GZMH, and CD8B **(Figure 4.5d**), suggesting that these cells - or at least the cells of Clonotype 4 - are degranulating. In addition, these cells are $CD28^{-}$ (as indicated by the transcriptional data) **(Figure 4.5d**); together with the CCR7 and CD45RA phenotype determined by ADT (**Figure 4.3e**), these observations suggest that that these cells belong to the T_{EMRA} T cell memory subset. T_{EMRA} cells are terminally differentiated and this phenotype is likely the result of recent antigen experience, which is reflected in the hyperexpanded nature of Clonotype 4. I therefore hypothesize that our “healthy” donor may have had an ongoing or recent infection that did not manifest physical symptoms. This may also, in part, explain the presence of some hyperexpanded “NA” (seemingly non-MR1-restricted) clonotypes from Donor 1. Thus, although the hyperexpanded clone did diminish the

recovery of non-cross-reactive MR1Ts (which are of more interest to us), it did provide an opportunity to assess the diversity of MAIT transcriptional states.

4.5 Tetramer reactivity reveals a more TCRα-diverse population of cross-reactive MR1T cells

Of the cells binned as being cross-reactive, the majority were - unsurprisingly - canonical MAITs, as indicated by their V(D)J rearrangements **(Supplemental Figure S4.7**). Nevertheless, I did observe several expanded TRAV1-2^{-}, cross-reactive MR1Ts **(Table 4.2**); indeed, of the unique clonotypes in each experiment, approximately 28% (Donor 1) and 13% (Donor 2) of those classified as "cross-reactive" were TRAV1-2^{-} (**Table 4.1**). Since MAIT cells have previously been shown to sometimes express two TCRα chains [174], I checked whether each of these cells expressed more than one TCRα or TCRβ chain with a specific emphasis on looking for conventional MAIT gene usage. Indeed, a few clonotypes (all singlets) were annotated with two TCRα chains, but none of them were TRAV1-2^{+} (and none of the TCRs shown in **Table 4.2** showed dual-chain expression) (data not shown). Altogether, at the surface, these data suggest that the 5-OP-RU and *M.smeg*/*E.coli*-loaded tetramers may bind a broader population of MR1T cells than we typically appreciate.

In assessing the specific V(D)J usage of these cells, I found that the usage and combinations of TRAV and TRAJ genes were much less biased than what is seen across all cross-reactive cells (including MAITs) **(Supplemental Figures S4.7, S4.8**), whether assessed as the proportion of all clonotypes or all cells. However, there were selective gene usages between the two donors; for

Donor	Clonotype ID	Frequency	TRAV	TRAJ	CDR3α	TRBV	TRBJ	CDR3β
D1	ct276	18	8-4	23	CAVISGPYNQGGKLIF	6-4	2-1	CASSNGLDWHEQFF
	ct192	11	12-1	27	CVVNRTGKSTF	4-2	2-5	CASSPDTERSAQYF
	ct134	6	19	52	CALNAGGTSYGKLTF	4-2	2-2	CASTGTSGDSTGELFF
D2	ct1064	20	1-1	31	CAVRDKGNNNARLMF	6-4	2-1	CASSDLAGAGEQFF
	ct1013	18	2	27	CAVINTNAGKSTF	20-1	2-1	CSASLERGGSDEQFF
	ct988	14	1-1	23	CAAHIYNQGGKLIF	6-1	2-6	CASSEAVSGANVLTF
	ct1251	13	8-3	21	CAVPYNFNKFYF	20-1	2-7	CSARDGGQFEGVEQY
	ct1141	13	1-1	36	CAVRDTGANNLFF	6-4	2-2	CASSDSYTGELFF
	ct1461	12	25	52	CAGKSVAGGTSYGKLTF	27	2-1	CASSQPPPYNEQFF
	ct1041	10	13-2	31	CAENDNARLMF	6-4	2-1	CASSDGTGGNEQFF

Table 4.2 Examples of gene usage and CDR sequences of expanded TRAV1-2^{-} XR MR1T clonotypes.

example, Donor 2 showed both clonotype bias toward and clonotype expansion of T cells using TRAV1-1, TRAV2, TRAV13-2, and TRAJ31 (**Figure 4.6**). Although there was some selective gene usage, the two donors did share a slight preference for TRAJ23 and TRAJ27 (**Figure 4.6, Table 4.2**). In contrast to the variety of TCRα rearrangements, the TCRβ chain shows a striking bias toward TRBV genes that are common among conventional MAIT cells **(Figure 4.6**). Indeed, nearly 90% of these cells used TRBV6-1, TRBV6-2, TRBV6-5, TRBV4-2, or TRBV20-1 and pair almost exclusively with TRBJ2 genes (*i.e.* TRBJ2-1, TRBJ2-2) **(Figure 4.6**), which is also a bias for MAIT TCRs **(Supplemental Figure S4.7**). This is also seemingly specific to these T cells, since the cells that were binned as "unclassified" did not show this bias (**Supplemental Figure S4.9**), although there were a few conventional MAIT rearrangements among the unclassified cells.

The reason for this bias toward MAIT TCRβ rearrangements is unclear, but it may be simply due to a limitation of the experiment in that although they have been categorized here as being TRAV1-2^{-}, there may have been dropout of the second TCRα chain by scRNA-seq. Indeed,

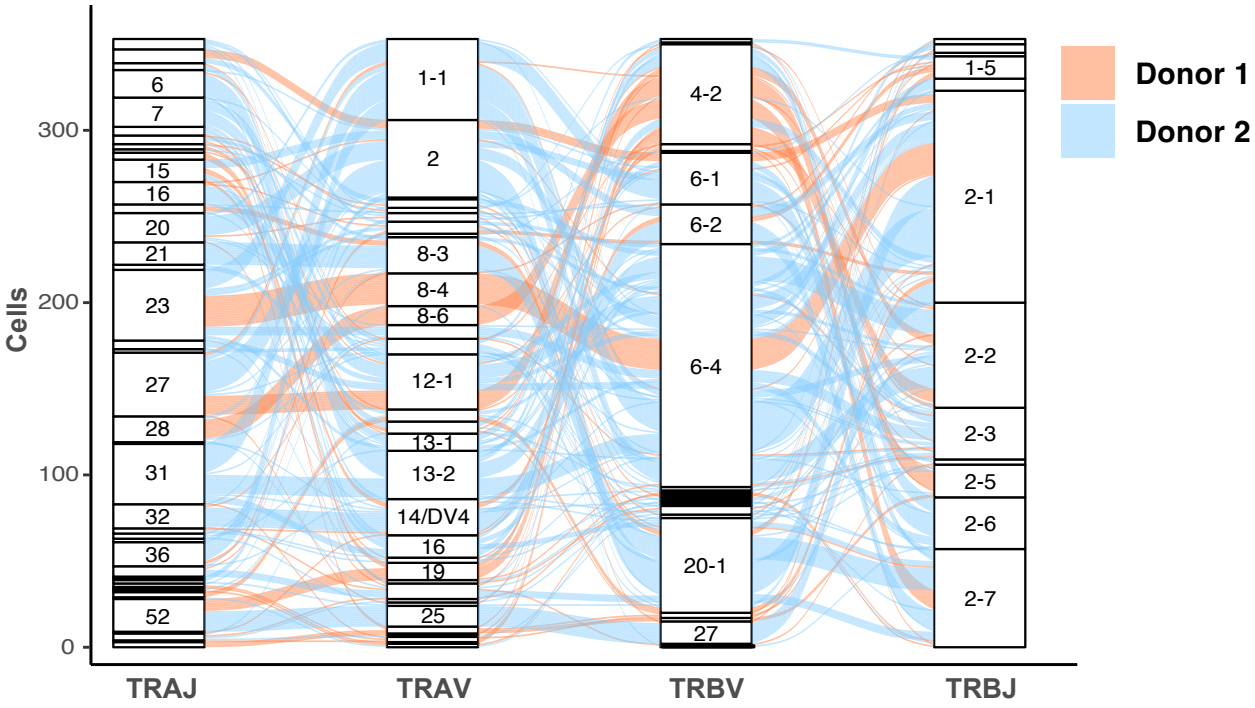


Figure 4.6 Gene usage by TRAV1-2⁻ XR MR1Ts. Only the more frequently-used genes are labeled for clarity.

when assessing the raw contig data, there were a few clonotype sequences that suggested a second (MAIT) TCRα rearrangement, but CellRanger and/or scRepertoire, the packages/ programs used for generating the contigs and appending those data to the Seurat object, had lower confidence in them and thus did not carry them over to the culled dataset. Since MAIT rearrangements are so abundant in these datasets (due to our specific enrichment of them by the MR1-5OPRU tetramer), it would be unsurprising to see contaminating MAIT sequences, but these also may be genuine secondary TCRα chains. Alternatively, if this is not an artifact, this result may allude to a stronger role for the MAIT/MR1T TCRβ chain than was previously appreciated. The only way to determine their true nature will be to validate them experimentally.

To investigate whether there were any common CDR motifs between the cells of the TRAV1-2⁻

XR subset, I then performed further hierarchical clustering on these sequences alone (**Figure 4.7**). Here, I annotated the dendrograms with appropriate donor information to determine whether the selective gene usage between donors skewed the CDR3 repertoire as well. Indeed, some of the TCRs from Donor 1 and Donor 2 formed their own clades (**Figure 4.7**), though there were several nodes with representation from both donors. Based on the CDR3 motifs, it seems that the clustering is dominated by the CDR3β loops, since these motifs show distinguishable patterns but the CDR3α loop motifs do not. Again, this is unsurprising considering the skewed usage toward the conventional MAIT TCRβ genes. In any case, it seems that there may be a slight preference for small aliphatic amino acids and acidic amino acids in the CDR3β, though this is just from a visual assessment of the motifs. Altogether, these results indicate that TRAV1-2$^-$ cross-reactive MR1Ts have a more invariant TCRα chain repertoire than MAIT cells, but may still be biased toward a MAIT TCRβ chain.

4.6 The transcriptional and cell surface phenotype of non-MAIT, cross-reactive MR1T cells

I next assessed whether there was any functional differences between the TRAV1-2$^+$ and TRAV1-2$^-$ MR1Ts in the GEX and ADT assays. To do this, I assessed only the cross-reactive and unclassified cells together, with the latter population serving as an out-group, though I also removed Clonotype 4 from the cross-reactive cells so as to compare these TRAV1-2$^-$ MR1Ts primarily to homeostatic MAIT cells. Both before and after subsetting and reclustering the remaining cells, the TRAV1-2$^-$ XR cells are distributed relatively evenly across all cross-reactive clusters, and are generally biased away from the NA (unclassified cell) clusters. These observations suggest that the TRAV1-2$^-$ MR1Ts do not differ significantly from canonical

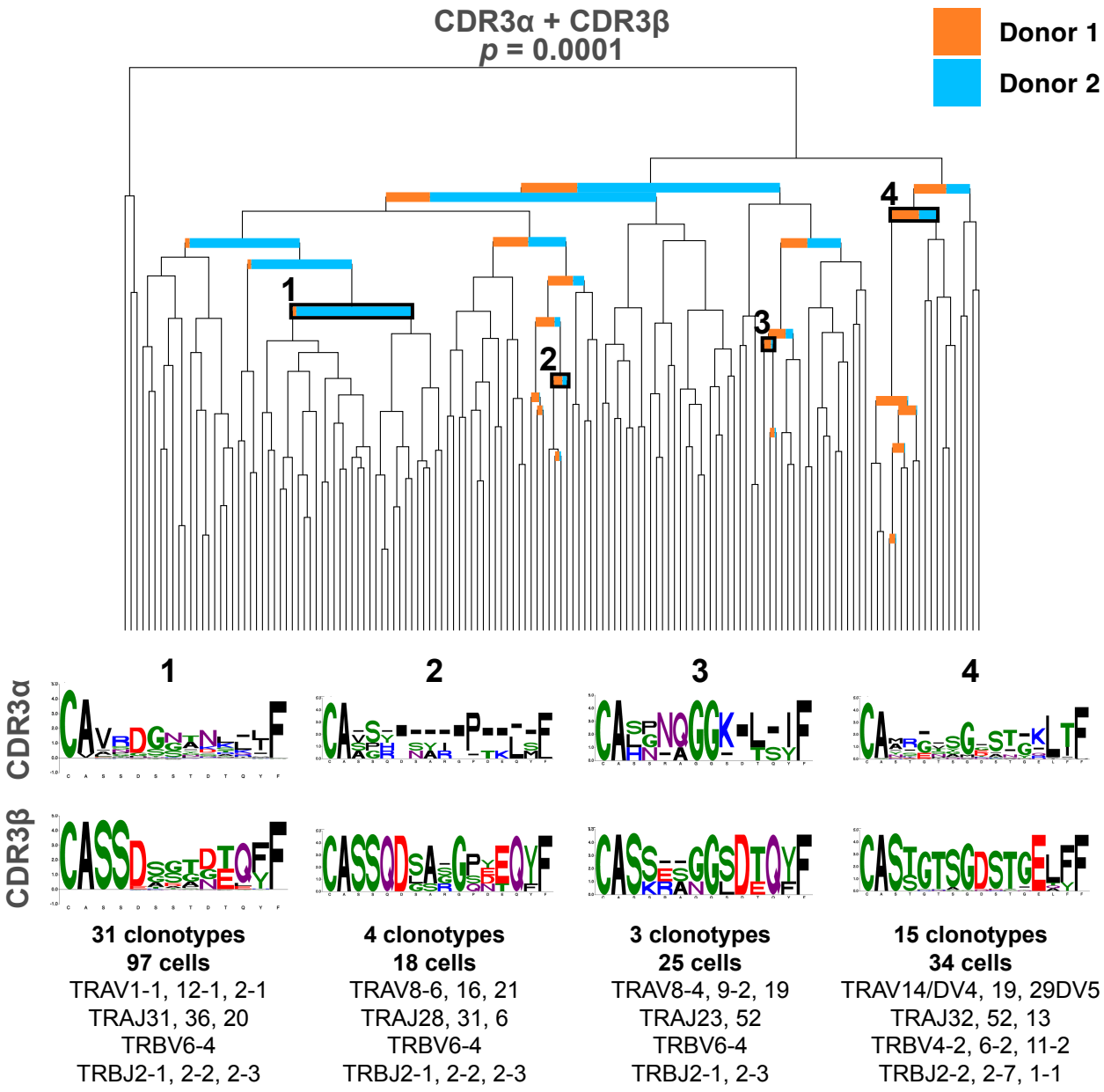


Figure 4.7 Hierarchical clustering of TRAV1-2$^-$ XR MR1Ts. The colored bars of the dendrogram indicate that the sequences found at that node are significantly similar to one another as compared with what would be expected of a random sequence. Interesting clusters are highlighted and their CDR sequence logos and the clonotype information and gene usage are provided.

MAITs at the transcriptional level (**Figure 4.8a**). Indeed, querying for DEG genes between TRAV1-2$^+$ and TRAV1-2$^-$ cross-reactive cells was unsuccessful at a minimum $\log_2$ fold change of 1.0. Given that the sample size of the TRAV1-2$^-$ cells was less than one tenth of the

TRAV1-2^{+} cells and that these cells are distributed about several clusters of the UMAP dimensional reaction, it is perhaps unsurprising that there is no intrinsic phenotypic difference. Further support of this finding lies in that the cell surface phenotype of TRAV1-2^{+} and TRAV1-2^{-} classes looked very similar, even when broken down by donor (**Figure 4.8b**).

In fact, the similarity among the cell surface phenotypes was interesting in that they showed similar binding to the α-TRAV1-2 antibody. In contrast, the TRAV1-2^{-} (as determined by the V(D)J data) "unclassified reactivity" (NA) cells were almost exclusively negative for this

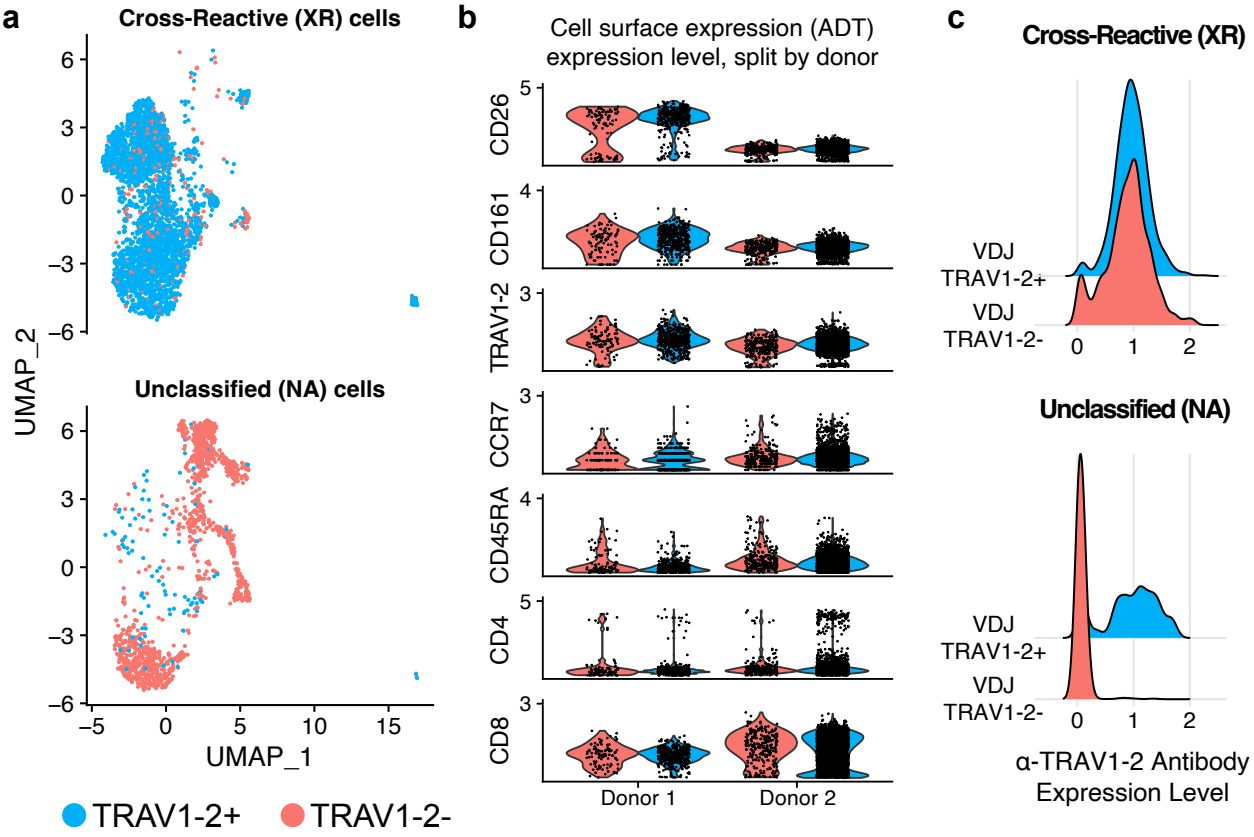


Figure 4.8 TRAV1-2^{+} and TRAV1-2^{-} MR1T cells do not differ strongly with respect to their transcriptome or cell surface phenotype. a) TRAV usage of MR1T cells mapped onto the dimensional reduction showing that the TRAV1-2^{-} MR1Ts are distributed relatively evenly across the same clusters as the TRAV1-2^{+} MR1Ts. b) Violin plots showing the average reads of the TRAV1-2^{+} and TRAV1-2^{-} MR1T cells for the antibodies of the ADT assay. c) Density plot (ridge plot) showing the average staining by the α-TRAV1-2 antibody for TRAV1-2^{-} and TRAV1-2^{+} subsets of XR and NA-classified cells.

antibody (**Figure 4.8c**). This evidence supports the hypothesis that the "TRAV1-2^{-}" XR MR1T cells may, indeed, be TRAV1-2^{+}, but that we are experiencing sequencing dropout of the conventional MAIT TCRα chain. Alternatively, it may be that the α-TRAV1-2 antibody promiscuously binds not only TRAV1-2 but also the MAIT TCRβ rearrangement. If our TRAV1-2^{-} XR MR1Ts are truly negative for TRAV1-2, perhaps the "residual" MAIT TCRβ chain is sufficient for at least partial α-TRAV1-2 antibody binding. This possibility would not be completely unfounded, as the antibody was generated by immunizing mice with a recombinant, heterodimeric TRAV1-2^{+} TRBV6-1^{+} MAIT TCR [175]. Evidence of this antibody's discrimination against other TRAV or TRBV chains was omitted from the published data describing its development, lending concern as to its specificity explicitly for TRAV1-2. This aside, these data suggest that the TRAV1-2^{-} cross-reactive MR1Ts (or at least those TRAV1-2^{+} cross-reactive cells bearing a second TCRα chain) do not occupy a unique transcriptional niche.

4.7 The phenotype and of a small population of putative 5-OP-RU-agnostic MR1Ts

In addition to the cross-reactive cells, I was able to assign MS, EC, DP, and FOP reactivity classes to a small proportion of all cells across both data sets. When I reclustered only these four bins of cells along with the heterogeneous XR cells, I found that these unique (non-cross-reactive) cells still fell mostly pooled together in cluster 5, with some cells also found in clusters 9 and 10 (**Figure 4.9a**); these cells were characterized by the enrichment of several transcripts and cell surface markers. Notably, cluster 5 showed showed a significant enrichment of TCF7, CCR7, FYB1, SELL (CD62L), and MAL transcripts. From the cell surface receptor data, it also seems that cluster 5 exhibits higher average expression of CD45RA and CCR7, the latter of

which is consistent with the transcriptomic data. Together, these markers indicate that at least cluster 5 consists of naive T cells. Intriguingly, a number of the cells in this cluster seem to be $CD4^+$, which is uncommon for conventional MR1T (MAIT) cells express CD4. Indeed, when assessed for CD4/CD8 usage, the MS, EC, and DP cells were approximately 50% $CD4^+$ and 50% $CD8^+$, with very few $CD4^-$ $CD8^-$ (DN) and no $CD4^+$ $CD8^+$ (DP) T cells (not shown). More work will need to be done to determine whether these are genuine MR1T cells. However, the fact that these cells all fell into a transcriptionally-similar population (independent of donor) made me more confident that this was a real phenotype and not just "noise" in the form of random contaminating clonotypes.

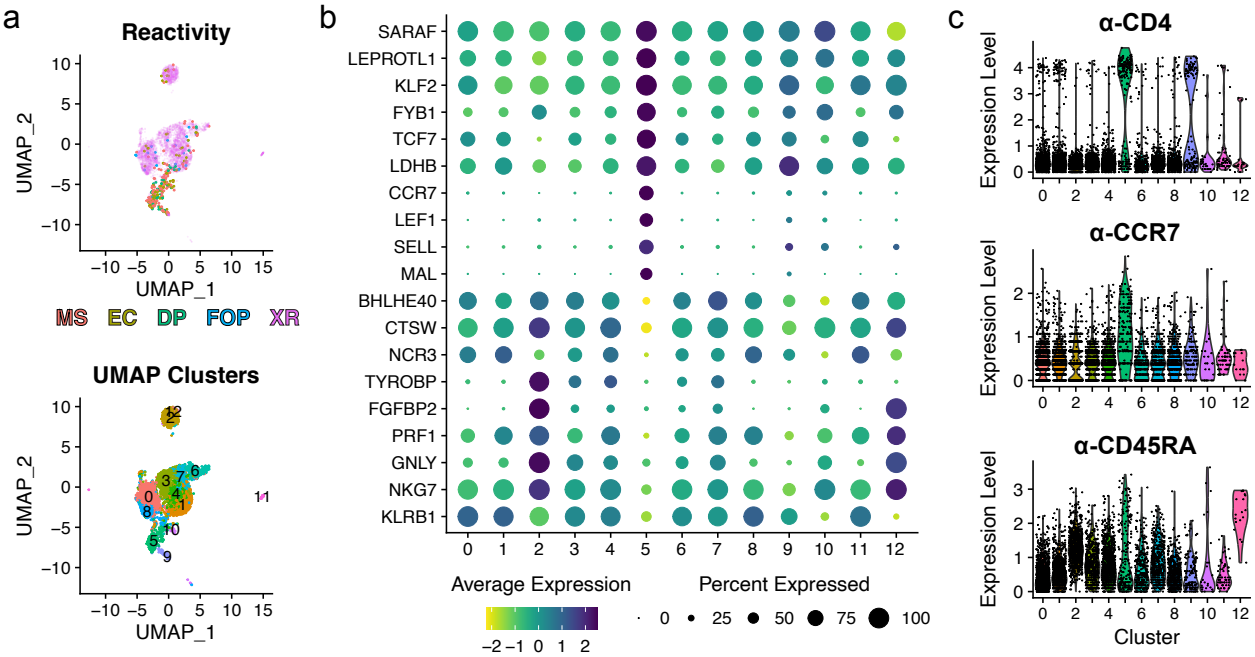


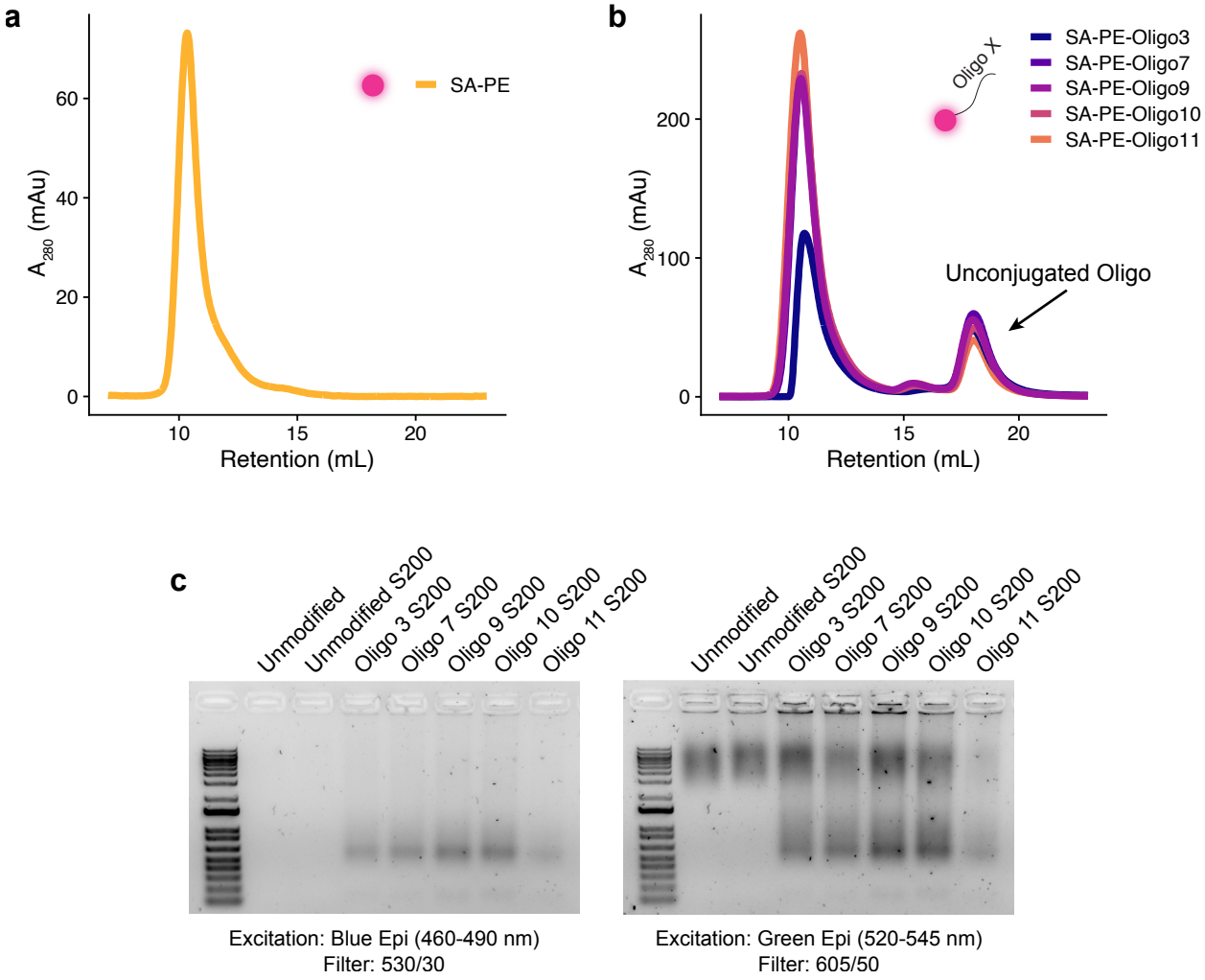
Figure 4.9 Unique properties of the gene expression and cell surface phenotype of non-XR MR1Ts with defined reactivity. a) Reactivity class projected onto the dimensional reduction; XR cells are partially transparent to highlight the location of MS, EC< DP, and FOP cells. b) Gene expression-derived clusters of the dimensional reduction. c) Differentially-expressed and relevant genes, shown by expression per cluster.

4.8 Discussion and ongoing work

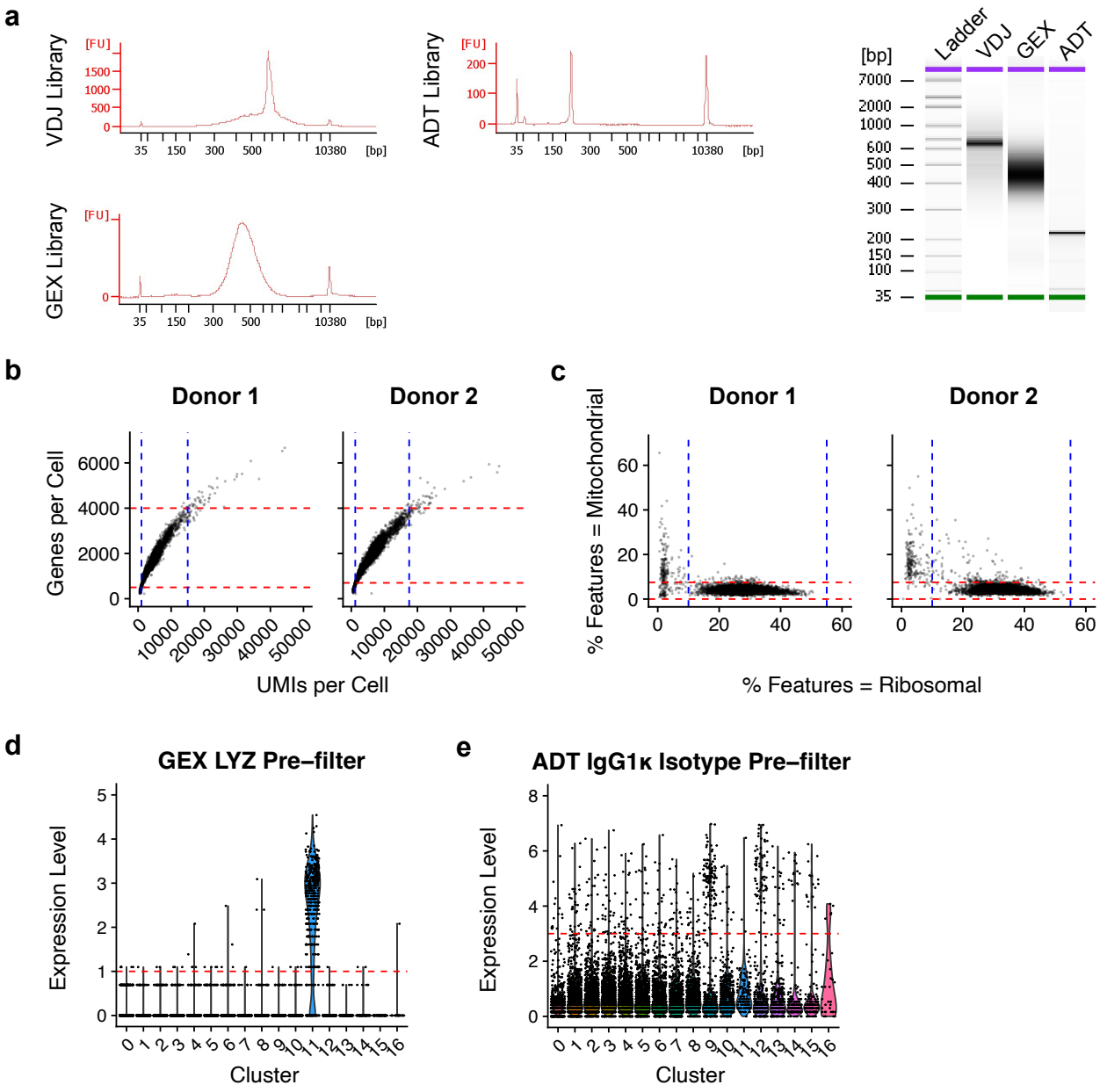
In this chapter, I described a pilot study for the multi-omics characterization of MR1T and MAIT cells in the peripheral blood of healthy donors. I showed that using custom oligo-conjugated tetramers, we can positively identify several unique populations of T cells, including putative TRAV1-2$^-$ MR1Ts which have been previously overlooked. I showed that about 10% of the "cross-reactive" cells - which bind the MR1-5OPRU and MR1-Ecoli and/or MR1-Msmeg tetramers - are TRAV1-2$^-$ and show a distinct bias toward MAIT TCRβ rearrangements, yet do not vary significantly from canonical MAITs with respect to their cell surface and transcriptional phenotypes. As discussed previously, however, this may be an artifact and limitation of sequencing. Further, the custom tetramer reagents revealed a population of putative 5-OP-RU-agnostic MR1T cells which bound our MR1-Ecoli and/or MR1-Msmeg tetramers. This population was - expectedly - smaller than the MAIT population, but it does seem transcriptionally and/or functionally unique in that the clusters dominated by these cells exhibited a more naïve phenotype. Importantly, although I have not described their TCR repertoire here in detail, they were overwhelmingly TRAV1-2$^-$ by both V(D)J sequence and their staining by the α-TRAV1-2 antibody, indicating that in a flow cytometric assay, they would likely have been gated out and ignored, again highlighting the strength of these novel reagents. Altogether, these data show that the oligo-conjugated MR1 tetramers are capable of identifying MAIT cells, but also may be able to define rare, non-MAIT MR1T cells. This reagent will strengthen the fidelity of multi-omics studies on MAIT cells, as it relinquishes the need to define MAIT cells entirely by phenotype and TCR gene usage.

One of the strengths of our lab is that we are able to perform these multi-omics workflows, but then also can recombinantly produce the TCRs that we have identified therefrom in order to determine their underlying molecular mechanisms governing their reactivity toward MHC-like molecules. Indeed, our current work in the lab has been focused on cloning TCRs from each reactivity bin into expression vectors such that they can be validated *in vitro*. Specifically, we are first validating their binding to these reagents by transiently co-transfecting HEK293 cells with full length, P2A-linked dual-chain TCR constructs and with a vector containing all components of the CD3 complex. This allows for the TCR to reach the cell surface, giving us a surrogate TCR-expressing cell in order to assess tetramer binding by flow cytometry. In the future, we will also generate lentivirus for each TCR to transduce Jurkat cells to determine whether we can induce TCR-mediated signaling. And, given that the MS, EC, and DP cells seem to be more biased toward the CD4 compartment, we plan to test co-transduction of these Jurkat cells with CD4 (or CD8) to test the effect of the co-receptor on T cell activation. Finally, upon successful validation and screening, we intend to clone these into baculovirus donor vectors such that they can be made recombinantly and tested *in vitro* for their binding to MR1 reagents by biophysical means. In complementing the single cell data with functional data, we will provide unique insights into the molecular basis of MR1T reactivity and the functional consequences thereof.

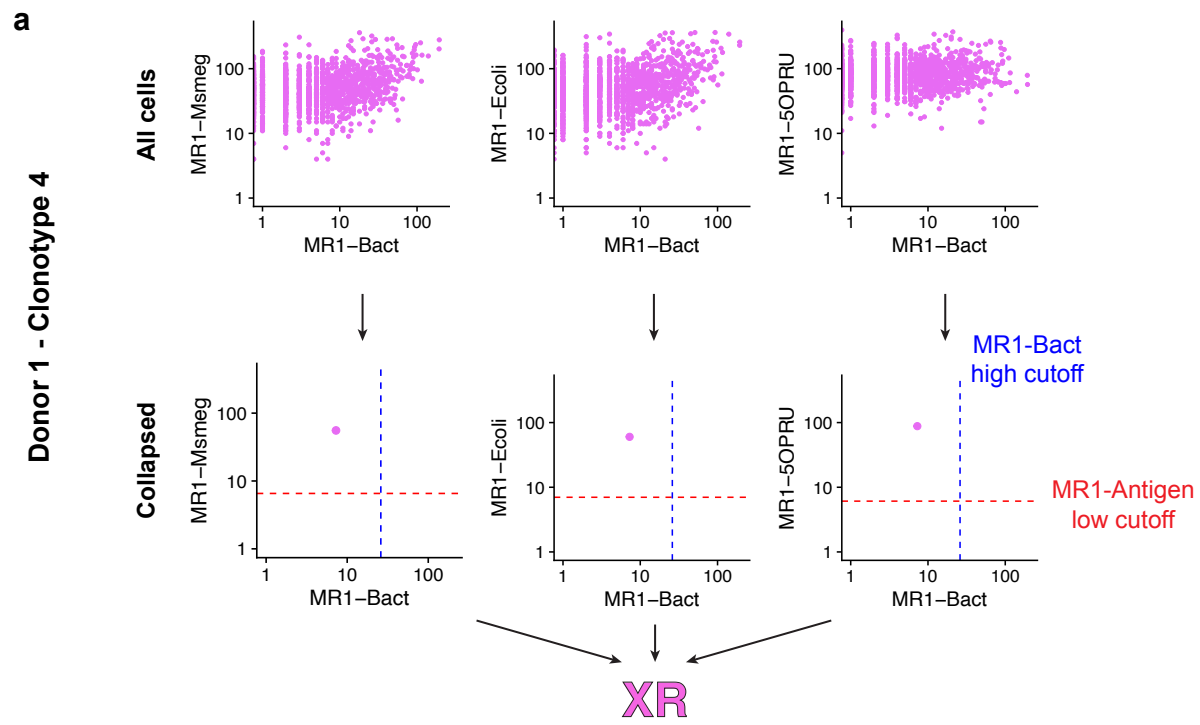
4.9 Supplemental information



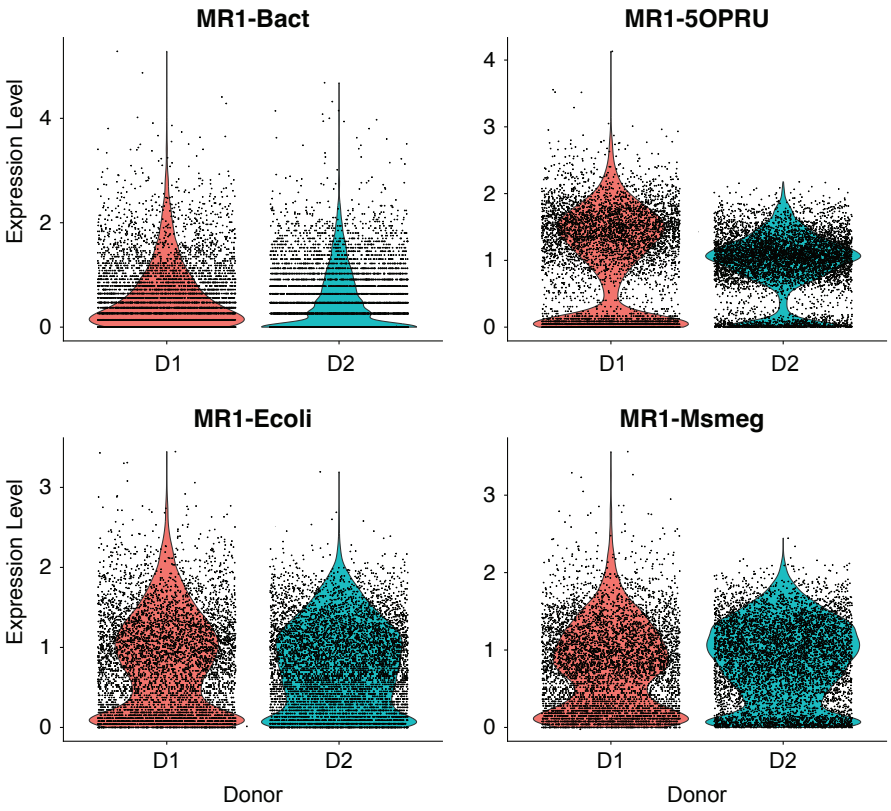
Supplemental Figure S4.1 Purification and quality control of the oligo-conjugated streptavidin-PE (from the second experiment, i.e. with Donor 2). a-b) Chromatograms from size-exclusion purification of the SA-PE alone (a) or SA-PE conjugated with oligo (b). SA-PE is ~300 kDa, so the protein elutes at the void volume of the S200i column, regardless of whether it's labeled with oligo or not. c) Fractions from the size exclusion chromatographic separation of SA-PE or the conjugated SA-PE were pooled and concentrated. A small sample of each variant (and a sample of SA-PE that was not purified by SEC and concentrated) was boiled at 95ºC for 10 minutes then loaded onto a DNA gel (doped with Sybr Safe; 1 kb Plus ladder included as a reference). Three species arise, one that does not migrate very far into the gel, another around 500-650 bp, and a third minor species around 100 bp. Based on these gels, the top band likely is unmodified SA-PE; its slow migration is likely due to the absence of negatively-charged oligo. The middle band is therefore the SA-PE conjugate, and the lowest, very faint band is likely oligo that has become unconjugated during the process of boiling (the HyNic-4FB bond is stable only to 92ºC).



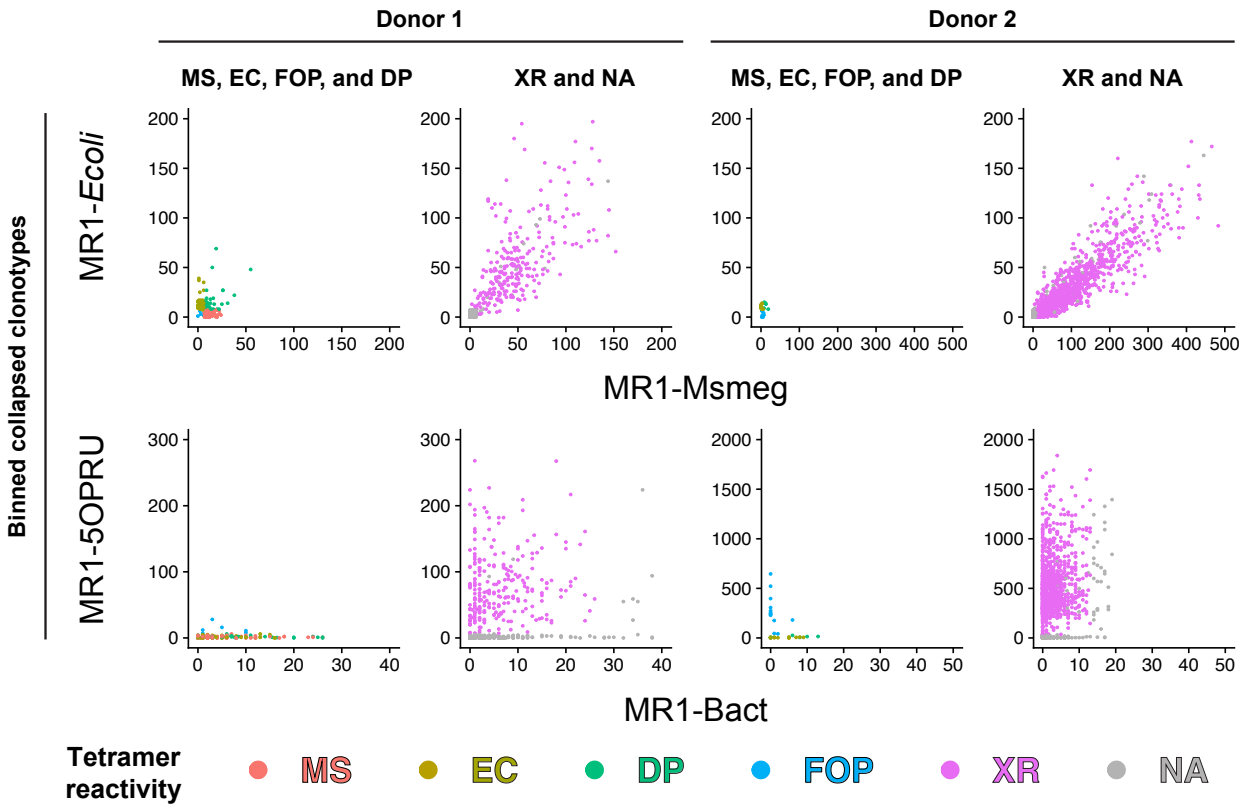
Supplemental Figure S4.2 Quality control of the DNA libraries and sequenced cells. a) Bioanalayzer (Agilent) chromatograms (left) and simulated "gels" (right) of the finalized V(D)J, ADT, and GEX libraries from the first donor. As expected, the V(D)J library shows a dominant peak around 600-700 base pairs, consistent with the transcript size for each TCR chain; likewise, the GEX library is broad due to the various transcript sizes of diverse genes, and the ADT library is very sharp due to the short barcodes appended to the antibodies/tetramer reagents used in this study. b-c) Cells were filtered based on the number of expressed genes, unique molecular indices (UMIs), and the percentage of expressed genes that were ribosomal and mitochondrial. Dashed lines indicate quality control cutoffs; cells falling within the box formed by each set of dotted lines were included in analysis. d-e) Feature-based filtering of cells. Cells with high reads (cutoff indicated by the dashed line) for the LYZ (lysosome, from the GEX assay) and the IgG1κ isotype control antibody (ADT assay) were removed from the analysis.



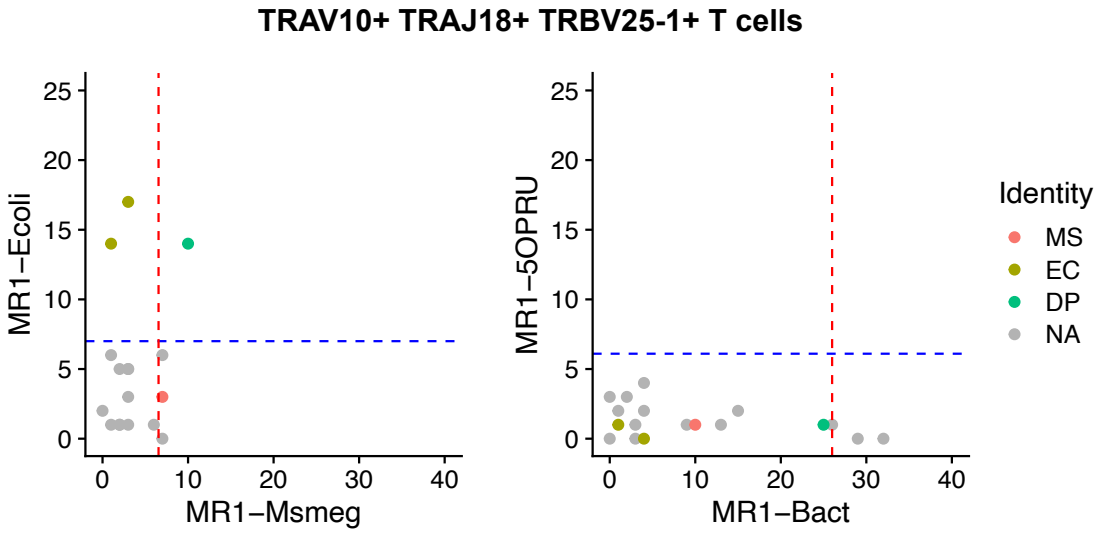
Supplemental Figure S4.3 Collapsed clonotype analysis. a) Example of this analysis for the hyperexpanded clone. The non-normalized reads of each antigen-loaded tetramer (y-axis) and the unloaded control tetramer (x axis) was averaged across all cells of "Clonotype 4" (from Donor 1) showing reads for at least one tetramer above background levels. After collapsing, the clonotype had to fall in the upper left quadrant (high reads for the antigen-loaded tetramer, low reads for the unloaded tetramer) to be considered reactive toward a particular tetramer. This clonotype met the criteria for all three tetramers and therefore was considered "cross-reactive" (XR). b) Empirically-determined cutoffs for each tetramer and each donor based on the distribution of tetramer reads for MAIT (TRAV1-2^{+}, TRAJ12/20/33^{+}) T cell clonotypes.



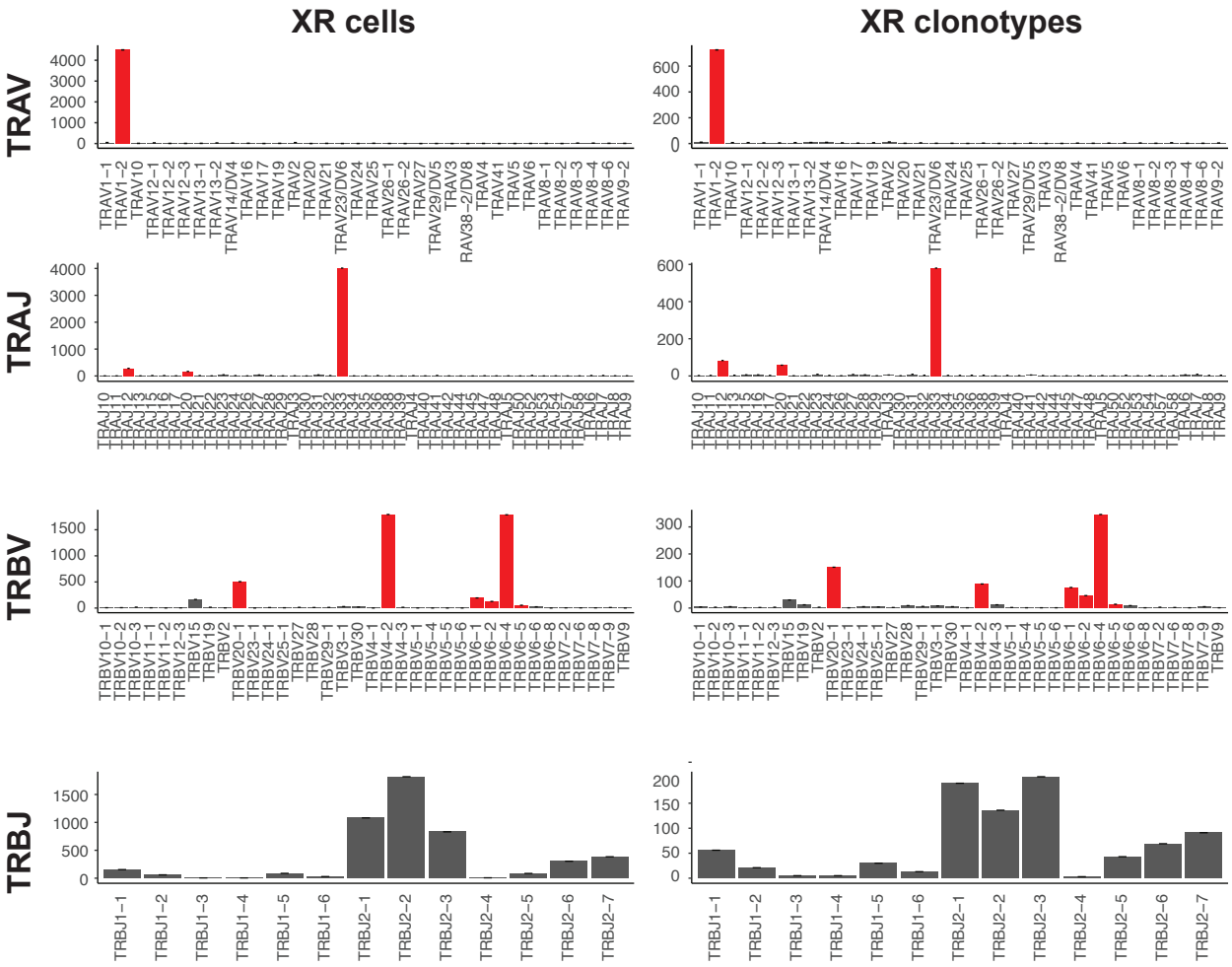
Supplemental Figure S4.4 Variability in tetramer reads across the two donors requires separate analysis of clonotype reactivity. a) Violin plots demonstrating different distributions of normalized (center-log ratio, margin 1) tetramer reads between the two donors (D1 and D2).



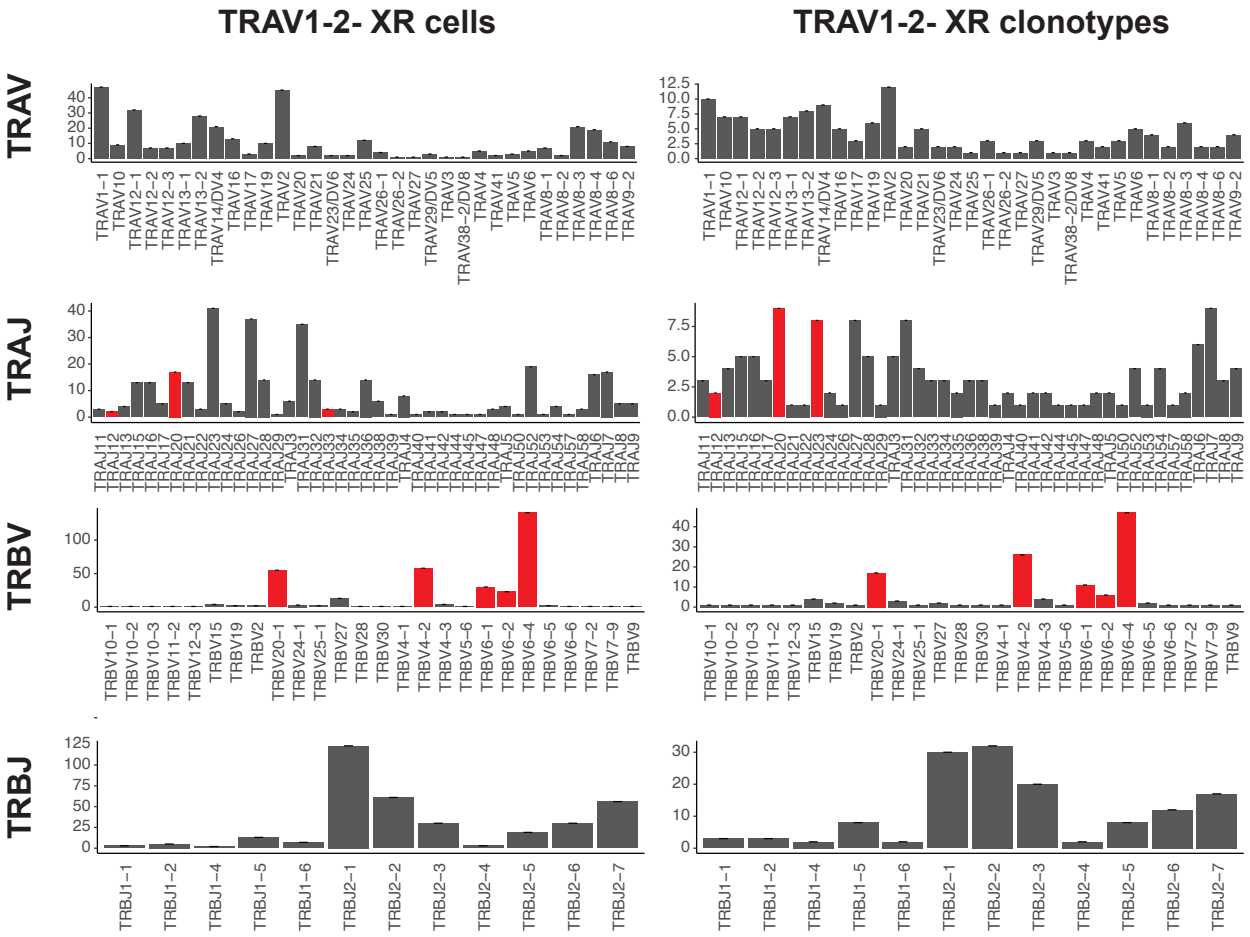
Supplemental Figure S4.5 Collapsed clonotypes binned into each reactivity category, shown as the non-scaled raw counts. The reactivities determined for each clonotype were re-appended onto each cell represented by the clonotype; these data are seen in **Figure 4.3**.



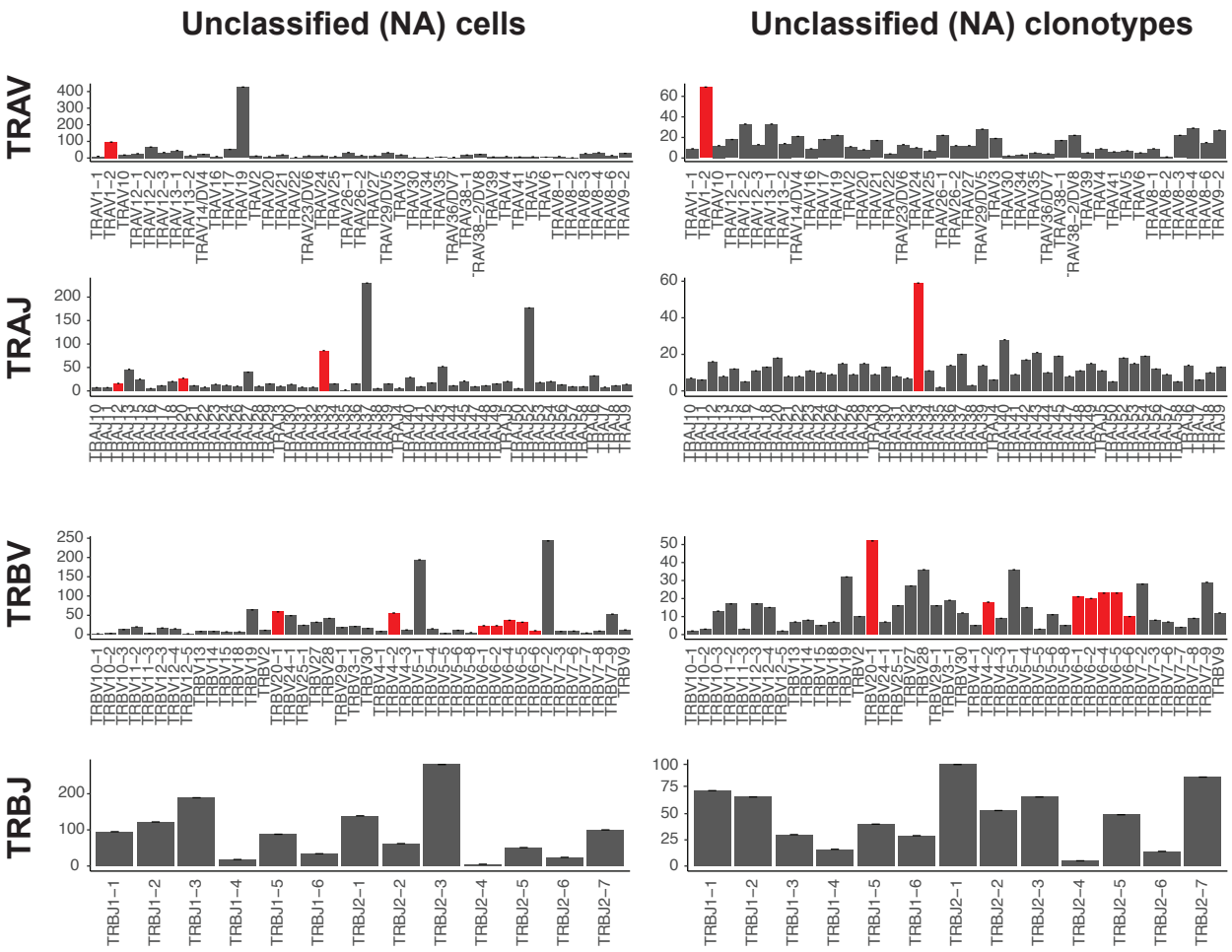
Supplemental Figure S4.6 Tetramer reactivity and binning of cells bearing the conventional Type I NKT rearrangement (TRAV10/TRAJ18/TRBV25-1).



Supplemental Figure S4.7 Gene usage by all XR cells and clonotypes. Common MAIT TCR genes are colored in red. TRBJ frequency is not typically assessed and therefore no bars have been shaded.



Supplemental Figure S4.8 Gene usage by TRAV1-2⁻ XR cells and clonotypes. Common MAIT TCR genes are colored in red. TRBJ frequency is not typically assessed and therefore no bars have been shaded.



Supplemental Figure S4.9 Gene usage by all unclassified cells and clonotypes. Common MAIT TCR genes are colored in red. TRBJ frequency is not typically assessed and therefore no bars have been shaded.

Chapter 5. Conclusions and future directions

5.1 Perspectives on the MR1 ligand repertoire and the dependence of MAIT cells on the riboflavin biosynthesis pathway (Relating Chapter 2)

It has long been suggested that the microbial ligand repertoire of MR1 is more diverse than is currently recognized. However, determining the "ligandome" for a protein that has evolved to bind small molecules it is an incredibly difficult question to address, even more so in an unbiased manner. Indeed, the work presented in **Chapter 2** is the result of a diverse team of microbiologists, immunologists, biochemists, mass spectrometrists, chemists, and cellular biologists, and yet it only begins to scratch the surface of what we believe to be the entire MR1 ligandome.

In **Chapter 2**, we showed that MR1 can bind a class of small indolyl-ribityllumazines belonging to a class we call "photolumazines" by characterizing the molecules bound to MR1 in a metabolomics pipeline. These molecules - PLV, PLVI, PLVII, PLIX, and PLX - in combination with the previously determined PLI and PLIII - add to the growing list of MR1 ligands arising from microbes. We know at least PLV to be capable of activating MAIT cells, though the potency of this ligand still does not account for the robust activation of our panel of MAIT clones and does not come near to the potency of the known ribitylpyrimidines. Interestingly, however, we did show that the position of the hydroxyl on the indole moiety affects its antigenicity, alluding further to the model of MR1T antigen selectivity set forth in **Chapter 3**.

Nevertheless, the potential of PLVI, PLVII, PLIX, and PLX to activate MR1T cells remain unknown, despite their high occupation of the MR1 ligand binding groove; while the purification of PLVI and PLIX is an ongoing process, PLVII and PLX remain elusive, due to what seems to be some kind of intrinsic chemical instability. It is therefore tempting to speculate that these two molecules may be ribitylpyrimidines bound to MR1 via a Schiff base, though we have no direct evidence to that end.

It also remains a mystery as to how these molecules are being formed by bacteria; they seem to be dependent on *Rib1ab* and *Rib2ab* for their biosynthesis, as indicated both by their absence from the MR1 pocket when MR1 was loaded with RBP knockout strains of *M. smegmatis.* Additionally, they are readily biosynthesized by *E. coli* only when both the *B. subtilis Rib1ab* and *Rib2ab* plasmids were co-transformed for overexpression. Nevertheless, as discussed in **Chapter 2**, there is no well-described metabolite that could be responsible for the condensation with 5-A-RU that could give rise to these metabolites. How the hydroxylation of this metabolite occurs, which presumably gives rise to at least PLV and PLVI, also remains unclear, as does the mechanism for the development of the putative double bond in the PLIX side chain. Altogether, there are very intriguing microbiological questions remaining as to how and why these molecules are being formed and whether they have a physiological role in bacteria or whether they are simply unintentional byproducts of normal metabolic processes. Further, it remains to be seen whether these molecules produced by *M. smegmatis* are also made by *M. tuberculosis* and whether - due to their enhanced stability over the ribitylpyrimidines - they (and other ribityllumazines) are reasonable candidates for immunotherapeutics.

One of the most intriguing open questions that arose from Chapter 2 is whether microbial molecules exist that are independent of the *Rib1ab* and yet can still activate MR1T/MAIT cells. The dogma in the field to this point has been that MR1T cells (or, at least, MAIT cells) depend absolutely on expression of the RBP pathway, and specifically by the presence of 5-A-RU, by microbes to be activated. Indeed, this is what we saw as well, when testing live bacteria of *M. smegmatis* RBP knockout strains. This was also supported by the complete inhibition of MAIT cells to be activated by MR1 loaded with the *M. smegmatis* Δ*Rib2ab* strain. However, when the *M. smegmatis* Δ*Rib1ab* strain was used to make lysate or to load MR1 in our co-infection system, there remained residual activation by MAIT cells toward these reagents. Our results suggest that there may be alternative substrates for *Rib2ab*, which implies that the key to MAIT cell antigens is not the metabolic intermediate 5-A-RU, but, rather, is the enzyme *Rib2ab*. This model does not overturn the decade of work that has been done on microbial MR1T-activating ligands; instead, it adds to the possibilities for the ligand diversity, and therefore the molecular mechanisms of MR1T cell antigen recognition, underlying this important immunological axis.

5.2 MR1T and MAIT cell ligand selectivity suggests that there is more plasticity underlying ligand recognition than once appreciated (Relating Chapter 3)

There have been several allusions to MR1T cell antigen selectivity over the past several years (reviewed in **Chapter 1**), but the case study that we present in **Chapter 3** is the first time that this phenomenon has been dissected with respect to antigen changes buried within the groove. Using a 5-deaza version of the well-characterized MR1 ligand DMRL (6,7-dimethyl-8-D-

ribityllumazine), we found that, indeed, MR1T and MAIT cells could discriminate between these two ligands despite the fact that the modification to the ligand is buried deep within the MR1 binding pocket and is presumably "invisible" to the TCR. Unfortunately, ligand availability made crystallization with the ligand (with or without a cognate TCR) infeasible, so we turned to molecular dynamics simulations to gain insight into any possible underlying molecular changes. Indeed, these simulations suggested that the differential recognition may be mediated by a change in the dynamics of the ribityl moiety, which provides the critical contact between the ligand and the TCR. However, given that the contact between Y95α of the canonical MAIT TCRα chain and the ribityl moiety is only responsible for around 0.5% of the buried surface area of these interactions, it is unlikely that the dynamics of the ribityl moiety alone are responsible for this discrimination. Rather, it is likely that there are allosteric changes to the entire molecular landscape of MR1, which could, in turn, influence TCR recognition. Determining the nature of these changes, however, is a difficult question to approach experimentally, since the best interpretation would be to visualize TCR-independent changes of the MR1 surface when complexed with different ligands. However, only one crystal structure of MR1 (of more than 50) is absent of a TCR, indicating that the TCR is likely acting in part as a crystallization chaperone. Going forward, however, we hope that we can use alternative crystallization chaperones that to not bind on the MR1 surface to answer some of these questions. Specifically, we hope that Fragment antigen binding (Fab) reagents generated by phage display in Anthony Kossiakoff's lab by Tomasz Slezak that were raised against *tn*-hpMR1^{-Bact} or *tn*-hpMR1^{+Msmeg} will help us overcome this limitation.

While the molecular details underlying ligand selectivity/discrimination are unclear, so too are the biological consequences of such a phenomenon. It remains a possibility that, in opposition to the current view of MAIT cell reactivity, certain antigens could promote selective expansions of MR1T cells with unique functional properties in a more adaptive-like manner than is currently thought possible for MAITs. Indeed, I have wondered whether the hyperexpanded clone we observed in Donor 1 of the single cell RNA sequencing experiment of **Chapter 4** was selected for by such a phenomenon, though we have no evidence that its expansion was MR1-dependent (as MAIT cells can also be activated *in vivo* by IL-12 and IL-18 produced in response to viruses, etc.). Nevertheless, paired with the findings of **Chapter 2**, where we showed that production of MR1 with different bacteria produced different repertoires of ligands in the MR1 groove, it seems as though both the ligand presentation and ligand recognition are more plastic than once described. As more microbial MR1 ligands are discovered in the coming years, it will be interesting to see whether the immune repertoire adapts to respond to each of these novel ligands in unique ways.

5.3 Single cell multi-omics investigations of MR1-restricted T cells reveal the potential for greater TCR diversity and 5-OP-RU-agnostic clonotypes (Relating Chapter 4)

In addition to exploring the variable responses of individual T cells toward diverse antigens - as well as the diverse antigens presented by MR1 - we have also been long interested in the true diversity of the MR1T cell population. While MAIT cells are interesting, we think that the tools available do not lend themself to explorations of anything other than these TRAV1-2^{+} MR1-restricted T cells. Indeed, one of the ways they're commonly defined is by the α-TRAV1-2

antibody. The recombinant MR1 reagents thus produced in our lab serve not only as a molecular fishing rod for MR1 antigens but also as a tool to isolate non-MAIT MR1T cells. In **Chapter 4**, I presented a pilot study using "multi-omics" approaches using oligo-labeled custom microbially-loaded tetramers to search for these cells. This yielded transcriptomic, cell surface protein expression, paired αβ TCR sequence, and tetramer reactivity data all at single-cell resolution. Unsurprisingly, many of the cells we recovered were MAIT cells, the majority of which bound all three of the microbially-loaded or 5-OP-RU-loaded tetramers (deemed "cross-reactive" cells). Even so, among cross-reactive cells, we did find a number of MR1Ts which seem to express TRAV1-2$^-$ TCRα chains yet are overwhelmingly paired with canonical MAIT TCRβ chains.

In addition to these TRAV1-2$^-$ cross-reactive cells, we also found a few non-5-OP-RU-reactive MR1Ts recovered from this exploration. Completely independent of their designation, they cluster together in the dimensional reductions, which are based on the transcriptomic data, indicating that in addition to a lack of reactivity toward the 5-OP-RU tetramer, they also share transcriptional properties. This phenotype, paired with the cell surface protein expression data, seem to imply that they are naïve or naïve-like cells, and many of them are CD4$^+$. The role of these cells in the periphery is completely unclear, but considering that the clonotypes are overwhelmingly unexpanded (even in the donor that may have had a recent antigen experience), I've speculated that perhaps these primarily TRAV1-2$^-$ cells are less competent to react toward agMR1 and therefore circulate the periphery without encountering an antigen capable of completing their peripheral training. In all, however, this investigation will require further experiments to validate our findings, especially with respect to the pairing of TCR sequence and

tetramer reactivity. In its entirety, this is the first study of its kind to shed light on the heterogeneity of the MR1T cell compartment *in vivo* using novel oligo-conjugated, heterogeneously-loaded MR1 tetramers in a single-cell RNA sequencing experiment and will hopefully reveal novel insights as to this unique immunological axis.

Chapter 6. Materials and methods

6.1 Materials and methods used in Chapter 2

Recombinant MR1 production and purification in insect and mammalian cell lines

For additional details regarding the production and purification of the MR1 reagents described here, reference Ladd *et al.* 2019. The insect and mammalian MR1 reagents, *tn*-hpMR1 and *hs*-hpMR1, were produced by baculoviral transduction of either *Trichoplusia ni* "Hi-5" or HEK293T/Expi293F cell lines. For each, a single chain construct of "human platform" MR1, consisting of the human α_1 and a_2 domains of MR1 and the bovine MR1 a_3 domain and bovine β_2m, was cloned into either pAcGP67a (insect cell transduction) or pVLAD6 (mammalian cell transduction). The baculovirus was generated by co-transfecting the virus with linearized baculovirus DNA into *Spodoptera frugiperda* (Sf9) cells and incubated for 1 week at 27ºC. This P1 virus was collected and used to transduce a larger culture of Sf9 cells to amplify the P2 virus, which was used to transduce host cells.

For the insect cell-generated protein, Hi-5 cells were transduced at $2x10^6$ cells/mL with a 1:500 hpMR1 baculovirus dilution. For biotinylated protein, baculovirus encoding for a cytoplasmically-expressed BirA ligase was added at 1:4000-1:2000 dilution and D-biotin was added to a final concentration of 10 μM. The culture was shaken at 27-28ºC at 120 rpm (1.5" orbital), then harvested 70-74 hours past baculoviral induction and cleared by centrifugation. The supernatant was filtered with glass fiber pre-filters and diluted 1:1 with HEPES-Buffered Saline

(1x HBS). A tangential flow concentrator (using one 30 kDa membrane per liter of culture) was then used to buffer-exchange the protein approximately 1:50 with 1x HBS, then PMSF was added to 1 mM, imidazole was added to 40 mM, and 0.5 mL Ni-NTA (bed volume) was added per liter of culture; this was stirred overnight at 4ºC. The resin was collected and washed with 75 mM imidazole in 1xHBS, then the protein was eluted sequentially with 200 mM imidazole in 1xHBS and 400 mM imidazole in 1xHBS. Fractions containing recombinant MR1 were pooled and quantified, then the pool was diluted with 1xHBS to achieve a concentration of 100 mM imidazole in 1xHBS. Recombinant HRV 3C protease (made in-house) was added 1:100 by mass to the diluted pool to cleave off the 12xHis tag; after mixing, this was incubated overnight at 4ºC. The next day, the protein was concentrated with a centrifugal concentrator to ~2 mL and diluted 1:1 with 1xHBS to further reduce the concentration of imidazole. The protein was then applied to a prepared column of new (or regenerated) Ni-NTA resin and the flow-through was collected. The column was further washed with 1xHBS with 40 mM imidazole and this was collected with the flow-through. The protein was then purified from any aggregate or dimer by size-exclusion chromatography (S200 Increase 10/300). For biotinylated proteins, tetramers were generated by quantifying the concentration of biotinylated protein (biotinylation efficiency was determined by a gel-shift assay) and adding the appropriate streptavidin variant to achieve a 1.1:4 molar ratio of streptavidin to biotinylated protein. The protein purity and appropriate pools were monitored by SDS-PAGE and stained with Coomassie G-250 for visualization.

For mammalian cell-generated protein, HEK293T cells were transduced at $2x10^6$ cells/mL typically with a 1:25 hpMR1 baculovirus dilution, or Expi293F cells were transduced at $3x10^6$

cells/mL typically with a 1:40 hpMR1 baculovirus dilution. Freshly-made 1M sodium butyrate solubilized in ultra pure water was added to the culture to a final concentration of 10 mM. The cultures were shaken at 130 rpm (1" orbital), 37 ºC, and 8% CO_2.The culture was harvested 70-74 hours past baculoviral induction and cleared by centrifugation, then dialyzed three times over 48 hours, each at a 1:20 dilution into 1xPBS pH 8.0, each at 4ºC. PMSF, imidazole, and Ni-NTA were then added and the mixture was stirred overnight at 4ºC as above; the resin was captured and washed with 1xPBS pH 8.0 with 25 mM imidazole, then eluted with 1xPBS pH 8.0 with 200 mM imidazole, then in the same buffer with 400 mM imidazole. The protein was then buffer-exchanged into 1xPBS pH 8.0 with 25 mM imidazole using a desalting column. The protein was quantified and HRV 3C protease was added 1:100 by mass to the protein, mixed by inversion, and incubated overnight at 4ºC. This protein was applied directly to a prepared (washed with 1xPBS with 25 mM imidazole) column of new Ni-NTA resin and the flow-through was collected. The column was washed further with 1xPBS pH 8.0 with 25 mM imidazole, and the washes were collected with the flow-through. For the biotinylated reagent, the protein was biotinylated with recombinant BirA ligase (made in-house) overnight at 4ºC. The protein was then purified by size-exclusion chromatography and, if appropriate, tetramers were made as described above.

Growth of mycobacterial strains for co-infection of insect and mammalian cells

All *M. smegmatis* strains were grown directly from frozen stocks directly into liquid culture. They were grown at 37ºC with a shaking speed of 112 rpm on a 1.5" orbital shaker and were not allowed to reach stationary phase (OD_{600} was not allowed to exceed 2). All lines were grown in

7H9 media with 10% OADC, 0.5% glycerol, and 0.05% Tween-80 and 0.22 μm-filtered. For experiments with *M. smegmatis* RBP knockout strains or relevant wild-type controls, riboflavin was added to the media to 85 μM and allowed to solubilize for 1-2 hours in the dark before filtering. Knockout strains were selected for with 50 μg/mL hygromycin.

Co-infection of cell lines to load MR1 with bacterial ligands

For co-infections, Hi-5 and HEK293T cells were transferred to antibiotic-free media at least 3 days before baculoviral transduction, and Expi293F cells were always grown in antibiotic-free media. For experiments comparing the wild-type *M. smegmatis* with gene knockouts of the riboflavin biosynthetic pathway, Expi293F cells were pelleted at 300 x g for 10 minutes and resuspended with Expi293 media with 85 μM exogenous riboflavin added. (The Expi293+RFN media was warmed at 37ºC to fully dissolve the riboflavin and kept in the dark as much as possible. When completely solubilized, the media was 0.22 μm-filtered.)

The time of baculoviral addition was considered "0 hours past induction" or "0 hpi." For co-infection of Hi-5 cells with *M. smegmatis*, a 50 mL overnight culture of *M. smegmatis* was inoculated and grown with shaking on the day before baculoviral induction. Immediately after Hi-5 cell transduction, the optical density of the culture was measured and converted into a cell density using the assumption that an OD_{600} of 1.0 is equivalent to $1.5x10^7$ cells/mL. *M. smegmatis* was then added (within the first 30 minutes after baculoviral transduction) to the culture medium at a final cell density of $4x10^3$ cells/mL.

For co-infection of Hi-5 cells with *E. coli*, on the day of baculoviral transduction, a 5 mL overnight culture of BL21 (DE3) pLys-S was inoculated from a glycerol stock into LB without antibiotic. At 17.5 hpi, the optical density of a 1:10 dilution of the overnight culture in LB broth was measured and converted into a cell density using the assumption that an OD_{600} of 1.0 is equivalent to $8x10^8$ cells/mL. *E. coli* was then added at 18 hpi to the culture medium at a final cell density of 40 cells/mL. This typically required several serial dilutions of the stock into sterile LB in sterile tubes to achieve a volumetric addition of less than 1 mL of *E. coli*.

For co-infection of Expi293F cells with *M. smegmatis*, an overnight culture of *M. smegmatis* was inoculated and grown with shaking on the day before baculoviral induction. At 30 hpi, 1 M HEPES pH 7.0 was added to the culture to a final concentration of 25 mM and the culture was transferred to a warm room (0% CO_2 supplementation, 34-37ºC) environment and shaken at 112 rpm (1.5" orbital). Immediately after transferring to the warm room, the optical density of the *M. smegmatis* culture was measured and converted as above, then the appropriate volume to add was calculated for a final cell density of $1.5x10^7$ cells/mL. This volume was transferred to sterile 50 mL conicals and the bacteria were pelleted by centrifugation at 500 x *g* for 10 minutes at 20ºC. The supernatant was decanted and the *M. smegmatis* was resuspended with 5 mL of the baculovirus-transduced cells, then everything was transferred to the Expi293F culture such that the time of co-infection was approximately 30.5 hpi.

Plate-bound tetramer ELISPOT (tetraSPOT) assay

Tetramer was added to the indicated concentration to 10 μg/mL anti-IFN-γ antibody (Mabtech,

clone 1-D1K) in 0.1 M $NaHCO_3/Na_2CO_3$, pH 9.6. Wells were coated overnight at 4ºC or for 6 hours at room temperature. Plates were washed washed three times with 1xPBS pH 7.2-7.4, then blocked with RPMI+10% human serum (HuS) for 1 hour. 0.5-1x10^4 MAIT cell clones were added to the wells overnight in RPMI + 10% HuS. PMA/ionomycin or PHA were used as positive controls. IFN-γ ELISPOTs were enumerated following development as previously described [109].

Flow cytometry of PBMCs and MAIT clones

Cryopreserved PBMCs or MAIT clones were thawed into RPMI + 10% fetal bovine serum (FBS), centrifuged, and resuspended in Fc receptor and internalization blocking media (1xPBS + 10% human serum + 100 nM dasatinib). After incubating for 30 minutes, the cells were aliquoted and stained for 30 minutes at 27.5ºC with varying concentrations of MR1 tetramers made with either streptavidin-APC or streptavidin-PE suspended in FACS buffer (PBS + 3% FBS). They were pelleted and resuspended in the appropriate cocktail of antibodies (or isotype controls) and live/dead stain. After antibody staining at 4ºC for 30 minutes, the cells were washed with FACS buffer twice and then analyzed by flow cytometry. Antibody clones used: TRAV1-2 (Vα7-2), clone 3C10; CD4, OKT4; CD8, BW135/80; CD3, UCHT1.

Recombinant expression and purification of MAIT TCRs for binding analysis

MAIT TCRs were produced as previously described [63]. Briefly, the RNA was extracted from each MAIT clones used in the activation panel and the V(D)J region of each TCR chain was amplified from the material. Each chain was cloned separately into the pAcGP67a vector after

the GP67 signal sequence and before either a “cysteine-optimized” TRAC or TRBC sequence [176]. To ensure heterodimer formation, the chains were tagged with cleavable C-terminal acidic and basic zippers each followed with a 6xHis tag [177]. The alpha chain construct contained a C-terminal AviTag before the zipper cleavage site. The P1 and P2 baculoviruses were produced in Sf9 cells [178], then the baculoviruses were added in tandem to Hi-5 cells and the cells were allowed to express the recombinant protein for 72 hours. The supernatant was harvested and the protein purified by Ni-NTA chromatography. Recombinant TCR was then treated with recombinant HRV 3C protease overnight and the cleaved zippers were removed with Ni-NTA. The protein was biotinylated *in vitro* using recombinant BirA ligase (made in-house) and the biotinylation efficiency was checked by a streptavidin gel-shift assay.

Biolayer interferometry (BLI)

For the binding between *tn*-hpMR1^{-Bact}, *tn*-hpMR1^{+Msmeg}, *hs*-hpMR1^{-Bact}, or *hs*-hpMR1^{+Msmeg} and the D481A9, D481C7 or D481F12 TCR, biotinylated *tn*-hpMR1 was immobilized on a streptavidin biosensor; binding by the recombinant TCR was analyzed as a titration series ranging from 0 to 32-64 μM, depending on protein availability. The *tn*-hpMR1 binding experiments with the D481A9, D481C7, and D481F12 TCRs were done by Dr. Cara Froyd. For binding between *tn*-hpMR1^{+Msmeg} and the D426G11 and D433H6 TCRs, biotinylated TCR was immobilized on a streptavidin biosensor; binding by MR1 was analyzed as a titration series from 0 to 100 μM. For all analysis, the presented curves are reference-subtracted, as indicated on the y-axis. For the equilibrium curve, the concentration of analyte was plotted against the response of the reference-subtracted curve 2 seconds before the disassociation step began.

Untargeted mass spectrometric analysis of recombinant MR1, synthetic standards, and bacterial supernatants

A total of 10 μg (5 μL) of hpMR1 was used for each injection for low pH reverse-phase nano-scale LCMS. Nano-LCMS was performed using an Eksigent NanoLC 415 RP-HPLC with an Eksigent autosampler (Sciex). Mobile phase solvents were: solvent A (0.1% formic acid in HPLC-grade water), and solvent B (0.1% formic acid in 95% acetonitrile, 4.9% HPLC-grade water; pH 3.0). Intact hpMR1 was loaded onto and desalted (using a trap-elute method) with a EXP 2 Nano Trap Halo C18 trap column (0.125 μL bed, 2.7 μm particle size) at 500 nL/min. After desalting, the trap was placed in line with a C18 separation column (75 μm [i.d.] × 15 cm long; 3 μm particle size, 120Å pore size; ChromXP) pre-equilibrated with 2% solvent B. Sample was then eluted at 300 nL/min using two linear gradients: 1. 10% solvent B to 40% solvent B in 70 minutes, 2. 40% solvent B to 80% solvent B in 7 minutes. Eluate was ionized using a PicoView ion source (New Objective) through an 11 cm long pulled glass emitter (20 μm i.d., tapered to 10 μm at the tip). Ion spectra were collected using an AB Sciex TripleTOF 5600 quadrupole time of flight mass spectrometer with source voltage set at -2600 V. Data was acquired in data-dependent acquisition mode with a survey m/*z* range of 150-1500 Da in negative ion polarity. Extracted ion chromatograms, MS1 survey spectra, and MS2 fragment spectra were made using PeakView 1.2 (Sciex). All EICs were drawn at the empirically-determined m/*z* with a mass tolerance of 0.05 Da. Pure PLV synthetic compounds and bacterial supernatants (fractionated or unfractionated) were run under the same conditions as the hpMR1 samples. Crude PLIII synthetic mixture was run as described in Harriff et al. 2018 [77]. When comparing

synthetic compounds with those eluted from MR1, the eluted ion and synthetic were verified to: 1. have retention times within 3 minutes of each other, 2. have MS1 ions within 3ppm of each other, and 3. have similar MS2 fragment spectra.

All EICs were determined with a 0.05 Da window and exported to be graphed in R.

Targeted mass spectrometric analysis of bacterial supernatants with multiple-reaction monitoring high resolution mass spectrometry

5µl of each fraction was subjected to MRM-HR MS. Samples were injected using the previously detailed Nano-LC and trap column. In addition to the Chrom XP C18 trap column by Optimize Technologies, a Waters ACQUITY M-Class HSS T3 75µm x 150mm with 1.8µm particle size was used for non-proteinaceous samples. A TOF-MS method of 24 minutes was used with the same parameters as the Nano LC-MS/MS injection method, except the samples were loaded onto the trap column over 4 minutes, and the gradient of 10% to 40% solvent B lasted from 2 minutes until 16 minutes, followed by the clearance gradient of 40% to 80% solvent B until 20 minutes. The following precursor ions were examined with a total TOF-MS cycle time of 1.8005 seconds, a total accumulation time of 0.249966 seconds, and a min and max TOF Mass range of 200 to 500 Da: 241.07, 275.10, 325.12, 327.09, 329.11, 362.10, 375.13, 385.10, 405.11, 428.12, 444.12, 460.11, 453.12, 488.10, and 485.12. The product ion scan accumulation times were 0.100037 seconds each, with a TOF mass range of 30 to 500 Da. Raw .wiff files were imported into Skyline [179], and data was examined searching for between 4 to 10 transitions for each target precursor ion, found in Table S2.2. Ions were then manually curated by making sure transitions

corresponded to precursor elution times, and the area under the curve was calc plated from the selected retention times using PeakView.

Comparative analysis of eluted hpMR1 ions

Raw .wiff files were converted to .mzxml files using mscovert from the ProteoWizard suite, with 64 bit compression. Files were analyzed with MarkerView for peak picking. A minimum retention time of 5.00 minutes and maximum retention time of 45.00 minutes was used to initially pare down ions, and then a minimum of 10.00 minutes and maximum of 30.00 minutes was used when most ions fell within this range. No subtraction offset was used. A minimum spectral peak width of 0.05 Da was used, as well as a minimum retention time peak width of three scans and a noise threshold of 1. Charge states were assigned, and the retention time tolerance was 6.00 minutes with a mass tolerance of 0.05 Da due to the variability of retention times due to injection of intact protein. A maximum number of peaks was set to 5,000 with no isotope filtering. Three biological replicates were examined, each with three technical replicates. Output data was examined in GraphPad Prism version 9.3.1 and plotted as a volcano plot.

Molecular Networking

.mzxml files used for MarkerView were also uploaded to the GNPS server using an FTP client. .mzxml files were subjected to the METABOLOMICS-SNETS (version 1.2.5) workflow. Files were loaded into their respective groups, with Precursor Ion Mass Tolerance set to 0.02 Da, and Fragment Ion Mass Tolerance set to 0.02 Da. The Min Pairs Cosine was set to 0.7, Network TopK set to 10, Maximum Connected Component Size to 100, Minimum Matched Fragment

Ions to 3, and Minimum Cluster size to 2, with MSCluster selected to run. Library Search Options included setting the Library Search Min Matched Peaks to 3, with Analog Searching enabled with a Score Threshold set to 0.6 and Maximum Analog Search Mass Difference at 50. All remaining settings in the workflow were set at Default. Networks were visualized in Cytoscape version 3.9.1.

Chemical synthesis of PLV isomers

General procedure for Photolumazine V. 5-A-RU HCl (14 mg, 0.045 mmol) was dissolved in 1 mL of H2O. Methyl hydroxyindoleglyoxylate (39 mg, 0.18 mmol) was dissolved in 1 mL of DMF. The solutions were combined, then the reaction was heated for 20 min at 120° C with microwave irradiation. The reaction was filtered, then purified by RP-HPLC (Waters Atlantis T3 30x150 mm column, water/acetonitrile, 20 mL/min flow rate, 0.9 mL injection volume, 20 min 10-40% B, 5 min 40% B). The fractions containing product were combined and lyophilized. All products were observed by mass spectrometry at the expected M/z of 444.

D-ribitylamine. D-ribose (3.6 g, 24 mmol) was added to 90 mL of saturated aqueous ammonia and 200 mL of saturated ammonium formate in ethanol. 4 mL of Raney nickel slurry were added, then the reaction was shaken under 45 psi of hydrogen for 18 hr. The reaction was filtered over Celite to remove the catalyst, then concentrated under vacuum. The crude green syrup was redissolved in water, then applied to an Amerlite 120 column. The flow through was reapplied 2X, then the column was washed 2X with water and eluted 3X with saturated aqueous ammonia. The eluant was concentrated with ethanol under vacuum. The crude product was redissolved in

water and acetonitrile, then lyophilized to give the product (3.4 g, 93% yield) as a beige, hygroscopic powder. The ribitylamine was dissolve in water and aliquoted at 91 mg/mL, then stored at -20° C.

5-N-RU. 6-chloro-5-nitropyrimidine-2,4(1H,3H)-dione (77 mg, 0.4 mmol) and D-ribitylamine (0.18 g, 1.2 mmol) in 2 mL of water were combined in a vial, which was sealed and heated to 190° C for 15 min. The reaction was diluted with 2 mL of MeOH, filtered, then purified by RP-HPLC (Waters Atlantis T3 30x150 mm column, water/acetonitrile, 20 mL/min flow rate, 0.9 mL injection volume, 5 min 0% B, 10 min 0-50% B, 5 min 50% B). The fractions containing product were combined and lyophilized to give 63 mg of product (52% yield).

5-A-RU HCl. 5-N-RU (56 mg, 0.18 mmol) was dissolved in 10 mL of methanol. 0.2 mL of 12.1 N aqueous HCl and 5 mg of platinum (IV) oxide were added, then the reaction was shaken under 45 psi of hydrogen, excluding light as much as possible, for 18 hr. The reaction was filtered over Celite to remove catalyst, then concentrated under vacuum. The crude product was redissolved in water and acetonitrile, then lyophilized to give 56 mg of product as a red syrup (quant. yield). General procedure for methyl hydroxyindoleglyoxylate. Hydroxyindole (0.1 g, 0.8 mmol) was loaded into a sealed vial, then dried under vacuum. 2 mL of dry THF were added and degassed, then the solution was cooled to 0o C. Oxalyl chloride (0.2 mL, 2.4 mmol) was added, then the reaction was stirred for 2 hr at RT. The reaction was quenched with 1 mL of 5:1 MeOH/TEA, then concentrated under vacuum. The crude was redissolved in 20 mL of EtOAc, washed 2X with 10 mL of 1N aq. HCl, 2X with 10 mL of sat. aq. NaHCO3, and 2X with 10 mL of brine,

then concentrated under vacuum to give the crude product, which was used without further purification.

NMR

Methyl 2-(4-hydroxy-1H-indol-3-yl)-2-oxoacetate. 1H NMR (d6-acetone) δ 8.53 (s, 1H); 7.20 (t, 1H); 7.03 (dd, 1H); 6.63 (dd, 1H); 3.97 (s, 3H).

Methyl 2-(5-hydroxy-1H-indol-3-yl)-2-oxoacetate. 1H NMR (d6-DMSO) δ 12.21 (s, 1H); 9.63 (s, 1H); 9.21 (s, 1H); 8.28 (d, 1H); 7.56 (s, 1H); 7.34 (dd, 1H); 6.77 (dd, 1H); 3.89 (s, 3H).

Methyl 2-(6-hydroxy-1H-indol-3-yl)-2-oxoacetate. 1H NMR (d6-acetone) δ 11.10 (s, 1H); 8.40 (s, 1H); 8.29 (t, 1H); 8.10 (d, 1H); 7.00 (d, 1H); 6.88 (dd, 1H); 3.91 (s, 3H).

Methyl 2-(7-hydroxy-1H-indol-3-yl)-2-oxoacetate. 1H NMR (d6-acetone) δ 11.45 (s, 1H); 9.10 (s, 1H); 8.37 (s, 1H); 7.80 (dd, 1H); 7.11 (t, 1H); 6.78 (dd, 1H); 3.93 (s, 3H).

PLV-4. 1H NMR (d6-DMSO) δ 13.28 (s, 1H); 11.23 (s, 1H); 8.50 (t, 1H); 6.88 (d, 1H); 6.69 (d, 1H); 6.49 (s, 2H); 6.24 (d, 1H); 4.34 (m, 1H); 4.21 (t, 1H); 3.92 (m, 1H); 3.51 (m, 2H); 3.35 (m, 2H).

PLV-5. 1H NMR (d6-DMSO) δ 11.02 (s, 1H); 8.66 (s, 1H); 8.39 (d, 1H); 8.13 (m, 1H); 7.12 (d, 1H); 6.62 (dd, 1H); 4.42 (d, 1H); 4.42 (m, 1H); 3.89 (m, 1H); 3.54 (m, 2H); 3.35 (m, 2H).

PLV-6. 1H NMR (d6-DMSO) δ 10.81 (s, 1H); 8.90 (s, 1H); 8.56 (d, 1H); 8.24 (s, 1H); 6.68 (s, 1H); 6.54 (d, 1H); 6.45 (s, 1H); 4.43 (dd, 1H); 4.23 (d, 1H); 3.89 (dd, 1H); 3.53 (m, 2H); 3.35 (m, 2H).

PLV-7. 1H NMR (d6-DMSO) δ 11.31 (s, 1H); 9.54 (s, 1H); 8.40 (s, 1H); 8.26 (d, 1H); 6.86 (t, 1H); 6.53 (d, 1H); 4.43 (d, 1H); 4.20 (s, 1H); 3.91 (s, 1H); 3.53 (m, 2H); 3.37 (m, 2H).

ELISPOT analysis of synthetic MR1 ligands, bacterial supernatants, or live bacteria

ELISPOT plates were coated with 10 μg/mL anti-IFN-γ antibody (Mabtech, clone 1-D1K) in 0.1 M NaHCO3/Na2CO3, pH 9.6 overnight at 4 ºC or for 6 hours at room temperature, then washed and blocked as described above. 0.5-1x10^4 human monocyte-derived dendritic cells were plated in the wells in RPMI+10% HuS. For synthetic ligands and bacterial supernatants, the sample was resuspended in RPMI+10% HuS at 2x the desired concentration, then added to each well as indicated such that their concentration was 1x. For co-incubation with *M. smegmatis*, each strain was added at an MOI of 10. After one hour, 0.5-1x10^4 MR1T cell clones were added and the plate was incubated overnight at 37 ºC. PMA/ionomycin, PHA, or *tn*-hpMR1^{+Msmeg} were used as positive controls. IFN-γ ELISPOTs were enumerated following development as previously described [109].

Cell surface stabilization of MR1 by synthetic ligands

BEAS-2B:MR1-GFP cells were grown in a 6-well tissue culture plate to ~70% confluency, and then incubated with synthetic compounds or vehicle at the indicated concentrations for 16 hours. Cells were harvested on ice and surface stained with the anti-MR1 26.5 antibody (1:100) conjugated to APC (BioLegend) for 40 minutes on ice in the presence of 2% human serum, 2% goat serum, and 0.5% FBS. Cells were washed and fixed, and subsequently analyzed with a BD FACS Symphony flow cytometer and FACS Diva software (BD).

Isolation of MR1 ligands from bacterial supernatants

B. subtilis RibIab in pET22b and *Rib2ab* in pET28a (or two unrelated proteins in the same two vectors) were transformed into the BL21(DE3)-pLysS strain of *E. coli* and plated onto LB agar with kanamycin, ampicillin, and chloramphenicol. 10 mL of an overnight liquid culture was used to inoculate each liter of LB for large-scale expressions. After growth to an optical density of 0.6-0.8, expression was induced with IPTG added to 1 mM and glucose was, variably, added to 20 g/L. (Glucose supplementation was found to divert 5-A-RU conversion toward DMRL in lieu of the photolumazines; see **Chapter 2.9**.) These proteins were expressed for 24 hours with shaking (225 rpm) at 37 ºC, then the supernatant was harvested and cleared by centrifugation and the pellets were stored at -80 ºC. The supernatant was rough filtered with glass fiber pre-filters and then 0.22 μm filtered. A tangential flow concentrator with two 10 kDa membranes was used to filter the supernatant (the flow-through) to remove large protein contaminants. A 100 mL bed volume of Florisil was then prepared by washing the gel with ultrapure water until the water ran clear, then transferring to a large glass buret. The supernatant was siphoned over the column and the flow-through collected. The column was then washed with 200 mL ultrapure water, which was collected. The photolumazines were then eluted with 10% acetone in ultrapure water and collected in fractions. The desired fractions (as determined by TLC) were pooled and lyophilized. The same fractions were selected for the negative control (two "unrelated" genes) and pooled separately. Both the control and *Rib1ab*/*Rib2ab* samples were then lyophilized.

The lyophilized material was weighed and resuspended to 100 mg/mL in solvent B (50:50 MeCN: 10 mM ammonium acetate; pH 9.0) and cleared by centrifugation. An XBridge BEH

Amide OBD Prep Column (130Å pore size, 5 μm particle size, 30 mm [i.d.] x 250 mm) was then equilibrated in 99% solvent B (90:10 MeCN: 10 mM ammonium acetate, pH 9.0). 1.5 mL of the soluble material (supernatant) was 0.22 μm filtered and injected onto the column. A decreasing acetonitrile gradient from 99% solvent B to 30% solvent B was immediately initiated to run over 120 minutes with a 5 mL/minute flow rate. The column was then held at 30% Solvent B for 16 minutes, transitioned to 10% Solvent B over 20 minutes, and then held at and then at 10 % B for 16 minutes. Approximately 115 fractions were collected (1.5 minutes or 7.5 mL each). Fractions were assessed by TLC.

Thin layer chromatography

All thin layer chromatography was performed on aluminum-backed, silica gel 60 matrix-coated (200 μm coating, 10-12 μm particle size) TLC plates impregnated with F_{254} fluorescent indicator (Millipore). The solvent system is indicated in the caption of relevant figures, but was typically 80:20 acetonitrile:water, 80:20 acetonitrile:10 mM ammonium acetate, or 80:5:15 acetonitrile:ethyl acetate:water. Bands were visualized by 254 nm, 312 nm, and/or 365 nm light.

6.2 Materials and methods used in Chapter 3

General chemical synthesis

Reagents were obtained from Sigma-Aldrich or Fisher Scientific. ^{1}H-NMR was recorded on a Bruker DPX spectrometer at 400 MHz. Chemical shifts are reported as parts per million (ppm) downfield from an internal tetramethylsilane standard or solvent references. High-resolution mass spectra were acquired on a ThermoElectron LTQ-Orbitrap Discovery high resolution mass spectrometer with a dedicated Accela HPLC system by Andrea DeBarber at the Bioanalytical MS facility, Portland State University. For air- and water-sensitive reactions, glassware was oven-dried prior to use and reactions were performed under argon. Dichloromethane, N,N-dimethylformamide, and tetrahydrofuran were dried using a solvent purification system manufactured by Glass Contour, Inc. (Laguna Beach, CA). All other solvents were of ACS chemical grade (Fisher Scientific) and used without further purification unless otherwise indicated. Analytical thin-layer chromatography was performed with silica gel 60 F_{254} glass plates (SiliCycle). Flash column chromatography was conducted with pre-packed normal or reversed phase columns (Biotage). High performance liquid chromatography (HPLC) was performed on an Agilent 1260 Infinity system with a flow rate of 1.0 mL/min using a Sunfire C18-A column 150 x 4.6 mm, 5 micron analytical column or a Sunfire 30 x 50 mm, 5 micron preparative column. HPLC analytical conditions: mobile phase (MP) A: 0.1% formic acid in water, B: 0.1% formic acid in acetonitrile (ACN), flow rate = 1.0 mL/min, gradient: 0% B for 2 min, 0-100% B over 13 min, 100% B for 2 min, UV-Vis detection at λ_1 = 254 nm and λ_2 = 220 nm. All final products were ≥95% purity as assessed by this method. Retention time (t_R) and purity refer to UV detection at 220 nm. Preparative HPLC conditions: mobile phase (MP) A:

0.1% formic acid in water, B: 0.1% formic acid in acetonitrile (ACN), flow rate = 10.0 mL/min, gradient **A**: 0-30% B over 7 min, 30-50% B over 2 min, 100% B for 1 min; gradient **B**: 30% B for 6 min, 30-50% B over 8 min, 100% B for 4 min, UV-Vis detection at λ_1 = 254 nm and λ_2 = 220 nm. Chemical/functional stability of DZ, as assessed by EliSpot with the DZ-reactive D481C7 clone showed that it was stable after freeze/thaw and storage at -80° C (**Figure S3.6**).

6-(((2S, 3S, 4R)-2,3,4,5-Tetrahydroxypentyl)amino)pyrimidine-2,4-(1H,3H)-dione (2a)

Commercially available ribitylamine [180] (150 mg, 1.0 mmol) and chlorouracil (42 mg, 0.28 mmol) were dissolved in 15 mL water in a 20 mL microwave vial and microwaved at 150° C for 2 h. After the reaction was complete, the desired ribitylaminouracil product (24 mg, 0.092 mmol, 33% yield) was isolated by preparative HPLC using gradient A. ^{1}H-NMR was consistent with reference spectra [57].

6-(((3S,4R)-3,4,5-Trihydroxypentyl)amino)pyrimidine-2,4-(1H,3H)-dione (2b)

Commercially available 1- amino- 1, 2- dideoxy D- *erythro*- pentitol [181] (135 mg, 1 mmol) and chlorouracil (42 mg, 0.28 mmol) were dissolved in 15 mL water in a 20 mL microwave vial and microwaved at 150° C for 2 h. After the reaction was complete, the desired aminouracil product (26 mg, 0.11 mmol, 39% yield) was isolated by preparative HPLC using gradient A and used without further purification.

6,7-Dimethyl-8-D-ribityldeazalumazine (DZ)

Ribitylaminouracil **2a** (26 mg, 0.10 mmol) and the sodium salt of commercially available 2-

methyl butan-3-one-ol [182] (26 mg, 0.20 mmol) were refluxed in 0.5 M HCl (1.0 mL) at 100° C for 2 h. After the reaction was complete, DZ (8.0 mg, 0.025 mmol, 25 % yield) was isolated by preparative HPLC using gradient B. ^{1}H-NMR was consistent with reference spectra (Al-Hassan 1980). ^{1}H-NMR (400 MHz, D_2O) δ 2.35 (s, 3H), 2.74 (s, 3H), 3.64 (m, 2H), 3.68 (m, 2H), 3.84 (m, 2H), 4.31 (m, 2H), 4.92 (m, 2H), 8.40(s, 1H); HRMS (ESI-TOF) Calculated for $C_{14}H_{20}N_3O_6$ [M +H^+] 326.1352, Found 326.1347.

6,7-Dimethyl-8-D-(2'-deoxyribityl)deazalumazine (2'-deoxy-DZ)

Compound **2b** (26 mg, 0.10 mmol) and the commercially available sodium salt of 2-methyl butan-3-one-ol (26 mg, 0.20 mmol) were refluxed in 0.5 M HCl (1.0 mL) at 100° C for 2 h. After the reaction was complete, 2'-deoxy-DZ (8.0 mg, 0.025 mmol, 25 % yield) was isolated by preparative HPLC using gradient B. ^{1}H-NMR (400 MHz, D_2O/CD_3OD) δ 1.93 (m, 2H), 2.19 (m, 2H), 2.41 (s, 3H), 2.74 (s, 3H), 3.34- 3.74 (m, 5H), 8.31(s, 1H); HRMS (ESI-TOF) Calculated for $C_{14}H_{20}N_3O_5$ [M +H^+] 310.1397, Found 310.1347.

3-(2,4,7-Trioxo-8-((2S,3S,4R)-2,3,4,5-tetrahydroxypentyl)-1,2,3,4,7,8-hexahydropyrido[2,3-d]pyrimidin-6-yl)propanoic acid (DZPLI)

Ribitylaminouracil **2a** (26 mg, 0.1 mmol) and commercially available 1,5-diethyl-2-formylpentanedioate (75 mg, 0.40 mmol) were refluxed in 0.5 M HCl (1.0 mL) at 100° C for 2 h. The crude mixture was then hydrolyzed in 2 M LiOH at 40° C overnight. After the reaction was complete, DZPLI (4.2 mg, 0.011 mmol, 11% yield) was isolated by preparative HPLC using gradient B. ^{1}H-NMR (400 MHz, CD_3OD) δ 2.61 (t, 2H), 2.82 (t, 2H), 3.67 (m, 2H), 3.83 (m,

2H), 4.14 (m, 1H), 4.35 (m,1H), 4.64 (d, 1H), 7.84(s, 1H); HRMS (ESI-TOF) Calculated for $C_{15}H_{20}O_9N_3$ [M + H^+] 386.1205, Found 386.1194.

4-Chloro-2, 6-dimethoxypyrimidine-5-carbaldehyde (4)

An oven-dried flask was charged with commercially available 4-chloro-2,6-dimethoxypyrimidine (0.95 g, 5.5 mmol) and then evacuated and backfilled with argon. Anhydrous THF (5 mL) was added through a rubber septum. The mixture was cooled to -78° C and a 1.6 M solution of n-butyllithium (*n*-BuLi) in hexanes (3.8 mL, 6.0 mmol) was added dropwise. The mixture was stirred for an additional 0.5 h, and DMF (1 mL, 13 mmol) was added and stirring was continued for 2 h at the same temperature. The reaction was quenched by addition of aqueous HCl (1.6 M, 25 mL), and the mixture was extracted with ether (3 x 40 mL). The combined organic layers were washed with aqueous HCl (1.6 M, 25 mL) and water (40 mL), dried (Na_2SO_4) and evaporated to dryness. The residue was purified by column chromatography on silica gel (ethyl acetate-toluene, 1:6) to afford 4-chloro-2,6-dimethoxypyrimidine-5-carbaldehyde **4** (0.80 g, 4.0 mmol, 73%). ^{1}H-NMR was consistent with reference spectra (Nencka 2007). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.11 (s, 3H), 4.15 (s, 3H), 10.34 (s, 1H).

(E)-4-(4-Chloro-2,6-dimethoxypyrimidin-5-yl)but-3-en-2-one (5)

To a solution of **4** (120 mg, 0.60 mmol) in toluene (4 mL) was added commercially available 1-(triphenylphosphoranylidene)-2-propanone (189 mg, 0.60 mmol). The reaction mixture was refluxed for 6 hours. After the reaction was complete, the organic solvent was removed *in vacuo*. The resulting residue was purified by column chromatography on silica gel (ethyl acetate-

hexane, 1:3) to afford the desired product **5**. ^{1}H-NMR (400 MHz, $CDCl_3$) δ2.37 (s, 3H), 4.01 (s, 3H), 4.09 (s, 3H), 7.03 (d, 1H, *J*=18Hz), 7.67 (d, 1H, *J*=18Hz).

(E)-4-(2,4-Dimethoxy-6-(((2S,3S,4R)-2,3,4,5-tetrahydroxypentyl)amino)pyrimidin-5-yl)but-3-en-2-one (6)

To a stirred solution of **5** (63 mg, 0.26 mmol) in DMF (4 mL) was added ribitylamine (120 mg, 0.80 mmol). The reaction mixture was refluxed for 12 h. After the reaction was complete, the organic solvent was removed in vacuo. The resulting residue was purified by preparative HPLC to deliver the desired product **6** (21 mg, 0.060 mmol, 23% over two steps). ^{1}H-NMR (400 MHz, CD_3OD) δ2.33 (s, 3H), 3.95-3.51 (m, 9H), 4.01 (s, 3H), 4.18 (s, 3H), 6.83 (d, 1H, *J*=16.4), 7.48 (d, 1H, *J*= 16.4).

7-Methyl-8-((2S,3S,4R)-2,3,4,5-tetrahydroxypentyl)pyrido[2,3-d]pyrimidine-2,4(3H,8H)-dione (monomethyl-DZ)

To a solution of **6** (20 mg, 0.050 mmol) in 6 M HCl (1 mL) was refluxed at 70° C overnight. After the reaction was complete, **monomethyl-DZ** (6.0 mg, 0.020 mmol, 39%) was purified by preparative HPLC. ^{1}H-NMR was consistent with reference spectra [157]. ^{1}H-NMR (400 MHz, D_2O) δ2.78 (s, 3H), 3.41-3.88 (m, 5H), 4.29 (m, 1H), 4.44-4.65 (m, 5H) 7.18 (d, 1H, *J*=10.3), 7.48 (d, 1H, *J*= 12.2)

Human subjects

This study was conducted according to the principles expressed in the Declaration of Helsinki.

Study participants, protocols and consent forms were approved by the Institutional Review Board at Oregon Health & Science University (OHSU) (IRB00000186).

Reagents and cells

Dendritic cells (DC) were derived from human peripheral blood monocytes as previously described [183, 184]. The bronchial epithelial cell line BEAS-2B (CRL-9609) was originally obtained from ATCC and was cultured in DMEM+10% heat inactivated fetal bovine serum (FBS). The BEAS-2B:ΔMR1 cell line was derived by CRISPR/Cas9 disruption of the MR1 gene, and MR1 expression was reconstituted in these cells [158]. Wild-type BEAS-2B cells overexpressing MR1 fused to GFP (BEAS-2B:MR1-GFP) were previously described [140]. MAIT cell clones were derived, expanded, and maintained as previously described [70, 109].

ELISpot assay

DC or BEAS-2B cells were harvested, counted and used in equivalent numbers, as indicated in the figure legends, as antigen presenting cells in an ELISpot assay with IFN-γ production by MAIT cell clones as the readout as previously described [109]. Synthetic compounds or positive controls (*M. smegmatis* supernatant or PHA) were added to the cells at concentrations indicated for one hour prior to addition of the MAIT cell clones, and the ELISpot plates were incubated for 18 hours prior to development. Phytohemagglutinin (PHA) was used at 10 mg/mL. Supernatant from *M. smegmatis* was prepared by collecting the <3 kDa fraction of supernatant from logarithmically-growing bacteria using a size exclusion column (Millipore). The volume of supernatant required for maximal response in the assay was determined empirically following

preparation. Blocking was performed using the a-MR1 26.5 clone (BioLegend) and an IgG2a isotype control, added at 2.5 mg/mL for 1 hour prior to the addition of ligand.

Flow cytometry

BEAS-2B:MR1-GFP cells were grown in a 6-well tissue culture plate to ~70% confluency, and then incubated with synthetic compounds or vehicle at the indicated concentrations for 16 hours. Cells were harvested on ice and surface stained with the anti-MR1 26.5 antibody (1:100) conjugated to APC (BioLegend) for 40 minutes on ice in the presence of 2% human serum, 2% goat serum, and 0.5% FBS. Cells were washed and fixed, and subsequently analyzed with a BD FACS Symphony flow cytometer and FACS Diva software (BD). Cells represented in the histograms on **Figure 3.3** were gated as BEAS-2B cells with high MR1-GFP expression. All analyses were performed using FlowJo software (TreeStar).

MR1 ligand docking

The original atomic coordinates of the β_2m/MR1 complex were taken from the x-ray crystal structures ("donor structures") of human MR1 and human β_2m in complex with either the ligand 6-FP only (PDB ID: 4GUP) [57] or in complex with a MAIT TCR and the ligand 7-methyl-6-hydroxy-8-D-ribityllumazine (HMRL) (PDB ID: 4L4V) [90]. Additionally, to confirm the accuracy of AutoDock in positioning ligands that do not form a Schiff base, we used donor structures from PDBs 4NQD, 4LCW and 4LCC [76, 89]. The crystallographic ternary complexes of 4NQD and 4LCW are of MAIT TCRs with $MR1^{K43A}$, an MR1 mutant which cannot form a Schiff base, with ligands 5-OP-RU and HMRL, respectively [76]. Importantly, the K43A

modification does not alter the binding mode of 5-OP-RU or HMRL despite the lack of the Schiff base. 4LCC contains coordinates for a crystal structure of chimeric human-bovine MR1/ β_2m in complex with rRL-6-CH_2OH (a reduced ribityllumazine) and a MAIT TCR [89].

For PDB 4GUP, chains A (MR1) and B (β_2m), which together compose one of the two conformations of MR1/β_2m in the asymmetric unit, were selected due to this conformation's similarity to that found in structures of agMR1-TCR complexes (PDB ID: 4L4V, 4IIQ, 6PUF) [63, 75, 90]. For PDB 4L4V, 4NQD, and 4LCW, chains A (MR1) and B (β_2m) were also chosen, though there was only minor structural heterogeneity between the two copies of MR1 in the asymmetric unit. For 4LCC, a structure of chimeric human-bovine MR1 in complex with rRL-6-CH_2OH and the same MAIT TCR as that of 4L4V, chain A (single chain β_2m-MR1) was chosen. The PDBs were stripped of the remaining polypeptide chains, the ligands, all waters, and crystallographically-resolved ions.

Separately, HMRL, DMRL, DZ, 2'-deoxy-DZ, monomethyl-DZ, rRL-6-CH_2OH, PLI, and DZPLI ligands were sketched with ChemDraw 18.2 and copied to BIOVIA Discovery Studio [185] for exporting as Mol2 files with 3D information. Ligand files were subsequently converted to PDB format using Pymol (Pymol Version 2) and visually inspected for appropriate geometry and bond angles. PDBQT files were prepared for both protein and ligand files using AutoDockTools suite [186, 187], which provides additional information regarding partial charge, atom type, and rotatable bonds. In order to restrict docking to the A' pocket, Vina [159] was run using an *x, y, z* box size of 18, 20, 14 Å centered at *x, y, z* coordinates -6.19, -9.44, -11.53. All

other Vina parameters were set to the default. Each ligand's top binding mode was selected as the representative structure, and a single PDB was prepared for each dock in Pymol by combining MR1 and ligand output files. Alignments between the two donor structures were performed by aligning the C_α (alpha carbons) in the β sheet of the α_1/α_2 domains.

CHARMM input generation and molecular dynamics

The β_2m/agMR1 complexes were solvated with the TIP3P water model using CHARMM-GUI Solution Builder [188-190], and neutralized with K^+ and Cl^- ions at physiological concentrations (0.15 M). To obtain missing parameters, ligands were parametrized by PDB coordinates using CGenFF [191]. Generated inputs were uploaded to the Midway compute cluster of the University of Chicago Research Computing Center to execute MD simulations. Each simulation was allowed 0.5 ns of equilibration followed by 80 ns of production, and each ligand was simulated in triplicate for a total of 15 simulations. Equilibrated systems use an NVT ensemble and production runs use an NPT ensemble, with the temperature kept constant at 300.15° K using Langevin dynamics [192]. The simulations were kept at constant pressure at one bar with the Nosé–Hoover Langevin piston by allowing the cell box size to change semi-isotropically [193]. van der Waals interactions were computed using a Lennard-Jones force-switching function over 10–12 Å while long-range electrostatics used particle mesh Ewald [194]. Production runs used a 2-fs time step and the SHAKE algorithm to constrain the bonds having hydrogen atoms [195].

Simulation analysis

Raw simulation data was processed using Bio3D [196], an R package with the ability to read,

write and process biomolecular structure and trajectory data. Root mean square fluctuation (RMSF) was calculated using included functions to determine the conformational variance of each atom with respect to their mean position. Structural visualizations and alignments were performed using VMD [197], and renders were generated with Tachyon internal-memory processes. Ribityl time lapse was accomplished by aligning dynamics by aromatic core, and displaying the initial ring structure while selecting the ribityl pose every 8 ns starting at 3 ns (to allow for equilibration). RMSD and RMSF were plotted using the *ggplot2* library in R [198].

Data analysis

Unless otherwise indicated, experimental data were plotted and analyzed for statistical significance using Prism 9 (GraphPad Software).

6.3 Materials and methods used in Chapter 4

Generation of custom oligo-labeled streptavidin PE and MR1 tetramerization

Oligos compatible with the Chromium Single Cell 5' v2 workflow (containing barcodes from the Feature Barcode Whitelist) were obtained from IDT with a 5' C12 Amino Modifier. Streptavidin-PE (SA-PE) (Agilent Technologies, PRJS34) was modified with each oligo using the Vector Laboratories Protein-Oligo Conjugation Kit per the kit instructions. The final modified SA-PE was either desalted twice with a 40 kDa-cutoff, 0.5 mL ZymoSpin filter (Donor 1) or purified by size-exclusion chromatography (Donor 2) to remove excess oligo. MR1 or CD1D (Donor 1, see **Chapter 4.2**) tetramers were produced by adding MR1 in eight parts to SA-PE to a final molar ratio of MR1:SA-PE of 4:1.1.

PBMC isolation, staining, and sorting to enrich tetramer$^+$ cells

For each of the experiments, PBMCs were originally isolated from fresh blood draws from anonymized, consented donors into heparin-coated tubes. They were separated from other blood components by Ficoll-Paque, washed several times with PBS, and strained with a 30 μm strainer before staining (Donor 1/Experiment 1) or freezing (Donor 2/Experiment 2). The Donor 2 sample was thawed as described into RPMI + 10% FBS, washed several times with PBS, blocked with human serum, and re-strained before staining. Cells were stained for 30 minutes in the dark on ice with a cocktail containing 200 nM of each tetramer in PBS + 2.5% FBS, then the appropriate CITE-SeqC antibodies (α-CD8, α-CD4, α-TRAV1-2 (Va7-2), α-PD1, α-CD26, α-CD161, α-CD45RA, α-CCR7, IgG1κ isotype, IgG2aκ isotype, and IgG2bκ isotype; BioLegend), Live/Dead Near-IR reagent (Thermo Fisher Scientific) and CD3-FITC (BioLegend) were added

at 2X concentration on top of the tetramer staining mixture (to a final concentration of 1X). The antibodies were allowed to bind for another 20 minutes on ice in the dark, then the cells were washed twice with PBS + 2.5% FBS and sorted as live, CD3-FITC$^+$, tetramer-PE$^+$ cells.

DNA library generation and sequencing

After sorting, the 10X Genomics Chromium Next GEM Single Cell 5' Reagent Kits v2 (Dual Index) with Feature Barcode Technology User Guide was followed exactly as written to generate DNA libraries (CG000330 Rev D). After verifying quality of the libraries, samples were submitted to the University of Chicago Genomics Facility for sequencing on the NovaSeq 6000. Paired end 26-10-10-90 bp runs with unique dual indices were executed, with samples pooled at a 1:10:1 ratio of the VDJ:GEX:ADT libraries.

Data processing

After sequencing, the GEX and ADT data were aligned and counted - and the V(D)J contigs were annotated - with 10X Genomics Cell Ranger 3.0.0. Using the Seurat package in R for analysis [199-202], the data sets were processed as follows: all TCR V, D, and J genes were removed from the gene expression matrix to prevent biasing GEX-based clustering due to the high expression of TRAV1-2 by MAIT cells (which were redundant since we sequenced the V(D)J library as well), then low-quality cells were removed based on the number of genes per cell and UMIs per cell (to remove doublet GEMs) and on the percent of expressed genes being of mitochondrial and ribosomal nature (to remove dead and dying cells) (**Supplemental Figure S4.2b**). Each GEX data set was normalized and scaled separately using SCTransform v2 [203,

204] and each ADT assay was normalized using a centered-log ratio method (margin = 1). The two data sets were merged, integrated using Harmony [205], and the V(D)J sequences were appended to the metadata using scRepertoire [206]. T cells were subset by removing cells showing high expression of LYZ and high binding by the IgG1κ antibody (from the ADT panel); finally, the cells were reclustered (**Supplemental Figure S4.2c**); the total population of analyzed cells are shown in **Figure 4.2**. Cell subsetting for relevant analyses is described in **Chapter 4**. Figures were generated using the Seurat, scRepertoire, and cowplot packages in R.

Bibliography

1. Janeway, C.A., Jr., The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol Today, 1992. **13**(1): p. 11-6.

2. Gonzalez, S., et al., Conceptual aspects of self and nonself discrimination. Self Nonself, 2011. **2**(1): p. 19-25.

3. Schluns, K.S. and L. Lefrancois, Cytokine control of memory T-cell development and survival. Nat Rev Immunol, 2003. **3**(4): p. 269-79.

4. Jarjour, N.N., D. Masopust, and S.C. Jameson, *T Cell Memory: Understanding COVID-19.* Immunity, 2021. **54**(1): p. 14-18.

5. Adams, E.J. and A.M. Luoma, The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol, 2013. **31**: p. 529-61.

6. Bennett, M.S., J.L. Round, and D.T. Leung, Innate-like lymphocytes in intestinal infections. Curr Opin Infect Dis, 2015. **28**(5): p. 457-63.

7. Godfrey, D.I., et al., The biology and functional importance of MAIT cells. Nat Immunol, 2019. **20**(9): p. 1110-1128.

8. Eckle, S.B., et al., A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. J Exp Med, 2014. **211**(8): p. 1585-600.

9. Huppa, J.B. and M.M. Davis, T-cell-antigen recognition and the immunological synapse. Nat Rev Immunol, 2003. **3**(12): p. 973-83.

10. Nakayama, M., Antigen Presentation by MHC-Dressed Cells. Front Immunol, 2014. **5**: p. 672.

11. Gaudino, S.J. and P. Kumar, Cross-Talk Between Antigen Presenting Cells and T Cells Impacts Intestinal Homeostasis, Bacterial Infections, and Tumorigenesis. Front Immunol, 2019. **10**: p. 360.

12. Flajnik, M.F., et al., Insight into the primordial MHC from studies in ectothermic vertebrates. Immunol Rev, 1999. **167**: p. 59-67.

13. Bjorkman, P.J., et al., Structure of the human class I histocompatibility antigen, HLA-A2. Nature, 1987. **329**(6139): p. 506-12.

14. Garcia, K.C., et al., An αβ T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science, 1996. **274**(5285): p. 209-19.

15. Birnbaum, M.E., et al., Deconstructing the peptide-MHC specificity of T cell recognition. Cell, 2014. **157**(5): p. 1073-87.

16. Germain, R.N., MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. Cell, 1994. **76**(2): p. 287-99.

17. Hewitt, E.W., The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology, 2003. **110**(2): p. 163-9.

18. Blum, J.S., P.A. Wearsch, and P. Cresswell, *Pathways of antigen processing.* Annu Rev Immunol, 2013. **31**: p. 443-73.

19. Vyas, J.M., A.G. Van der Veen, and H.L. Ploegh, The known unknowns of antigen processing and presentation. Nat Rev Immunol, 2008. **8**(8): p. 607-18.

20. Janeway, C.A., Jr., The T cell receptor as a multicomponent signalling machine: CD4/ CD8 coreceptors and CD45 in T cell activation. Annu Rev Immunol, 1992. **10**: p. 645-74.

21. Li, Y., Y. Yin, and R.A. Mariuzza, Structural and biophysical insights into the role of CD4 and CD8 in T cell activation. Front Immunol, 2013. **4**: p. 206.

22. Persson, G., et al., HLA class Ib in pregnancy and pregnancy-related disorders. Immunogenetics, 2017. **69**(8-9): p. 581-595.

23. Djaoud, Z. and P. Parham, HLAs, TCRs, and KIRs, a Triumvirate of Human Cell-Mediated Immunity. Annu Rev Biochem, 2020. **89**: p. 717-739.

24. Barker, D.J., et al., *The IPD-IMGT/HLA Database.* Nucleic Acids Res, 2023. **51**(D1): p. D1053-D1060.

25. Bjorkman, P.J., et al., The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature, 1987. **329**(6139): p. 512-8.

26. Oyarzun, P., et al., PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. BMC Bioinformatics, 2013. **14**: p. 52.

27. Brown, J.H., et al., Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature, 1993. **364**(6432): p. 33-9.

28. Garcia, K.C. and E.J. Adams, How the T cell receptor sees antigen--a structural view. Cell, 2005. **122**(3): p. 333-6.

29. Allison, J.P., B.W. McIntyre, and D. Bloch, Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. J Immunol, 1982. **129**(5): p. 2293-300.

30. Hedrick, S.M., et al., Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature, 1984. **308**(5955): p. 149-53.

31. Yanagi, S., et al., Evaluation of a histologic classification of mouse liver tumors based on pyruvate kinase isozymes and status of host lipids. J Natl Cancer Inst, 1984. **73**(6): p. 1311-7.

32. Lythe, G., et al., How many TCR clonotypes does a body maintain? J Theor Biol, 2016. **389**: p. 214-24.

33. Alt, F.W., et al., *VDJ recombination.* Immunol Today, 1992. **13**(8): p. 306-14.

34. Bassing, C.H., W. Swat, and F.W. Alt, The mechanism and regulation of chromosomal V(D)J recombination. Cell, 2002. **109 Suppl**: p. S45-55.

35. Giudicelli, V., X. Brochet, and M.P. Lefranc, IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb. Protoc., 2011. **2011**.

36. Davis, M.M. and P.J. Bjorkman, T-cell antigen receptor genes and T-cell recognition. Nature, 1988. **334**(6181): p. 395-402.

37. Gellert, M., V(D)J recombination: RAG proteins, repair factors, and regulation. Annu Rev Biochem, 2002. **71**: p. 101-32.

38. Matthews, A.G. and M.A. Oettinger, *RAG: a recombinase diversified.* Nat Immunol, 2009. **10**(8): p. 817-21.

39. Malissen, M., et al., Mouse T cell antigen receptor: structure and organization of constant and joining gene segments encoding the beta polypeptide. Cell, 1984. **37**(3): p. 1101-10.

40. Davis, M.M. and Y. Chien, Topology and affinity of T-cell receptor mediated recognition of peptide-MHC complexes. Curr Opin Immunol, 1993. **5**(1): p. 45-9.

41. Katayama, C.D., et al., Predicted complementarity determining regions of the T cell antigen receptor determine antigen specificity. EMBO J, 1995. **14**(5): p. 927-38.

42. Hennecke, J. and D.C. Wiley, T cell receptor-MHC interactions up close. Cell, 2001. **104**(1): p. 1-4.

43. Rudolph, M.G., R.L. Stanfield, and I.A. Wilson, *How TCRs bind MHCs, peptides, and coreceptors.* Annu Rev Immunol, 2006. **24**: p. 419-66.

44. Smith-Garvin, J.E., G.A. Koretzky, and M.S. Jordan, *T cell activation.* Annu Rev Immunol, 2009. **27**: p. 591-619.

45. Barry, M. and R.C. Bleackley, Cytotoxic T lymphocytes: all roads lead to death. Nat Rev Immunol, 2002. **2**(6): p. 401-9.

46. Kortekaas Krohn, I., et al., T-cell subsets in the skin and their role in inflammatory skin disorders. Allergy, 2022. **77**(3): p. 827-842.

47. Reinherz, E.L., et al., Separation of functional subsets of human T cells by a monoclonal antibody. Proc Natl Acad Sci U S A, 1979. **76**(8): p. 4061-5.

48. Broere, F. and W. van Eden, *T Cell Subsets and T Cell-Mediated Immunity*, in *Nijkamp and Parnham's Principles of Immunopharmacology*, M.J. Parnham, F.P. Nijkamp, and A.G. Rossi, Editors. 2019, Springer International Publishing: Cham. p. 23-35.

49. Germain, R.N., T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol, 2002. **2**(5): p. 309-22.

50. Parel, Y. and C. Chizzolini, CD4+ CD8+ double positive (DP) T cells in health and disease. Autoimmun Rev, 2004. **3**(3): p. 215-20.

51. Velikkakam, T., K.J. Gollob, and W.O. Dutra, Double-negative T cells: Setting the stage for disease control or progression. Immunology, 2022. **165**(4): p. 371-385.

52. Wu, Z., et al., CD3(+)CD4(-)CD8(-) (Double-Negative) T Cells in Inflammation, Immune Disorders and Cancer. Front Immunol, 2022. **13**: p. 816005.

53. Layre, E., A. de Jong, and D.B. Moody, Human T cells use CD1 and MR1 to recognize lipids and small molecules. Curr Opin Chem Biol, 2014. **23**: p. 31-8.

54. Mori, L., M. Lepore, and G. De Libero, The Immunology of CD1- and MR1-Restricted T Cells. Annu Rev Immunol, 2016. **34**: p. 479-510.

55. Reinink, P. and I. Van Rhijn, *Mammalian CD1 and MR1 genes.* Immunogenetics, 2016. **68**(8): p. 515-23.

56. Bendelac, A., et al., CD1 recognition by mouse NK1+ T lymphocytes. Science, 1995. **268**(5212): p. 863-5.

57. Kjer-Nielsen, L., et al., MR1 presents microbial vitamin B metabolites to MAIT cells. Nature, 2012. **491**(7426): p. 717-23.

58. Corbett, A.J., et al., T-cell activation by transitory neo-antigens derived from distinct microbial pathways. Nature, 2014. **509**(7500): p. 361-5.

59. Hashimoto, K., M. Hirai, and Y. Kurosawa, A gene outside the human MHC related to classical HLA class I genes. Science, 1995. **269**(5224): p. 693-5.

60. Riegert, P., V. Wanner, and S. Bahram, Genomics, isoforms, expression, and phylogeny of the MHC class I-related MR1 gene. J Immunol, 1998. **161**(8): p. 4066-77.

61. Boudinot, P., et al., Restricting nonclassical MHC genes coevolve with TRAV genes used by innate-like T cells in mammals. Proc Natl Acad Sci U S A, 2016. **113**(21): p. E2983-92.

62. Huang, S., et al., MR1 antigen presentation to mucosal-associated invariant T cells was highly conserved in evolution. Proc Natl Acad Sci U S A, 2009. **106**(20): p. 8290-5.

63. Lopez-Sagaseta, J., et al., The molecular basis for Mucosal-Associated Invariant T cell recognition of MR1 proteins. Proc Natl Acad Sci U S A, 2013. **110**(19): p. E1771-8.

64. Bacher, A., et al., Biosynthesis of vitamin b2 (riboflavin). Annu Rev Nutr, 2000. **20**: p. 153-67.

65. Mansoorabadi, S.O., C.J. Thibodeaux, and H.W. Liu, The diverse roles of flavin coenzymes--nature's most versatile thespians. J Org Chem, 2007. **72**(17): p. 6329-42.

66. Mosegaard, S., et al., Riboflavin Deficiency-Implications for General Human Health and Inborn Errors of Metabolism. Int J Mol Sci, 2020. **21**(11).

67. Baba, T., et al., Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol, 2006. **2**: p. 2006 0008.

68. Mondot, S., P. Boudinot, and O. Lantz, MAIT, MR1, microbes and riboflavin: a paradigm for the co-evolution of invariant TCRs and restricting MHCI-like molecules? Immunogenetics, 2016. **68**(8): p. 537-48.

69. Garcia-Angulo, V.A., Overlapping riboflavin supply pathways in bacteria. Crit Rev Microbiol, 2017. **43**(2): p. 196-209.

70. Meermeier, E.W., et al., Human TRAV1-2-negative MR1-restricted T cells detect S. pyogenes and alternatives to MAIT riboflavin-based antigens. Nat Commun, 2016. **7**: p. 12506.

71. Averianova, L.A., et al., Production of Vitamin B2 (Riboflavin) by Microorganisms: An Overview. Front Bioeng Biotechnol, 2020. **8**: p. 570828.

72. Sa, N., et al., Identification and characterization of the missing phosphatase on the riboflavin biosynthesis pathway in Arabidopsis thaliana. Plant J, 2016. **88**(5): p. 705-716.

73. Haase, I., et al., Enzymes from the haloacid dehalogenase (HAD) superfamily catalyse the elusive dephosphorylation step of riboflavin biosynthesis. Chembiochem, 2013. **14**(17): p. 2272-5.

74. Mak, J.Y., et al., Stabilizing short-lived Schiff base derivatives of 5-aminouracils that

activate mucosal-associated invariant T cells. Nat Commun, 2017. **8**: p. 14599.

75. Awad, W., et al., The molecular basis underpinning the potency and specificity of MAIT cell antigens. Nat Immunol, 2020. **21**(4): p. 400-411.

76. Reantragoon, R., et al., Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. J Exp Med, 2013. **210**(11): p. 2305-20.

77. Harriff, M.J., et al., MR1 displays the microbial metabolome driving selective MR1-restricted T cell receptor usage. Sci Immunol, 2018. **3**(25).

78. Gherardin, N.A., et al., Diversity of T Cells Restricted by the MHC Class I-Related Molecule MR1 Facilitates Differential Antigen Recognition. Immunity, 2016. **44**(1): p. 32-45.

79. Keller, A.N., et al., Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. Nat Immunol, 2017. **18**(4): p. 402-411.

80. De Libero, G., A. Chancellor, and L. Mori, Antigen specificities and functional properties of MR1-restricted T cells. Mol Immunol, 2021. **130**: p. 148-153.

81. Lepore, M., et al., Functionally diverse human T cells recognize non-microbial antigens presented by MR1. Elife, 2017. **6**.

82. Crowther, M.D., et al., Genome-wide CRISPR-Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. Nat Immunol, 2020. **21**(2): p. 178-185.

83. Vacchini, A., et al., Nucleobase adduct-containing metabolites are MR1 ligands that stimulate self-reactive MR1T cells. bioRxiv, 2022: p. 2022.08.26.505411.

84. Porcelli, S., et al., Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4-8- alpha/beta T cells demonstrates preferential use of several V beta genes and an invariant TCR alpha chain. J Exp Med, 1993. **178**(1): p. 1-16.

85. Bendelac, A., P.B. Savage, and L. Teyton, *The biology of NKT cells*. Annu Rev Immunol, 2007. **25**: p. 297-336.

86. Krovi, S.H. and L. Gapin, Invariant Natural Killer T Cell Subsets-More Than Just Developmental Intermediates. Front Immunol, 2018. **9**: p. 1393.

87. Treiner, E., et al., Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature, 2003. **422**(6928): p. 164-9.

88. Tilloy, F., et al., An invariant T cell receptor alpha chain defines a novel TAP-independent major histocompatibility complex class Ib-restricted alpha/beta T cell subpopulation in

mammals. J Exp Med, 1999. **189**(12): p. 1907-21.

89. Lopez-Sagaseta, J., et al., MAIT recognition of a stimulatory bacterial antigen bound to MR1. J Immunol, 2013. **191**(10): p. 5268-77.

90. Patel, O., et al., Recognition of vitamin B metabolites by mucosal-associated invariant T cells. Nat Commun, 2013. **4**: p. 2142.

91. Le Bourhis, L., et al., Antimicrobial activity of mucosal-associated invariant T cells. Nat Immunol, 2010. **11**(8): p. 701-8.

92. Dusseaux, M., et al., Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161hi IL-17-secreting T cells. Blood, 2011. **117**(4): p. 1250-9.

93. Kreslavsky, T., et al., TCR-inducible PLZF transcription factor required for innate phenotype of a subset of gammadelta T cells with restricted TCR diversity. Proc Natl Acad Sci U S A, 2009. **106**(30): p. 12453-8.

94. Koay, H.F., et al., A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. Nat Immunol, 2016. **17**(11): p. 1300-1311.

95. van Wilgenburg, B., et al., MAIT cells are activated during human viral infections. Nat Commun, 2016. **7**: p. 11653.

96. Gherardin, N.A., et al., Human blood MAIT cell subsets defined using MR1 tetramers. Immunol Cell Biol, 2018. **96**(5): p. 507-525.

97. Gibbs, A., et al., MAIT cells reside in the female genital mucosa and are biased towards IL-17 and IL-22 production in response to bacterial stimulation. Mucosal Immunol, 2017. **10**(1): p. 35-45.

98. Souter, M.N.T., et al., CD8 coreceptor engagement of MR1 enhances antigen responsiveness by human MAIT and other MR1-reactive T cells. J Exp Med, 2022. **219**(9).

99. Huang, S., et al., Evidence for MR1 antigen presentation to mucosal-associated invariant T cells. J Biol Chem, 2005. **280**(22): p. 21183-93.

100. Gold, M.C., et al., MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. J Exp Med, 2014. **211**(8): p. 1601-10.

101. Global tuberculosis report 2022. 2022: Geneva.

102. Seung, K.J., S. Keshavjee, and M.L. Rich, Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. Cold Spring Harb Perspect Med, 2015. **5**(9): p. a017863.

103. Chakaya, J., et al., Global Tuberculosis Report 2020 - Reflections on the Global TB burden, treatment and prevention efforts. Int J Infect Dis, 2021. **113 Suppl 1**(Suppl 1): p. S7-S12.

104. Cho, T., et al., A review of the BCG vaccine and other approaches toward tuberculosis eradication. Hum Vaccin Immunother, 2021. **17**(8): p. 2454-2470.

105. Scriba, T.J., M.G. Netea, and A.M. Ginsberg, Key recent advances in TB vaccine development and understanding of protective immune responses against Mycobacterium tuberculosis. Semin Immunol, 2020. **50**: p. 101431.

106. Guo, T., K. Chamoto, and N. Hirano, Adoptive T cell therapy targeting CD1 and MR1. Front. Immunol., 2015. **6**.

107. Van Rhijn, I. and D.B. Moody, Donor Unrestricted T Cells: A Shared Human T Cell Response. J Immunol, 2015. **195**(5): p. 1927-32.

108. Seshadri, C., et al., A polymorphism in human MR1 is associated with mRNA expression and susceptibility to tuberculosis. Genes Immun, 2017. **18**(1): p. 8-14.

109. Gold, M.C., et al., Human mucosal associated invariant T cells detect bacterially infected cells. PLoS Biol, 2010. **8**(6): p. e1000407.

110. Malka-Ruimy, C., et al., Mucosal-Associated Invariant T Cell Levels Are Reduced in the Peripheral Blood and Lungs of Children With Active Pulmonary Tuberculosis. Front Immunol, 2019. **10**: p. 206.

111. Greene, J.M., et al., MR1-restricted mucosal-associated invariant T (MAIT) cells respond to mycobacterial vaccination and infection in nonhuman primates. Mucosal Immunol, 2017. **10**(3): p. 802-813.

112. Darrah, P.A., et al., Prevention of tuberculosis in macaques after intravenous BCG immunization. Nature, 2020. **577**(7788): p. 95-102.

113. Gela, A., et al., Effects of BCG vaccination on donor unrestricted T cells in two prospective cohort studies. EBioMedicine, 2022. **76**: p. 103839.

114. Hedrick, P.W., T.S. Whittam, and P. Parham, Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. Proc Natl Acad Sci U S A, 1991. **88**(13): p. 5897-901.

115. Creech, A.L., et al., The Role of Mass Spectrometry and Proteogenomics in the Advancement of HLA Epitope Prediction. Proteomics, 2018. **18**(12): p. e1700259.

116. Seidler, J., et al., De novo sequencing of peptides by MS/MS. Proteomics, 2010. **10**(4): p. 634-49.

117. Dancik, V., et al., De novo peptide sequencing via tandem mass spectrometry. J Comput Biol, 1999. **6**(3-4): p. 327-42.

118. Bocker, S. and F. Rasche, Towards de novo identification of metabolites by analyzing tandem mass spectra. Bioinformatics, 2008. **24**(16): p. i49-i55.

119. Dunn, W.B., et al., The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. Bioanalysis, 2012. **4**(18): p. 2249-64.

120. Hu, Y.B., et al., The endosomal-lysosomal system: from acidification and cargo sorting to neurodegeneration. Transl Neurodegener, 2015. **4**: p. 18.

121. Kjer-Nielsen, L., et al., An overview on the identification of MAIT cell antigens. Immunol Cell Biol, 2018. **96**(6): p. 573-587.

122. Ladd, N.A., C.A. Froyd, and E.J. Adams, *Production of MR1 Tetramers Loaded with Microbial Ligands*, in *MAIT Cells: Methods and Protocols*, H. Kaipe and I. Magalhaes, Editors. 2020, Springer US: New York, NY. p. 191-207.

123. Dukkipati, A., et al., BacMam system for high-level expression of recombinant soluble and membrane glycoproteins for structural studies. Protein Expr Purif, 2008. **62**(2): p. 160-70.

124. Illarionov, B., et al., Rapid preparation of isotopolog libraries by in vivo transformation of 1)C-glucose. Studies on 6,7-dimethyl-8-ribityllumazine, a biosynthetic precursor of vitamin B2. J Org Chem, 2004. **69**(17): p. 5588-94.

125. McNutt, W.S. and I. Takeda, A naturally occurring indolylpteridine. Biochemistry, 1969. **8**(4): p. 1370-6.

126. Wang, M., et al., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol, 2016. **34**(8): p. 828-837.

127. Aron, A.T., et al., Reproducible molecular networking of untargeted mass spectrometry data using GNPS. Nat Protoc, 2020. **15**(6): p. 1954-1991.

128. Long, Q., et al., Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets. Chem Biol Drug Des, 2010. **75**(4): p. 339-47.

129. Ashoori, M. and A. Saedisomeolia, Riboflavin (vitamin B_2) and oxidative stress: a review. Br J Nutr, 2014. **111**(11): p. 1985-91.

130. Jabeen, H., et al., Investigating the scavenging of reactive oxygen species by antioxidants

via theoretical and experimental methods. J Photochem Photobiol B, 2018. **180**: p. 268-275.

131. Khan, S., et al., ROS mediated antibacterial activity of photoilluminated riboflavin: A photodynamic mechanism against nosocomial infections. Toxicol Rep, 2019. **6**: p. 136-142.

132. Reddy, T.B., et al., TB database: an integrated platform for tuberculosis research. Nucleic Acids Res, 2009. **37**(Database issue): p. D499-508.

133. Pedrolli, D., et al., The ribB FMN riboswitch from Escherichia coli operates at the transcriptional and translational level and regulates riboflavin biosynthesis. FEBS J, 2015. **282**(16): p. 3230-42.

134. Vikram, et al., Riboswitch-mediated regulation of riboflavin biosynthesis genes in prokaryotes. 3 Biotech, 2022. **12**(10): p. 278.

135. Matsuoka, T., et al., The effects of 5-OP-RU stereochemistry on its stability and MAIT-MR1 axis. Chembiochem, 2021. **22**(4): p. 672-678.

136. Bellmaine, S., A. Schnellbaecher, and A. Zimmer, Reactivity and degradation products of tryptophan in solution and proteins. Free Radic Biol Med, 2020. **160**: p. 696-718.

137. Iten, P.X., et al., Isolierung und Struktur von Pteridinen (Lumazinen) aus Russula sp. (Täublinge; Basidiomycetes). Helvetica Chimica Acta, 1984. **67**(2): p. 550-569.

138. Märki-Danzig, H. and C.H. Eugster, *Synthesen von Aminolumazinen.* Helvetica Chimica Acta, 1984. **67**(2): p. 570-573.

139. Daniels, B.J., et al., Naturally Occurring Lumazines. J Nat Prod, 2019. **82**(7): p. 2054-2065.

140. Harriff, M.J., et al., Endosomal MR1 Trafficking Plays a Key Role in Presentation of Mycobacterium tuberculosis Ligands to MAIT Cells. PLoS Pathog, 2016. **12**(3): p. e1005524.

141. McWilliam, H.E., et al., The intracellular pathway for the presentation of vitamin B-related antigens by the antigen-presenting molecule MR1. Nat Immunol, 2016. **17**(5): p. 531-7.

142. Hümbelin, M., et al., GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone 4-phosphate synthase are rate-limiting enzymes in riboflavin synthesis of an industrial Bacillus subtilis strain used for riboflavin production. Journal of Industrial Microbiology and Biotechnology, 1999. **22**(1): p. 1-7.

143. Plaut, G.W. and G.F. Maley, The isolation of 6-methyl-7-hydroxy-8-ribityllumazine from Ashbya gossypii and further evidence for its chemical structure. J Biol Chem, 1959. **234**: p. 3010-3.

144. Klein, G. and A. Bacher, Biosynthesis of Riboflavin. Enzymatic Deamination of 2,5-Diamino-6-hydroxy-4-ribitylaminopyrimidine-5′-phosphate. Zeitschrift für Naturforschung B, 1980. **35**(4): p. 482-484.

145. Bacher, A., et al., Biosynthesis of riboflavin. Incorporation of 13C-labeled precursors into the xylene ring. J Biol Chem, 1983. **258**(22): p. 13431-7.

146. Akira, S. and G. Miki, Isolation and Characterization of Pteridines from Pseudomonas ovalis. Bulletin of the Chemical Society of Japan, 1971. **44**(7): p. 1869-1872.

147. Akira, S., M. Toshiaki, and G. Miki, The Structure of a New Pteridine Compound Produced by Pseudomonas ovalis. Bulletin of the Chemical Society of Japan, 1972. **45**(7): p. 2198-2199.

148. Li, X., et al., New Insights Into Gut-Bacteria-Derived Indole and Its Derivatives in Intestinal and Liver Diseases. Front Pharmacol, 2021. **12**: p. 769501.

149. Sugimori, D., et al., Microbial hydroxylation of indole to 7-hydroxyindole by Acinetobacter calcoaceticus strain 4-1-5. Biosci Biotechnol Biochem, 2004. **68**(5): p. 1167-9.

150. Ma, Q., X. Zhang, and Y. Qu, Biodegradation and Biotransformation of Indole: Advances and Perspectives. Front Microbiol, 2018. **9**: p. 2625.

151. Braganza, C.D., et al., The effect of MR1 ligand glyco-analogues on mucosal-associated invariant T (MAIT) cell activation. Org Biomol Chem, 2019. **17**(40): p. 8992-9000.

152. Corbett, A.J., et al., Antigen Recognition by MR1-Reactive T Cells; MAIT Cells, Metabolites, and Remaining Mysteries. Front Immunol, 2020. **11**: p. 1961.

153. Mukherji, B., et al., Induction of antigen-specific cytolytic T cells in situ in human melanoma by immunization with synthetic peptide-pulsed autologous antigen presenting cells. Proc Natl Acad Sci U S A, 1995. **92**(17): p. 8078-82.

154. Butz, E.A. and M.J. Bevan, Massive expansion of antigen-specific CD8+ T cells during an acute virus infection. Immunity, 1998. **8**(2): p. 167-75.

155. Veerapen, N., et al., Chemical insights into the search for MAIT cells activators. Mol Immunol, 2021. **129**: p. 114-120.

156. Ladenstein, R., M. Fischer, and A. Bacher, The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. FEBS J, 2013. **280**(11): p. 2537-63.

157. Al-Hassan, S.S., Specific enzyme inhibitors in vitamin biosynthesis. Part 3. The synthesis

and inhibitory properties of some substrates and transition state analogues of riboflavin synthase. J. Chem. Soc. Perkin Trans., 1980. **1**.

158. Narayanan, G.A., et al., Alternative splicing of MR1 regulates antigen presentation to MAIT cells. Sci Rep, 2020. **10**(1): p. 15429.

159. Trott, O. and A.J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem., 2010. **31**.

160. Jin, H., et al., Deaza-modification of MR1 ligands modulates recognition by MR1-restricted T cells. Sci Rep, 2022. **12**(1): p. 22539.

161. Hollingsworth, S.A. and R.O. Dror, *Molecular Dynamics Simulation for All.* Neuron, 2018. **99**(6): p. 1129-1143.

162. Braun, E., et al., Best Practices for Foundations in Molecular Simulations [Article v1.0]. Living J Comput Mol Sci, 2019. **1**(1).

163. Baker, B.M., et al., Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism. Immunol Rev, 2012. **250**(1): p. 10-31.

164. Wang, H., et al., MAIT cells protect against pulmonary Legionella longbeachae infection. Nat Commun, 2018. **9**(1): p. 3350.

165. Wang, H., et al., IL-23 costimulates antigen-specific MAIT cell activation and enables vaccination against bacterial infection. Sci Immunol, 2019. **4**(41).

166. Zhao, Z., et al., Francisella tularensis induces Th1 like MAIT cells conferring protection against systemic and local infection. Nat Commun, 2021. **12**(1): p. 4355.

167. Yu, H., et al., Artificially induced MAIT cells inhibit M. bovis BCG but not M. tuberculosis during in vivo pulmonary infection. Sci Rep, 2020. **10**(1): p. 13579.

168. Sakai, S., et al., Functional inactivation of pulmonary MAIT cells following 5-OP-RU treatment of non-human primates. Mucosal Immunol, 2021. **14**(5): p. 1055-1066.

169. Legoux, F., et al., Microbial metabolites control the thymic development of mucosal-associated invariant T cells. Science, 2019. **366**(6464): p. 494-499.

170. Ma, K.-Y., et al., High-throughput and high-dimensional single-cell analysis of antigen-specific CD8+ T cells. Nature Immunology, 2021. **22**(12): p. 1590-1598.

171. Garner, L.C., et al., Human MAIT cells show clonal diversity but transcriptional and functional homogeneity. bioRxiv, 2022: p. 2022.02.26.482031.

172. Leeansyah, E., et al., Activation, exhaustion, and persistent decline of the antimicrobial MR1-restricted MAIT-cell population in chronic HIV-1 infection. Blood, 2013. **121**(7): p. 1124-35.

173. Sobkowiak, M.J., et al., Tissue-resident MAIT cell populations in human oral mucosa exhibit an activated profile and produce IL-17. Eur J Immunol, 2019. **49**(1): p. 133-143.

174. Suliman, S., et al., Dual TCR-alpha Expression on Mucosal-Associated Invariant T Cells as a Potential Confounder of TCR Interpretation. J Immunol, 2022. **208**(6): p. 1389-1395.

175. Martin, E., et al., Stepwise development of MAIT cells in mouse and human. PLoS Biol, 2009. **7**(3): p. e54.

176. Boulter, J.M., et al., Stable, soluble T-cell receptor molecules for crystallization and therapeutics. Protein Eng, 2003. **16**(9): p. 707-11.

177. Chang, H.C., et al., A general method for facilitating heterodimeric pairing between two proteins: application to expression of alpha and beta T-cell receptor extracellular segments. Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11408-12.

178. *Baculovirus and Insect Cell Expression Protocols*, D.W. Murhammer, Editor. 2007, Humana Totowa: New Jersey, USA. p. 470.

179. MacLean, B., et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics, 2010. **26**(7): p. 966-8.

180. Peterson, K.E., et al., Alcohol-, diol-, and carbohydrate-substituted indenoisoquinolines as topoisomerase I inhibitors: investigating the relationships involving stereochemistry, hydrogen bonding, and biological activity. J Med Chem, 2011. **54**(14): p. 4937-53.

181. Dangerfield, E.M., et al., Protecting-group-free synthesis of amines: synthesis of primary amines from aldehydes via reductive amination. J Org Chem, 2010. **75**(16): p. 5470-7.

182. Kuwano, R., et al., Catalytic asymmetric hydrogenation of 2,3,5-trisubstituted pyrroles. J Am Chem Soc, 2008. **130**(3): p. 808-9.

183. Lewinsohn, D.M., et al., Mycobacterium tuberculosis-reactive CD8+ T lymphocytes: the relative contribution of classical versus nonclassical HLA restriction. J Immunol, 2000. **165**(2): p. 925-30.

184. Harriff, M.J., et al., Human lung epithelial cells contain Mycobacterium tuberculosis in a late endosomal vacuole and are efficiently recognized by CD8(+) T cells. PLoS One, 2014. **9**(5): p. e97515.

185. BIOVIA Discovery Studio, in Release 2019. 2020: Dassault Systemes.

186. Morris, G.M., et al., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.

187. Morris, G.M., et al., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem, 2009. **30**(16): p. 2785-91.

188. Jo, S., et al., CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem, 2008. **29**(11): p. 1859-65.

189. Lee, J., et al., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput, 2016. **12**(1): p. 405-13.

190. Lee, J., et al., CHARMM-GUI supports the Amber force fields. J Chem Phys, 2020. **153**(3): p. 035103.

191. Vanommeslaeghe, K., et al., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem, 2010. **31**(4): p. 671-90.

192. Hoover, W.G., Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A Gen Phys, 1985. **31**(3): p. 1695-1697.

193. Nosé, S. and M.L. Klein, A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. The Journal of Chemical Physics, 1983. **78**(11): p. 6928-6939.

194. Darden, T.A. and L.G. Pedersen, Molecular modeling: an experimental tool. Environ Health Perspect, 1993. **101**(5): p. 410-2.

195. Ryckaert, J.P., G. Ciccotti, and H.J. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys., 1977. **23**.

196. Grant, B.J., et al., Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics, 2006. **22**(21): p. 2695-6.

197. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

198. Wickham, H., *ggplot2*, in *Elegant Graphics for Data Analysis*. 2016, Springer-Verlag:

New York.

199. Satija, R., et al., Spatial reconstruction of single-cell gene expression data. Nat Biotechnol, 2015. **33**(5): p. 495-502.

200. Butler, A., et al., Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol, 2018. **36**(5): p. 411-420.

201. Stuart, T., et al., Comprehensive Integration of Single-Cell Data. Cell, 2019. **177**(7): p. 1888-1902 e21.

202. Hao, Y., et al., Integrated analysis of multimodal single-cell data. Cell, 2021. **184**(13): p. 3573-3587 e29.

203. Hafemeister, C. and R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol, 2019. **20**(1): p. 296.

204. Choudhary, S. and R. Satija, Comparison and evaluation of statistical error models for scRNA-seq. Genome Biol, 2022. **23**(1): p. 27.

205. Korsunsky, I., et al., Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods, 2019. **16**(12): p. 1289-1296.

206. Borcherding, N., N.L. Bormann, and G. Kraus, scRepertoire: An R-based toolkit for single-cell immune receptor analysis. F1000Res, 2020. **9**: p. 47.