THE UNIVERSITY OF CHICAGO

MOLECULAR INSIGHTS INTO UNCONVENTIONAL IMMUNE RECOGNITION: A CASE OF COMMENSAL SUPERANTIGENS, POLYREACTIVE ANTIBODIES AND BUTYROPHILIN SIGNALING

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

MARTA TERESA BOROWSKA

CHICAGO, ILLINOIS

DECEMBER 2020

To my parents and sister for always having my back.

Table of Contents

List of Figures

List of Tables

Acknowledgment

It would not have been possible to complete this doctoral thesis without the help and support of all the kind people I encountered at the University of Chicago, and whose entire contribution is impossible to list in this format.

Above all, I want to thank my advisor, Dr. Erin Adams, for many lessons on science and life. Thank you for your guidance, patience and doing everything you can to turn my graduate work into success. I grew impressed how Erin mastered the balance of keeping students grounded but adventurous and feeling supported but also challenged.

Erin has also assembled a tight and diverse community of young scientists who are involved in each other's projects with positivity and support. Specifically, I want to thank Dr. Caitlin Castro, Dr. Kristof Nolan, Dr. Sobhan Roy and Christopher Boughter for abundant discussions, bright ideas and a full array of skillsets I learned from them. All current and past members of Adams Lab and Keenan Lab also played an important role in supporting me through challenging times.

I always looked forward to my biannual committee meetings thanks to my dream thesis committee members: Drs. Margaret Gardel, Bana Jabri, and Engin Özkan. Their theoretical and technical input was tremendous and often groundbreaking for my projects. Margaret, with her fervent and glowing attitude, would supply a unique perspective influenced by her biophysics background. Bana, with her piercing knowledge of immunology that she would calmly boil down and translate onto my project. Engin, for his biochemistry perspective and also for being my second mentor, who would lay it all out in a good ol' European way. Additionally, I want to acknowledge Dr. Albert Bendelac who was always a helping mind to look after this research.

A special thank you to all collaborators who expanded this research with their expertise. Members of Bendelac Lab: Dr. Jeffrey Bunker whose research planted the seeds for the superantigen and polyreactivity projects and Dr. Christoph Drees who greatly complemented the superantigen project. Dr. Alex Yarawsky and Dr. Andrew Herr from Cincinnati Children's Hospital Medical Center for their Ibp stoichiometry measurements with analytical ultracentrifugal studies. Jason Krawic and Dr. Curtis McMurtrey from Oklahoma University Health Sciences Center, for their expansive expertise in mass spectrometry that helped to determine complexes in both superantigen and polyreactivity projects. Our local collaborators, Dr. Man Pan, Jane Lodwick and Dr. Minglei Zhao for their hard work and help with cryo-EM structure elucidation of butyrophilin-3. Jeffrey Tarrasch and Dr. Georgios Skiniotis from Stanford University who did initial negative stain-EM work on butyrophilin-3 in nanodiscs.

Additionally, Dr. Tian Li, Dr. Ahmed Rohaim, Dr. Peter Chung, Dr. Somnath Mukherjee, each contributed amazing ideas and altruistic assistance with some experiments in this research work. And thank you to all my companions on trips to Argonne including members of Perozo, Kossiakoff, Ara*ç*, Özkan and Rice Labs who made the whole experience fun and fruitful.

Finally, I've learned and acquired many important techniques during my graduate career, thanks to the guidance and advice from directors of our core facilities in the Biological Sciences Division. I also wanted to thank scientists at 23ID and 24ID beamlines at the Advanced Proton Source (APS) and SIBYLS HT-SAXS at the Advanced Light Source (ALS) for their help and advice in collecting and solving diffraction datasets.

Abstract

Our body responds to self and foreign agents through a myriad collection of effector cells and molecules that together form the immune system. The innate response is broad and occurs rapidly, whereas the adaptive response is slower, and it relies on highly specialized immune recognition of specific pathogens. Adaptive immune cells, like T cell and B cell lymphocytes possess a vast repertoire of antigen receptors called T cell receptors (TCR) and B cell receptors (BCR), respectively, where BCRs are also found in a secreted form as antibodies (Ab). The delicate balance of immune recognition lies in details, where minute changes in amino acid composition affect structure and function of antigen receptors. In this work, I dissect molecular details of three unconventional systems of immune recognition. Specifically, I present my molecular dissection of a newly identified protein from commensal microbes that engages host antibodies and has structural features of known superantigens but does not contribute to a pathogenic response. This interaction may provide an important function in the maintenance of host/microbiome gut balance. Additionally, I characterize homeostatic IgA and antigen-experienced IgG antibodies that naturally bind multiple diverse targets with low affinity and are thus called polyreactive. This work attempts to find the molecular basis for polyreactivity in natural antibodies and to emphasize the importance of embracing positive polyreactivity in antibody-based drug or vaccine design. Finally, I present my work on identifying the molecular mechanism behind butyrophilin-3 - γδTCR-dependent antigen recognition. Contrary to the canonical view of extracellular antigen presentation and direct recognition by T cells, the BTN3-γδTCR system involves intracellular sensing of molecules during cell stress. All findings in this thesis not only deepen our understanding of the myriad ways our immune system works, but also expand on the classical definitions of antigen-antibody and antigen-receptor interactions.

Chapter 1 – Introduction

1.1 Host-commensal superantigen-like recognition

1.1.1 Pathogens are potent stimulators of the host immune system

The relationship between host and microbiota can be viewed as a spectrum between symbiosis, commensalism, and pathogenicity (Round et al. 2018, Hooper et al. 2001, Geva-Zatorsky et al. 2017). In certain diseases, some commensal microbiota can shift toward pathogenicity, thus eliciting broader immune activation. Since most microbial antigens are able to stimulate only a small fraction of lymphocytes, certain pathogenic organisms evolved molecules to hijack the host system by harvesting and re-directing T or B cell stimulatory potential of hosts against themselves. These molecules are called superantigens (SAgs); until recently this term was reserved for mostly pathogenic *Staphylococci* or *Streptococci* species and some viruses (Proft and Fraser, 2003). For T cell SAgs, their predominant mechanism of activation is through crosslinked binding to a germline region of a variable beta chain of the T cell receptors (TCRs) and outside of the peptide-binding groove of a class II major histocompatibility complex (MHC) (Choi et al. 1990, Dellabona et al. 1990). By bringing these two molecules together, T cell SAgs activate a vast number of T cells leading to release of a large and sudden storm of cytokines which causes acute toxic shock for the host (Miethke et al. 1992, Fast et al. 1989). Similarly, B cell SAgs, i.e. *Staphylococcal* Protein A (SpA) and *Streptoccoccal* Protein G, bind outside of the complementarity determining region (CDR) of variable heavy (VH) of an antigen binding Fragment (Fab) as well as antibody crystallizable Fragment (Fc) (Graille et al. 2000, Derrick et al. 1992). The ability of SAgs to bind to a large number of B cells or serum immunoglobulins (Ig) elicits vast activation of B cells and a production of excess of cytokines, resulting in a cytokine-mediated suppression or deletion of

activated lymphocytes. This inflammatory response leads to an array of diseases including food poisoning, toxic shock and scalded skin syndrome in humans and other mammals. Additional examples of superantigenic-like interactions have been uncovered, which expand on classical definition of conventional superantigens i.e. protein Fv (a human liver sialoprotein), protein L (a coat protein of *Peptostreptococcus magnus*), and *Staphylococcal* enterotoxin D (Patella et al. 1993, Domiati-Saad et al. 1998, Graille et al. 2001).

1.1.2 Molecular details of common B cell superantigens: Protein A/ G and L

Staphylococci and *Streptococci* are highly successful pathogens. These two species produce a myriad of cell surface and secreted virulence factors, among these being B cell superantigens. While many T cell superantigens are studied, only a handful of B cell superantigens are known, among them are Protein A, Protein G and Protein L and are described below.

Protein A (SpA) is a 42kDa protein from a common bacterial pathogen *Staphylococcus aureus* and contains five homologous repeat domains that together exist as soluble and surface expressed variants during virulence of *S. aureus* (Pauli et al. 2014, Kim et al. 2010) (Fig. 1.1A). *S. aureus* is one of the deadliest known pathogens as it frequently colonizes human nasopharynx and gastrointestinal tract which is a key factor for invasive disease (Krismer et al. 2017, van Belkum et al. 2009, Lowy 1998). SpA can bind two distinct areas on a single immunoglobulin with separate, non-competing sites: crystallizable fragment (Fc) and antigen binding fragment (Fab) (Boyle et al. 1990, Graille et al. 2001). These two binding sites lead to two different outcomes for the host: binding of SpA to Fc blocks the effector functions of that immunoglobulin, whereas binding to Fab is characterized with variable region specificity confined to human variable heavy

chain domain 3 (VH3), which does not compete with antigen binding (Young et al. 1984). VH3+ Igs, which represent 30-60% of human peripheral B cells are induced by SpA leading to a large supraclonal VH3+ targeted B-cells depletion by an activation-induced cell-death (AICD) pathway in mice (Goodyear et al. 2003, Silverman et al. 2000).

Structurally, Protein A is a three-helix bundle, using helix II and III to complex with Fab and helix I to complex with Fc (Fig. 1.1B, left). The interaction with Fab is mediated by 11 residues on helix II and III of SpA and 13 residues on framework regions of VH3 Fab. This interface maps outside of typical antigen binding complementarity determining region (CDR) and involves core residues of VH3: Arg/Lys19, Gly65, Arg66, Thr68, Ser70, Gln81 and Asn82a. A salt bridge and extensive electrostatic interactions are responsible for the germline preference as other nonbinding human and murine VH families usually contain two or more residue differences at positions identified as SpA contacts. This Fab-SpA interface is mostly polar as opposed to the predominantly hydrophobic interaction of SpA with Fc.

Another commonly studied superantigen called Protein G (SpG) is a 65kDa protein expressed by *Streptococcus sp. group G*. It contains only two repeat domains and its structure is different from SpA, as it consists of four β strands forming a β sheet tied with an α helix (Fig. 1.1A). It forms an antiparallel interaction with the last β strand in the constant heavy chain domain CH1 of the Fab, thus forming a single β sheet between Ig and Protein G (Fig. 1.1B). The CH1 domain is relatively invariant between different Fabs derived from IgG and therefore represents the most attractive binding site for virulent *Streptococcus sp.,* as it maintains wide specificity that can elicit a broad immune response. The site of this interaction involves predominantly backbone contributions from

the CH1, but also use both main chain and side chain atoms from Protein G. This interface provides a continuous hydrophobic core which is shielded from solvent. Analogous to SpA, Protein G has an additional binding site to Fc that does not inhibit binding to the Fab (Erntell et al. 1988).

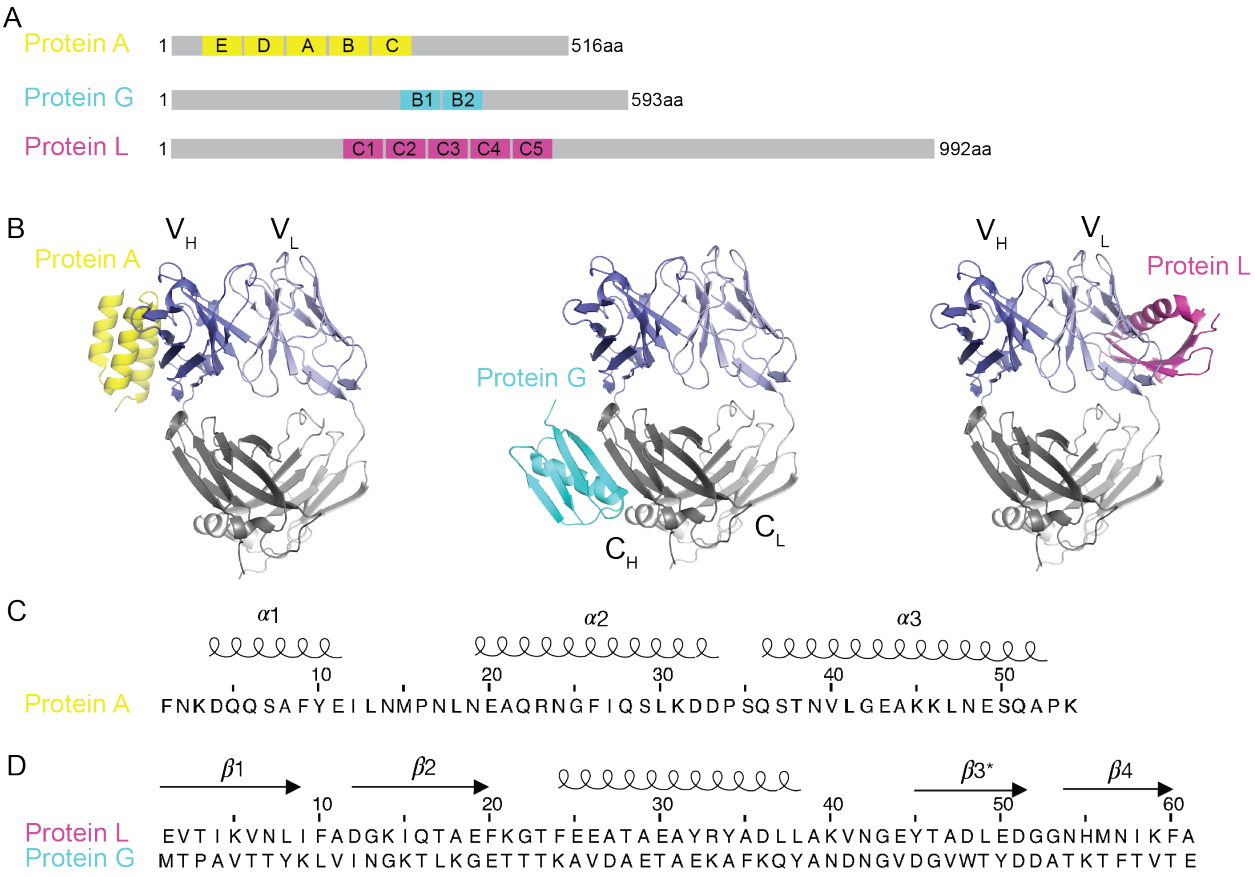

Figure 1.1 Characterization of B cell superantigens SpA, SpG and PpL
(A) Cartoon representation of repeat regions in full length sequence of superantigens Protein A, G and L. (B) Crystal structures show docking sites and superantigens bound to Fab fragments: Protein A (left), Protein G (middle) and Protein L (right). (C and D) Amino acid sequence alignment and secondary structure annotations of individual repeat domains: Protein A (C) and Protein G (cyan) and Protein L (magenta) (D).

Protein L (PpL) was isolated from *Peptostreptococcus magnus* and this 76kDa protein is present at the surface of only about 10% of *P. magnus* strains. It has been described as a virulence factor of bacterial vaginosis in different clinical specimens (Kastern et al. 1990). Each of up to five homologous immunoglobulin binding domains can bind to the light chain variable domain outside

of the hypervariable loops and it does not bind to Fcγ (Fig. 1.1A). PpL is known to have two binding sites on the variable light chain domain (VL) with markedly different affinities. The first VL chain interface involves 13 residues with PpL, the second has 15 residues from VL chain (and 10 in common between the two VL), but none of PpL residues overlap between the two binding interfaces. Both sites have high affinities for the human VκI, VκIII and VκIV subtypes, but not the VκII or lambda subtype (Fig. 1.1B). One possible explanation of that bias is based on the sequence of VκII subgroup, where a Pro12 residue introduces a major steric hindrance by pointing its imine ring toward PpL. Protein L, despite 15% sequence similarity, has an almost identical fold to Protein G and also forms β zipper interactions with their target sites on a Fab (Fig. 1.1C). This unusual 1:2 (Protein L:Fab) stoichiometry could potentially bridge two Igs anchored at the membrane of B cells. Finally, given many similarities between Protein G and Protein L and the fact that *Streptococci* and *Peptostreptococci* are found in the same habitats, including the human intestinal and genital tracts, it has been proposed that a gene transfer may have occurred between these two species (Goward et al. 1993).

1.1.3 Identification of commensal Immunoglobulin binding protein (Ibp)

Recently, the evidence emerged about novel proteins with remarkably similar domain organizations and structural homology to mentioned pathogenic superantigens, but so far, these novel proteins do not appear to initiate or contribute to disease pathology. Previous studies of antibodies found in mammalian mucosal surfaces, mainly immunoglobulins A (IgA) were identified to bind to multiple bacterial taxa and are commonly polyreactive (Bunker et al. 2017). To extend these observations, Bunker et al. screened a panel of microbiota-reactive IgA monoclonal antibodies (mAbs) for binding to fecal microbiota derived from human infants by

bacterial flow cytometry. The unusual pattern of reactivity was observed, where microbiota bound mAbs expressing mainly murine VH5/6/7 and human VH3 variable regions. Further confirmation of this result came from a similar VH preference in binding by mAbs and observed with isolated strains of unclassified Lachnospiraceae family, specifically *Ruminococcus gnavus* and *Coprococcus comes*. A similar pattern of variable region preference is common for known superantigens namely Protein A, although no Fcγ binding was noted (Bunker et al. 2019). This observation prompted Bunker et al. to isolate candidate proteins from a lysed bacterial extract using VH5 mAb as a Western blot probe. Further sequencing by mass spectrometry revealed a candidate protein that did not show any sequence homology to known B cell superantigens by basic local alignment search tool (BLAST), but the recombinantly expressed and purified candidate protein showed consistent binding specificity by enzyme-linked immunoabsorbent assay (ELISA). This led to the discovery of novel uncharacterized genes from *R. gnavus* that are now designated as Immunoglobulin binding protein A and B (IbpA and IbpB).

IbpA and IbpB have 83% amino acid sequence similarity and are likely the product of a gene duplication event. Even though the consequences of Ibp expression for commensal and host are poorly understood, these proteins include several structural and functional features reminiscent of other bacterial immunoglobulin-binding proteins including common B cell superantigens such as *Staphylococcus aureus* Protein A (SpA) or *Streptoccoccal* Protein G (Silverman et al. 2006). For example, IbpA/B contain four 67 amino acid long repeat domains which are reminiscent of the repeat domains in other mentioned B cell superantigens that were shown to mediate binding to Fab fragments of immunoglobulins. Finally, an N-terminal signal peptide and C-terminal SPKTG sortase motif indicate cell wall anchoring, reminiscent of Protein A tethering on surface of virulent

strains of *S. aureus* (Bunker et al. 2019, Schneewind et al. 2012). Despite the functional and structural similarity of Ibp to B cell SAgs, it is unclear whether the Ibp proteins are involved in pathogenicity or whether they are involved in host/microbe homeostasis.

1.1.4 Transient expansions of commensal microbiont *R. gnavus* correlate with disease outbreaks

In this work I specifically focus on the IbpA/B proteins from *Ruminococcus gnavus* species, which is a Gram-positive anaerobe that inhabits ~90% of humans yet represents ~0.1% of the gut microbiota (Qin et al. 2010). *R. gnavus* is well adapted to the gut by using unique sialic acid metabolism pathway, which provides a competitive strategy to colonize a niche in the gut mucosal layer (Bell et al. 2019, Tailford et al. 2015). Despite being a small percentage of the healthy gut microbiota, *R. gnavus* is disproportionately represented in diseases. Notably, some disease flare ups were associated with transient expansions and strain-specificity in *R. gnavus* in a range of inflammatory conditions such as inflammatory bowel disease (IBD), Lupus nephritis or atopy (Lloyd-Price et al. 2019, Hall et al. 2017, Azzouz et al. 2019, Chua et al. 2018, Zheng et al. 2016). Specifically, Hall et al., 2017 found that the strongest, yet transient, deviation in the microbial composition of IBD patients was the bloom of *R. gnavus* from an average of 0.1% in controls to up to 69% in IBD patients with active disease. However, only limited molecular association have been described and it is currently unknown how, or if at all, *R. gnavus* is contributing to either disease or dysbiosis or if it is just expanding due to a competitive advantage during inflammation. Some of the potential reasons for this disproportionate correlation with inflammatory diseases include factors such as a higher capacity in utilization of mucin glycans or expression of unique pro-inflammatory polysaccharides by *R. gnavus* (Crost et al. 2013, Henke et al. 2019). Specifically, Henke et al. described a potent, toll-like receptor 4 (TLR4) - dependent inflammatory

glucorhamnan polysaccharide made by *R. gnavus*. Polysaccharides are mediators of host-microbe communications including evasion of the host immune system (Mazmanian et al. 2008, Comstock et al. 2006), however they are difficult to study since they are enzymatically, not genetically, encoded. It is therefore speculated that the dramatic change in *R. gnavus* abundance affects the delicate equilibrium of the mucus layer, due to the dramatic increase in the fraction of the community utilizing the mucus layer. Subsequently, this could in turn influence the gut barrier integrity, potentially increasing the intestinal permeability of IBD patients.

Considering the need for better understanding of host/microbiome interactions, the discovery by Buker et al. of these new protein effectors, IbpA and B, led me to investigate the molecular and structural features of these proteins. In this thesis work I sought to provide detailed molecular and functional information about these *R. gnavus* immunoglobulin binding proteins and their superantigen-like properties *in vitro* to give new insights of how *R. gnavus* interacts with the host and utilizes its surface proteins to bind a significant group of antibodies.

1.2 Mechanism of recognition in polyreactive antibodies

Various immune mechanisms are deployed in the mucosa to maintain balance between host and commensal microbiota. Previously, I discussed an example of specialized molecules expressed by commensal microbiota that bind to a large number of host immunoglobulins. In this part I will focus on host IgA antibodies that were found to bind to multiple bacterial taxa and are called polyreactive (Bunker et al. 2017). In fact, polyreactive antibodies are not only found to maintain microbiota balance in the gut, but are also widely common, circulating in the periphery i.e. anti-

flu broadly neutralizing IgG antibodies. This part will focus on characterizing polyreactive antibodies and their mechanism of recognition of an array of ligands.

1.2.1 Specificity vs. Promiscuity in immune receptors

One of the fundamental features in the human immune response by which we protect ourselves against pathogens is the antigen-antibody interaction. Antibodies are secreted forms of B cell receptors composed of Fc domain that dictates the homing and function of an antibody and two identical Fabs that bind to their antigen. Each Fab is composed of heavy and light polypeptide chains and antigen binding is mediated through three hypervariable loops or complementarity determining regions (CDRs) on each chain. To be able to respond to foreign antigens with high specificity and affinity, B cells evolve their antibody specificity for antigens by introducing amino acid mutations in their CDR loops through a process called somatic hypermutations (SHM). This process of maturing antibody's affinity for their specific antigen, leads to targeted removal of that pathogen. This high specificity comes at a cost, since every interaction in biology is related to three fundamentally important issues as proposed by Wucherpfenning et al. 2007: one is uncertainty about the target, another is finite capacity for the recognition, and the third is noise in the recognition process. Uncertainty about the target relates to a high level of target possibilities in terms of their sequence and structure. Finite capacity relates to the limited number of receptors produced in a body that can recognize these distinct epitopes. And noise means that even with the perfect receptor-epitope pairing, recognition and activation might not take place. A possible solution to accommodating these diverse number of targets, without unnecessary overproduction of receptors, is called "degeneracy" or promiscuity, by promoting molecular overlap among receptors.

The term promiscuity (includes cross-reactivity and polyreactivity) is an inherent and common property of recognition among immune molecules including TCR and antibodies, and relevant to many aspects of T and B cell biology i.e. TCR triggering is a result of low affinity ligands; peptide binding motifs recognized by MHC class II molecules often tolerate multiple amino acid substitutions, some TCRs recognize an array of peptide/MHC ligands that leads to a range of different responses (Chen et al. 2003, Udaka et al. 1992, Wucherpfenning et al. 1995, Wucherpfenning et al. 1994). This specificity/promiscuity trade-off is a fine line in immunology, where too low binding affinity gets lost in the noise and too high affinity gets selected against, but in this thesis, I aim to propose that the lower affinity interactions are intended and valuable part of the immune recognition. In this thesis I will focus specifically on polyreactivity in antibodies.

1.2.2 Identification of naturally polyreactive antibodies

Despite a recent review describing the positive effects of polyreactive antibodies in certain complex pathological conditions (Dimitrov 2020), in a clinical scenario, all antibody drug candidates rely on highly specific recognition of their target antigen *in vitro* and *in vivo* as off-target effects are highly undesirable. Additionally, polyreactive antibodies tend to have considerably shorter circulatory half-life via FcγRn-mediated recycling and poor pharmacokinetics (Hotzel et al. 2012, Kelly et al. 2015, Kelly et al. 2018, Datta-Mannan et al. 2015). Therefore, single cell–derived mAbs are routinely screened through antigen-binding promiscuity, binding to autoantigens, and propensity for self-binding (homophilic interaction) assays. A commonly used ELISA for polyreactivity uses a 4-7 variety of structurally and biochemically diverse antigens including DNA, insulin, lipopolysaccharide (LPS), flagellin, albumin, cardiolipin, and keyhole limpet hemocyanin (KLH), commonly utilized in the literature

(Mouquet et al. 2010, Prigent et al. 2016, Bunker et al. 2017, Planchais et al. 2019, Mouquet et al. 2011, Prigent et al. 2018, Andrews et al. 2015, Jain et al. 2017, Neu et al. 2019, Wrammert et al. 2011). These ligands represent a diverse sampling of polar and amphipathic as well as large and small sized molecules. The major difference between polyreactive and monoreactive antibodies is that a polyreactive mAb can bind to multiple diverse ligands with very low affinity. This does not preclude it also binding to a specific target with very high affinity and specificity as well (Mouquet et al. 2010, Prigent et al. 2016). Polyreactivity is also critically different from self-reactive antibodies, or autoantibodies, where these antibodies bind with high affinity to self-targets instead of foreign, pathogenic targets.

Polyreactivity is a common feature found in antibodies, where B cells express polyreactive B cell receptors and secreted polyreactive Igs that comprise 6% of the naïve B cell repertoire and ~25% of the IgG memory B cell pool (Koelsch et al. 2007, Tiller et al. 2007, Wardemann et al. 2003). It is proposed that to overcome extreme antigenic variation, BCRs may have evolved to harbor polyreactivity to accommodate hard to reach, rapidly mutating or shielded by glycosylation epitopes. However, the precise role of polyreactivity in B cells and antibody responses remains unclear. Interestingly, broadly neutralizing antibodies (bnAbs, binding to the same epitope of multiple strains of the same virus) against HIV and influenza viruses are commonly polyreactive. Nearly 60-80% of bnAbs besides recognizing their high affinity viral target also recognize other diverse epitopes with lower affinity (Chen et al. 2014, Dennison et al. 2011, Haynes et al. 2005, Mouquet et al. 2005, Andrews et al. 2015, Bajic et al. 2019). It is hypothesized that polyreactivity augments overall antibody binding strength to its viral epitope by utilizing the binding surroundings for increased avidity. I.e. HIVs envelope protein is so sparse on the surface of the

virus that it allows only one arm of the antibody to bind. With a proposed model of "heteroligation", a single antibody can bind the primary target with one Fab domain while the other Fab domain is binding to surroundings (membrane, other receptors) in a polyreactive manner (Mouquet et al. 2010). Another model was proposed for denser viral surfaces, that are harder to reach i.e. in influenza, where polyreactivity can accommodate the molecular crowding at the membrane, surface charges, glycosylations or other receptors while searching for its epitope. In the face of an ever-changing and often glycosylated viral epitope, polyreactivity allows for adaption to imperfectly conserved or poorly accessible epitopes. Attempts at elimination of polyreactivity in broadly neutralizing antibodies does come at a sacrifice of broad neutralization (Mouquet et al. 2010, Prigent et al. 2016, Guthmiller et al. in print). This poses an interesting question whether specificity and promiscuity in antibodies are similar mechanistically and share residues involved in binding to both: high and low affinity ligands.

In contrast to anti-viral Igs which also derive from the adaptive immune system, previously mentioned homeostatic intestinal IgA antibodies do not have known high affinity ligands and additionally possess innate-like recognition properties that may facilitate adaptation to the broad and dynamic array of gut microbiota and dietary antigens. This poses an interesting question about the origin of polyreactivity, whether it is acquired through affinity maturation or polyreactive naïve B cells are recruited into the antigen-antibody response.

1.2.3 Recent reports on biochemical, structural and biophysical characterization of polyreactivity

In the past, high affinity antigen-antibody interactions have been described in terms of the "lock and key" hypothesis, where the antigen and antigen-binding pocket exist as rigid, perfectly

complementary pieces. This theory has been challenged multiple times, including a high-throughput analysis of thousands of CDR-H3 models that shows no clear delineation in the flexibility between naïve and antigen-experienced antibodies (Jeliazkov et al. 2018, Burnett et al. 2020). Several crystal structures have demonstrated substantial conformational differences for antibodies binding to two or more diverse ligands by adopting induced-fit conformational changes (James et al. 2003, Arevalo et al. 1994, Foote and Milstein 1994, Wilson and Stanfield 1994). These observations indicate that flexibility of Ig CDR loops allows them to accommodate diverse ligands within its single binding pocket and it has been therefore proposed that polyreactive or germline antibodies are more flexible than classic monoreactive or antigen-experienced antibodies (Prigent et al. 2018, Fernandez-Quintero et al. 2019). This hypothesis however is often limited to an individual CDR-H3 loops, which are longer than the germline encoded CDR1 and CDR2 loops and often contain multiple glycine residues, enhancing the rotational freedom of the loop.

Additionally, such flexibility/plasticity may not be the only structural explanation for polyreactivity and so far, no proof has been shown that CDR-H3 is responsible for polyreactive interactions. Alternative structural studies revealed that light chains can modulate the neutralizing spectrum of antibodies by affecting the local conformation of heavy chains, instead of direct interaction with the antigen (Xiao et al. 2019). Moreover, examples from other receptor systems suggest that binding of structurally diverse ligands does not always require large conformational changes of the receptor. I.e. NKG2D, a conserved activating receptor binds to a vast number of polymorphic ligands MHC-class-I-polypeptide-related sequence A and B (MICA and MICB) and cytomegalovirus UL16-binding protein (ULBP) in a multispecific manner (Eagle and Trowsdale, 2007). NKG2D uses two key tyrosine residues that are rigidly held in the NKG2D structure but

can form interactions with different residues on different ligands in a largely distinct way (McFarland and Strong 2003).

More factors proposed to be involved in the polyreactivity of antibodies include CDR3 lengths, net hydrophobicity and charge (Prigent et al. 2016, Prigent et al. 2018, Rabia et al. 2018, Lecerf et al. 2019). Other studies have found an increased prevalence of arginine and tyrosine in polyreactive antibodies (Kelly et al. 2018, Birtalan et al. 2008). While these studies represent substantial advances in the study of polyreactivity they have often been limited in scope, focusing on a singular antibody source.

This area provides a multitude of interesting questions: What is the direct affinity of polyreactive antibodies to its polyreactive ligands and what affinity threshold differentiates between physical binding and noise? How do polyreactive antibodies bind to multiple ligands – using its antigen binding site or germline sequence? Is this binding mechanism different between different polyreactive clones? Or do polyreactive antibodies bind their target ligands in the same way? In this work I explore experimental techniques to determine the affinity of polyreactive antibodies to their polyreactive ligands. Additionally, I seek to obtain structures of polyreactive Fabs apo and in complex to investigate biophysical properties involved in polyreactivity. Finally, I use classical molecular dynamics (MD) simulations of solved structures to analyze the dynamic nature of polyreactive antibodies.

1.3 A "Inside-out" mechanism of antigen recognition in γδ T cells

1.3.1 γδ T cells and nonconventional mechanisms of antigen recognition

The adaptive immune system is composed of specialized cells and processes that eliminate pathogens and react to cellular distress. The conventional T cells that carry out the acquired response, bear a TCR composed of α and β chains forming a heterodimer. There also exists a unique subset of T cells that are more innate-like in their function and express heterodimeric TCRs consisting of γ and δ chains. As opposed to αβ, γδ T cells dominate in peripheral tissues such as the skin, gastrointestinal and reproductive tracts and are present in human blood in smaller numbers than αβ t cells (Gober et al. 2003, Morita et al. 2007, Carding and Egan, 2002). Despite their significance and often critical role in various kinds of immune responses to infections, tissue damage, and cancers γδ T cells are one of the most enigmatic T cell types and very little is known about their development, differentiation and signals they respond to (Silva-Santos et al. 2015, Vantourout and Hayday, 2013).

The majority of γδ T cells do not follow the conventional peptide/MHC-dependent antigen presentation and do not require antigen processing as do αβ T cells. They instead follow diverse mechanisms of TCR-dependent antigen recognition. Recent studies show γδ T cells can respond to peptidic, non-peptidic antigens including mostly studied phosphoantigens (pAgs) or whole proteins found in both infected and tumor cells. As recent studies add more layers of complexity to enigmatic γδ T cells, little is understood about of the mechanism of antigen recognition (Melandri et al. 2018, Willcox et al. 2019). This thesis focuses on a small subset called Vγ9Vδ2 T cells and its mechanism of recognition of pAgs, present in human blood. Vγ9Vδ2 T cells were

shown to have potent reactivity towards certain microbial infections and cancers and therefore possess significant therapeutic potential in immunotherapies.

1.3.2 Vγ9Vδ2 T cells, Btn3A and an "inside-out" signal stimulation

2-10% of T cells in human blood are Vγ9Vδ2 T cells that can expand to ~40-60% of total T cells upon stimulation with small molecules containing a pyrophosphate group called phosphoantigens (Morita et al. 2000, Chen et al. 2013). Mostly studied example of pAg is isopentenyl pyrophosphate (IPP), an intermediate of the mevalonate pathway involved in cholesterol synthesis and it accumulates intracellularly in many tumor cells (Gober et al. 2003, Tanaka et al. 1994). Another study identified the structurally similar compound hydroxy-methyl-butyl-pyrophosphate (HMBPP), a microbial metabolite produced through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway found in most eubacteria and apicomplexan protozoa, as an even more potent stimulant than IPP (Belmant et al. 1999, Hintz et al. 2001, Morita et al. 2007) (Fig. 1.2A).

Our collaborators, Drs. Olive, Bonneville and Scotet, pioneered the breakthrough in our understanding of Vγ9Vδ2 T cell activation, by identifying the butyrophilin-3 (Btn3) protein as a key mediator in this process (Harly et al. 2012). They revealed through the knockout studies, that Btn3 is indispensable to mediate pAg-induced T cell activation. As physiological pAgs are known to be present inside target cells therefore Btn3 was proposed as an intracellular sensor for these small molecules. It is proposed that in stressed or infected cells, accumulated pAg binds to the intracellular domain of Btn3, which then transduces the signal through its extracellular Ig domains leading to Vγ9Vδ2 T cell stimulation (Fig. 1.2B). The details of this novel "inside-out" recognition and activation is still unknown and while cell-to-cell contact is required, no direct interaction

between the Vγ9Vδ2 TCR and Btn3 or pAg has been demonstrated (Sandstrom et al. 2014, Bukowski et al. 1995, Tanaka et al. 1995, Lang et al. 1995). Furthermore, by introducing Btn3A1 into murine target cells it was demonstrated that Btn3A1 is *necessary* for pAg-induced Vγ9Vδ2 T cell stimulation, but not sufficient (Riano et al. 2014, Sandstrom et al. 2014). This suggests that additional mediator protein is required for Vγ9Vδ2 T cell activation that would link intracellular pAg and Btn3 expressed on the surface of tumor/ infected cell with Vγ9Vδ2 TCR present on a T cell. Most recently two groups reported that another butyrophilin, Btn2A1 associates with Btn3A1, and act together to initiate responses to pAg through direct binding of Btn2A1 to germline encoded Vγ9 regions of Vγ9Vδ2 TCR (Rigau et al. 2020, Karunakaran et al. 2020). Additionally, earlier reports identified two cytoskeletal mediators that are involved with Btn3A in pAg-induced T cell activation: periplakin - a cytoskeletal adaptor protein, and RhoB GTPase - a regulator for multiple actin-associated proteins (Rhodes et al. 2015, Sebestyen et al. 2016). Despite interaction with Btn3A1, neither of these molecules directly contacts Vγ9Vδ2 TCR and were shown to play more of a regulatory role in this mechanism.

1.3.3 Btn3A structural information

Another important aspect of this field is focused on molecular details of pAg binding to the intracellular domain of surface expressed Btn3 and how this signal is mediated outside of the cell. Btn3 (CD277) is a family of type I membrane proteins ubiquitously expressed on the surface of a broad variety of cell types including immune cells and has three isoforms A1, A2 and A3 (Arnett and Viney, 2014). Each isoform has two immunoglobulin extracellular domains, IgV and IgC with over 95% sequence identity and a single pass transmembrane helix (Abeler-Dorner et al. 2012). Btn3A1 and A3 both have an additional juxtamembrane (JTM) region of unknown structure and

intracellular B30.2 domain with 87% sequence identity between the two (Gu et al. 2014). Our structural and biophysical data show that the intracellular B30.2 domain of BTN3A1 is the direct sensor for pAg, with the pAg binding pocket lined with basic (positively charged) residues. Of note, BTN3A3 also has a B30.2 domain that does not bind known pAgs, this was mapped to a single difference between BTN3A3 and BTN3A1 in this domain, H351R (Sandstrom et al. 2014). While the BTN3A2 and A3 isoforms do not bind pAg directly, there is evidence that they can still stimulate Vγ9Vδ2 T cells when treated with a 20.1 mAb that mediates cross-linking between extracellular domains without the requirement for pAg (Harly et al. 2012, Palakodeti et al. 2012).

Our lab determined the three-dimensional structures of the Btn3 extracellular Ig domains with and without agonist (20.1) and antagonist (103.) mAb and two types of dimers were identified in the crystal lattice: a "V-shaped" and "head-to-tail" dimer structure (Palakodeti et al. 2012) (Fig. 1.2C, D). The agonist mAb 20.1, when added to target cells is capable of activating Vγ9Vδ2 T cells in the absence of pAgs, binds exclusively to the V-shaped dimer, whereas agonist mAb 103.2 binds to both. This suggests that the "V-shaped" form of the Btn3A1 dimer is hypothesized to be the active form, however an alternative model has been proposed (Yang et al. 2019).

More recently, the attention of the butyrophilin-3 field has shifted to the juxtamembrane region of this protein, which is predicted to be the coiled-coil but remains structurally and functionally elusive (Wang et al. 2019) (Fig. 1.2E). Therefore, I sought to understand what are the structural events that follow pAg binding to the B30.2 domain? Whether intracellular binding of pAg triggers conversion from resting to activating state that leads to Vγ9Vδ2 T cell stimulation? To determine

which dimer is prevalent in native environment, I sought to understand the native structural orientation of full-length Btn3 dimers during my doctoral work.

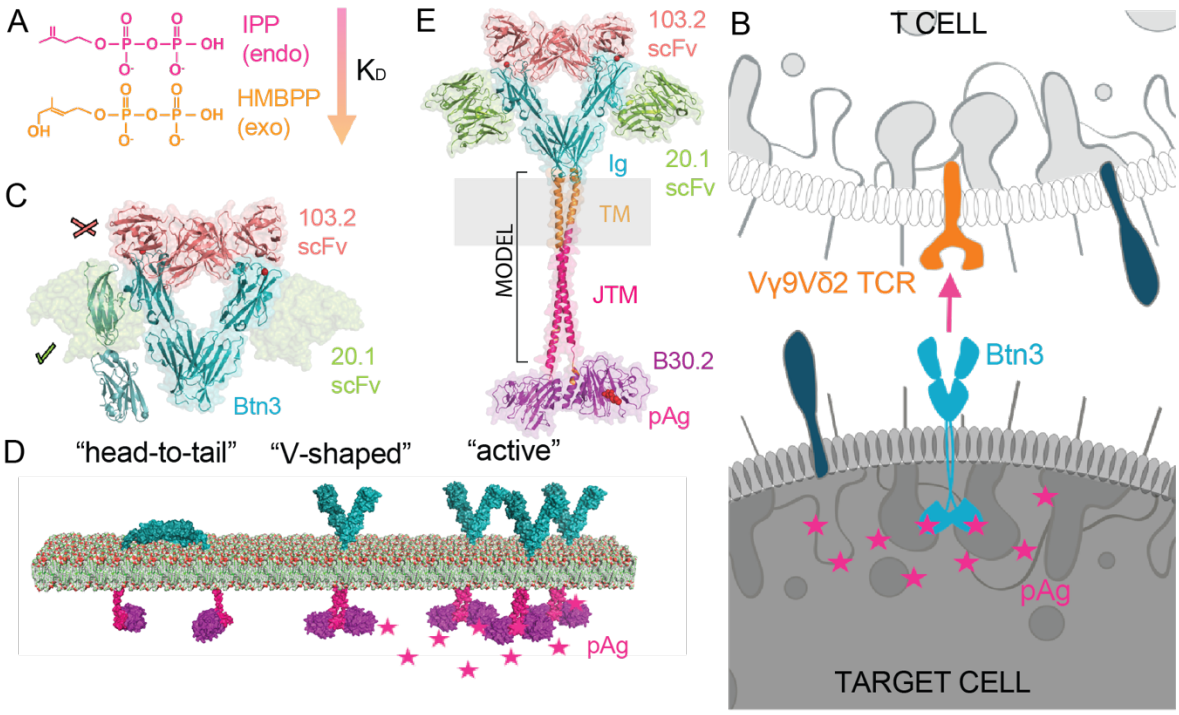

Figure 1.2 Btn3A is pAg-sensing molecule, inducing Vγ9Vδ2 T cell activation via "inside-out" mechanism
(A) Examples of commonly studied phosphoantigens: an endogenous isopentenyl pyrophosphate (IPP) and 1000-fold more potent exogenous hydroxy-methyl-butyl-pyrophosphate (HMBPP). (B) Model of "inside-out" Vγ9Vδ2 T cell stimulation (orange) with cell surface expressed butyrophilin-3 (blue) binding to phosphoantigen (pink) accumulated intracellularly in a target cell. (C) Crystal structure of extracellular Ig domain in complex with agonist 20.1 scFv or antagonist 103.2 scFv reveals two types of dimer: "head-to-tail" and "V-shaped". (D) Proposed models *in vivo*, where "V-shaped" dimer is the active conformation that clusters together upon pAg presence. (E) Full length Btn3 model using known structures of Ig and B30.2 domains (PDB ID: 4F9P, 4F9L and 4N7I) and modelled alpha helix as TM and JTM.

Chapter 2 – Molecular mechanism of a novel host-commensal microbiota interaction

2.1 Investigation of the function and structure of Ibp as a novel superantigen-like molecule

Superantigens are expressed on the surface by invading microbes and can result in devastating effects on the immune system, even leading to host death. While all known superantigens induce pathogenic immune proliferation and come from opportunistic pathogens, the recently discovered Immunoglobulin binding proteins (Ibp) proteins showed similar specificity to antibody variable (V) domain families but were identified in commensal microbiota, without seemingly pathogenic functions. This posed an interesting question about the role of these supeantigen-like proteins in commensal *R. gnavus* and led me to investigate their functional and structural features. I sought to provide more detailed information about commensal superantigen-like proteins to gain more insight at host/ microbiota interactions.

2.1.1 IbpA and IbpB bind to a broad repertoire of human and mouse B cells and initiate BCR signaling

Bunker et al. 2019 previously tested an archive of mouse and human derived monoclonal antibodies for Ibp reactivity and observed preferential binding of mouse VH5/6/7 and human VH3 variable regions. However, together with Christoph Drees from Albert Bendelac's Lab, we also detected a considerable fraction of antibodies expressing other VH families which were also bound by Ibp. To test if full length IbpA and IbpB also bind to membrane-associated immunoglobulins and to more precisely define the repertoire of Ibp-reactive mouse and human B cells we labeled full length IbpA and IbpB with fluorophores and used them as staining reagents

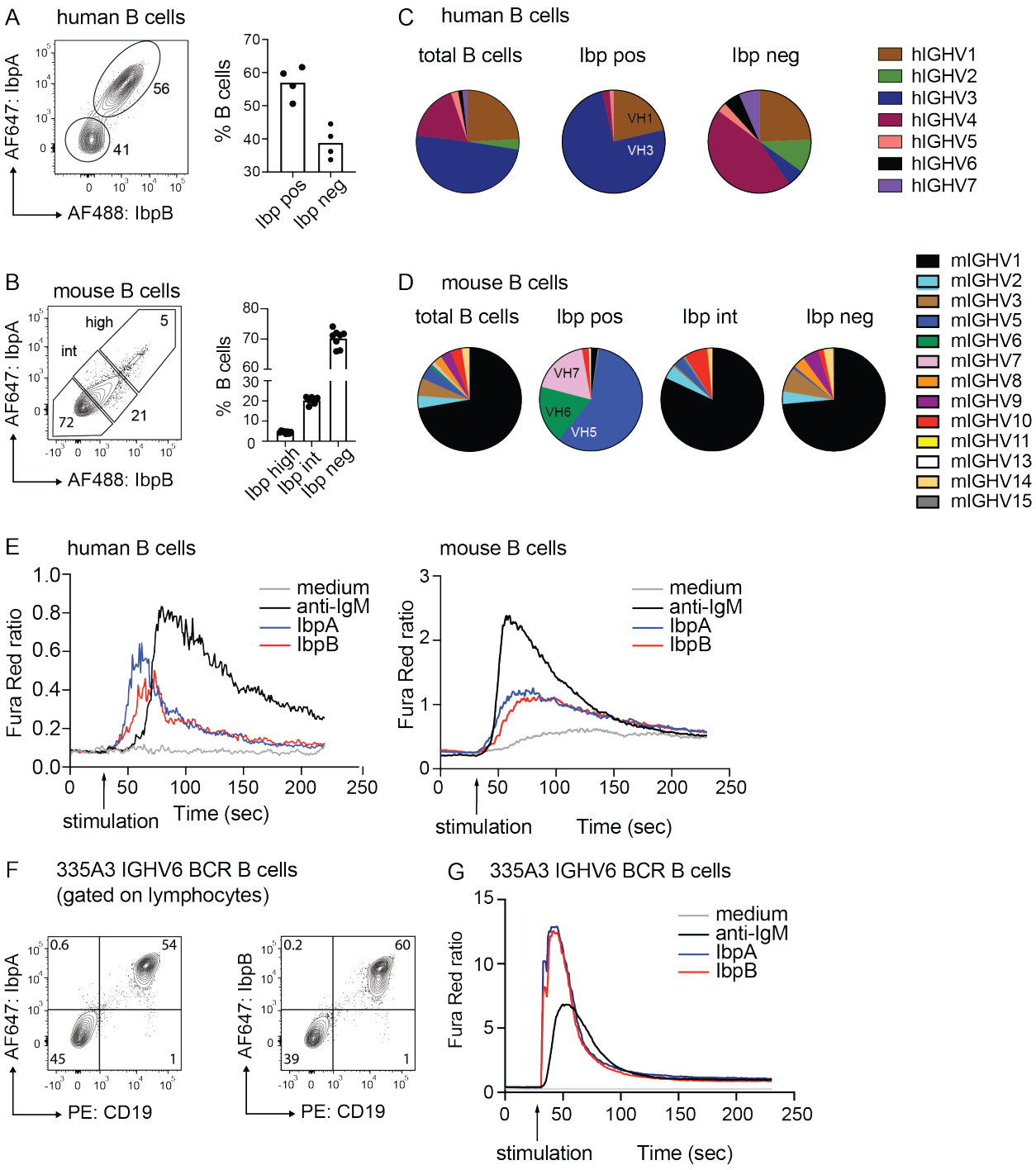

Figure 2.1 IbpA and IbpB activate a defined repertoire of human and mouse B cell receptors (A) Flow cytometry plots show human CD19pos B cells stained with fluorescently labeled IbpA and IbpB (left) and bar graph quantification (right). (B) Analogous flow cytometry plots for mouse CD19pos B cells stained with fluorescently labeled IbpA or IbpB (left) and bar graph quantification (right). (C) Immunoglobulin heavy chain repertoire of sorted human B cells. Each symbol represents the repertoire of the corresponding sorted B cell population. Total B cells indicate repertoire of all CD19pos B cells regardless of Ibp binding. *(continued on the next page)*

Figure 2.1, continued: (D) Analogous immunoglobulin heavy chain repertoire of sorted mouse B cells and mouse B cells. (E) Ca2+ flux analysis of human peripheral blood CD19pos B cells (left) and mouse CD19pos splenocytes (right) after *in vitro* stimulation with IbpA, IbpB or anti-IgM. (F) Representative contour plots of transgenic B cells expressing a VH6 BCR show bright staining of CD19 B cells with IbpA (left) or IbpB (right). (G) Ca2+ flux analysis mouse CD19pos splenocytes expressing a transgenic VH6 BCR after *in vitro* stimulation with IbpA, IbpB or anti-IgM.

for flow cytometry. 50-60% of human peripheral blood B cells were co-stained with IbpA and IbpB (Fig. 2.1A), whereas the frequency of brightly labeled splenic mouse B cells was lower (Ibp high, 5%) (Fig. 2.1B). No discernable difference could be observed between IbpA and IbpB suggesting binding to a shared repertoire. In addition, a dose dependent population of mouse B cells with intermediate binding pattern could be detected (Fig. 2.1B). We next sorted Ibp-bound, Ibp-unbound as well as total human and mouse B cell populations and performed repertoire sequencing. The majority of Ibp positive human B cells (75%) expressed VH3 gene families, whereas 25% showed expression of VH1 genes (Fig. 2.1C). Ibp high murine B cells exclusively expressed VH5/6/7 genes (Fig. 2.1D). The repertoire of Ibp int B cells was more heterogenous but mostly consisted of VH1 families. There was also a partial overlap between the heavy chain repertoire of Ibp neg and Ibp int B cells (Fig. 2.1D). A restricted light chain repertoire was not found in either dataset, suggesting Ibp binding to antibodies is solely mediated through the heavy chain variable regions. These data confirm previous findings by Bunker et al. and indicate that IbpA and IbpB bind a shared and broad repertoire of human and mouse antibodies (Bunker et al. 2019).

We then sought to determine whether recombinant IbpA and B also initiate BCR signaling. We therefore stimulated B cells *in vitro* and examined intracellular calcium flux by flow cytometry. A strong calcium flux signal could be detected in bulk human B cells (Fig. 2.1E, left) whereas the

signal of bulk mouse B cells was lower, consistent with Ibp's engagement of a more restricted segment of the mouse Ig repertoire (Fig. 2.1E, right). However, in transgenic B cells expressing an IGHV6 BCR (Erickson et al., unpublished) both IbpA and IbpB induced calcium flux considerably faster and stronger than even anti-IgM stimulated cells (Fig. 2.1F, G). Thus, Ibp binds to a broad repertoire of human and mouse BCRs and, in a soluble form, can activate BCR signaling.

2.1.2 A truncated IbpA construct is functional and shows higher affinity to antibodies than wild-type IbpA

To analyze the molecular details on how Ibp binds to antibodies and maintains its B cell repertoire specificity, I pursued structural studies of full length Ibp in complex with a mouse VH5 Fab. Unfortunately, structural trials with wild-type Ibp in complex with Fab either by X-ray crystallography or by negative stain electron microscopy were unsuccessful. I have therefore obtained an optimized IbpA (from now on called truncated IbpA) construct that bears only one of the four 67 amino acid repeat domains that might reduce sample heterogeneity. This construct contains the fourth repeat (Domain D) and extends to the C-terminus of the full length IbpA (Fig. 2.2A). The size-exclusion chromatography profile of the truncated IbpA in complex with Fab shows a shifted peak compared to that of the full length IbpA and Fab complex. Christoph Drees also tested the ability of truncated IbpA to bind and to activate mouse BCRs *in vitro* and found that the truncated Ibp activated as well as the full length (Fig. 2.2B). Interestingly, direct affinity measurement by biolayer interferometry (BLI) showed three times higher K_D and slower dissociation rate (k_{off}) for the truncated IbpA construct compared to the full length when tested for binding with mouse VH5 IgA 338E6 Fab (Fig. 2.2C). I have also observed similar trends in K_D and k_{off} for human VH3 IgG F16 clone (Fig. 2.2D). Altogether this data demonstrates that truncated

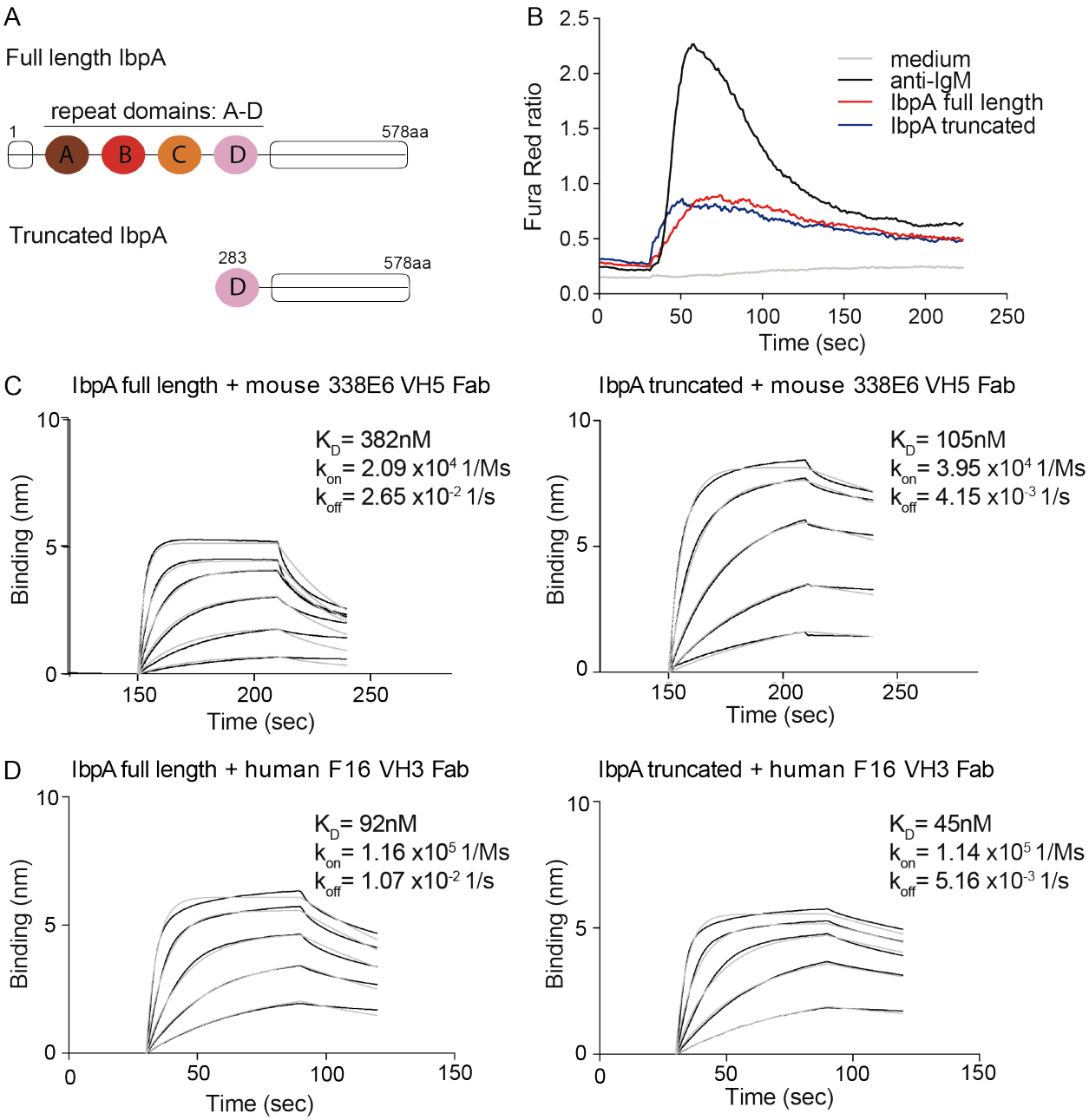

Figure 2.2 Truncated IbpA maintains binding to Fab fragment with higher affinity
(A) Cartoon representation shows domains of full length IbpA protein (top) in comparison with truncated IbpA construct (bottom) which was used for crystallization and *in vitro* B cell activation assays. (B) Ca2+ flux analysis comparing full length and truncated IbpA after *in vitro* stimulation of mouse splenic B cells. (C) Binding studies using Bio-layer interferometry (BLI) comparing affinity of full length (left) versus truncated IbpA (right) to mouse 338E6 VH5 Fab. (D) Binding studies using BLI comparing affinity of full length (left) versus truncated IbpA (right) to human F16 VH3 Fab.

IbpA can activate mouse splenic B cells as well as wild-type IbpA, additionally it binds to tested mouse and human Fabs with higher affinity and slower dissociation, suggesting that full length IbpA might adopt a conformation that modifies its interaction with Fab.

2.2 Characterization of the global structure of Ibp in complex

2.2.1 A truncated IbpA structure reveals two superantigen-like fold domains that engage a Fab: the Domain D and the C-terminal heavy chain binding domain (HCBD)

In contrast to full length IbpA, the optimized Ibp construct proved successful in my crystallization attempts in complex with an engineered mouse VH5 Fab (clone 338E6), where the variable domains derive from a mouse gut IgA Fab and the constant C_L and C_H1 domains are from a mutagenized human IgG1 scaffold. I solved the resulting structure and refined to 2.1A resolution with R_{work}/R_{free} 0.2377/ 0.2817 (Table 1). The truncated IbpA structure reveals three equal size domains: Domain D (lightpink), middle Domain E (gray) and a C-terminal domain, from now on referred to as the Heavy Chain Binding Domain (HCBD, teal) (Fig. 2.3A). Domain D comprises one of the four repeat domains (A, B, C and D) present in the full length structure of Ibp.

Strikingly, both Domain D and HCBD, despite having no sequence homology (and low DALI server z score =2.4, data not shown), both adopt a fold reminiscent of that of Protein G and L (Fig. 2.3B, C). Domain D has three instead of four complete beta strands (ß1, ß2, and ß4), with a truncated ß3* likely due to a disulfide bond between Cys317-Cys338 bringing the two chains closer to each other and away from the ß4 strand (Fig. 2.3A, yellow). This S-S bond is present in all repeat domains of full length IbpB and only in the last repeat of four in exchange of Tyr317 in the full length IbpA. The alpha helix between ß2 and ß3 is shorter and turned ~40-60 degrees

Table 1 Data collection and refinement statistics of truncated IbpA + mouse 338E6 VH5 Fab complex

Statistics for the highest-resolution shell are shown in parentheses.

	338E6 VH5 Fab + IbpA 283-578aa
Wavelength energy (keV)	12.68
Wavelength range (Å)	0.98
Resolution range	83.36 - 2.102 (2.177 - 2.102)
Space group	P 1 21 1
Unit cell	85.2401 102.191 118.28 90 102.05 90
Total reflections	3855839 (373260)
Unique reflections	115284 (11391)
Multiplicity	33.4 (32.5)
Completeness (%)	99.11 (98.16)
Mean I/sigma(I)	21.04 (1.12)
Wilson B-factor	34.89
R-merge	0.6793 (3.98)
R-meas	0.6898 (4.046)
R-pim	0.1178 (0.7121)
CC1/2	0.971 (0.31)
CC*	0.993 (0.688)
Reflections used in refinement	114354 (11299)
Reflections used for R-free	1998 (198)
R-work	0.2377 (0.3581)
R-free	0.2817 (0.4015)
CC(work)	0.833 (0.376)
CC(free)	0.810 (0.334)
Number of non-hydrogen atoms	9930
macromolecules	9476
solvent	454
Protein residues	1258
RMS(bonds)	0.009
RMS(angles)	1.17
Ramachandran favored (%)	89.84
Ramachandran allowed (%)	7.72
Ramachandran outliers (%)	2.44
Rotamer outliers (%)	0
Clashscore	12.98
Average B-factor	51.51
macromolecules	51.55
solvent	50.55
Number of TLS groups	1

compared to protein G and L respectively and accommodates a long loop not present in Protein G/L (Fig. 2.3A, black arrow). HCBD also resembles a superantigen-like fold despite being located

outside of the repeat region of IbpA, as no additional antibody binding sites were reported for other known superantigens. While composed of two (and shorter ß4*) beta strands, additionally HCBD has a metal binding site mediated by three Asp in DGDGDG motif, the role of which is not clear. The middle Domain E fold (Fig. 2.3D) resembles (but not closely) other adhesion type proteins (mucus binding proteins, PDB ID: 3LYY and 4MT5) according to DALI server (data not shown) and links other domains with unstructured regions, suggesting lack of coordination between Domain D and HCBD.

To evaluate potential flexibility between the three domains of the truncated IbpA, I used small-angle X-ray scattering (SAXS) analysis to investigate the flexibility of truncated IbpA structure when not bound to Fab. The initial fit of the experimental data with a rigid body IbpA apo model from our crystal structure did not yield a good score using the FoXS server, suggesting that the solution structure may adopt additional conformations to that of our crystal structure. Using MultiFoXS server, which allows for conformational sampling and multi-state modelling, I stipulated flexibility in the two linker regions connecting the three domains. Using this analysis, I have determined six distinct topologies from that of our crystal structure with an average χ^2= 0.64 (Fig. 2.3E). Taken together, my structure of the truncated IbpA revealed three new domains, two of which (Domain D and HCBD) adopt a fold similar to Protein G and L, despite no sequence conservation. The middle domain serves as a linker between these two independent superantigen-like domains allowing truncated IbpA to sample multiple conformations.

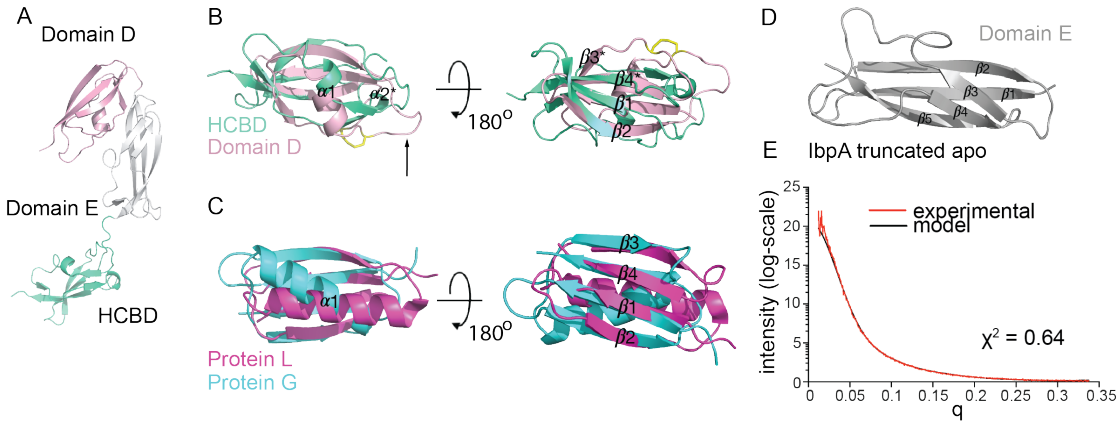

Figure 2.3 Crystal structure of truncated IbpA reveals two superantigen-like domains and one adhesion-like novel domain
(A) Truncated IbpA structure from complex crystal structure showing three equal size domains: repeat domain D (lightpink) with disulfide bridge (yellow) and black arrow pointing at the unstructured loop, middle domain E (gray) and heavy chain binding domain (HCBD, greenteal). (B) Close up analysis of superimposed structures of Domain D (lightpink) and HCBD (greenteal). (C) Close up analysis of superimposed structure of related superantigen fold of Protein L (magenta) and Protein G (cyan). (D) Structure of novel adhesion-like domain of middle domain E (gray). (E) Experimental fit of SAXS data of truncated IbpA apo analyzed by multi-state modelling with MultiFoXS and plotted by I(q).

2.2.2 Truncated IbpA in complex with mouse VH5 Fab shows contacts between two distinct IbpA domains and both chains of the Fab

The asymmetric unit of my crystal contains two Fabs and two IbpAs, where each IbpA establishes contacts between the heavy chain of one Fab and the light chain of the other (Fig. 2.4A), producing a 2:2 stoichiometric complex. The primary Fab/IbpA interface is between the heavy chain and the HCBD (Fig. 2.4B) and is composed of 44 contacts between the HCBD and the D, C'' and E beta strands of the VH Fab with 15 H-bonds between seven residues of the HCBD and 12 residues of the Fab (Table 2). This interface is discussed more in the following section. A secondary interface was also noted between the variable region of the light chain of the Fab and the Domain D of IbpA (Fig. 2.4C; Table 2). In contrast to Protein L/G and HCBD, which maintain interactions with the

Fab through formation of strong backbone-mediated beta-zippers with their ß2 or ß3 strands, Domain D contact site in this interface maps to two

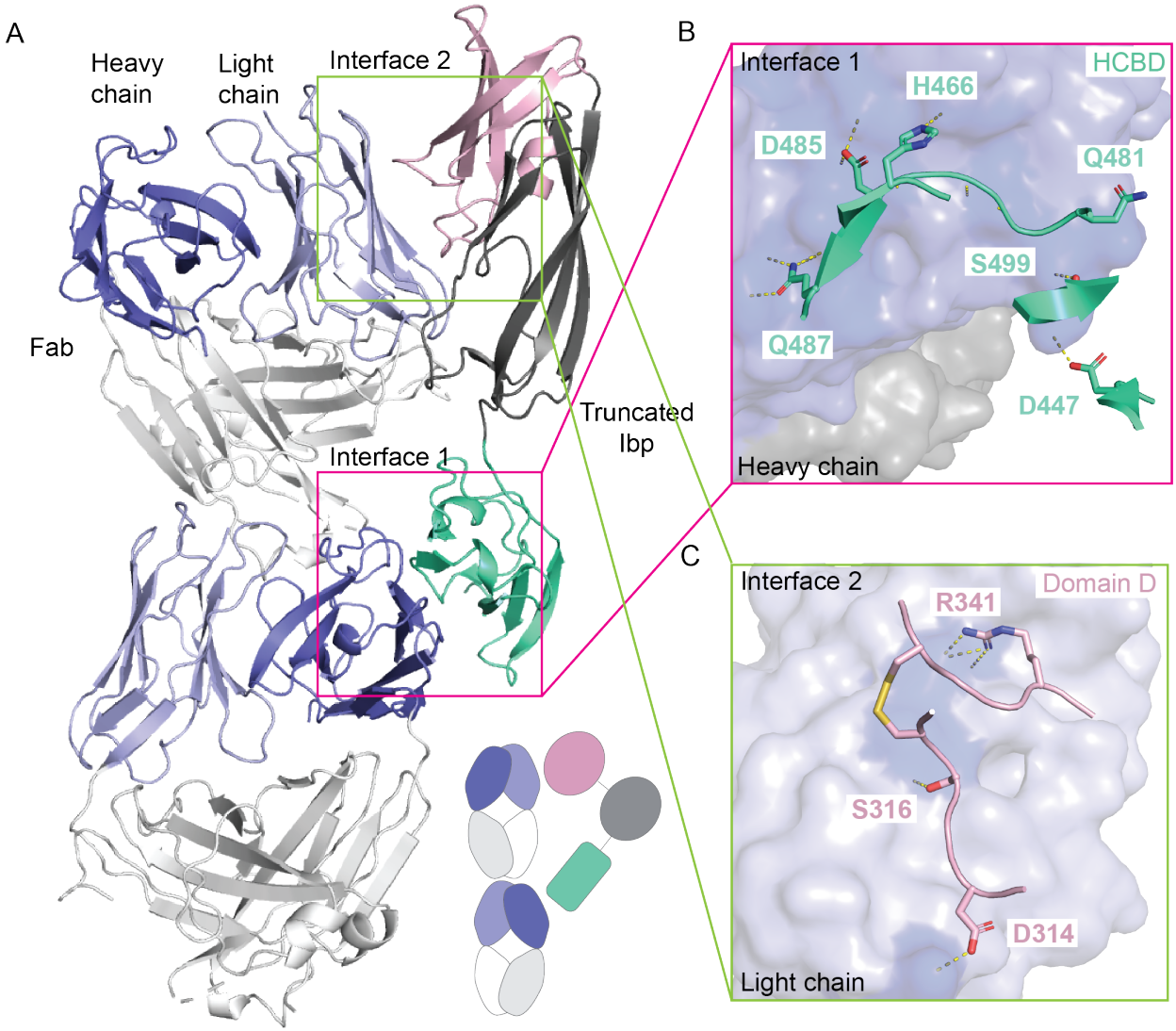

Figure 2.4 Complex crystal structure of functional, truncated IbpA shows unconventional binding to mouse VH5 Fab clone 338E6 involving heavy and light chains
(A) Crystal structure showing two distinct interfaces between truncated IbpA and heavy and light chains of two mouse VH5 Fab fragments. Simplified cartoon present in lower right corner. (B) Close up view of interface 1 between heavy chain (blue, surface representation) and participating residues of the heavy chain binding domain (HCBD, greenteal, cartoon representation). (C) Close up view of interface 2 between light chain (light blue, surface representation) and participating residues of Domain D (pink, cartoon representation).

unstructured loops between ß2 and ß4. The middle domain E, between Domain D and the HCBD, only maintains three vdW contacts with the Fab (Table 2). Taken together, both interfaces dock on the side of the variable domain of the Fab, outside of the conventional CDR binding region, reminiscent of the binding of known B cell superantigens i.e. Protein A and L. My complex structure has revealed that, the VH domain of the Fab is contacted exclusively through the HCBD while a smaller contact exists between the VL and Domain D of IbpA. These interfaces are investigated in more detail below.

2.2.3 The C-terminal HCBD of IbpA confers VH specificity

The primary interface between IbpA and Fab revealed in the crystal structure involves interactions between the C-terminal HCBD and the VH domain of the Fab. This result is consistent with our B cell repertoire sequencing studies, where IbpA binding was heavily biased towards Fabs containing particular VH family members. This interface includes one salt bridge held by Lys76 and seven backbone H-bonds involving Gly55 in CDR2 region and Arg67, Ile70, Arg72, Ala75 and Gln82 in FR3 region (Fig. 2.5A; Table 2). Notably, the contact area of this primary interface is ~633 $Å^2$ and spans not only the hypervariable region 4 (HV4), as typically seen for Protein A superantigens, but also involves the CDR2 region of the Fab. Despite Protein A and HCBD sharing similar VH germline bias, only three contact residues are commonly recognized on Ig (Lys/Thr58, Tyr60 and Asn/Ser84), suggesting differential molecular recognition of the VH (Fig. 2.5B). To understand how this Ig binding interface is recognized by HCBD, I have mutated individual contact residues on HCBD to alanine and tested for binding by BLI. Given the affinity between HCBD and Fab is 37nM, slow dissociation and the vast interface between HCBD and Fab, no individual alanine mutation on HCBD site was able to break that interaction and only few mutations had a slight

Table 2 Contact residues between individual HCBD, Domain D and Domain E of truncated IbpA and mouse 338E6 VH5 Fab

338E6 Fab Heavy chain	HCBD	Contact
Ile51	Val483, Gly484	vdW
Gly55 O	Val483 N	H***
Gly55	His466, Val483	vdW
Thr58 OG1	Asp485 OD2	H***
Thr58	Asp485	vdW
Tyr60 OH	Asp485 OD2	H***
Tyr60	Asp485	vdW
Gly66	Ile464	vdW
Arg67 O	Gln487 NE2	H***
Arg67	Gln487	vdW
Phe68	Gln487	vdW
Thr69 OG1	Gln487 NE2	H***
Thr69	Ile464, Asp485, Gln487	vdW
Ile70 O	Asp485 N	H***
Ile70	Asp485	vdW
Ser71	Val483, Asp485, Phe486	vdW
Arg72 N	Val483 O	H***
Arg72 O	Glu482 CA	H*
	Val483 N	H***
Arg72	Glu482, Val483	vdW
Asp73 OD1	Ser499 OG	H*
OD2	Ser499 OG	H***
Asp73	Glu482, Ser499	vdW
Asn74	Gln481, Glu482	vdW
Ala75 N	Gln481 OE1	H*
Ala75	Ala443, Gln481	vdW
Lys76 NZ	Asp447 OD2	SB
Lys76 NZ	Asp447 OD2	H***
Lys76	Asp447	vdW
Gln82 O	Gln487 NE2	H*
Gln82	Gln487, Leu490	vdW
Ser84 OG	Gln487 OE1	H***
Ser84	Gln487	vdW

338E6 Fab Light chain	Domain D	Contact
Ser7	Asp342	vdW
Met11	Ser316	vdW
Ser12	Asp314	vdW
Val19	Ser316	vdW
Ser20 N	Ser316 OG	H***
Ser20	Ser316, Ile339	vdW
Thr22	Ile340	vdW
Asp70	Arg341	vdW
Phe71 O	Arg341 NH1	H*
Phe71	Arg341	vdW
Thr72 OG1	Arg341 NH1	H***
	NH2	H***
Thr72	Ile339, Arg341	vdW
Lys107 NZ	Asp314 OD2	SB
Lys107 CD	Asp314 OD2	H*
Lys107 CE	Asp314 OD2	H*
Lys107	Asp314	vdW

338E6 Fab Light chain	middle Domain E	Contact
Asp60	Asp408	vdW
Ser76	Thr411	vdW
Asn77	Leu403	vdW

vdW: van der Waal contact (distance < 4 Å);
H: hydrogen bond (distance < 3.4 Å, *** and * denote high and low possibilities for bond formation);
SB: salt bridge (distance > 3.4 and < 4.5 Å, polar interaction between oppositely charged atoms).

effect, including V383A slightly increasing the affinity and Q385A and S499A slightly decreasing it (Fig. 2.5C, D).

To investigate how this interaction translates across the repertoire of VH family members, I selected representative mouse and human Fabs from different VH families and tested for binding with HCBD. Consistent with our repertoire sequencing studies, only mouse VH5, but not VH3 or VH14 and only human VH1 and VH3, but not VH4 germline specificities, could be observed. Intriguingly, no discernable differences in key residues in the aligned amino acid sequences from tested VH families were noticeable, therefore I investigated the structural details between available atomic models of these Fabs. Although no major structural differences are seen between binders and non-binders Ig, I hypothesized IbpA germline bias could be attributed to interface complementarity in this high-affinity interaction as many interactions are mediated by backbones. The binding energy is distributed across this vast interface such that mutagenesis of individual

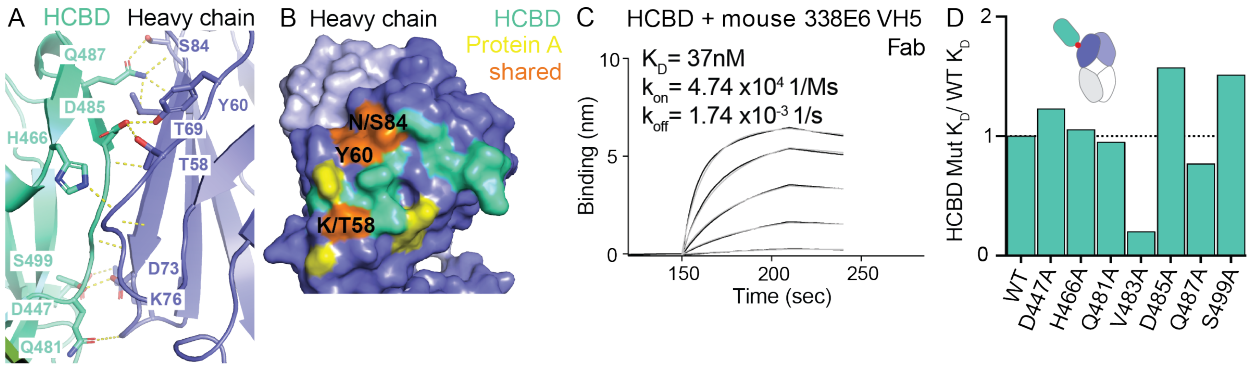

Figure 2.5 IbpA heavy chain binding domain interacts with framework region of mouse VH5 heavy chain through extensive contact sites
(A) Close up view of interface 1 between HCBD and mouse VH5. (B) Binding footprint of HCBD (greenteal), Protein A (yellow) and three shared sites between the two (orange), mapped on the surface representation of heavy (dark blue) and light chains (light blue) of the Fab. (C) BLI binding plot between immobilized wild-type HCBD and analyzed mouse VH5 Fab vs. time. (D) Alanine mutational analysis of measured K_D of HCBD contact residues using bar graph, normalized to HCBD WT.

residues result in only minor decrease in binding energy. Taken together, I show individual HCBD is able to mediate Fab strong interactions with heavy chain preference, but I cannot determine precisely how this specificity is conferred.

2.2.4 IbpA Domain D contacts the light chain of the Fab in complex structure

All previously described B cell superantigens mediate their binding to antibodies through 50-70 amino acid long domains that are repeated in their structures, which for Protein L was shown to confer an enhanced avidity towards their targets. To explore the role of Domain D, which has three other highly homologous repeats in the full length IbpA (Domains A, B and C), in Fab binding, I further examined its interface with the Fab in our crystal structure. While our repertoire data with full length IbpA suggests the bias in binding is mediated solely through heavy chain contacts, I wanted to establish whether this interface was real, or an artifact of the crystal packing lattice.

First, to determine any light chain repertoire preference for Domain D, Christoph Drees sorted mouse B cells stained with individual Domain D and again observed a restricted binding to VH5 and VH7 gene families, as was seen with the full length protein, but no preference for light chains. (Fig. 2.6A). I have also validated Domain D VH specificity to selected mouse and human Fab clones by direct binding measurement by BLI, and similarly to HCBD, in these experiments binding to mouse Fabs containing VH5 and human Fabs containing VH1 and VH3 domains was observed (Fig. 2.6B).

To further evaluate the Domain D binding interface with light chain, I generated hybrid Fabs that carried VK6-15 chain from Fab 338E6 used in complex crystal structure, where Domain D formed an interface with the light chain and swapped heavy chains from other VH families. Consistent

with our repertoire and binding studies, Domain D was bound only to Fabs containing the VH5 domain (Fig. 2.6C). These data in conjunction with the lack of kappa and lambda light-chain bias in IbpA-binding versus non-binding mouse and human B cells strongly suggest that Domain D binds to a heavy chain restricted B cell repertoire and interface between Domain D and the light chain is a crystal artifact.

Next, Christoph Drees investigated whether Domain D competes with HCBD for binding to VH by testing if staining of wild-type mouse B cells with labeled Domain D or HCBD could be blocked

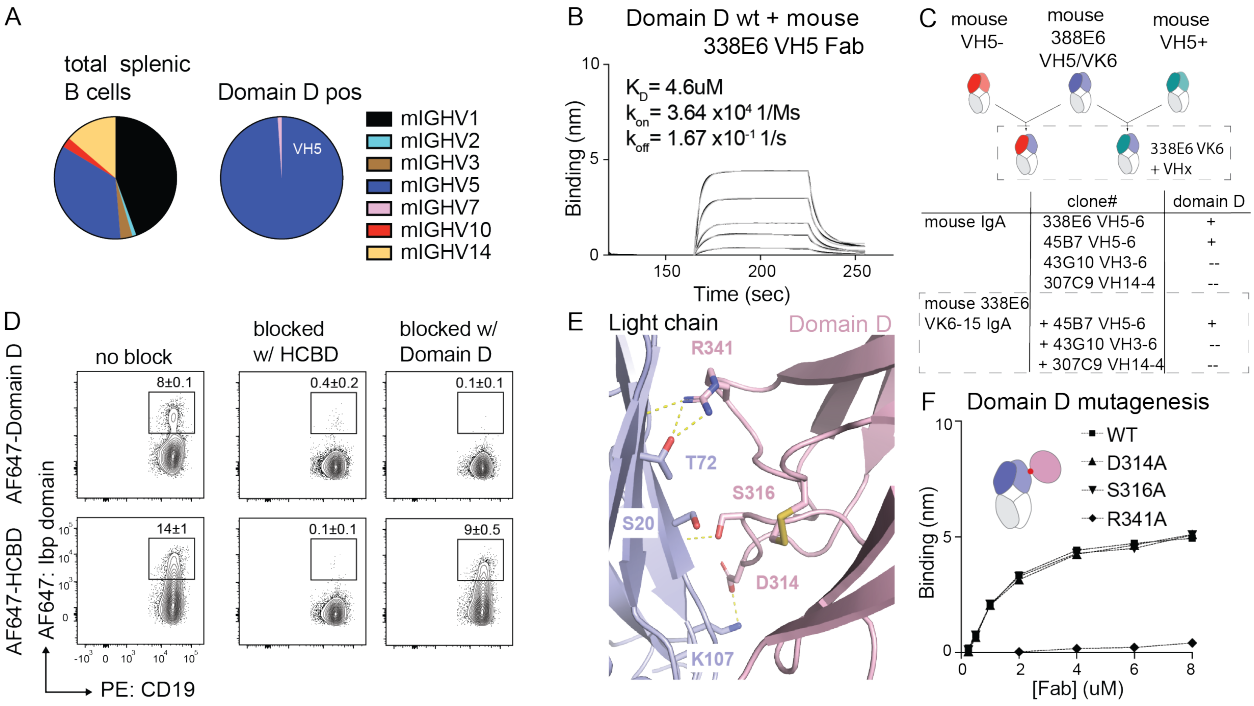

Figure 2.6 Characterization of IbpA repeat domain D contacts with Fab
(A) Heavy chain repertoire of sorted mouse CD19pos B cells binding to single repeat domain D. (B) BLI binding plot between immobilized wild-type Domain D and analyzed mouse VH5 Fab vs. time. (C) Experimental scheme of mouse Fabs with constrained light chain and swapped heavy chains (top), table reflecting measure binding between wild-type Domain D and swapped Fabs; "+" = 10^{-6}, "--" = no binding (bottom). (D) Representative contour plots show wild-type mouse B cells stained with HCBD or domain with and without preincubation with unlabeled HCBD (middle column) or Domain D (right column). (E) Close up view of crystal contacts between IbpA Domain D (lightpink) and VL of a Fab (lighblue). (F) Comparison plot of alanine substitutions in Domain D, directly involved residues in binding to Fab fragment vs. Fab concentration determined by BLI.

by preincubation with either purified Domain D or HCBD. If Domain D indeed binds light chains, we would expect a signal even after blocking with heavy-chain specific HCBD. However, Domain D staining could be blocked by both HCBD and Domain D indicating binding to a shared, heavy chain specific binding site (Fig. 2.6D). HCBD staining on the other hand was only reduced after preincubation with Domain D, consistent with lower affinity (K_D) and faster dissociation (k_{off}) of Domain D than HCBD to Fab (Fig. 2.6D). In addition, we saw that Domain D exclusively binds to IbpA-positive but not IbpA-negative human B cells. Interestingly, I mutated each of the three Domain D contact residues in the artifactual light chain interface: Asp314, Ser316 and Arg341 to Ala and observed that the R341A mutant abrogated binding to the Fab (Fig. 2.6E, F). Altogether, this data suggests that individual Domain D binds the same heavy chain repertoire as HCBD but with lower affinity. Exact identification of residues involved in that binding is unknown but interestingly, since this binding was almost completely abrogated with R341A mutated construct, this would suggest that minimally Arg341 is involved.

2.3 Stimulation with individual IbpA domains is sufficient to induce BCR signaling

Next, Christoph Drees compared the frequency of human and mouse B cells binding full length and truncated IbpA versus single IbpA domains. The frequency of wild-type mouse and human B cells labeled with truncated IbpA was only slightly diminished compared to wild-type IbpA (Fig. 2.7A). Single domains however showed decreased staining and even with higher concentrations Domain D would only stain 5-10% of total mouse B cells. Furthermore, staining was almost completely abrogated with R341A mutated construct (Fig. 2.7A). The same was true when binding to a homogenous VH6 expressing B cell population was examined. B cells could be brightly stained with full length and truncated IbpA, but the signal decreased from HCBD to Domain D

and was completely gone when in samples stained with Domain D R341. Furthermore, HCBD not only showed binding to a larger fraction of mouse B cells compared to Domain D but also induced calcium flux in wild-type and transgenic B cells (Fig. 2.7B, C). Domain D only induced a detectable calcium flux response when VH6 B cells were analyzed. The signal however was slower

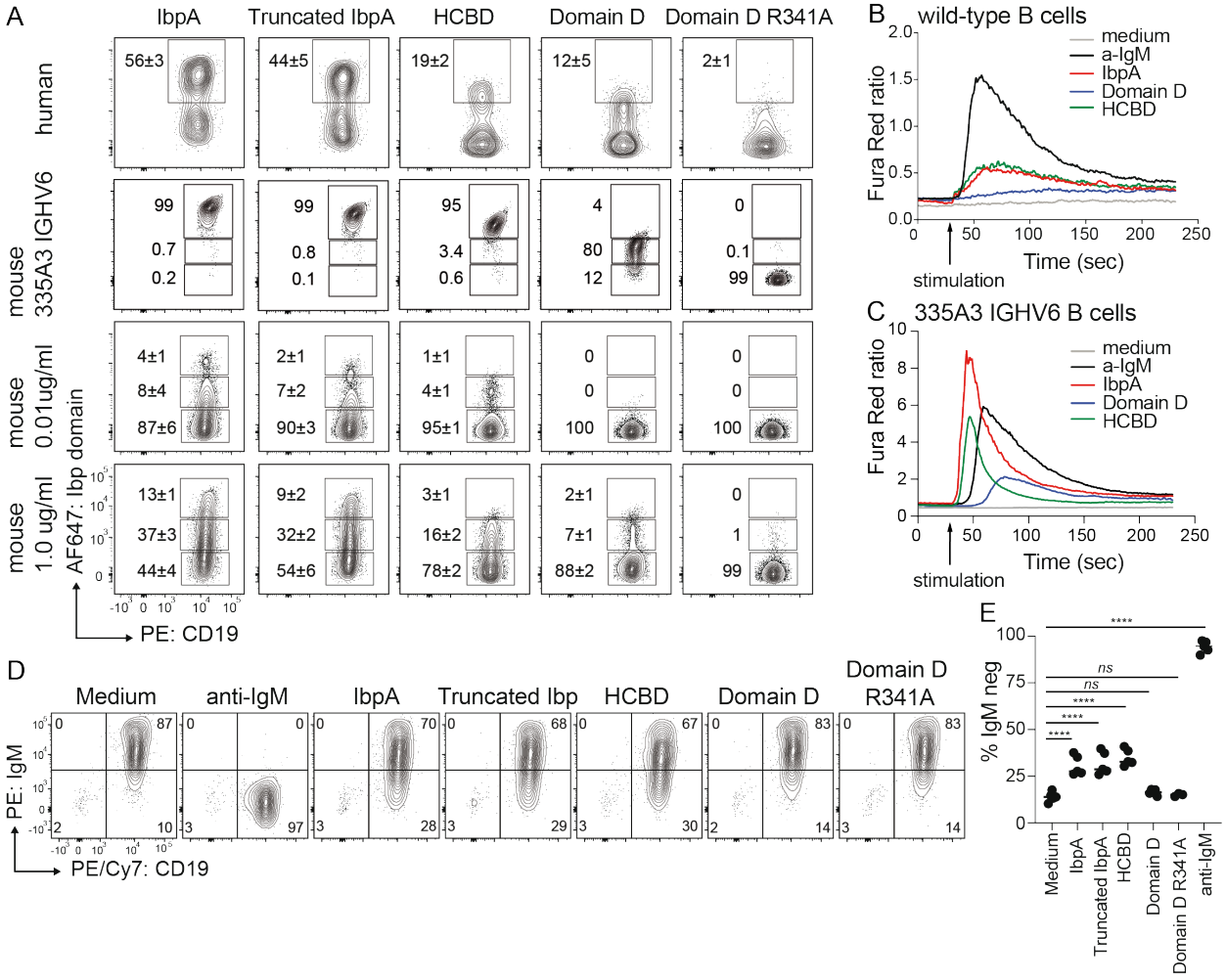

Figure 2.7 Functional characterization of single IbpA domains
(A) Contour plots show frequency of human and mouse B cells binding to fluorescently labeled full length, truncated and single IbpA domains. (B, C) Ca2+ flux analysis of wild-type (B) or VH6 transgenic (C) mouse B cell after *in vitro* stimulation with full length IbpA versus single IbpA domains. (D) Analysis of B cell receptor downregulation of MACS purified mouse B cells after two hours *in vitro* stimulation with single IbpA domains compared with full length and truncated IbpA. (E) Quantification of B cell downregulation with tested IbpA constructs.

and diminished compared to HCBD (Fig. 2.7C). This could be contributed to longer dwelling of HCBD on BCR as observed by higher affinity and slower dissociation in BLI studies (Fig. 2.5C). In addition, surface IgM downregulation occurred to a similar degree on wild-type B cells stimulated with full length IbpA, truncated IbpA and HCBD but not with single Domain D (Fig. 2.7D, E).

2.4 Role of Domain D and HCBD in IbpA interaction with Immunoglobulins

The convention for most receptor signaling models involves cross-linking of receptors, either through engagement of a multivalent or multimerized ligand. Here, we have shown that a single domain of IbpA, the HCBD, can activate B cells through their BCR. To explore whether the HCBD itself can exist as a multimer, I investigated the multimeric states of this domain, as well as for truncated and full length IbpA. Using Multi-Angle Light Scattering (MALS), I show that the HCBD exists exclusively in a monomeric state, even at high concentrations (Fig. 2.8A). To explore the stoichiometry of HCBD binding to Fab, I employed Isothermal Titration Calorimetry (ITC), which directly measures the thermodynamics of binding but can also estimate the stoichiometry of binding. My results approximate a 1:1 binding stoichiometry (N=1.2, Fig. 2.8B) between the HCBD and 338E6 VH5 Fab from our structure. This supports a model whereby soluble HCBD is monomeric yet is sufficient to induce signaling in B cells.

To further explore the stoichiometry of Fab binding of the full length IbpA, in collaboration with Alex Yarawsky from Dr. Andrew Herr's Lab, we performed a series of sedimentation velocity analytical ultracentrifugation (AUC) experiments at increasing molar ratios between full length IbpA and 338E6 VH5 Fab. Alex's titration experiments reveal that full length IbpA mixed with

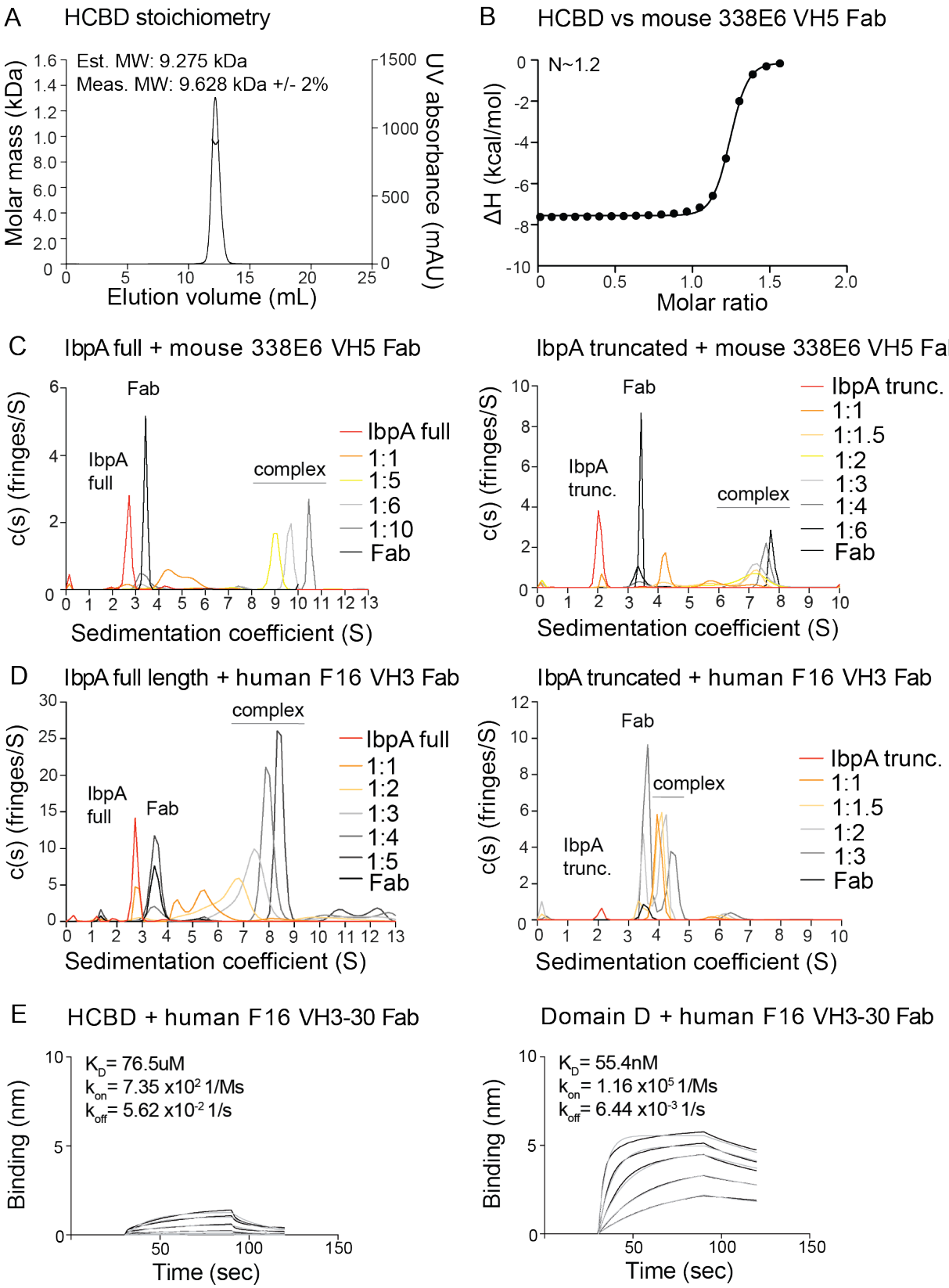

Figure 2.8 Role of multiple binding sites in full length and truncated IbpA *(continued on the next page)*

Figure 2.8, continued: (A) Stoichiometry of individual HCBD by MALS. (B) Stoichiometry of HCBD and Fab complex by ITC. (C) Stoichiometry of full length (left) or truncated (right) IbpA and mouse 338E6 VH5 Fab complex by AUC. (D) Stoichiometry of full length (left) or truncated (right) IbpA and human F16 VH3 Fab complex by AUC. (E) BLI binding plot between immobilized wild-type HCBD (left) or Domain D (right) and analyzed human F16 VH3 Fab vs. time.

five molar excess of the Fab results in essentially a single peak (Fig. 2.8C, left). To determine molecular weight and stoichiometry of that complex, Alex collected data for full length IbpA:Fab at 301nm wavelength. After calculating the absorbance per concentration of Fab (the absorbance from IbpA was negligible) and the area under the peak of ~9.7 sedimentation coefficient S, we determined an estimate of the molecular weight to be ~343 kDa with a frictional ratio of 1.95 (a frictional ratio of 1.2-1.4 is expected for globular proteins). This MW estimate lies between a 1:5 and a 1:6 complex, consistent with four available repeat domains and fifth HCBD binding site. By contrast, the truncated IbpA preincubated with three molar excess of the Fab forms a peak at ~7.2 S (~90% of the total signal) (Fig. 2.8C, right) and using similar analysis to the full length complex, the weight-averaged frictional ratio was estimated to be 1.58 and molecular weight of ~190 kDa – similar to the estimated molecular weight for a 1:3 complex based on amino acid sequence. Thus, truncated IbpA may potentially have an additional Fab binding site not apparent in the full length IbpA or form a heterogenous complex, both hypotheses need further investigation. We have additionally tested stoichiometry of full length and truncated IbpA with human F16 VH3 Fab. Conversely to our results with mouse Fab, we observed 1:4 and 1:1 stoichiometries of F16 Fab to full length and truncated IbpA respectively (Fig. 2.8D). This result supports my BLI binding studies, where we observed Domain D and not HCBD binding to F16 Fab with higher affinity. Taken together, I hypothesize that both Domain D and HCBD are complementary in binding to mouse and human Fabs and given potential flexibility of the truncated IbpA, the observed higher

stoichiometry for mouse Fab could be indicative of crosslinked binding, rather than formation of a homogenous complex.

Due to the confounding results generated by the AUC measurements I sought to investigate, at low-resolution, the solution structure of the apo full length IbpA by small-angle X-ray scattering (SAXS). Similar to earlier studies of truncated IbpA, I used MultiFoXS, stipulating flexibility in the six linker regions inbetween all domains (average $\chi^2 = 0.49$). Although some flexibility was present in the more external domains (the unknown N-term domain in black and HCBD in greenteal) of the full length IbpA, SAXS analysis of the entire IbpA supported a mostly extended conformations ($f/f_0 \sim 1.7$) with a more arched architecture, suggesting some cooperativity between the linked domains. Only one of the predicted full length IbpA conformations was able to accommodate all five Fab molecules concurrently, forming a tightly coordinated complex. Taken together, both truncated and full length IbpA structures show some level of flexibility, but the truncated construct maintains its flexibility while binding to Fabs.

I sought to investigate how different affinities of Domain D and HCBD affect Ibp binding toward a full length antibody (mAb). I tested for oligomerization between Ibp variants and two studied mAb: mouse 338E6 and human F16 by BLI. I immobilized mAb on native Protein A tip followed by alternating analytes Ibp (truncated or full length) then mAb for three rounds. My previous BLI and AUC binding studies for 338E6 Fab showed that HCBD dominates over Domain D, but still some low affinity participation from Domain D is present. I have therefore observed continuous oligomerization for truncated Ibp and 338E6 mAb, but only initial oligomerization that would reach saturation for the full length Ibp (Fig. 2.9A). This was in contrast with F16 mAb, where

Domain D is a dominant binding site for Fabs without support from HCBD according to our BLI and AUC data. Here I see no oligomerization for truncated Ibp and mAb consistent with a singular binding site coming from Domain D and continuous oligomerization for the full length Ibp and mAb, consistent with four available repeat domains (Fig. 2.9B). My structural modelling simplified with cartoons in Fig. 2.9 show proposed models of full length antibodies binding to Ibp variants and forming varying degrees of oligomerization that is dependent on binding strengths of Domain D (and other repeats) and HCBD.

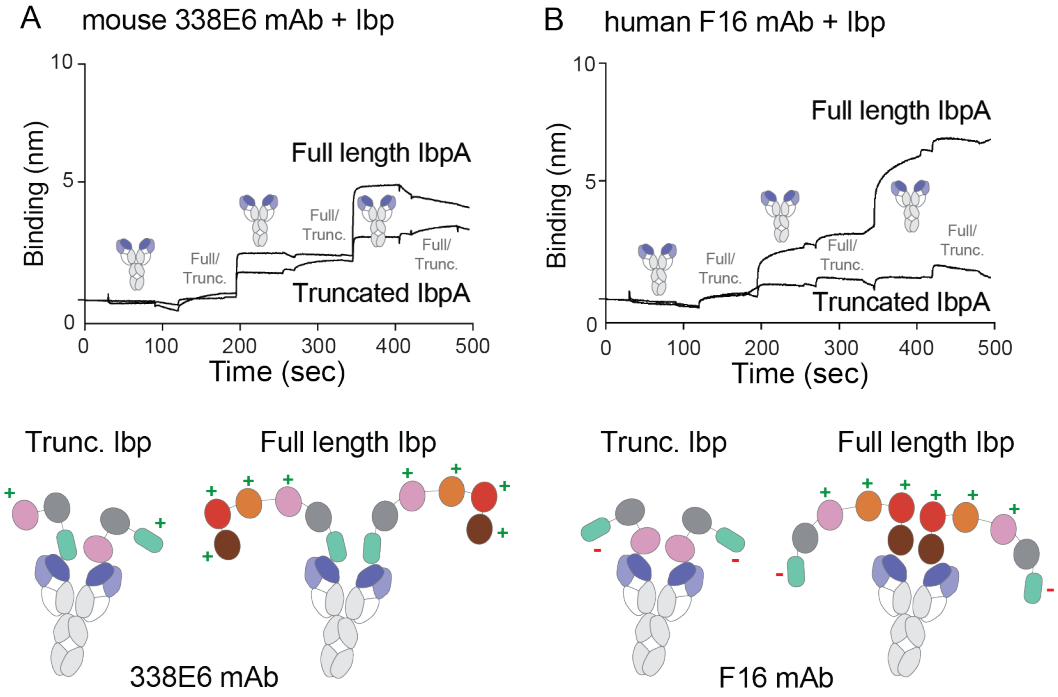

Figure 2.9 Proposed model of binding of truncated and full length Ibp to two discrete clones

(A) Multimerization experiment by BLI of mouse 338E6 full length antibody binding to either truncated or full length Ibp (top); proposed cartoon model of 338E5 mAb shows high levels of binding to both Ibp variants (bottom). (B) Multimerization experiment by BLI of human F16 full length antibody binding to either truncated or full length Ibp (top); proposed cartoon model of F16 mAb shows high levels of binding only to full length but not truncated Ibp (bottom).

2.5 Discussion

In this work, with help from functional data by Christoph Drees, I characterize a novel immunoglobulin binding protein expressed by a commensal *R. gnavus* with superantigen-like properties *in vitro*. On a population level, we provide detailed repertoire analysis of sorted IbpA-binding mouse and human B cells confirming high affinities to selected VH families: mouse VH5/6/7 and human VH3 families, and do not detect any associated light chain bias. However, we also detected a range of mouse BCRs, mostly expressing VH1 families, that bound IbpA with lower affinities.

To uncover the molecular basis for this binding, I determined the structure of IbpA bound to Fab revealing that truncated IbpA primarily engages mouse VH5 Fab with two separate domains. First, a C-terminal domain called HCBD that is able to strongly bind to VH region. Second, Domain D which binds weakly to potentially overlapping VH site. Even though crystal interface suggested interaction between Domain D and Fab light chain, further investigations revealed a heavy chain restricted binding pattern. In concert with the number of contacts at each interface, Fab binding was abrogated by mutating a single Domain D residue R314A, whereas binding to HCBD was intact by targeted mutagenesis of individual HCBD residues. The exact molecular basis for VH germline recognition could not be determined in this study as no amino acid sequences or structural differences between different VH families were apparent. Additionally, Domain D and HCBD having inconsistent affinities for different mouse and human clones, sequence and structure-based analysis of Ibp germline bias should be further expanded and tied with more binding studies. Interestingly, our findings for mouse system emphasize, HCBD alone, with its stronger affinity and slow dissociation rate, outcompetes the Domain D binding and elicits vast mouse BCR

activation *in vitro* without self-multimerization. This remarkable monovalent ligation of BCR without crosslinking has been also observed for a certain HEL specific BCR (Kim et al. 2006).

Having identified two binding sites on a single truncated IbpA molecule in my structure, I examined its binding capabilities by BLI and through collaborators also AUC and compared it with wild-type IbpA. I have observed that truncated IbpA has higher affinity and can bind three mouse Fabs which is more than its number of available binding sites, whereas wild-type IbpA binds five, equal to four repeat domains and fifth HCBD site. This greater valency could be explained by greater flexibility of truncated IbpA in complex with Fab as compared to full length IbpA-Fab complex as shown by our SAXS data. Same experiment with human Fab gave 1:1 stoichiometry with truncated IbpA and 1:4 for the full length, consistent with engagement of repeat domains and not HCBD. Whether there is a possible coordination between all repeat domains that constrains movements of full length IbpA remains to be investigated. My oligomerization experiment with mAbs suggests varying degrees of daisy-chaining that is tuned by binding strengths of Domain D (and other repeats) and HCBD in Ibp. This subtle, but interesting observations, could be a factor for *R. gnavus* to disrupt the commensal balance, by changing its magnitude of IgA coating. As IgA can influence microbial populations through many mechanisms, further studies should asses metagenomic heterogeneity of Ibp genes (especially mutations withing repeat and HCBD domains) in *R. gnavus* and its functional consequences. For example, I hypothesize IbpA/B to be sortase-mediated surface expressed, but presence of secreted forms of IbpA/B cannot be excluded. The soluble IbpA would also provide an interesting possibility for it to reach and stimulate immune cells typically excluded from mucosal surfaces in healthy individuals.

Chapter 3 – Biochemical and biophysical characterization of natural polyreactivity in antibodies

3.1 Low affinity interactions between polyreactive Fabs and diverse ligands

The specificity/promiscuity trade-off of immune receptors is a fine line in immunology. It is generally thought that development of high affinity, and high specificity, antibody response to a target is the main goal of the adaptive immune response, but recently more research is focused on a large range of antibodies that also have lower affinity, more promiscuous binding. Typically, promiscuity in antibodies are associated with undesired off-target effects, but in this chapter, I aim to propose that the lower affinity interactions in antibodies are very much intended and valuable part of the immune recognition. Therefore, I have investigated the molecular basis of the polyreactivity.

3.1.1 Direct affinity between polyreactive Fab and diverse ligands

The major difference between polyreactive and monoreactive antibodies is that polyreactive mAbs are able to bind, with very low affinity, to multiple diverse ligands. Low affinity interactions are difficult to detect by standard biochemistry techniques, especially when measuring direct monomeric binding events. Therefore, there have not been any publications, to my knowledge, that reported direct affinity measurements for polyreactive antibodies, posing an interesting question: what is the affinity threshold for polyreactive antibodies?

For my binding studies I focused on selected, highly polyreactive mouse IgA mAbs identified from a thousand of clones screened through a high throughput polyreactive ELISA in *Bunker et al. 2017*

(Fig. 3.1A). For monomeric binding studies, I designed scFv and Fab constructs from selected IgA mAb clones, but only one in six expressed in bacteria. New constructs for expression in insect cells

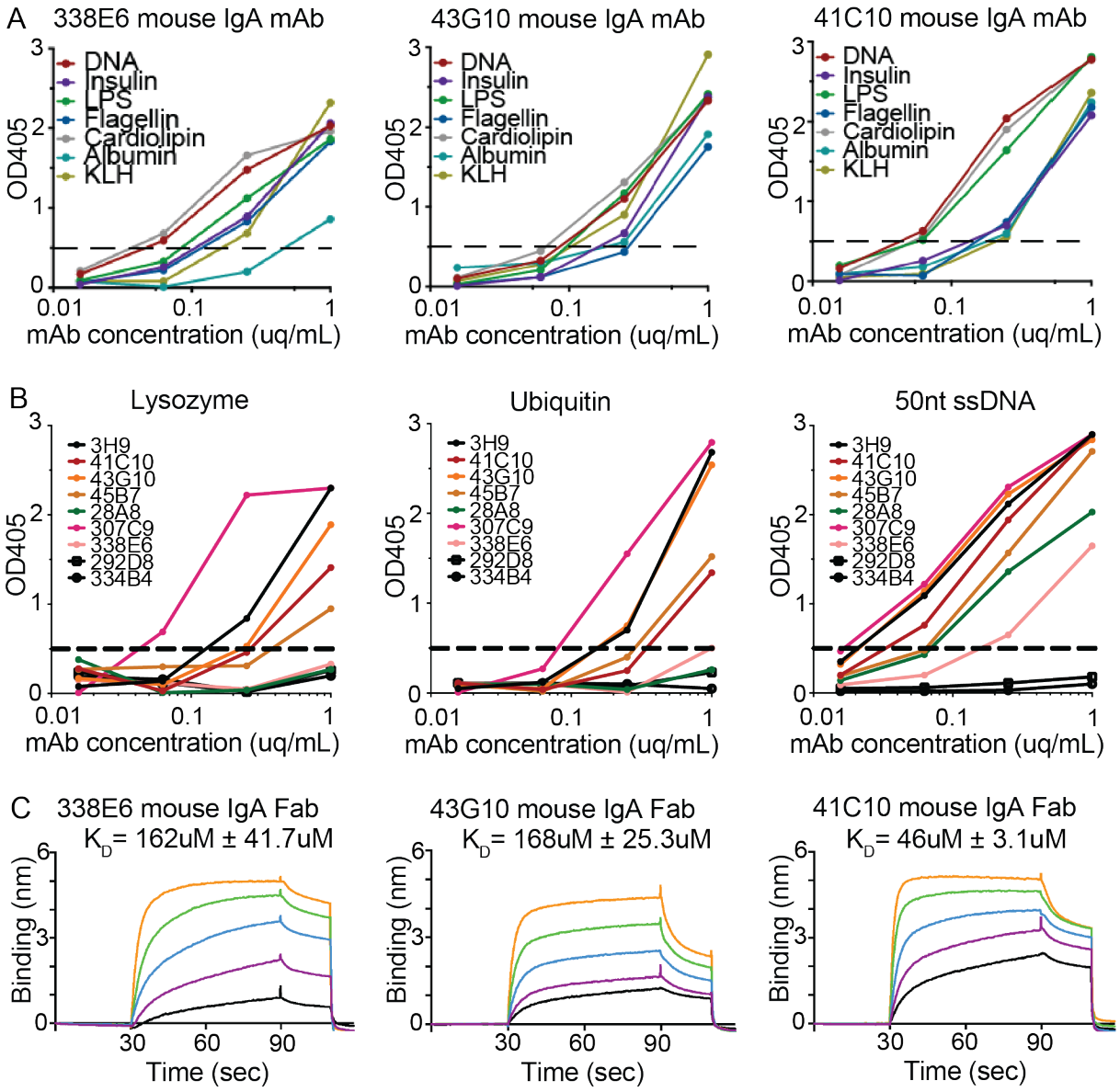

Figure 3.1 Binding of selected polyreactive mouse IgA to polyreactive ligands
(A) Polyreactivity ELISA OD405 values of selected mouse IgA mAbs tested at different concentrations. (B) New ligands: lysozyme, ubiquiting and 50nt ssDNA, tested with Polyreactivity ELISA showing OD405 values for selected mouse IgA mAbs tested at different concentrations. Marked in black are positive (3H9) and negative controls (292D8, 334B4). (C) BLI binding plot between immobilized biotinylated 25nt ssDNA and sselected polyreactive mouse IgA Fabs vs. time.

as secreted Fabs, were stabilized with charge complementary leucine zippers and yielded between 2-20mg/L culture for each of six clones. For ligands, I chose to study well characterized and smaller molecules, that are also good targets for crystallization studies, this included insulin, lysozyme, ubiquitin and short ssDNA. Our collaborator Jeff Bunker confirmed binding of our candidate mAb clones to new ligands, lysozyme, ubiquitin and homogenous 50nt ssDNA, using standard mAb ELISA and as expected we observed similar level of binding to new ligands as we have seen for polyreactive ligand panel (Fig. 3.1B). My attempts to form complexes between these ligands and polyreactive Fabs by size exclusion chromatography were unsuccessful and evaluating the binding of these Fabs to immobilized polyreactive ligands by BLI was quite challenging. No binding was observed between the polyreactive Fabs and lysozyme, ubiquitin or even di-ubiquitin ligands using this method. However, I observed binding to biotin immobilized 25nt long ssDNA with six different polyreactive mouse Fab clones and no binding for control non-polyreactive mouse Fab (Fig. 3.1C). A range of affinities between 50-170uM was observed, which is comparable to an average TCR-MHC affinity.

3.1.2 Single-alanine mutagenesis of polyreactive Fabs

Next, I was interested what residues in polyreactive antibodies are involved in binding to diverse ligands. This experiment was done by Jeff Bunker with help from other members of Dr. Albert Bendelac's Lab. Because I could only confirm direct binding of Fab to one ligand, we reverted to mutate single CDR residues to alanine in the full length polyreactive mAbs of the previously mentioned six Fab clones using ELISA. Previously mentioned reports suggested a use of tyrosine, arginine, heavy chain and long CDR3 loops to mediate polyreactive interactions, but no such patterns were observed (Fig. 3.2A-C). Some tyrosine or arginine single mutations are shown to

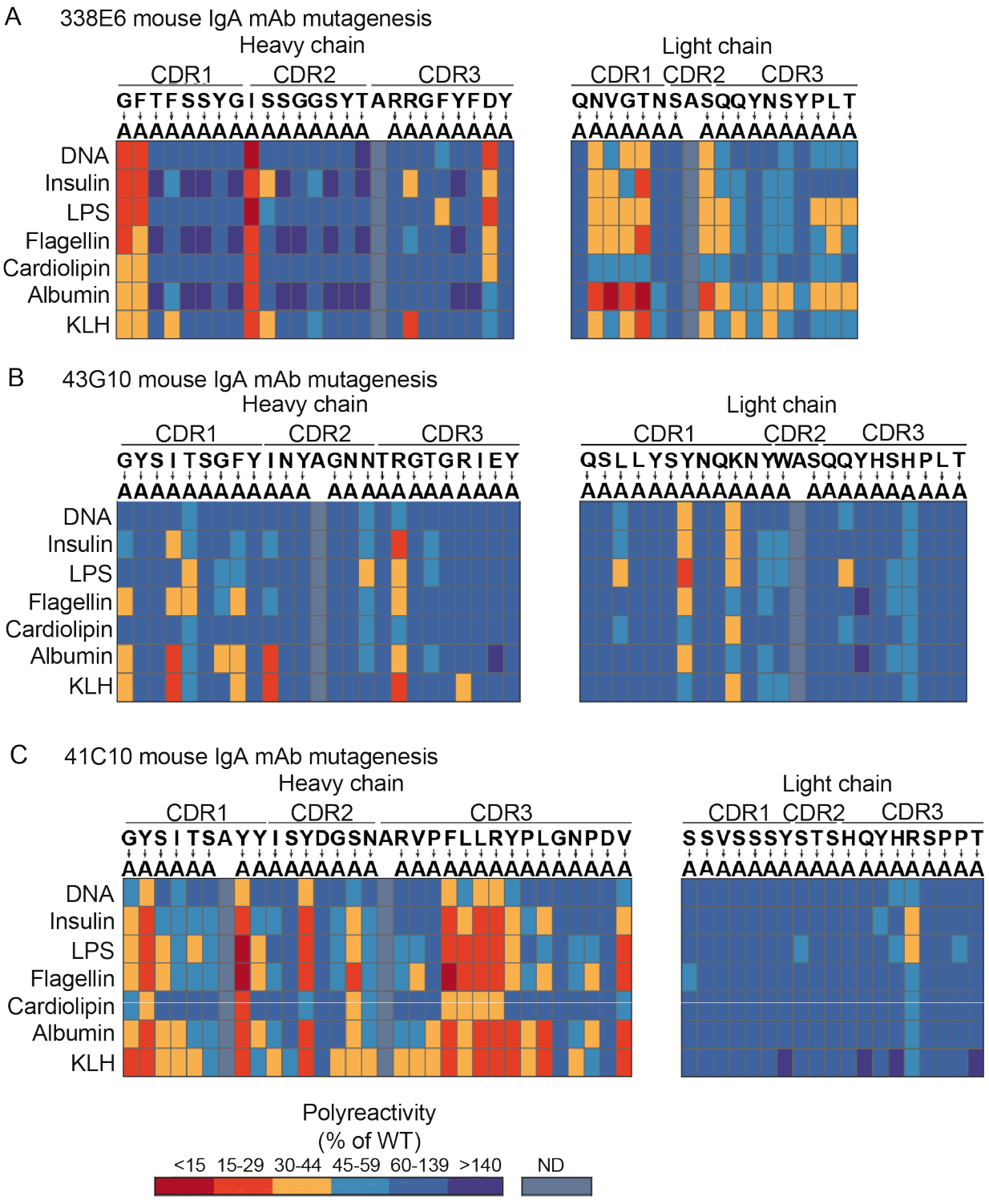

Figure 3.2 Single alanine scanning of polyreactive mouse IgA mAbs and their reactivity to ligands

Summary of polyreactivity ELISA of selected mouse IgA mAbs that were single-alanine mutated, shown as % of wild-type mAb: (A) 338E6 mAb, (B) 43G10 mAb and (C) 41C10 mAb.

disrupt binding to all ligands for certain clones, but overall different clones would utilize different mutations to disrupt their polyreactive binding. In other words, no individual loop or individual chain or preference in residues is responsible for polyreactive binding across tested clones. However, the same residues for individual clones, are most likely involved in binding to multiple ligands. It should be noted, that this analysis doesn't test for backbone interactions as well as interactions that require multiple simultaneous mutations to abrogate the binding.

3.2 Structure based analysis of polyreactive antibodies

3.2.1 Conventional structural analysis of polyreactive antibodies

Although machine learning studies that characterize polyreactivity proved to be promising, conventional sequence-based comparison between polyreactive and monoreactive antibodies was unsuccessful as shown by multiple groups (Mouquet et al. 2010, Bunker et al. 2017, Boughter et al. in review). Therefore, I searched for clues in the structures of polyreactive antibodies, which included the aforementioned polyreactive mouse IgA's as well as human anti-Flu broadly neutralizing IgG obtained from our collaborators in Dr. Patrick Wilson's Lab, that were also shown to be polyreactive (Andrews 2015, Guthmiller in print). I solved a total of five high resolution crystal structures: three of polyreactive Fabs (left) and two monoreactive Fabs (right) by X-ray crystallography (Fig. 3.3A, Table 3). It is important to compare polyreactive Fabs to their monoreactive counterparts, coming from closely related source (human and mouse) and specificity (anti-flu blood IgG and gut homeostatic IgA). Interestingly, the Fab structures ranged from having short CDRs forming a flat antigen binding surface (polyreactive 338E6 and monoreactive 3B03) to longer CDRs, extending the antigen binding site (polyreactive 43G10 and 2G02).

Table 3 Data collection and refinement statistics of polyreactive and monoreactive Fabs apo

	338E6 Fab	43G10 Fab	2G02 Fab	3B03 Fab	4C05 Fab
Wavelength					
Resolution range	40.4 - 1.828 (1.893 - 1.828)	57.88 - 1.634 (1.692 - 1.634)	61.22 - 1.67 (1.73 - 1.67)	64.12 - 1.632 (1.69 - 1.632)	39.54 - 1.769 (1.832 - 1.769)
Space group	P 1 21 1	P 21 21 21	P 1 21 1	P 1	C 1 2 1
Unit cell	40.92 59.65 82.2502 90 99.1 90	62.2399 108.47 115.75 90 90 90	70.1598 59.44 99.2799 90 98.8301 90	52.4198 66.17 90.2383 104.09 93.31 91.6	79.8391 70.6089 87.6797 90 108.95 90
Total reflections	1092228 (106303)	2239885 (187748)	175059 (16892)	903633 (90390)	44162 (4399)
Unique reflections	34044 (3308)	97569 (9583)	92960 (9082)	132793 (12999)	
Multiplicity	32.1 (31.8)	23.0 (19.5)	1.9 (1.9)	6.8 (6.8)	
Completeness (%)	97.36 (95.41)	99.33 (95.12)	96.38 (87.99)	88.47 (88.66)	98.04 (98.77)
Mean I/sigma(I)	21.54 (4.96)	18.90 (0.55)	7.22 (2.10)	27.55 (1.42)	
Wilson B-factor	14.93	26.58	16.28	21.81	18.35
R-merge	0.7018 (1.132)	0.7845 (7.964)	0.1468 (1.662)	0.5797 (1.552)	
R-meas	0.7131 (1.15)	0.8025 (8.182)	0.2076 (2.35)	0.6379 (1.698)	
R-pim	0.1241 (0.2031)	0.1657 (1.844)	0.1468 (1.662)	0.2612 (0.6786)	
CC1/2	0.914 (0.721)	0.741 (0.0942)	0.943 (0.171)	0.648 (0.318)	
CC*	0.977 (0.915)	0.923 (0.415)	0.985 (0.541)	0.887 (0.695)	
Reflections used in refinement	33881 (3308)	96977 (9162)	90525 (8203)	129061 (12973)	44160 (4399)
Reflections used for R-free	1696 (168)	4874 (445)	4495 (406)	2000 (196)	2104 (193)
R-work	0.2114 (0.2585)	0.2197 (0.4006)	0.2564 (0.2926)	0.2087 (0.3306)	0.1858 (0.2207)
R-free	0.2465 (0.3138)	0.2378 (0.4037)	0.2961 (0.3662)	0.2355 (0.3350)	0.2135 (0.2463)
CC(work)	0.748 (0.549)	0.796 (0.372)	0.920 (0.537)	0.797 (0.623)	
CC(free)	0.659 (0.517)	0.797 (0.251)	0.878 (0.399)	0.778 (0.641)	
Number of non-hydrogen atoms	3794	3564	6498	7475	3689
macromolecules	3292	3391	6498	6523	3281
solvent	502	173		952	408
Protein residues	432	441	859	863	433
RMS(bonds)	0.009	0.007	0.009	0.008	0.007
RMS(angles)	1.34	0.92	1.05	1.3	1.29
Ramachandran favored (%)	94.37	97.46	95.96	93.8	97.2
Ramachandran allowed (%)	5.16	2.31	2.97	4.91	2.33
Ramachandran outliers (%)	0.47	0.23	1.07	1.29	0.47
Rotamer outliers (%)	2.13	0.77	0.28	0	0
Clashscore	4.01	1.94	7.63	7.79	5.56
Average B-factor	16.71	37.71	22.38	30.4	25.2
macromolecules	16.37	37.52	22.38	29.3	24.26
solvent	18.88	41.52		37.96	32.75
Number of TLS groups	28	1	1		

(Statistics for the highest-resolution shell are shown in parentheses.)

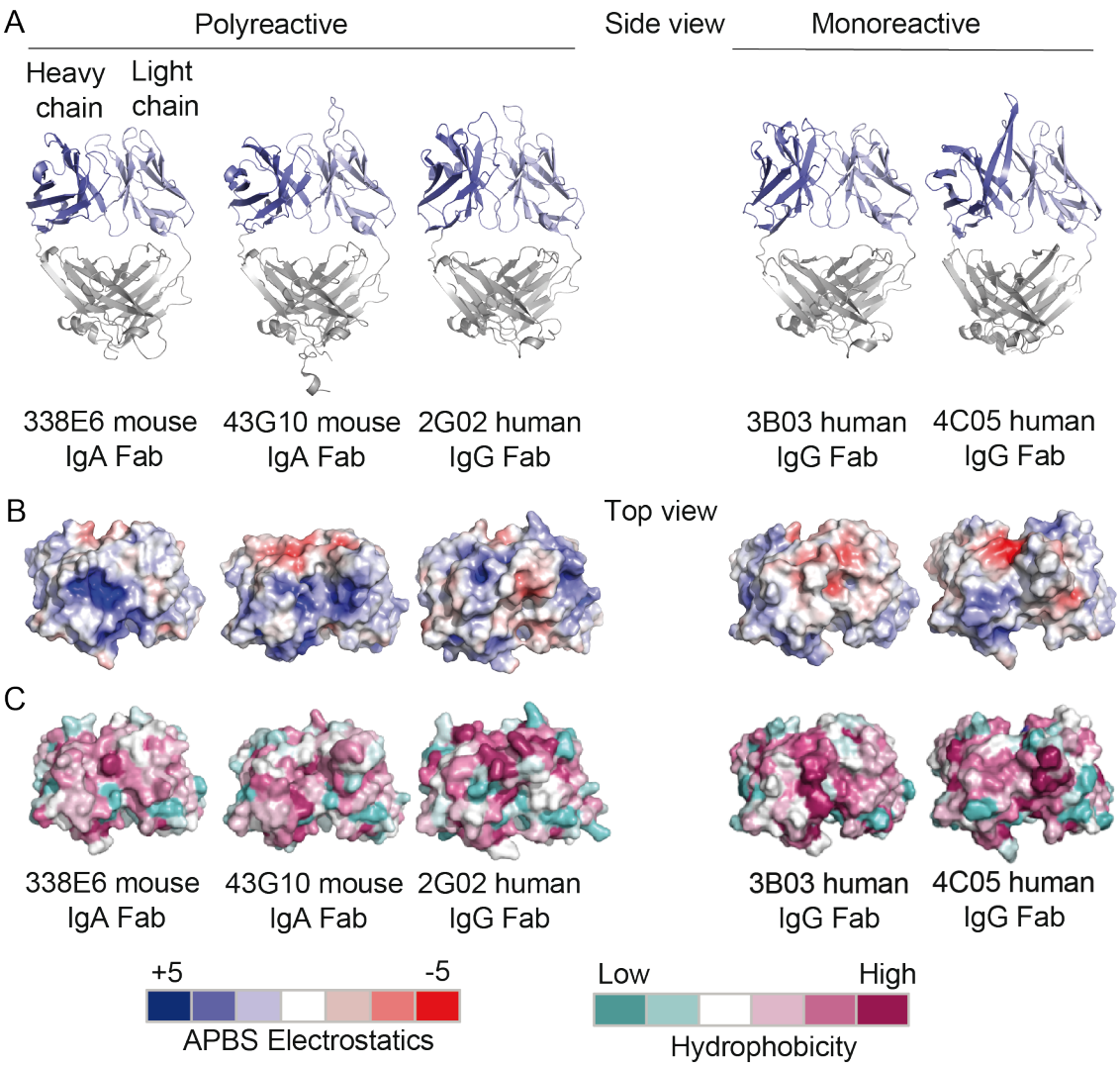

Figure 3.3 Charge and hydrophobicity analysis of crystal structures of polyreactive and monoreactive Fabs
(A) Cartoon crystal structures of polyreactive mouse 338E6, mouse 43G10, human 2G02 and monoreactive human 3B03, human 4C05 Fabs. (B) Top surface view of five Fab structures highlighting APBS electrostatics of their antigen binding site. (C) Top surface view of five Fab structures highlighting Eisenberg scale of hydrophobicity of their antigen binding site.

Interestingly, CDR H3 on monoreactive 4C05 shows a structured β ribbon similar to those seen in a subset of bovine Fabs with long stalk atop which sits a knob domain (Stanfield et al. 2016). To investigate previous suggestions that polyreactive Fabs have overwhelming positive antigen binding surface, I inspected the antigen binding site of solved crystal structures for any

hydrophobic or electrostatic patches in a surface view (Lercef, 2019). No predominant pattern was present in terms of charge or hydrophobicity when compared between poly- and closely related monoreactive Fabs (Fig. 3.3B, C). Neither was any pattern observed in the ten additional structures of polyreactive anti-Flu, selected from the protein structure database. In conclusion, polyreactivity cannot be simply determined by just looking at sequence or a static structure.

3.2.2 Structural determination of polyreactive Fab in complex with its ligands by crystallography and SAXS

All published reports on polyreactivity in antibodies focus exclusively on unligated antibodies, and more specifically on CDR H3 as a region defining high-affinity and by extent promiscuity. However, not knowing where the ligand is binding on polyreactive antibodies can lead to misinterpretation of the results. Therefore, I believe it is crucial to study molecular basis of polyreactivity in regard to the antibody-ligand complex. Most experimental techniques to find the binding site i.e. ligand docking, NMR fails short because of low affinity of this interaction. I have therefore attempted to solve crystal structures of complexes between polyreactive Fabs and small ligands: DNA, ubiquitin, insulin and lysozyme. As stoichiometric complexes could not be formed on size exclusion chromatography, my initial attempts included mixing both components in 1:1 molar ratio, but highly crystallizable ligands lysozyme and insulin prevented me from obtaining complex crystals. Further efforts included soaking polyreactive Fab crystals with 5nt ssDNA ligand that is small enough to penetrate solvent channels formed within a crystal, however, this strategy requires antigen binding site on Fab to be accessible to solvent and free from crystal contacts. Ultimately, I covalently fused lysozyme or insulin (ubiquitin would act as a signal for degradation in insects, therefore was excluded) ligands with 15nt long glycine-serine linker to the N-terminus of the heavy chain of the Fab, which resulted in all constructs successfully expressed. This approach was successful in producing diffracting crystals and crystal structures of two

Table 4 Data collection and refinement statistics of polyreactive Fabs in complex with ligands

	scLYS33	scLYS43
Wavelength		
Resolution range	61 - 1.607 (1.664 - 1.607)	79.43 - 1.685 (1.745 - 1.685)
Space group	P 21 21 21	P 21 21 21
Unit cell	59.844 65.3 171.029 90 90 90	62.706 107.545 117.824 90 90 90
Total reflections	581212 (53763)	598123 (54509)
Unique reflections	87746 (8484)	89035 (8253)
Multiplicity	6.6 (6.3)	6.7 (6.6)
Completeness (%)	99.32 (96.00)	98.05 (92.21)
Mean I/sigma(I)	9.44 (0.84)	15.70 (1.74)
Wilson B-factor	23.15	27.84
R-merge	0.1581 (1.614)	0.07144 (0.981)
R-meas	0.1716 (1.758)	0.07748 (1.064)
R-pim	0.06598 (0.6871)	0.02963 (0.4067)
CC1/2	0.981 (0.472)	0.996 (0.873)
CC*	0.995 (0.801)	0.999 (0.965)
Reflections used in refinement	87504 (8367)	88835 (8243)
Reflections used for R-free	2000 (199)	2000 (186)
R-work	0.2159 (0.2975)	0.2452 (0.3090)
R-free	0.2596 (0.3270)	0.2659 (0.3022)
CC(work)	0.928 (0.710)	0.935 (0.872)
CC(free)	0.916 (0.572)	0.961 (0.870)
Number of non-hydrogen atoms	4827	4453
macromolecules	4310	4453
solvent	517	
Protein residues	564	578
RMS(bonds)	0.006	0.007
RMS(angles)	0.88	1.22
Ramachandran favored (%)	96.77	96.15
Ramachandran allowed (%)	2.69	3.5
Ramachandran outliers (%)	0.54	0.35
Rotamer outliers (%)	0	0.2
Clashscore	2.36	2.96
Average B-factor	28.6	53.21
macromolecules	27.89	53.21
solvent	34.57	
Number of TLS groups	1	

(Statistics for the highest-resolution shell are shown in parentheses.)

complexes were solved: lysozyme+338E6 Fab and lysozyme+43G10 Fab (Table 4). Both complex structures showed multiple ligands in proximity of the antigen binding site of a single Fab molecule, making the correct ligand molecule hard to identify (Fig. 3.4A). Given the low affinity of these complexes, I sought to further validate these interfaces. Among different biophysical techniques i.e. NMR, EPR and availability of the equipment, I focused on using SAXS to verify interfaces in the crystal complexes by fitting experimental and model values and assessing χ^2 based on I(q) plots that were then fit with experimental values collected from single-chain complex

samples at concentrations ranging from 10-20mg/mL. I created multiple models with 1:1 (Fab:ligand) stoichiometries based on complex. First analyzed complex, 43G10 Fab + Lysozyme, contacts two lysozyme molecules in the crystal, positioned on almost mirror opposites of each other on the Fab. However, SAXS analysis gave light chain associated Lys2 molecule a slightly better score (χ^2 = 0.49). The second analyzed complex, 338E6+Lysozyme, shows three possible ligand positions in the structure. However, Lys1 molecule gave the highest score, χ^2 = 1.28, and positions the lysozyme molecule around the heavy chain of 338E6 Fab. It should be noted that,

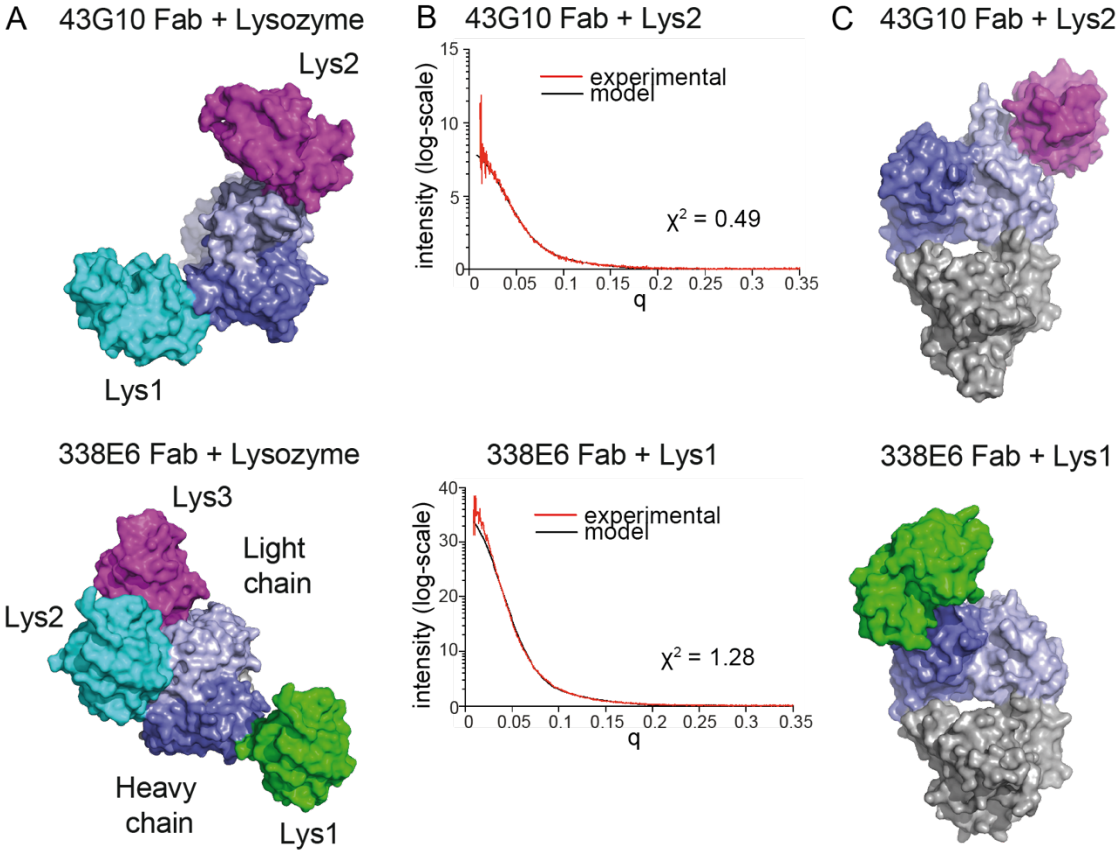

Figure 3.4 SAXS analysis of complex crystal structures between polyreactive Fabs and different ligands

(A) Top surface view of complex crystal structures of: 43G10 Fab with two Lysozyme molecules (top) and 338E6 Fab with lysozyme (bottom). (B) Experimental fit of SAXS data of both Fab complexes using FoXS and plotted by I(q). (C) Side surface view of two Fab + lysozyme complex structures that show the highest χ^2 fit scores.

although 338E6 Fab is highly polyreactive and protein concentration within crystal is much higher than used in our assays, it did not show binding to lysozyme on ELISA within tested concentrations (Fig. 3.1B).

Altogether, the complex crystal structures together with SAXS low resolution validation suggest different binding modes between different clones and their ligands (Fig. 3.4C). Both structures show Lysozyme locating to either the light chain (in 43G10) or heavy chain (in 338E6) and contact germline region of the Fab. However, more precise techniques should be employed to validate studied complex structures as well as identify exact residues that participate in the binding. This analysis supports structures and our earlier mutagenesis data, where we see different polyreactive clones engage their ligands with different residues and parts of the antibody.

3.3 Dynamic analysis of polyreactive antibodies

3.3.1 All-atom Molecular Dynamics simulations of polyreactive and monoreactive Fabs

As mentioned in the introduction, multiple groups reported polyreactive antibodies to be more flexible as opposed to rigid nature of monoreactive antibodies. These studies are limited to the CDR3 loop in the heavy chain and a single source of polyreactive antibodies (i.e. therapeutic or human) and mostly rely on *in silico* predicted models of structures, for which computational predictions of CDR loops still remain unreliable. I have therefore explored a more comprehensive molecular dynamic analysis of polyreactive Fabs alone using apo crystal structures I solved (in Chapter 3.2) as well as verified polyreactive Fab crystal structures from the protein data bank. Together with a fellow graduate student in the Adams lab, Mr. Christopher Boughter, we ran all-atom molecular dynamics (MD) simulations in parallel using different architectures; I used CPU based NAMD using Midway supercomputer cluster and Mr. Boughter used the GPU based

AMBER using Dr. Benoit Roux's Lab cluster for 500ns. This strategy cross-validates our results, given the stochastic nature of all-atom simulations and limited computational resources available. Our analysis follows a workflow for each Fab, where we use a python package called pyEMMA that includes time-independent component analysis (tICA), which is able to reduce ~120 dimensionalities in our simulations to only 4, and this way sort all different structural conformations of a simulated Fab into multiple clusters (Fig. 3.5A). We further pick five random clusters, by dropping the pin in the center of it and visually sampling representative conformations from these clusters. Additionally, we inspect root-mean-square-fluctuations (RMSF) of backbone and side-chain movements in individual CDRs, to observe how long these different conformations persist.

3.3.2 Analysis of MD simulations by PyEmma and VMD

In our dynamic analysis, we observed that structures of the polyreactive Fabs are overall quite rigid compared with monoreactive Fabs. tICA analysis for polyreactive Fabs shows consistently less clusters of distinct conformations suggesting overall more static behavior (Fig. 3.5B, C). Occasionally, a global shift is present in polyreactive Fabs, which comes from global scaffold movements of the Fab, but the CDR loops themselves do not change much of its conformations. For example, a single conformational change is present in polyreactive 2G02 Fab in the CDR L1 that persist for the remainder of the simulations based on RMSF and visualizations (Fig. 3.5B, grey RMSF plot). However, the overall shape of that loop is conserved, which could suggest that the original state that was captured in the crystal structure might be different than in the relaxed version of the protein in solution. On the contrary, monoreactive Fabs show more distinct CDR conformations and higher RMSF. The CDR H2 in the monoreactive 3B03 Fab adopts distinct

conformations and many events of exchange between these conformations (based on RMSF), which could indicate flexibility (Fig. 3.5C, green RMSF plot). Similarly, CDR H3 in monoreactive 3B03 Fab is adopting many conformations as opposed to the polyreactive 2G02 Fab. This is an especially intriguing observation, given that both Fabs evolved toward their Flu target and use CDR H3 to mediate that high affinity interaction. These dynamic differences, although relating to selected CDR loops only, reveal overall a more rigid nature of antigen binding surface in serve as a uniform platform that is broadly permissive and accessible to polyreactive ligands.

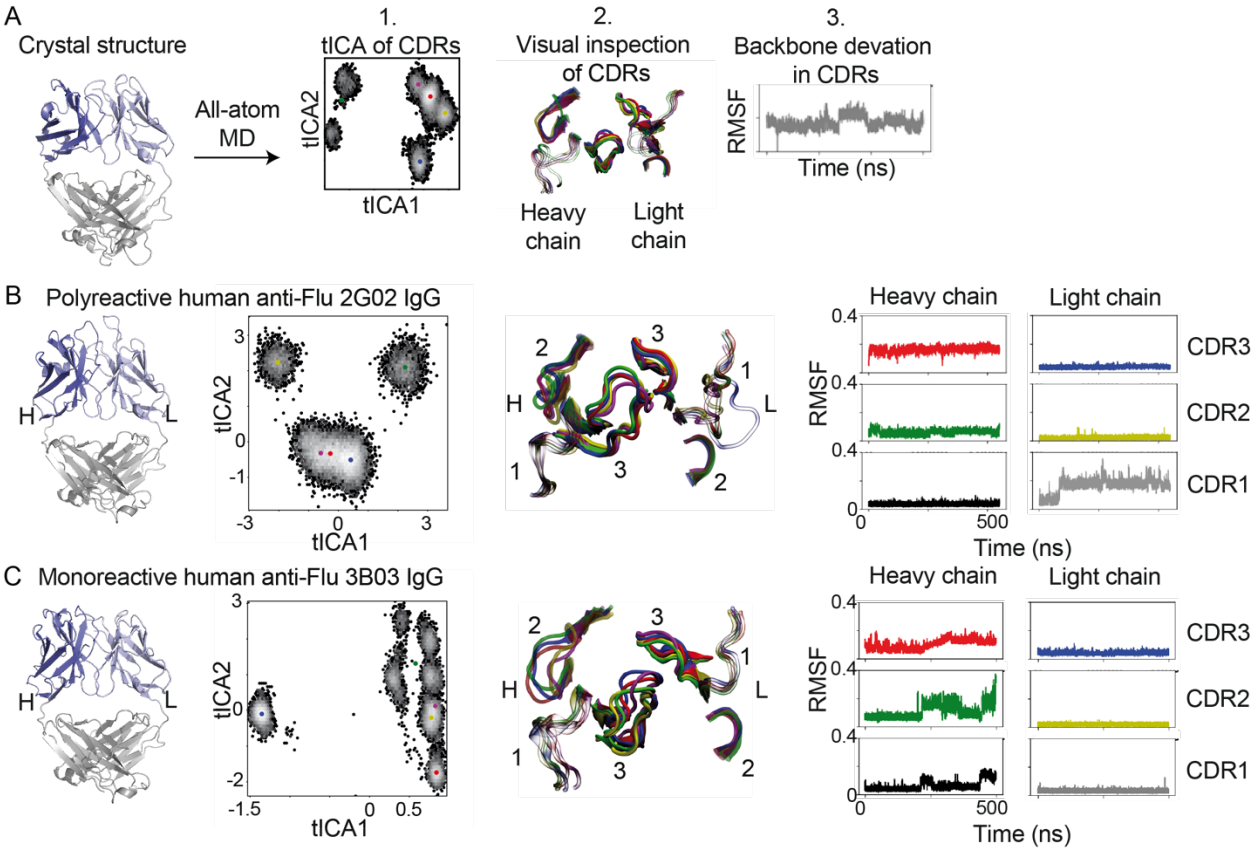

Figure 3.5 All-atom MD simulations and analysis on example human anti-Flu polyreactive vs. monoreactive IgG
(A) Workflow of all-atom MD analysis, including clustering of CDR conformations via time-Independent Component Analysis (1), Visual inspection of CDR clusters (2) and RMSF of backbone movements in CDR loops (3). (B) Cartoon structure and MD analysis of human anti-Flu 2G02 polyreactive IgG clone. (C) Cartoon structure and MD analysis of human anti-Flu 3B03 monoreactive IgG clone.

polyreactive antibodies. Although, currently this data does not directly explain how would polyreactive ligands bind to that surface, I hypothesize, it is very likely that rigidity can play a supporting role in polyreactivity.

3.3.3 Closer inspection of rigid loops in polyreactive and monoreactive Fabs

I further investigated the molecular basis of selected CDRs displaying increased rigidity in polyreactive antibodies and compared them with increased flexibility in monoreactive Fabs. I was specifically interested in how extended CDR L1 in polyreactive 43G10 maintains its rigidity despite no stabilizing secondary structures and how CDR H3 in monoreactive 4C05 shows flexibility despite forming structured β ribbon (Fig. 3.6A, B). Upon closer inspection of the intrachain interactions of the two mentioned CDRs, both showed a striking hydrogen bond network mediated by CDR residues, but in two distinct ways. Long, unstructured CDR L1 loop in polyreactive 43G10 is extensively supported by seven H-bonds mediated by side chains, which together form a three-dimensional stable network. In contrast, structured CDR H3 of monoreactive 4C05 forms its vast H-bonds between backbones of the ribbon (except two H- bonds at the tip, between Ser104 and Asp106) allowing only for two-dimensional stability. This raises an interesting hypothesis that 43G10 CDR L1 could possibly shield any polar or hydrophobic residues from intercepting a binding to polyreactive ligands, by engaging its side chains for stability. Conversely, exposed side chains in CDR H3 of 4C05 along with their biochemical properties, could show selectivity in binding to promiscuous ligands. I further looked for other examples of rigidity in CDRs in polyreactive antibodies and noticed similar, mostly sidechain mediated, network in CDR H3 of polyreactive 2G02 (Fig. 3.6C). Here, seven H-bonds stabilize this unstructured loop, where positively charged Arg100 faces inward the loop. Interestingly, that same

Arg100 and neighboring Arg101 when both mutated to Lys (along with five other germline residues) in 2G02, show reduced polyreactivity on ELISA (Guthmiller et al. in print). Consistent with our hypothesis, in our dynamic studies we see increased flexibility in CDR H3 of 2G02 double Lys mutant compared with wild-type 2G02 (Fig. 3.6D). This change in rigidity could come from

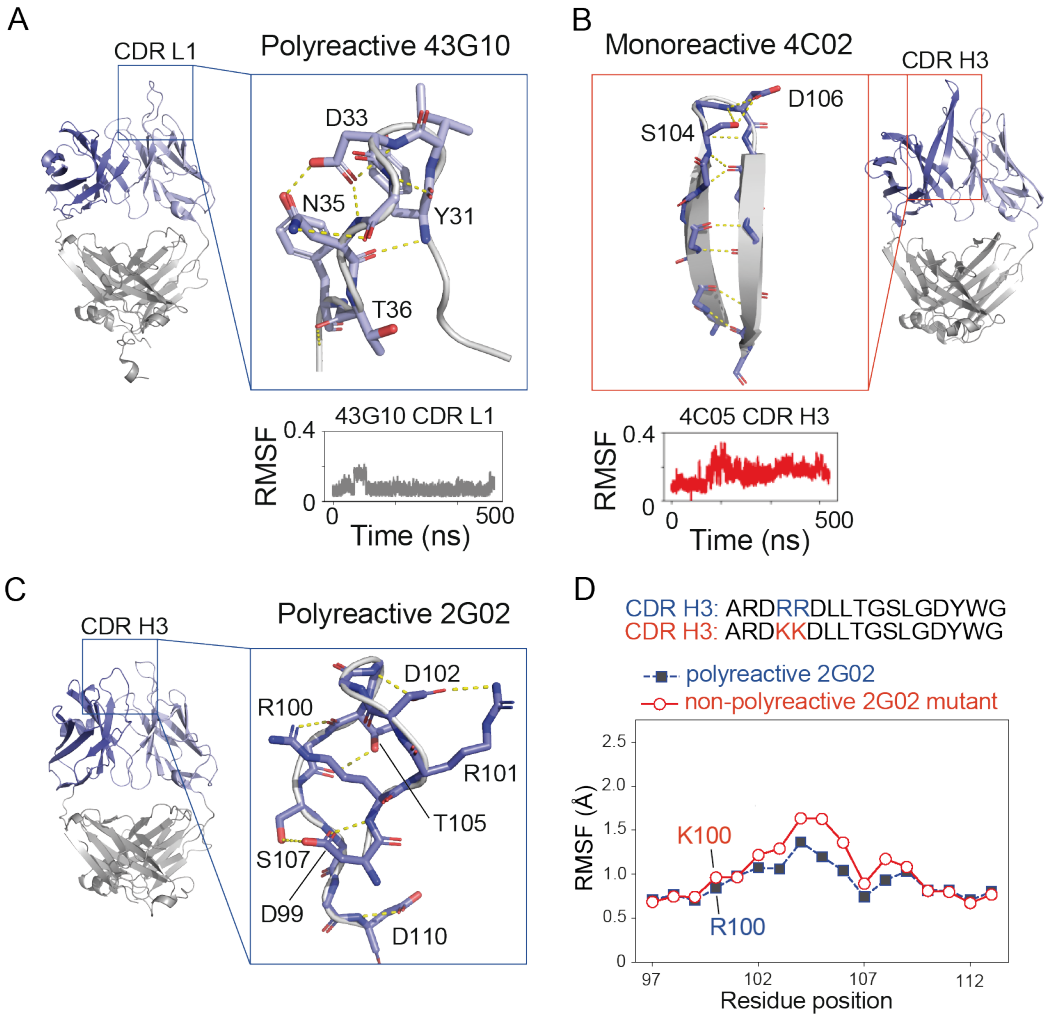

Figure 3.6 Examples of hydrogen bond network within CDRs loops of polyreactive vs monoreactive Fabs

(A) Polyreactive 43G10 CDR L1 maintains extensive hydrogen bond network mostly mediated by side chains throughout 500ns of MD simulations. (B) Monoreactive 4C05 CDR H3 forms β ribbon with H-bond network mostly mediated by backbone throughout 500ns of MD simulations. (C) Polyreactive 2G02 CDR H3 keep stability through vast H-bond network mostly mediated by side chains. (D) Stability and therefore polyreactivity within 2G02 CDR H3 can be disrupted with R100K and R101K mutations.

losing guanidinium group on Arg that can form three-dimensional H-bond network, in lieu of two-dimensional ε-ammonium group in Lys. These subtle observations could add up to maintaining overall rigidity in polyreactive antibodies.

3.4 Discussion

Altogether, I was able to determine the affinity of multiple polyreactive Fabs to one of the polyreactive ligands, ssDNA, (K_D= 30-160uM), which is on the low end of protein-protein interactions. To put this in context, these K_Ds are similar to those measured for conventional αβ TCRs to their cognate agonist ligands presented by MHC molecules (Corr et al. 1994, Matsui et al. 1994, Alam et al. 1996). Considering that antibodies have increased valences and can form higher order self-complexes depending on the Fc region utilized, the final avidity of a polyreactive interaction can be significantly amplified, therefore could play a more important function than previously thought. Additionally, despite other groups reporting specific motifs and patterns characteristic of polyreactive antibodies, no patterns in sequences and structures between polyreactive and non-polyreactive Fabs were identified in this analysis. This observation opens a potential hypothesis that polyreactive antibodies can utilize different chains and loops to bind to its polyreactive ligands and polyreactive binding can be clone specific, rather than ligand specific. However, often single alanine substitution lead to reduced binding across multiple ligands, suggesting possibly many ligands can bind to the same region on a particular Fab.

Important observations come from complex crystal structures that show (in most cases) multiple ligands decorating the binding site. Given the antigen binding site is accessible to ligands, I

hypothesized one of them will likely be placed correctly. Therefore, I used low resolution SAXS analysis of these complexes, which suggested that most ligands bind to the germline region on the light chain of polyreactive Fabs. This result would also exclude the hypothesis of a polyreactive antibody engaging many molecules of the same ligand with multiple binding sites. However, this work needs further studying as low affinity interactions require additional validation with multiple techniques.

In our work on dynamics of polyreactive Fabs, we noticed that polyreactive antibodies tend to be more rigid as they adopt less conformations in their CDRs. We also found multiple examples of events that enhance rigidity in polyreactive Fabs including formation of stabilizing structures like hairpins and increased network of stabilizing H-bonding mediated by sidechain in CDRs. To hypothesize, in addition to antigen binding surface forming a rigid platform, by engaging polar or hydrophobic side chains in intrachain network, it could also organize into inoffensive surface that is exposed on the antigen binding surface making it available for ligands to collide. I further conclude, that this result alone does not explain polyreactivity, but it could assist in forming a stable interface supporting the binding to diverse ligands. More experimental techniques should be employed to explore the binding models for polyreactive, low-affinity interactions (see Conclusions & Perspectives).

During this research I came up with many questions and ideas that could not be tested experimentally. Through our collaborators from Bendelac, Wilson and Nussenzweig Labs I gained access to large databases with thousands of sequences of antibodies which were tested for polyreactivity. This gave rise to a collaboration with my labmate and biophysics student, Mr.

Christopher Boughter, that led to the development of a bioinformatic tools to identify biophysical characteristics of polyreactivity and build a classifier to identify polyreactivity based on amino acid sequence (Boughter et al. 2020). In his work, he found that charge and hydrophobicity of CDR loops are in fact important determinants of polyreactivity. Although, this characteristic feature appears to be a move towards neutrality of the binding interface, consistent with my observations of side chains of CDRs being engaged in stabilization rather than exposed to solvent. Moreover, he also found an increase in inter- and intra-loop crosstalk between residues in polyreactive antibodies (using mutual information theory), which we translate as increase in stability that we observed in our all-atom MD simulations.

Chapter 4 – Structural dissection of Btn3 involved in γδ T cell stimulation

4.1 Strategies for structure determination of full length Btn3A

Activation of Vγ9Vδ2 T cells is dependent on a molecule called Butyrophilin-3 and early studies excluded the requirement for peptide-MHC complexes for activation (Morita et al. 1995, O'Brien et al. 1989). Further it was suggested that high intracellular concentration of phosphoantigens, mediated by transmembrane Btn3, stimulates Vγ9Vδ2 T cells. Given this novel "inside-out", Btn3-mediated mechanism of signal transduction, the structural events that follow pAg binding to the intracellular Btn3 domain are still elusive. How and whether intracellular signal triggers conversion from resting to activating state of Btn3 is unclear. To determine which dimer is prevalent in native environment and what happens to the full length Btn3 structure after binding pAg, I sought to understand the native structural orientation of Btn3 dimers during my doctoral work.

4.2 Crystallographic optimization of Btn3A in detergents

To identify the conformations of Btn3A1 and further explore how intracellular pAg binding triggers Btn3A1 conversion from resting to activating state, I pursued the structure of full-length Btn3A1 in the presence or absence of pAg or agonist 20.1 antibody using the complementary structural methods: X-ray crystallography and electron microscopy (EM).

First, my approach involved expression of the full length butyrophilin-3 in insect cells, for which all three isoforms A1, A2 and A3 were designed with a C-term 12xHis construct by Wioletta Nawrocka. From a small-scale detergent screen, I determined by anti-His Western blot candidate

detergents including Triton X-100 (TX-100), *n*-dodecyl β-D-maltopyranoside (DDM), n-dodecylphosphocholine (FC-12) and n-dodecyl-N,N-dimethylamine-N-Oxide (LDAO) that support Btn3 extraction from insect membranes and stabilization in solution. Interestingly, initial attempts to express and purify Butyrophilin 3 in detergent, yielded more protein for Btn3A heterodimers (A1+A2, A1+A3) and A2/A3 homodimers than the Btn3A1 homodimer. For large scale purifications, I proceeded with DDM, FC-12 and LDAO as these detergents are the most compatible with crystallography. My initial efforts in crystallizing the full length Btn3 membrane protein focused on the heterodimer of Btn3A1 (C-terminal His- tag) and Btn3A2 in complex with 103.2 single-chain Fv (scFv) chaperone and in the presence of DDM detergent. Initial unverified crystals were identified, but these likely formed from an excess of DDM in the crystallization solution and diffracted poorly. As a result, I pursued a larger scale purification of the protein to increase yield and avoid sample concentration (and thus detergent concentration). However, when 103.2 scFv chaperone is added, the Btn3 hetero- and homodimers precipitate, whereas 20.1 scFv remains in solution. This could be attributed to possible conformational changes due to 103.2 binding or inherent 103.2 scFv instability at high concentrations. This observation triggered me to switch to 20.1 scFv as the primary co-crystallization chaperone.

Following this, I optimized the purification and solubility conditions for the Btn3 heterodimers using a new class of neopentyl glycol detergents, *2,2*-didecylpropane-1,3-bis-β-D-maltopyranoside (LMNG). LMNG has a substantially lower critical micelle concentration (CMC) than DDM, thus allowing Btn3 to be soluble with a smaller amount of detergent, therefore reducing the interference from detergent in crystallization process. Additionally, LMNG monomers show slower dissociation kinetics, allowing me to purify Btn3 via size-exclusion chromatography (SEC)

in buffers without any detergent added, Indeed, Btn3 solubilized initially with LMNG remains soluble in standard SEC buffers, even after repeated freeze-thaw cycles. With these improvements in mind, I set up Btn3 A1His+A3+20.1sc in LMNG in multiple high-throughput crystal trays using the ARI Gryphon crystallography robot and identified a new condition yielding crystals that did not diffract. I have further constructed and expressed 20.1 and 103.2 as Fabs in insect cells as larger and more stable crystallization chaperones. Btn3 in complex with 20.1 Fab proves to be stable for at least two weeks in 4C at high concentration, whereas complex with 103.2 Fab cannot be concentrated above 3mg/mL consistent with 103.2 scFv behavior described above. I also focused on expressing Btn3 homodimers A2 and A3, which can be produced in larger yields, supplemented with DDM detergent and cholesterol hemisuccinate (CHS), a well-established trick to aid in transmembrane stability. I set up all available crystal screens using the ARI Gryphon crystallography robot, however, 80-90% conditions were precipitated, excluding the MIDAS screen, where ~50% precipitation was observed. The trend in non-precipitated conditions solely contained PEG's of various MW's with some even showing crystal-like morphologies. One particular condition, 55% polypropylene glycol 400 from the MIDAS screen, yielded needle-like crystals that were reproducible only when Btn3 was present and not in buffer or Fab controls (Fig. 4.1A). Interestingly, the A2 homodimer showed crystals and precipitation in this same condition. Furthermore, the A2 homodimer complexed with the 20.1 Fab also crystallized but did not precipitate. I was further encouraged to see protein when the crystals were analyzed on SDS-PAGE, confirming protein complex within crystals (Fig. 4.1B). More than 30 crystals of A2 or A3 homodimers in complex with 20.1 were harvested and frozen in different cryos, but no protein diffraction was observed at the synchrotron. These frozen crystals were further washed, dissolved, and run again on SDS- PAGE followed by Silver staining, which again showed protein. Altogether,

despite vast efforts, my attempts to crystallize full length Btn3 did not yield many crystals, which could suggest that additional strategies should be taken to further stabilize protein sample.

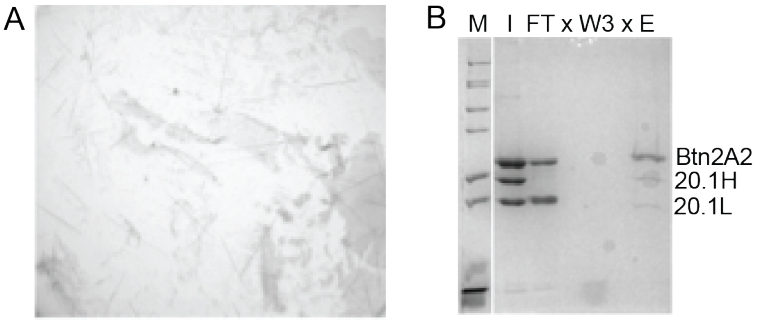

Figure 4.1 Attempts to crystallize full length Btn3 in detergents
(A) Needle-like crystals formed in sample of Btn3A2+20.1Fab with DDM:CHS. (B) Washed crystals run on SDS-PAGE confirmed the protein complex in the crystal content, however these crystals did not diffract. M-marker, I-input, FT-flow through, W3-third wash, E-elution.

4.3 Negative-stain EM of Btn3A in native-like lipid environment

Butyrophilin-3 has a single transmembrane and might not be stable enough in detergents for structural studies. Therefore, I explored using nanodiscs as an environment that provides more native-like lipid stabilization for Btn3A dimers. The extraction and purification of Btn3A1+A2 heterodimers in TX-100, followed by reconstitution into two types of nanodiscs 1D1 and E3D1 yielded a single peak on Superdex 200 15/300 column. The sample was sent to our collaborators Jeffrey Tarasch from Dr. Georgios Skiniotis Lab, then at the University of Michigan. Initial visualization by negative stain electron microscopy followed by single-particle analysis to examine the conformation, revealed the V-shaped conformation in several two-dimensional class averages. In contrast, the head-to-tail conformation could not be unequivocally detected, even though it could be theoretically accommodated by the size of nanodiscs. This observation is fully consistent with previous cellular data suggesting that in the absence of other cellular components and/or

exogenously added pAgs, the extracellular domains of full-length BTN3A in a lipid bilayer form V-shaped dimer. This EM data was included in the PNAS article by another graduate student in the Adams Lab – May Siyi Gu (Gu et al., 2017).

4.4 Cryo-EM optimization of Btn3A in amphipatic polymer

I further used cryoEM to try to obtain a high-resolution structure of Btn3 dimers. EM does not consume large amounts of protein and does not require extended stability of the sample so I revisited a search for a detergent condition that would support homogenous and stable Btn3 sample. I also managed to optimize Btn3A1 homodimer expression in insect cells with a new baculovirus that yielded sufficient amount of protein for electron microscopy. Initially, detergent screening was done by negative stain electron microscopy (NS-EM) as a quality control step, including DDM:CHS, FC-12 and LMNG conditions, which elute as single peaks on Superdex 200 10/300 size-exclusion column (Fig. 4.2A, B). Interestingly, DDM:CHS yielded uniform particles that looked in size like small clusters of Btn3 oligomers. Cymal-6 solubilized Btn3 elutes as two peaks on size-exclusion chromatography, likely corresponding to higher and lower oligomeric states. The lower oligomeric state showed more homogenous particles on NS-EM compared to previously studied DDM:CHS sample. Additionally, a zwitterionic Fos-Choline-14 solubilized material was not compatible with negative stain EM, but appeared as three peaks on SEC. Both Cymal-6 and FC-14 samples were screened for multiple crystallization conditions without success. Ultimately, a monodisperse sample came from a new approach using an amphipathic polymer called Amphipole A8-35, which once reconstituted into, allows the sample to be kept in solution without additional supplemental detergents (Fig. 4.2C, D). In short, once solubilized from the cell membrane with TX-100 detergent, the protein is isolated from other contaminants by binding to

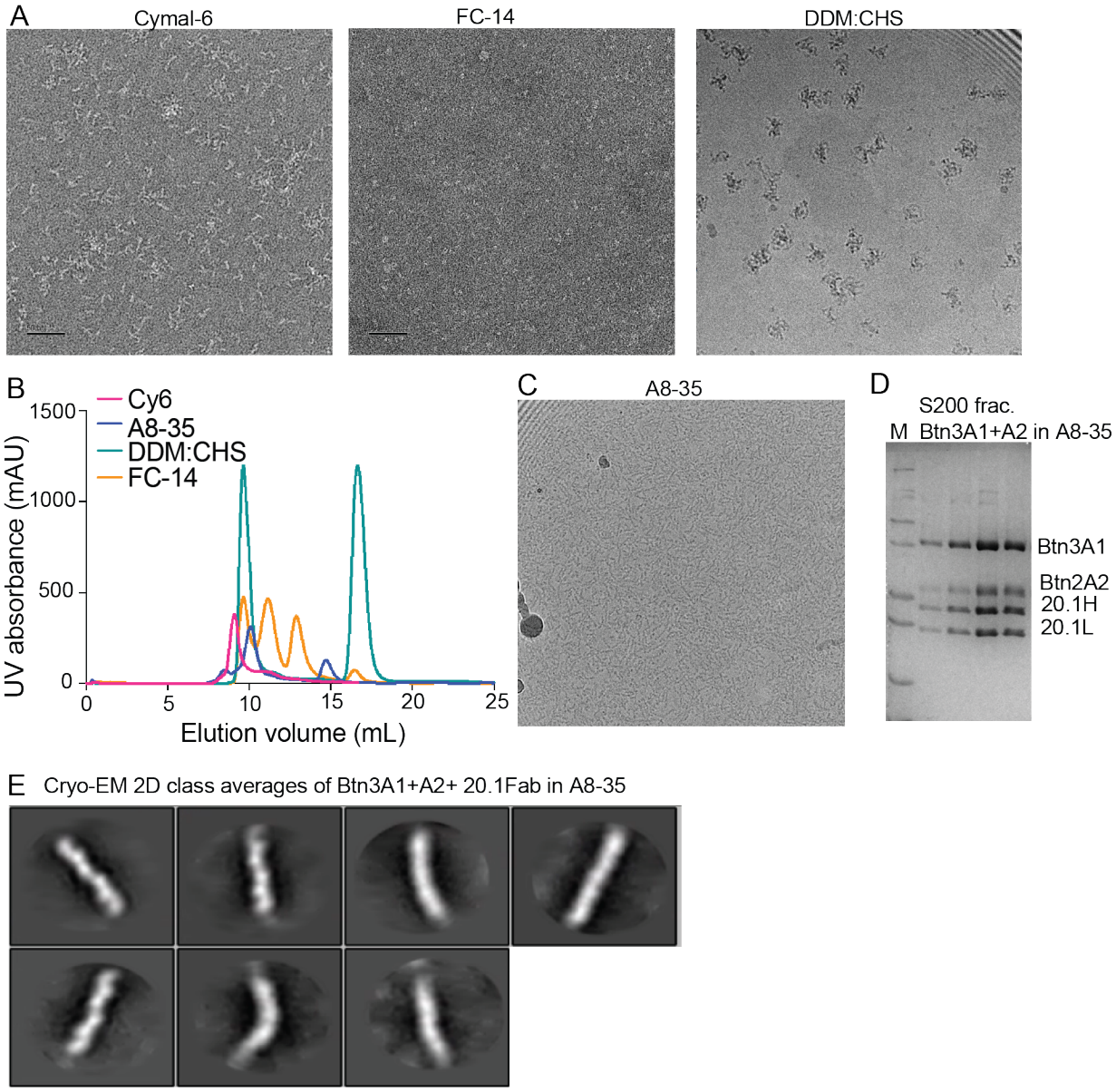

Figure 4.2. Btn3A detergent optimization for cryo-EM
(A) Negative stain EM micrographs of Cymal-6 and FC-14 (left and middle) and cryoEM micrograph of DDM:CHS (right), show different detergent conditions of Btn3A sample. (B) Size-exclusion chromatography shows Btn3A sample forming different oligomeric states in different detergents. (C) Cryo-Em micrograph of monodisperse Btn3A1+A2+20.1Fab in A8-35 condition. (D) SDS-PAGE from SEC Btn3A in A8-35 sample tested on cryo EM. (E) 2D class averages formed from a small dataset collected on A8-35 sample.

Co-TALON resin and then eluted with imidazole in TX-100. Concentration of the sample was avoided and directly eluted sample was further reconstituted into A8-35 by incubating with Bio-

Beads o/n and additional 20.1 Fab for stability. With the help of our departmental EM collaborators Man Pan and Jane Lodwick from Dr. Minglei Zhao Lab, we collected encouraging cryo-EM dataset with 700 micrographs for Btn3A1 homodimer in complex with 20.1 Fab. Using the data processing software Relion, we generated 7 2D class averages from each sample group using over 5,000 particles. Although only the top view orientation was visible in 2D classes, clear and encouraging C2 symmetry was visible (Fig. 4.2E). Further effort to generate more views of the sample included optimization of vitrification procedure, i.e. longer blotting time to attempt to reduce ice thickness and therefore force the protein complex to orient sideways, yielded a lot of protein denaturation due to protein being more exposed to water-air interface. To alleviate that process we have added OMF detergent that forms a layer at the water-air interface but does not interfere with the membrane protein/ micelle integrity. We observed better contrast of the sample, but still orientation preference was observed. This shows cryo-EM as a promising strategy to obtain a structure of the full length Btn3, although more effort should be put toward finding a condition that alleviates orientation preference when Btn3 is placed on grids.

4.5 Discussion

The conformational change of Btn3 that transduces the signal via "inside-out" mechanism to be sensed by Vγ9Vδ2 T cell has not been addressed structurally. Two distinct dimer conformations have been proposed, namely a "V-shaped" and a "head-to-tail" form, which were observed in the structures of extracellular domains of Btn3 (Palakodeti et al., 2012). In this chapter I examine the presence of dimer conformations in the context of the full-length BTN3A1 protein in the detergent and reconstituted lipid environments. Original efforts to crystallize full length Butyrophilin 3 did not yield diffracting crystals, despite testing multiple constructs, chaperones and detergent

conditions. The conventional crystallography is impeded by dynamic proteins, as Btn3 is thought to adopt multiple conformations. Therefore, using a stabilizing chaperone i.e. targeting juxtamembrane of Btn3 or B30.2 domain, can facilitate conformational stability. Additionally, for single-pass transmembrane proteins like Btn3, detergent sample preparation for crystallography is very challenging. Therefore, further efforts in crystallizing Btn3 should focus on trying lipid assisted crystallization (e.g. bicelle, lipidic cubic phase or sponge lipid mesophase). The quantity and homogeneity of protein material for EM studies is much less stringent than crystallography, which encouraged me to pursue cryoEM studies on Btn3. A broad screen of different detergent conditions revealed striking heterogeneity in oligomeric states of Btn3, with A8-35 condition stabilizing the most monodisperse sample. Two-dimensional analysis from that condition was consistent with the previous data, where we found evidence for the existence of the "V-shaped" dimer form. Despite limited orientations of Btn3 in complex with 20.1 Fab, I speculate that this form presents an active state of the Btn3A1 molecule. Further analysis using cryoEM should focus on more native environment i.e. Btn3-nanodisc reconstituted complex, as well adding full length Btn2 to the system.

Chapter 5 – Conclusions and perspectives

5.1 Molecular mechanism of a novel host-commensal microbiota interaction (Relating Chapter 2)

In this chapter I provide molecular and functional studies revealing the basis of this Ibp/Ig interaction. The crystal structure and biochemical assays of a truncated IbpA construct in complex with mouse VH5 Fab shows a non-canonical binding of Ig heavy chain framework residues to the Ibp Domain D and the C-terminal Heavy Chain Binding Domain (HCBD). I used targeted mutagenesis and affinity measurements of contact residues as well as performed stoichiometry studies of Fab-IbpA complex to reveal a complex stoichiometry between Ibp and VH domains, suggesting Ibp may serve to cluster full length antibodies *in vivo*. Furthermore, *in vitro* stimulation experiments indicate binding of the Ibp HCBD alone is sufficient to activate responsive murine B cell receptors (BCRs). Together, I demonstrate mechanistically how these commensal derived Ibp proteins bind entire families of immunoglobulins via their heavy chains and compare them with known superantigen interactions. The presence of these proteins in a commensal microbe suggest that broad engagement of immunoglobulins, particularly in the gut/microbiome environment, may provide an important function in the maintenance of host/microbiome homeostasis contrasting with the pathogenic role of structurally homologous superantigens expressed by pathogens.

This work provides more detailed molecular and functional information about *R. gnavus* Ibp, defining the *in vitro* properties of full length Ibp and single Ibp domains. Overall these results give new insights of how *R. gnavus* interacts with the host and utilizes its surface proteins to attract a wide range of antibodies. However, the *in vivo* functional significance of having different number of binding sites in mouse and human Igs, is currently unclear. Whether truncated IbpA has a hidden

site for mouse 338E6 Fab or higher propensity to crosslink should be further investigated. Having multiple, even weaker binding domains could be potentially beneficial for *R. gnavus*. Unlike T cell SAgs, B cell SAgs have to additionally accommodate high levels of circulating immunoglobulins. Paradoxically, Protein A, with its ability to also bind Fc-regions of soluble IgG, has been shown to play critical roles for efficient BCR cross-linking and induction of AICD. Despite the single HCBD of Ibp being able to stimulate B cells *in vitro*, our data suggests that having multiple available binding sites enhances avidity and B cell activation as shown by our binding and cellular studies. Therefore, in a physiological environment where IgA dominates as dimers, IbpA/B having the potential to bind up to five antibodies is consistent with high IgA coating of *R. gnavus in vivo*. Moreover, since IbpA does not bind the Fc portion of antibodies, it poses an important question whether highly IgA coated *R. gnavus* will be still accessible to Fc receptors of phagocytes.

Additionally, my observations on different levels of oligomerization by different antibody clones, suggests a possible modulation of the number of antibodies bound by surface expressed Ibp. This could enable *R. gnavus* to vary its magnitude of IgA coating and therefore modulate its relationship to the host. As IgA can influence microbial populations through many mechanisms, further studies should asses metagenomic heterogeneity of Ibp genes (especially mutations withing repeat and HCBD domains) in *R. gnavus* and its functional consequences.

Besides IbpA being the first to our knowledge identified immunoglobulin binding protein with *in vitro* superantigen-like properties that is expressed by a commensal, it is also the first protein binding to VH with two distinct domains. Despite Domain D and HCBD having a superantigenic-like fold, they don't have any sequence similarity or superantigenic potential when compared with

either Protein G or L. A possible gene transfer or co-evolution might have happened between this commensal and other pathogenic species, which opens the need to search for other commensals expressing similar immunoglobulin binding proteins as a novel mechanism of host-microbiota communication.

5.2 Molecular basis of polyreactivity in antibodies (Relating Chapter 3)

In this chapter, I aim to understand the basis of polyreactivity in antibodies and describe their low affinity interactions. I was able to measure the direct affinity between polyreactive Fabs and a ssDNA ligand. While no specific patterns have been found in my analysis of static structures of polyreactive antibodies, dynamic studies suggest increased rigidity in CDR region of polyreactive Fabs. Finally, I was able to obtain multiple structures of the complex between polyreactive Fabs and different ligands. Despite low resolution, my SAXS analysis of these structures further added some clarification in terms of where ligands might bind on Fab. While these structures provide models upon which to test future hypotheses, they need confirmation through other validation techniques i.e. NMR, EPR or mass spectrometry of the complex. Further experiments should focus on a more comprehensive study, that would include high-throughput sequence-, structure- and dynamic-based analysis. Finally, a more functional understanding of the selection of polyreactivity *in vivo* poses interesting area for improvement of antibody-based vaccines.

To further expand on this thesis's work, the search for the binding model between polyreactive antibodies and multiple ligands is necessary, especially by linking dynamic studies with polyreactive recognition. My finding of the rigid binding surface in polyreactive antibodies and therefore high-energy state (low entropy) of CDRs, challenge the conventional lock-and-key and

induced-fit models. The concept of both of these models assumes a passive adaptation, with a low-energy (high entropy) state of an antibody in apo and a loss of conformational freedom for both antibody and ligand upon its binding (loss of entropy). Therefore, the polyreactive high-energy conformation will not easily shift the equilibrium toward the complexed state, according to these models. The low affinity of polyreactive interaction, might not be in fact sufficient for the polyreactive ligands to displace water and make contacts with the antibody. However, hidden charges on the surface of polyreactive antibody might affect the level of solvation of the antibody molecule and therefore be more energetically favorable to form the complex, for that, the exact energy calculations will be required.

Therefore, alternative binding models should be explored. I.e. Greives and Zhou demonstrated that the conformational selection binding model dominates for the slow conformational transition of the protein and the low ligand concentration (Greives and Zhou, 2014). In the conformational-selection binding model, protein can preexist in the "active" conformation (high-energy) and the binding occurs when ligand comes in contact. Alternatively, all three mechanisms of binding may exist both in a simultaneous or in a sequential manner to explain polyreactive binding.

Another hypothesis could point to allostery, where rigidity in the CDR region may transfer information to remote parts (i.e. germline) of the antibody to induce binding. However, even though recent advances in solution nuclear magnetic resonance (NMR) make it possible to directly measure the allostery in protein structure, these experiments focus on fluctuations of selected atoms, traditionally backbone amides, that might not correctly reflect overall dynamics (Wand et al. 2013).

Finally, another possibility for polyreactive binding to occur is by slow isomerization of the ligand to a high-affinity complex as was shown for other low-affinity complex (James and Tawfik, 2015).

5.3 Structural elucidation of Btn3 signaling (Relating Chapter 4)

In this work, I explore structural techniques to address key questions about the mechanism of Btn3-mediated Vγ9Vδ2 T cell stimulation that centers on Btn3 active and resting conformations. My initial crystallographic attempts did not yield diffracting crystals. However, further techniques including lipid assisted crystallization (e.g. bicelle, lipidic cubic phase or sponge lipid mesophase) or use of additional conformation specific chaperones can facilitate stability and crystal formation. I further use the cryoEM technique to identify distinct Btn3 conformations. Specifically, I show a promising condition that stabilizes Btn3 active conformation by complexing with 20.1 agonist Fab and supplementing with A8-35 polymer. Alternatively, I also show promising results using Btn3 reconstituted in nanodiscs.

Perhaps the most central question that should be addressed in this field is what molecules are directly recognized by Vγ9Vδ2 T cell receptor. Most recently two groups reported that another butyrophilin, Btn2A1 associates with Btn3A1, and act together to initiate responses to pAg through direct binding of Btn2A1 to germline encoded Vγ9 regions of Vγ9Vδ2 TCR (Rigau et al. 2020, Karunakaran et al. 2020). This work can therefore be expanded to include Btn2A1 and other players shown to mediate Vγ9Vδ2 T cell stimulation.

Chapter 6 – Experimental procedures

6.1 Experimental procedures used in Chapter 2

Constructs

For mouse Fab clones: variable domain of both heavy and kappa chains of mouse IgA clones were amplified from plasmids provided by Albert Bendelac. Similarly, both chains of modified IgG constant domain amplified from plasmids gifted by Anthony Kossiakoff were fused with corresponding variable domains by overlap extension PCR. Both chains were cloned with sequence and ligation independent (SLIC) method in a pACgp67a vector with a 3C protease sequence, either basic or acid zippers and a C-terminal 6x histidine tag.

Human Fab clones were obtained and unmodified from Patrick Wilson.

Full length and truncated Ibp plasmids were obtained from Albert Bendelac. Individual domains were cloned from these plasmids into pET301b with N-terminal 6x histidine tag. For alanine scanning mutagenesis, contacting residues were selected based on PISA analysis of Ibp-VH5 Fab structure. Alanine mutants were introduced by quickchange mutagenesis using most common alanine codons.

Proteins expression, purification

Baculoviruses with all hybrid mouse Fab chains were prepared by transfecting and amplifying into Sf9 cells. Mouse Fabs were expressed in the Hi5 cells by co-infecting with baculoviruses containing both chains and incubating up to 3 days. Fabs were purified from the supernatant using

Ni-NTA agarose in buffer A, o/n 4C 3C protease cleavage, second Ni-NTA subtractive step and size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare).

Human Fabs were transfected into suspension Expi293 cells (ThermoFisher) using Expifectamine and according to manufacturer's protocol. Supernatant was harvested after 5 days, spun down for 15min at 4°C/3,000 ×g and incubated o/n with 1mL bv Ni-NTA agarose in buffer A with f.c. 20mM imidazole 7.2. Next day after washing the column with high salt buffer A supplemented with 20 mM imidazole, then eluted in 5mL of buffer A with 350mM imidazole. Samples were further purified by size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) in buffer A and single peak fraction was used for binding studies or concentrated for AUC studies using in 30-kDa MWCO Amicon Ultra centrifugal filter (Millipore).

Both wild-type and alanine mutants of all Ibp constructs were expressed in the same way. DNA constructs were heat shock transformed into BL21(DE3)pLysS and single colony was used to inoculate 25mL terrific broth (TB) starter culture supplemented with ampicillin, grown overnight at 37C and 250rpm shaking. 1L TB/Amp culture was inoculated with 1% of starter culture, induced with 1mM IPTG at OD600 and continued to grow for 4hrs/ 37C with 250rpm. Cells were harvested by centrifugation and resuspended in 50 ml buffer A with high salt (10 mM Hepes pH 7.2, 500 mM NaCl supplemented with 1 mM PMSF and 2 mM DNase I). Cells were lysed by passing three times through a high-pressure microfluidizer (Avastin), soluble fraction was collected by centrifugation for 40 min at 4°C/18,500 ×g and 2mL bv Ni-NTA beads, f.c. 20mM imidazole 7.2 were added for o/n incubation. Next day after washing the column with high salt buffer A supplemented with 20 mM imidazole, protein was eluted in buffer A with 10 c.v. 350mM

imidazole. Typically, the protein was further purified by size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) in buffer A. Peak fractions were pooled, concentrated to 5 mg/ml in a 3-kDa or 30-kDa MWCO Amicon Ultra centrifugal filter (Millipore).

Protein concentrations were determined by A280 using a calculated extinction coefficient. Purified IbpA constructs were mixed with 1.2-fold molar excess of smaller component and separated from excess by size exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) in buffer A. Peak fractions were pooled, concentrated in a 50-kDa MWCO Amicon Ultra centrifugal filter (Millipore), large aggregates were removed by 0.22um spin filter columns, and either used immediately for crystallization or stored 4°C up to a month.

Crystal structure determination and analysis

Crystals of the truncated IbpA-mouse VH5 Fab complex were grown at room temperature by hanging drop vapor diffusion. Equal molar volumes of a protein solution containing 15 mg/ml complex was mixed with a reservoir solution containing 0.2 M calcium chloride and 18% PEG 4000. Crystals were cryoprotected in 0.2 M calcium chloride, 18% PEG 3350, 20% ethylene glycol, and flash frozen in liquid nitrogen. One dataset was collected from a single region of a single crystal at 100 K at APS beamline 24ID-E (λ = 0.9791 Å) on an Eiger detector. The data were processed, scaled, and merged using iMosfilm, with diffraction limits of 2.1Å. Data collection and processing statistics are listed in Suppl. Table 1.

VH5 Fab, with its complementarity determining regions (CDR) and V-C hinges omitted, was used as the search model using PHASER as molecular replacement tool[44]. Two VH5 Fabs packed

against each other were determined per asymmetric unit. Since no search model was known for IbpA, remaining electron density was fitted with secondary structures using Buccaneer, an auto-build package in CCP4 i2. Cycles of manual model building and refinement (torsional non-crystallographic symmetry restraints applied throughout) with COOT and PHENIX allowed us to connect helices and strands and trace the backbone of truncated IbpA[45,46]. Electron density for the most C-terminal ~40 amino acids of IbpA is not visible, which suggest disordered structure, which was also confirmed by structure prediction with Phyre server[47]. Refinement and validation statistics are listed in Suppl. Table 1. Structure figures were generated with PyMOL (http://www.pymol.org).

Sizing analysis

For stoichiometry analysis 1mg of HCBD was injected onto pre-conditioned size exclusion chromatography column (Superdex 200 10/300 GL, GE Healthcare) with Hepes buffer saline (10 mM HEPES pH-7.2, 150 mM NaCl) at a flow rate of 0.5 mL/min. The effluent of the SEC column flowed through an inline UV/Vis detector (Waters Corporation), the MALS/DLS detector and an Optilab UT-rEX dRI detector (Wyatt Technology Corporation) at room temperature. Data collection and analysis was performed with ASTRA software, version 7.1.1 (Wyatt Technology Corporation). ASTRA software utilizes a patented method for the inter-detector delay and band broadening correction[48]. First-order fit Zimm formalism is used for all molar mass calculation.

Binding studies

All the binding studies were carried out using Bio-layer interferometry (BLI) (either Forte bio or Octet K2 from Pall Life Sciences) at room temperature (22°C). Concentrations of different IbpA

constructs were individually immobilized on Ni-NTA sensor to reach ~2-3 RU for binding with mouse Fab clones at different concentrations and responses (in nm) were recorded. Hepes buffer saline (10 mM HEPES pH 7.2, 150 mM NaCl) was used in all the measurements and sensor tips were recharged with 350mM imidazole after each round. For binding studies with human Fab clones, Ibp constructs were non-specifically labelled with NHS-SS-Biotin at 1:1 ratio. Excess of biotin was removed by size exclusion chromatography and level of biotinylation was assessed by SDS-PAGE shift assay with Trap protein. Biotinylated Ibp constructs were immobilized with Streptavidin sensor tips and recharged with 0.1M Glycine pH 2.5. Traces were reference subtracted and kinetic parameters (K_D, k_{on}, k_{off}) were calculated by fitting the data with Forte-bio build-in analysis software.

PBMC and splenic B cells isolation

Peripheral blood was diluted 1:1 with 1xPBS, overlaid on Ficoll-Paque PLUS density gradient media (GE Healthcare) and centrifuged at room temperature for 30 min and 400g with low brakes. Mononuclear cells were harvested from interlayer and washed with PBS before flow cytometry staining. Whole blood was obtained from healthy volunteers with approval from the University of Chicago Institutional Review Board.

For isolation of mouse B cells, single cell suspensions were prepared by grinding spleens from C57BL/6 mice between glass slides. Cell suspension in HBSS, 0.25% BSA were next filtered through 70 um cell strainer (Fisher). Cells were pelleted by 5 min centrifugation at 400g, resuspended in 1 x RBC lysis buffer (ebioscience) and incubated for 5 min on ice for RBC lysis. To purify untouched mouse B cells, cells were incubated with anti-mouse CD43 (S7) and MACS

purified using Streptavidin MicroBeads (Miltenyi) and autoMACS Separator with a purity of ≥90%.

Ibp labeling and lymphocyte flow cytometry staining

Recombinant Ibp protein solutions were buffer exchanged using Zeba™ Spin Desalting Columns (Fisher) and eluted in 1xPBS before labeling with Alexa Fluor 488 and Alexa Fluor 647 using commercially available Alexa Fluor labeling kits (Fisher). Ibp concentration after labeling was measured by Nanodrop.

For flow cytometry staining, human and mouse B cells were first incubated with Zombie dye life/dead stain (BioLegend) and either Human TruStain FcX Fc Receptor Blocking solution (Biolegend) or anti-mouse CD16/CD32 (Fisher). Next, cells were labeled for 20 min at 4° to 8° C with Ibp at typically 10 – 100 ng/ml or as indicated in figures or text. After washing with HBSS, 0,25% BSA, cells were incubated with fluorophore conjugated mAbs purchased from Biolegend, eBioscience and BD: anti-mouse CD19 (6D5), anti-human CD19 (HIB1) and anti-mouse IgM (Il/41) for 20 min at 4° to 8°C. Cells were washed again, resuspended in HBSS, 0.25% BSA and analyzed using a LSR II cytometer (BD).

Repertoire analysis of sorted Ibp-binding mouse and human B cells

Cells were stained with labeled full length IbpA and IbpB and indicated B cell populations were sorted using an AriaIIIu cell sorter (BD). Post sort analysis showed purity of ≥ 95%. Cells were washed and RNA purified using RNA easy mini kits (Qiagen) following the manufacturers recommendations. Samples were sent to iRepertoire, Inc. for sequencing heavy and light chain repertoires by Next Generation Sequencing (NGS).

Ibp HCBD and repeat domain blocking experiments

To block mouse B cell staining with labeled repeat domain, cells were incubated with high dose (100 ug/ml) unlabeled HCBD for 20 min at 4° to 8° C, wash and stained with labeled repeat domain at indicated concentrations for 20 min at 4° to 8° C. Next, cells were stained with fluorophore labeled mAbs before acquisition on LSR II cytometer (BD).

Mouse B cell in vitro stimulation

MACS purified untouched B cells were counted using a Neubauer counting chamber and cells in in RPMI medium supplemented with 10% FCS (Gibco)were transferred to 96 well U-bottom plate (150,000 per well) in RPMI medium supplemented with 10% FCS (Gibco). For analysis of BCR downregulation cells were stimulated for 2 hours at 37°C with indicated Ibp constructs at 10 ug/ml or F(ab')2 -goat anti-mouse (10 ug/ml). Before stimulation with Ibp constructs possible endotoxin contamination was minimized using High Capacity Endotoxin Removal Spin Columns (Fisher).

ITC

HCBD and VH5 Fab pure protein samples in matching buffers and protein concentrations measured by BCA were analyzed for binding stoichiometry using Malvern MicroCal iTC200. 200uL of 30uM HCBD were injected into the measured cell kept at 30°C, then 40μL of 1mM of VH5 Fab was titrated into the temperature regulated cell in one preinjection of 0.4μL and 20 subsequent injections of 2μL. Data were analyzed via the MicroCal ITC-Origin Analysis software to extract the binding stoichiometry.

Analytical ultracentrifugation

Experiments were performed using a Beckman Coulter XL-I analytical ultracentrifuge with an An-60 Ti rotor. Data were collected using both interference and absorbance optics (wavelength was varied depending on sample). Experiments were run at a speed of 36,000 rpm for 16 - 20 hours or until sedimentation became insignificant. Proteins were purified by size exclusion chromatography into 10 mM HEPES pH 7.2, 150 mM NaCl and diluted to desired concentrations in the same buffer prior to loading the samples. Volumes of 450 ul were loaded into two-sector epon-charcoal 1.2 cm centerpieces with sapphire windows. Samples were equilibrated at 20 °C for at least one hour before beginning experiments. Data were analyzed using the continuous c(s) distribution model in SEDFIT version 16.1[49]. Partial specific volumes, buffer density, and buffer viscosity were estimated using SEDNTERP[50] (20130813 beta release).

The $S_{20,w}$ value is the sedimentation coefficient of a single species corrected for protein partial specific volume, buffer density, viscosity and temperature. The $S_{20,w}$ value, frictional ratio (f/f_0), and experimental molecular weight (MW_{exp}, kDa) were reported by SEDFIT's c(s) distribution analysis. The sequence-based MW (MW_{seq}) was calculated using Protparam.

SAXS analysis

Small Angle X-ray Scattering data was collected at the SIBYLS 12.3.1 beamline at the Advanced Light Source using High Throughput Mail-In program. Scattering measurements were performed on 20-µl samples at concentrations ranging between 0.1-2mg/mL of truncated or full length IbpA. Collected datasets were processed using ATSAS and Scatter softwares by trimming low resolution data and adjusting R_g, d_{max} based on Guinier Peak Analysis and Real Space plots. Multi-state modelling was done in MultiFoXS by indicating flexible regions in the models, analyzing

scattering profiles and fitting them to the experimental data. Full length IbpA models were generated in Coot by using a combination of duplicating repeat domains and Phyre server predictions. We defined the linker residues (79-83, 157–163) in truncated and (85-89, 151-155, 218-222, 285-289, 352-356, 430-436) in full length IbpA models as flexible. The server sampled over 10 000 conformations, calculated their SAXS profiles, enumerated and scored multi-state models. The quality of fit between models and experimental SAXS data was assessed by the Chi^2 value <1 based on I(q) plots. Frictional ratio (f/f_0) of models were calculated using HullRad.

Statistical analysis

Statistical analysis by unpaired student's t test was performed using GraphPad Prism 6.

6.2 Experimental procedures used in Chapter 3

Constructs design and expression

Mouse and human Fab clones were obtained, cloned and expressed as explained in experimental procedures described above. For single-chain Fab+ligand constructs, N-terminal ligand (lysozyme or insulin) were fused with 15x GS linker to the N-terminus of the havy chain of the Fab, with the rest of the construct intact. N-terminally linked ubiquiting was fused in the same way into Kossiakoff Lab pIPTG vector for Fab periplasmic secretion in bacteria.

ELISA

Polyreactivity assays were performed as described (Andrews et al. 2015, Mouquet et al. 2010, Wardemann et al. 2005). ELISA plates (Thermo) were coated overnight with 50 ml/well antigen diluted in carbonate buffer (Bethyl) with the exception of cardiolipin, which was coated overnight

in 100% ethanol left uncovered to allow evaporation. The following antigens and concentrations were used for coating: calf thymus DNA (Life Technologies), 10 mg/ml; human insulin (Fitzgerald), 5 mg/ml; LPS from E. coli (Sigma), 10 mg/ml; flagellin from S. typhimurium (InvivoGen), 2 mg/ml; cardiolipin (Sigma), 10 mg/ml; albumin from human serum, low endotoxin (Sigma), 10 mg/ml; and KLH, endotoxin free (Millipore), 10 mg/ml. Plates were washed 4x with nanopure H2O with an ELISA plate washer (BioTek) and blocked with 150 ml/well blocking buffer [1X TBS-T (Thermo), 1 mM EDTA (Boston BioProducts)] for 1 hour at 37°C. Plates were washed 4x with H2O using an ELISA plate washer (BioTek) and 50 ml/well of mAbs pre-diluted in TBS pH 7.4 were added. Different mutants of mAbs were assayed at 1 mg/ml and three additional 1:4 dilutions. Plates were incubated at 37°C for one hour then washed 4x with H2O. 75 ml/well of goat anti-human IgG HRP (Southern Biotech) diluted in blocking buffer were added and plates were incubated one hour at 37°C then washed 4x with H2O. 150 ml/well of blocking buffer was added and plates incubated 5 min at room temperature. Plates were washed 4x with H2O and 100 ml/well of developing reagent was added (Super AquaBlue ELISA Substrate, eBioscience). Plates were monitored at OD405 using an ELISA plate reader (BioTek) and reading was stopped when the positive control reached an OD405 of ~3.0. Average OD405 of TBS-only wells was subtracted from mAb containing wells for analysis.

Binding studies

All the binding studies were carried out using Bio-layer interferometry (BLI) (either Forte bio or Octet K2 from Pall Life Sciences) at room temperature (22°C). Biotinylated 25nt long ssDNA (from IDT) was immobilized on Streptavidin sensor to reach ~2-3 RU for binding with mouse IgA Fab clones at different concentrations and responses (in nm) were recorded. Hepes buffer saline

(10 mM HEPES pH 7.2, 150 mM NaCl) was used in all the measurements and sensor tips were recharged with 0.1M Glycine, pH 2.5 after each round. Traces were reference subtracted and kinetic parameters (K_D, k_{on}, k_{off}) were calculated by fitting the data with Forte-bio build-in analysis software.

Crystallography

Crystals of different Fab clones and single-chain ligand-Fab complexes were grown at room temperature by sitting drop vapor diffusion. Equal molar volumes of a protein solution containing 10-15 mg/ml complex was mixed with a reservoir solution containing different salts and PEGs. Crystallization conditions, cryoprotection, beamline and date of collection is shown in Table 5.

Table 5 Crystallization conditions and data collection

Crystal	Crystallization condition	Cryoprotection	Beamline	Date of collection
338E6 Fab	29% PEG 3000, 0.1M Sodium Sulfate, 0.1M Tris pH 8.5 (optimization)	20% Glycerol	APS 23IDB	02/06/2018
43G10 Fab	16% PEG 4000, 0.2M Ammonium sulfate, 0.1M HEPES pH 7.5, 10% Isopropanol (optimization, seeding)	20% Ethylene Glycol	APS 23IDC	08/13/2018
2G02 Fab	20% PEG 3350, 0.2M Magnesium formate (directly from HT screen)	40% PEG 3350	APS 23IDD	02/29/2019
3B03 Fab	25% PEG4000, 0.2M Ammonium sulfate, 0.1M Sodium acetate pH 4.6 (optimization)	20% Glycerol	APS 24IDE	07/17/2019
4C05 Fab	30% PEG 5000 MME, 0.2M Ammonium sulfate, 0.1M MES pH 6.5 (directly from HT screen)	20% Ethylene Glycol	APS 24IDE	07/17/2019
338E6+ Lysozyme	PEG 4000, Ammonium sulfate, Sodium Acetate pH 4.6 (optimized)	20% Ethylene Glycol	APS 23ID	12/03/2019
43G10+ Lysozyme	16% PEG 4000, 0.2M Ammonium sulfate, 0.1M HEPES pH 7.5, 10% Isopropanol (directly from HT)	20% Ethylene Glycol	APS 24IDC	12/17/2019

One dataset was collected from a single region of a single crystal for all protein samples at 100 K at APS beamlines 23ID and 24ID (λ = 0.9791 Å) on an Eiger detector. The data were processed, scaled, and merged using iMosfilm, truncated with ccp4 and search model using PHASER as molecular replacement tool was done with PHENIX. Cycles of manual model building and refinement (torsional non-crystallographic symmetry restraints applied throughout) were done with COOT and PHENIX. Data collection and processing statistics are listed in Table 3 and 4. Structure figures were generated with PyMOL (http://www.pymol.org).

SAXS

Small Angle X-ray Scattering data was collected at the SIBYLS 12.3.1 beamline at the Advanced Light Source using High Throughput Mail-In program. Scattering measurements were performed on 20-µl samples at concentrations ranging between 10-20mg/mL of various single-chain-ligand-Fab constructs. Collected datasets were processed using ATSAS and Scatter softwares by trimming low resolution data and adjusting R_g, d_{max} based on Guinier Peak Analysis and Real Space plots. Fitting of experimental profile to theoretical scattering profile from models was done using FoXS. Different models were designed based on crystal structures of Fab with ligands, where multiple models were generated from single complex to reflect single ligand and single Fab complex and sample different binding site of the ligand. The quality of fit between models and experimental SAXS data was assessed by the Chi^2 value <1 based on I(q) plots.

All-atom MD simulations

All simulations performed were prepared using the CHARMM-GUI Input Generator (Jo et al. 2007, Lee et al. 2016, Jo et al. 2008). All Fabs structures, were fully hydrated with TIP3P water

molecules and neutralized with 0.15 M KCl. All simulations were carried out in simulation boxes with periodic boundary conditions (20) using the additive PARAM36 force field from the CHARMM (Chemistry at HARvard Macromolecular Mechanics) (Jo et al. 2008). Simulations run for ~500ns, with a 2fs time step at 293.15K and used a combination of NAMD, and AMBER (Lee et al. 2016, Phillips et al. 2005, Friedrichs et al. 2009, Salomon-Ferrer et al. 2013). For all simulated systems run on the Midway Computing Cluster at the University of Chicago at least two replicas were run to confirm independence of results on initial velocity assignments. Data were analyzed using customized pyEMMA package installed within Jupyter notebook.

6.3 Experimental procedures used in Chapter 4

Expression and purification of full length Btn3A protein using insect cell system

The full-length Btn3A1 was subcloned into the pAcGP67A baculovirus transfer vector (BD Biosciences) encoding C-terminal 12x-polyhistidine-tag (12xHis). The full-length Btn3A2 was subcloned into a variant of pAcGP67A and in an additional construct a stop codon introduced upstream of the 12xHis allowing for the expression of tagless form of Btn3A2. Constructs were used for the production of recombinant baculoviruses using BestBac linearized baculovirus DNA (Expression Systems). Proteins were expressed through the infection or co-infection of High Five insect cells and incubation for 72 hours at 27°C and 125 rpm.

High Five insect cell culture was co-infected with baculoviruses encoding either Btn3A1-12xHis or Btn3A2 or a combination of both, incubated as described above and spun down for 15 minutes at 4°C and 1700 xg. Cells were washed in 20 mM HEPES buffer, pH 7.2 with 150 mM NaCl and spun down for 15 minutes at 4°C and 1700 xg. Pellet was resuspended in 10 mM Tris buffer, pH 8 with 1 mM EDTA and protease inhibitors cocktail (PIC; Sigma). Cells were lysed in a glass

homogenizer. Lysate was spun down for 30 minutes at 4°C and 40 000 xg and the pellet containing membrane fraction was collected. Btn3A1-12xHis/A2 or Btn3A1 only was extracted from the membrane with 50 mM Tris buffer, pH 8, with 300 mM NaCl, 1% (v/v) Triton X-100 (TX-100; Acros Organics) and PIC, rotated at 4°C for 2 hours. The suspension was spun down for 30 minutes at 4°C and 40 000 xg. Detergent soluble fraction was collected and incubated with Co-TALON resin (Qiagen) in the presence of 30 mM imidazole for 1 hour at 4°C. Depending on the detergent used downstream (for crystallography: DDM, LDAO, LMNG, FC14 or EM: A8-35), the resin was washed with 50 mM Tris buffer, pH 8 with 300 mM NaCl, 30 mM imidazole, 0.1% (v/v) TX-100 for EM. Alternatively resin was washed with a supplement of 0.04% DDM/ DDM:CHS (10:1) or 0.1% LDAO/ FC-14 or 0.008% LMNG for crystallography. Protein was eluted with 50 mM Tris buffer, pH 8, 150 mM NaCl, 350 mM imidazole, and appropriate detergent (as for wash step). For crystallography, sample containing Btn3A1-12xHis/A2 homo- or heterodimer was concentrated using Amicon Ultra filter with 100 kDa molecular cut-off (Millipore) and purified by gel filtration on Superdex 200 10/300 GL column (GE Healthcare) in 10 mM Tris buffer, pH 8 with 50 mM NaCl and appropriate detergent (as described before).

For cryo-EM, Btn3A1 elution from resin was reconstituted into Amphipole A8-35 as following. 50uL of 200ug/mL stock of A8-35 was added to 1mL of diluted (0.5-2.5mg/mL) Btn3A1 sample, incubated 5min on ice, then 0.5mL of concentrated (2-5mg/mL) 20.1Fab was added for stability, incubated 5min on ice. Excess of TX-100 detergent was removed by overnight incubation at 4°C with Bio-Beads SM-2 Adsorbents (Bio-Rad). For EM, Btn3A1+20.1Fab complex was lightly concentrated (2x) using Amicon Ultra filter with 100 kDa molecular cut-off (Millipore) and

purified by gel filtration on Superdex 200 10/300 GL column (GE Healthcare) in 10 mM Tris buffer, pH 8 with 350 mM NaCl.

Btn3A1/2 nanodisc production

Reconstitution of detergent solubilized his-tagged Btn3A1/A2 dimers was performed according to the protocols described by the Sligar laboratory with minor modifications (Ritchie et al., 2009). Btn3A1/A2 dimer was reconstituted into nanodiscs using either MSP 1D1 or E3D1 (Addgene) and POPC: TX-100 mixed micelles described in the SI Materials and Methods. Excess of detergent was removed by overnight incubation at 4°C with Bio-Beads SM-2 Adsorbents (Bio-Rad). Btn3A nanodiscs were separated from proteoliposomes and empty nanodiscs first by Ni-NTA resin pull-down (Qiagen) followed by Superdex 200 10/300 GL column (GE Healthcare) in 50 mM HEPES pH 7.5 with 50 mM NaCl. As a control, empty nanodiscs were prepared using the same procedure. Gel filtration fractions corresponding to Btn3A1/A2 nanodiscs were concentrated and an aliquot was shipped on ice to G. Skiniotis for negative stain electron microscopy. The remainder sample was supplemented with 5% (w/v) sucrose, flash-frozen in liquid nitrogen and stored at -80°C.

Negative stain electron microscopy of Btn3A1/2 nanodiscs

For Btn3A1/A2 nanodiscs produced as described above, samples were prepared for electron microscopy using the conventional negative staining protocol (Filizola, 2015), and imaged at room temperature with a Tecnai T12 electron microscope operated at 120 kV using low-dose procedures. Images were recorded at a magnification of 71,138x and a defocus value of ~1.5 μm on a Gatan US4000 CCD camera. All images were binned (2 × 2 pixels) to obtain a pixel size of 4.16 Å on the specimen level. Particles were manually excised using e2boxer (part of the EMAN 2 software

suite) 200 (Tang et al., 2007). 2D reference-free alignment and classification of particle projections was performed using ISAC (Yang et al., 2012). 4359 projections of Btn3A1/A2 nanodisc were subjected to ISAC producing 70 classes consistent over two-way matching and accounting for 3968 particle projections.

Cryo-electron microscopy of Btn3A1+20.1Fab in A8-35

For cryo-EM, just before grid preparation 1uL of 0.1% OMF was added to 10uL of Btn3A1 homodimer in complex with 20.1 at 0.35 mg/ml and was incubated for 30s on glow discharged Quantifoil 1.2/1.3 200 mesh EM grids, blotted for 5 s at 100% humidity and plunge frozen in liquid ethane using an FEI Vitrobot. Cryo-EM samples were imaged using a FEI Titan Krios at 300 KV, equipped with a K2 direct electron detector camera (Gatan). Images were collected manually at a nominal magnification of 30,000x with a pixel size of 1.19A˚ and defocus range of 1-3 μm. Total exposure time was 4 s with an accumulated dose of 48 e- /A˚ 2. Briefly, about 5000 particles were manually picked and used to generate a set of 7 reference-free 2D class averages using as templates for automatic particle picking from ~700 micrographs. Particle selection, CTF correction, 2D class averaging, and measurements were performed in Relion (Scheres, 2012) on the University of Chicago Midway computing cluster.

References

1. Round, J. L. & Palm, N. W. Causal effects of the microbiota on immune-mediated diseases. *Sci Immunol* **3,** eaao1603 (2018).

2. Hooper, L. V. & Gordon, J. I. Commensal Host-Bacterial Relationships in the Gut. *Science* **292,** 1115–1118 (2001).

3. Geva-Zatorsky, N., Sefik, E., Kua, L., Pasman, L., Tan, T. G., Ortiz-Lopez, A., Yanortsang, T. B., Yang, L., Jupp, R., Mathis, D., Benoist, C. & Kasper, D. L. Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell* **168,** 928-943.e11 (2017).

4. PROFT, T. & FRASER, J. D. Bacterial superantigens. *Clin Exp Immunol* **133,** 299–306 (2003).

5. Choi, Y., Herman, A., DiGiusto, D., Wade, T., Marrack, P. & Kappler, J. Residues of the variable region of the T-cell-receptor β-chain that interact with S. aureus toxin superantigens. *Nature* **346,** 471–473 (1990).

6. Dellabona, P., Peccoud, J., Kappler, J., Marrack, P., Benoist, C. & Mathis, D. Superantigens interact with MHC class II molecules outside of the antigen groove. *Cell* **62,** 1115–1121 (1990).

7. Miethke, T., Gaus, H., Wahl, C., Heeg, K. & Wagner, H. T-cell-dependent shock induced by a bacterial superantigen. *Chem Immunol* **55,** 172–84 (1992).

8. Fast, D. J., Schlievert, P. M. & Nelson, R. D. Toxic shock syndrome-associated staphylococcal and streptococcal pyrogenic toxins are potent inducers of tumor necrosis factor production. *Infect Immun* **57,** 291–4 (1989).

9. Graille, M., Stura, E. A., Corper, A. L., Sutton, B. J., Taussig, M. J., Charbonnier, J.-B. & Silverman, G. J. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human IgM antibody: Structural basis for recognition of B-cell receptors and superantigen activity. *Proc National Acad Sci* **97,** 5399–5404 (2000).

10. Derrick, J. P. & Wigley, D. B. Crystal structure of a streptococcal protein G domain bound to an Fab fragment. *Nature* **359,** 752–754 (1992).

11. Patella, V., Florio, G., Petraroli, A. & Marone, G. HIV-1 gp120 Induces IL-4 and IL-13 Release from Human FcεRI + Cells Through Interaction with the V H 3 Region of IgE. *J Immunol* **164,** 589–595 (2000).

12. Domiati-Saad, R. & Lipsky, P. E. Staphylococcal enterotoxin A induces survival of VH3-expressing human B cells by binding to the VH region with low affinity. *J Immunol Baltim Md 1950* **161,** 1257–66 (1998).

13. Graille, M., Stura, E. A., Housden, N. G., Beckingham, J. A., Bottomley, S. P., Beale, D., Taussig, M. J., Sutton, B. J., Gore, M. G. & Charbonnier, J.-B. Complex between Peptostreptococcus magnus Protein L and a Human Antibody Reveals Structural Convergence in the Interaction Modes of Fab Binding Proteins. *Structure* **9,** 679–687 (2001).

14. Pauli, N. T., Kim, H. K., Falugi, F., Huang, M., Dulac, J., Dunand, C. H., Zheng, N.-Y., Kaur, K., Andrews, S. F., Huang, Y., DeDent, A., Frank, K. M., Charnot-Katsikas, A., Schneewind, O. & Wilson, P. C. Staphylococcus aureus infection induces protein A-mediated immune evasion in humans. *J Exp Medicine* **211,** 2331–9 (2014).

15. Kim, M. W., Greenfield, B. K., Snyder, R. E., Steinmaus, C. M. & Riley, L. W. The association between community-associated Staphylococcus aureus colonization and disease: a meta-analysis. *Bmc Infect Dis* **18,** 86 (2018).

16. Krismer, B., Weidenmaier, C., Zipperer, A. & Peschel, A. The commensal lifestyle of Staphylococcus aureus and its interactions with the nasal microbiota. *Nat Rev Microbiol* **15,** 675–687 (2017).

17. van Belkum, A., Verkaik, N. J., de Vogel, C. P., Boelens, H. A., Verveer, J., Nouwen, J. L., Verbrugh, H. A. & Wertheim, H. F. L. Reclassification of Staphylococcus aureus Nasal Carriage Types. *J Infect Dis* **199,** 1820–1826 (2009).

18. Raeder, R. & Boyle, M. D. Association between expression of immunoglobulin G-binding proteins by group A streptococci and virulence in a mouse skin infection model. *Infect Immun* **61,** 1378–84 (1993).

19. Goodyear, C. S. & Silverman, G. J. Death by a B Cell Superantigen. *J Exp Medicine* **197,** 1125–1139 (2003).

20. Silverman, G. J., Cary, S. P., Dwyer, D. C., Luo, L., Wagenknecht, R. & Curtiss, V. E. A B Cell Superantigen–Induced Persistent "Hole" in the B-1 Repertoire. *J Exp Med* **192,** 87–98 (2000).

21. ERNTELL, M., MYHRE, E. B., SJØBRING, U. & KRONVALL, G. Non-immune Fab- and Fc- mediated interactions of avian Ig with S. aureus and group C and G streptococci. *Apmis* **96,** 239–249 (1988).

22. Goward, C. R., Scawen, M. D., Murphy, J. P. & Atkinson, T. Molecular evolution of bacterial cell-surface proteins. *Trends Biochem Sci* **18,** 136–140 (1993).

23. Bunker, J. J., Erickson, S. A., Flynn, T. M., Henry, C., Koval, J. C., Meisel, M., Jabri, B., Antonopoulos, D. A., Wilson, P. C. & Bendelac, A. Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science* **358,** eaan6619 (2017).

24. Bunker, J. J., Drees, C., Watson, A. R., Plunkett, C. H., Nagler, C. R., Schneewind, O., Eren, A. M. & Bendelac, A. B cell superantigens in the human intestinal microbiota. *Sci Transl Med* **11,** eaau9356 (2019).

25. Silverman, G. J. & Goodyear, C. S. Confounding B-cell defences: lessons from a staphylococcal superantigen. *Nat Rev Immunol* **6,** 465–475 (2006).

26. Schneewind, O. & Missiakas, D. M. Protein secretion and surface display in Gram-positive bacteria. *Philosophical Transactions Royal Soc Lond Ser B Biological Sci* **367,** 1123–39 (2012).

27. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Paslier, D. L., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K. L., Leclerc, M., Maguin, E., Minardi, R. M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S. D. & Wang, J. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).

28. Bell, A., Brunt, J., Crost, E., Vaux, L., Nepravishta, R., Owen, C. D., Latousakis, D., Xiao, A., Li, W., Chen, X., Walsh, M. A., Claesen, J., Angulo, J., Thomas, G. H. & Juge, N. Elucidation of a sialic acid metabolism pathway in mucus-foraging Ruminococcus gnavus unravels mechanisms of bacterial adaptation to the gut. *Nat Microbiol* **4,** 2393–2404 (2019).

29. Tailford, L. E., Owen, C. D., Walshaw, J., Crost, E. H., Hardy-Goddard, J., Gall, G. L., Vos, W. M. de, Taylor, G. L. & Juge, N. Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nat Commun* **6,** 7624 (2015).

30. Investigators, I., Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R. A., Braun, J., Denson, L. A., Jansson, J. K., Knight, R., Kugathasan, S., McGovern, D. P. B., Petrosino, J. F., Stappenbeck, T. S., Winter, H. S., Clish, C. B., Franzosa, E. A., Vlamakis, H., Xavier, R. J. & Huttenhower, C. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569,** 655–662 (2019).

31. Hall, A. B., Yassour, M., Sauk, J., Garner, A., Jiang, X., Arthur, T., Lagoudas, G. K., Vatanen, T., Fornelos, N., Wilson, R., Bertha, M., Cohen, M., Garber, J., Khalili, H., Gevers, D., Ananthakrishnan, A. N., Kugathasan, S., Lander, E. S., Blainey, P., Vlamakis, H., Xavier, R. J. & Huttenhower, C. A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Med* **9,** 103 (2017).

32. Azzouz, D., Omarbekova, A., Heguy, A., Schwudke, D., Gisch, N., Rovin, B. H., Caricchio, R., Buyon, J. P., Alekseyenko, A. V. & Silverman, G. J. Lupus nephritis is linked to disease-activity associated expansions and immunity to a gut commensal. *Ann Rheum Dis* **78,** 947–956 (2019).

33. Chua, H.-H., Chou, H.-C., Tung, Y.-L., Chiang, B.-L., Liao, C.-C., Liu, H.-H. & Ni, Y.-H. Intestinal Dysbiosis Featuring Abundance of Ruminococcus gnavus Associates With Allergic Diseases in Infants. *Gastroenterology* **154,** 154–167 (2018).

34. Zheng, H., Liang, H., Wang, Y., Miao, M., Shi, T., Yang, F., Liu, E., Yuan, W., Ji, Z.-S. & Li, D.-K. Altered Gut Microbiota Composition Associated with Eczema in Infants. *Plos One* **11,** e0166026 (2016).

35. Crost, E. H., Tailford, L. E., Gall, G. L., Fons, M., Henrissat, B. & Juge, N. Utilisation of Mucin Glycans by the Human Gut Symbiont Ruminococcus gnavus Is Strain-Dependent. *Plos One* **8,** e76341 (2013).

36. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–625 (2008).

37. Comstock, L. E. & Kasper, D. L. Bacterial Glycans: Key Mediators of Diverse Host Immune Responses. *Cell* **126,** 847–850 (2006).

38. Wyatt, P. J. Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation. *J Colloid Interf Sci* **197,** 9–20 (1998).

39. Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys J* **78,** 1606–1619 (2000).

40. Ribaric, S., Peterec, D. & Sketelj, J. Computer aided data acquisition and analysis of acetlycholinesterase velocity sedimentation profiles. *Comput Meth Prog Bio* **49,** 149–156 (1996).

41. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

42. Wucherpfennig, K. W., Allen, P. M., Celada, F., Cohen, I. R., Boer, R. D., Garcia, K. C., Goldstein, B., Greenspan, R., Hafler, D., Hodgkin, P., Huseby, E. S., Krakauer, D. C., Nemazee, D., Perelson, A. S., Pinilla, C., Strong, R. K. & Sercarz, E. E. Polyspecificity of T cell and B cell receptor recognition. *Semin Immunol* **19,** 216–224 (2007).

43. Chen, J., Eisen, H. N. & Kranz, D. M. A model T-cell receptor system for studying memory T-cell development. *Microbes Infect* **5,** 233–240 (2003).

44. Udaka, K., Tsomides, T. J. & Eisen, H. N. A naturally occurring peptide recognized by alloreactive CD8+ cytotoxic T lymphocytes in association with a class I MHC protein. *Cell* **69,** 989–998 (1992).

45. Wucherpfennig, K. W. & Strominger, J. L. Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80,** 695–705 (1995).

46. Wucherpfennig, K. W., Sette, A., Southwood, S., Oseroff, C., Matsui, M., Strominger, J. L. & Hafler, D. A. Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. *J Exp Medicine* **179,** 279–290 (1994).

47. Dimitrov, J. D. Harnessing the Therapeutic Potential of "Rogue" Antibodies. *Trends Pharmacol Sci* **41,** 409–417 (2020).

48. Kelly, R. L., Sun, T., Jain, T., Caffry, I., Yu, Y., Cao, Y., Lynaugh, H., Brown, M., Vásquez, M., Wittrup, K. D. & Xu, Y. High throughput cross-interaction measures for human IgG1 antibodies correlate with clearance rates in mice. *Mabs* **7,** 770–777 (2015).

49. Kelly, R. L., Le, D., Zhao, J. & Wittrup, K. D. Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries. *J Mol Biol* **430,** 119–130 (2018).

50. Datta-Mannan, A., Lu, J., Witcher, D. R., Leung, D., Tang, Y. & Wroblewski, V. J. The interplay of non-specific binding, target-mediated clearance and FcRn interactions on the pharmacokinetics of humanized antibodies. *Mabs* **7,** 1084–1093 (2015).

51. Mouquet, H., Scheid, J. F., Zoller, M. J., Krogsgaard, M., Ott, R. G., Shukair, S., Artyomov, M. N., Pietzsch, J., Connors, M., Pereyra, F., Walker, B. D., Ho, D. D., Wilson, P. C., Seaman, M. S., Eisen, H. N., Chakraborty, A. K., Hope, T. J., Ravetch, J. V., Wardemann, H. & Nussenzweig,

M. C. Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature* **467,** 591–5 (2010).

52. Prigent, J., Lorin, V., Kök, A., Hieu, T., Bourgeau, S. & Mouquet, H. Scarcity of autoreactive human blood IgA(+) memory B cells. *Eur J Immunol* **46,** 2340–2351 (2016).

53. Planchais, C., Kök, A., Kanyavuz, A., Lorin, V., Bruel, T., Guivel-Benhassine, F., Rollenske, T., Prigent, J., Hieu, T., Prazuck, T., Lefrou, L., Wardemann, H., Schwartz, O., Dimitrov, J. D., Hocqueloux, L. & Mouquet, H. HIV-1 Envelope Recognition by Polyreactive and Cross-Reactive Intestinal B Cells. *Cell Reports* **27,** 572-585.e7 (2019).

54. Mouquet, H., Klein, F., Scheid, J. F., Warncke, M., Pietzsch, J., Oliveira, T. Y. K., Velinzon, K., Seaman, M. S. & Nussenzweig, M. C. Memory B Cell Antibodies to HIV-1 gp140 Cloned from Individuals Infected with Clade A and B Viruses. *Plos One* **6,** e24078 (2011).

55. Prigent, J., Jarossay, A., Planchais, C., Eden, C., Dufloo, J., Kök, A., Lorin, V., Vratskikh, O., Couderc, T., Bruel, T., Schwartz, O., Seaman, M. S., Ohlenschläger, O., Dimitrov, J. D. & Mouquet, H. Conformational Plasticity in Broadly Neutralizing HIV-1 Antibodies Triggers Polyreactivity. *Cell Reports* **23,** 2568–2581 (2018).

56. Andrews, S. F., Huang, Y., Kaur, K., Popova, L. I., Ho, I. Y., Pauli, N. T., Dunand, C. J. H., Taylor, W. M., Lim, S., Huang, M., Qu, X., Lee, J.-H., Salgado-Ferrer, M., Krammer, F., Palese, P., Wrammert, J., Ahmed, R. & Wilson, P. C. Immune history profoundly affects broadly protective B cell responses to influenza. *Sci Transl Med* **7,** 316ra192-316ra192 (2015).

57. Jain, T., Boland, T., Lilov, A., Burnina, I., Brown, M., Xu, Y. & Vásquez, M. Prediction of delayed retention of antibodies in hydrophobic interaction chromatography from sequence using machine learning. *Bioinformatics* **33,** 3758–3766 (2017).

58. Neu, K. E., Guthmiller, J. J., Huang, M., La, J., Vieira, M. C., Kim, K., Zheng, N.-Y., Cortese, M., Tepora, M. E., Hamel, N. J., Rojas, K. T., Henry, C., Shaw, D., Dulberger, C. L., Pulendran, B., Cobey, S., Khan, A. A. & Wilson, P. C. Spec-seq unveils transcriptional subpopulations of antibody-secreting cells following influenza vaccination. *J Clin Invest* **129,** 93–105 (2018).

59. Wrammert, J., Koutsonanos, D., Li, G.-M., Edupuganti, S., Sui, J., Morrissey, M., McCausland, M., Skountzou, I., Hornig, M., Lipkin, W. I., Mehta, A., Razavi, B., Rio, C. D., Zheng, N.-Y., Lee, J.-H., Huang, M., Ali, Z., Kaur, K., Andrews, S., Amara, R. R., Wang, Y., Das, S. R., O'Donnell, C. D., Yewdell, J. W., Subbarao, K., Marasco, W. A., Mulligan, M. J., Compans, R., Ahmed, R. & Wilson, P. C. Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection Human B cell responses to pandemic H1N1 influenza. *J Exp Medicine* **208,** 181–193 (2011).

60. Koelsch, K., Zheng, N.-Y., Zhang, Q., Duty, A., Helms, C., Mathias, M. D., Jared, M., Smith, K., Capra, J. D. & Wilson, P. C. Mature B cells class switched to IgD are autoreactive in healthy individuals. *J Clin Invest* **117,** 1558–1565 (2007).

61. Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C. & Wardemann, H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* **329,** 112–124 (2008).

62. Wardemann, H., Yurasov, S., Schaefer, A., Young, J. W., Meffre, E. & Nussenzweig, M. C. Predominant Autoantibody Production by Early Human B Cell Precursors. *Science* **301,** 1374–1377 (2003).

63. Chen, J., Frey, G., Peng, H., Rits-Volloch, S., Garrity, J., Seaman, M. S. & Chen, B. Mechanism of HIV-1 Neutralization by Antibodies Targeting a Membrane-Proximal Region of gp41. *J Virol* **88,** 1249–1258 (2014).

64. Dennison, S. M., Anasti, K., Scearce, R. M., Sutherland, L., Parks, R., Xia, S.-M., Liao, H.-X., Gorny, M. K., Zolla-Pazner, S., Haynes, B. F. & Alam, S. M. Nonneutralizing HIV-1 gp41 Envelope Cluster II Human Monoclonal Antibodies Show Polyreactivity for Binding to Phospholipids and Protein Autoantigens. *J Virol* **85,** 1340–1347 (2011).

65. Haynes, B. F., Fleming, J., Clair, E. W. St., Katinger, H., Stiegler, G., Kunert, R., Robinson, J., Scearce, R. M., Plonk, K., Staats, H. F., Ortel, T. L., Liao, H.-X. & Alam, S. M. Cardiolipin Polyspecific Autoreactivity in Two Broadly Neutralizing HIV-1 Antibodies. *Science* **308,** 1906–1908 (2005).

66. Bajic, G., Poel, C. E. van der, Kuraoka, M., Schmidt, A. G., Carroll, M. C., Kelsoe, G. & Harrison, S. C. Autoreactivity profiles of influenza hemagglutinin broadly neutralizing antibodies. *Sci Rep-uk* **9,** 3492 (2019).

67. Jeliazkov, J. R., Sljoka, A., Kuroda, D., Tsuchimura, N., Katoh, N., Tsumoto, K. & Gray, J. J. Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification. *Front Immunol* **9,** 413 (2018).

68. Burnett, D. L., Schofield, P., Langley, D. B., Jackson, J., Bourne, K., Wilson, E., Porebski, B. T., Buckle, A. M., Brink, R., Goodnow, C. C. & Christ, D. Conformational diversity facilitates antibody mutation trajectories and discrimination between foreign and self-antigens. *Proc National Acad Sci* 202005102 (2020). doi:10.1073/pnas.2005102117

69. James, L. C., Roversi, P. & Tawfik, D. S. Antibody Multispecificity Mediated by Conformational Diversity. *Science* **299,** 1362–1367 (2003).

70. Arevalo, J. H., Hassig, C. A., Stura, E. A., Sims, M. J., Taussig, M. J. & Wilson, I. A. Structural Analysis of Antibody Specificity Detailed Comparison of Five Fab′-Steroid Complexes. *J Mol Biol* **241,** 663–690 (1994).

71. Foote, J. & Milstein, C. Conformational isomerism and the diversity of antibodies. *Proc National Acad Sci* **91,** 10370–10374 (1994).

72. Wilson, I. A. & Stanfield, R. L. Antibody-antigen interactions: new structures and new conformational changes. *Curr Opin Struc Biol* **4,** 857–867 (1994).

73. Fernández-Quintero, M. L., Loeffler, J. R., Bacher, L. M., Waibl, F., Seidler, C. A. & Liedl, K. R. Local and Global Rigidification Upon Antibody Affinity Maturation. *Frontiers Mol Biosci* **7,** 182 (2020).

74. Xiao, H., Guo, T., Yang, M., Qi, J., Huang, C., Hong, Y., Gu, J., Pang, X., Liu, W. J., Peng, R., McCauley, J., Bi, Y., Li, S., Feng, J., Zhang, H., Zhang, X., Lu, X., Yan, J., Chen, L., Shi, Y., Chen, W. & Gao, G. F. Light chain modulates heavy chain conformation to change protection profile of monoclonal antibodies against influenza A viruses. *Cell Discov* **5,** 21 (2019).

75. Eagle, R. A. & Trowsdale, J. Promiscuity and the single receptor: NKG2D. *Nat Rev Immunol* **7,** 737–744 (2007).

76. McFarland, B. J. & Strong, R. K. Thermodynamic Analysis of Degenerate Recognition by the NKG2D Immunoreceptor Not Induced Fit but Rigid Adaptation. *Immunity* **19,** 803–812 (2003).

77. Rabia, L. A., Zhang, Y., Ludwig, S. D., Julian, M. C. & Tessier, P. M. Net charge of antibody complementarity-determining regions is a key predictor of specificity. *Protein Eng Des Sel* **31,** 409–418 (2019).

78. Lecerf, M., Kanyavuz, A., Lacroix-Desmazes, S. & Dimitrov, J. D. Sequence features of variable region determining physicochemical properties and polyreactivity of therapeutic antibodies. *Mol Immunol* **112,** 338–346 (2019).

79. Birtalan, S., Zhang, Y., Fellouse, F. A., Shao, L., Schaefer, G. & Sidhu, S. S. The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *J Mol Biol* **377,** 1518–1528 (2008).

80. Jo, S., Kim, T. & Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *Plos One* **2,** e880 (2007).

81. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **29,** 1859–1865 (2008).

82. Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B. & Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12,** 405–413 (2016).

83. Gober, H.-J., Kistowska, M., Angman, L., Jenö, P., Mori, L. & Libero, G. D. Human T Cell Receptor γδ Cells Recognize Endogenous Mevalonate Metabolites in Tumor Cells. *J Exp Medicine* **197,** 163–168 (2003).

84. Morita, C. T., Jin, C., Sarikonda, G. & Wang, H. Nonpeptide antigens, presentation mechanisms, and immunological memory of human Vγ2Vδ2 T cells: discriminating friend from foe through the recognition of prenyl pyrophosphate antigens. *Immunol Rev* **215,** 59–76 (2007).

85. Carding, S. R. & Egan, P. J. γδ T cells: functional plasticity and heterogeneity. *Nat Rev Immunol* **2,** 336–345 (2002).

86. Silva-Santos, B., Serre, K. & Norell, H. γδ T cells in cancer. *Nat Rev Immunol* **15,** 683–691 (2015).

87. Vantourout, P. & Hayday, A. Six-of-the-best: unique contributions of γδ T cells to immunology. *Nat Rev Immunol* **13,** 88–100 (2013).

88. Melandri, D., Zlatareva, I., Chaleil, R. A. G., Dart, R. J., Chancellor, A., Nussbaumer, O., Polyakova, O., Roberts, N. A., Wesch, D., Kabelitz, D., Irving, P. M., John, S., Mansour, S., Bates, P. A., Vantourout, P. & Hayday, A. C. The γδTCR combines innate immunity with adaptive immunity by utilizing spatially distinct regions for agonist selection and antigen responsiveness. *Nat Immunol* **19,** 1352–1365 (2018).

89. Willcox, C. R., Vantourout, P., Salim, M., Zlatareva, I., Melandri, D., Zanardo, L., George, R., Kjaer, S., Jeeves, M., Mohammed, F., Hayday, A. C. & Willcox, B. E. Butyrophilin-like 3 Directly Binds a Human Vγ4+ T Cell Receptor Using a Modality Distinct from Clonally-Restricted Antigen. *Immunity* **51,** 813-825.e4 (2019).

90. Morita, C. T., Mariuzza, R. A. & Brenner, M. B. Antigen recognition by human γδ T cells: pattern recognition by the adaptive immune system. *Springer Semin Immun* **22,** 191–217 (2000).

91. Chen, C. Y., Yao, S., Huang, D., Wei, H., Sicard, H., Zeng, G., Jomaa, H., Larsen, M. H., Jacobs, W. R., Wang, R., Letvin, N., Shen, Y., Qiu, L., Shen, L. & Chen, Z. W. Phosphoantigen/IL2 Expansion and Differentiation of Vγ2Vδ2 T Cells Increase Resistance to Tuberculosis in Nonhuman Primates. *Plos Pathog* **9,** e1003501 (2013).

92. Tanaka, Y., Morita, C. T., Tanaka, Y., Nieves, E., Brenner, M. B. & Bloom, B. R. Natural and synthetic non-peptide antigens recognized by human γδ T cells. *Nature* **375,** 155–158 (1995).

93. Belmant, C., Espinosa, E., Halary, F., Tang, Y., Peyrat, M.-A., Sicard, H., Kozikowski, A., Buelow, R., Poupot, R., Bonneville, M. & Fournié, J.-J. A chemical basis for recognition of nonpeptide antigens by human γδ T cells*. *Faseb J* **14,** 1669–1670 (2000).

94. Hintz, M., Reichenberg, A., Altincicek, B., Bahr, U., Gschwind, R. M., Kollas, A.-K., Beck, E., Wiesner, J., Eberl, M. & Jomaa, H. Identification of (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate as a major activator for human γδ T cells in Escherichia coli. *Febs Lett* **509,** 317–322 (2001).

95. Sandstrom, A., Peigné, C.-M., Léger, A., Crooks, J. E., Konczak, F., Gesnel, M.-C., Breathnach, R., Bonneville, M., Scotet, E. & Adams, E. J. The Intracellular B30.2 Domain of Butyrophilin 3A1 Binds Phosphoantigens to Mediate Activation of Human Vγ9Vδ2 T Cells. *Immunity* **40,** 490–500 (2014).

96. Bukowski, J. F., Morita, C. T., Band, H. & Brenner, M. B. Crucial role of TCR gamma chain junctional region in prenyl pyrophosphate antigen recognition by gamma delta T cells. *J Immunol Baltim Md 1950* **161,** 286–93 (1998).

97. Lang, F., Peyrat, M. A., Constant, P., Davodeau, F., David-Ameline, J., Poquet, Y., Vié, H., Fournié, J. J. & Bonneville, M. Early activation of human V gamma 9V delta 2 T cell broad cytotoxicity and TNF production by nonpeptidic mycobacterial ligands. *J Immunol Baltim Md 1950* **154,** 5986–94 (1995).

98. Rigau, M., Ostrouska, S., Fulford, T. S., Johnson, D. N., Woods, K., Ruan, Z., McWilliam, H. E. G., Hudson, C., Tutuka, C., Wheatley, A. K., Kent, S. J., Villadangos, J. A., Pal, B., Kurts, C., Simmonds, J., Pelzing, M., Nash, A. D., Hammet, A., Verhagen, A. M., Vairo, G., Maraskovsky, E., Panousis, C., Gherardin, N. A., Cebon, J., Godfrey, D. I., Behren, A. & Uldrich, A. P. Butyrophilin 2A1 is essential for phosphoantigen reactivity by γδ T cells. *Science* **367,** eaay5516 (2020).

99. Karunakaran, M. M., Willcox, C. R., Salim, M., Paletta, D., Fichtner, A. S., Noll, A., Starick, L., Nöhren, A., Begley, C. R., Berwick, K. A., Chaleil, R. A. G., Pitard, V., Déchanet-Merville, J.,

Bates, P. A., Kimmel, B., Knowles, T. J., Kunzmann, V., Walter, L., Jeeves, M., Mohammed, F., Willcox, B. E. & Herrmann, T. Butyrophilin-2A1 Directly Binds Germline-Encoded Regions of the Vγ9Vδ2 TCR and Is Essential for Phosphoantigen Sensing. *Immunity* **52,** 487-498.e6 (2020).

100. Rhodes, D. A., Chen, H.-C., Price, A. J., Keeble, A. H., Davey, M. S., James, L. C., Eberl, M. & Trowsdale, J. Activation of Human γδ T Cells by Cytosolic Interactions of BTN3A1 with Soluble Phosphoantigens and the Cytoskeletal Adaptor Periplakin. *J Immunol* **194,** 2390–2398 (2015).

101. Sebestyen, Z., Scheper, W., Vyborova, A., Gu, S., Rychnavska, Z., Schiffler, M., Cleven, A., Chéneau, C., Noorden, M. van, Peigné, C.-M., Olive, D., Lebbink, R. J., Oostvogels, R., Mutis, T., Schuurhuis, G. J., Adams, E. J., Scotet, E. & Kuball, J. RhoB Mediates Phosphoantigen Recognition by Vγ9Vδ2 T Cell Receptor. *Cell Reports* **15,** 1973–85 (2016).

102. Arnett, H. A. & Viney, J. L. Immune modulation by butyrophilins. *Nat Rev Immunol* **14,** 559–569 (2014).

103. Abeler-Dörner, L., Swamy, M., Williams, G., Hayday, A. C. & Bas, A. Butyrophilins: an emerging family of immune regulators. *Trends Immunol* **33,** 34–41 (2012).

104. Gu, S., Nawrocka, W. & Adams, E. J. Sensing of Pyrophosphate Metabolites by Vγ9Vδ2 T Cells. *Front Immunol* **5,** 688 (2014).

105. Palakodeti, A., Sandstrom, A., Sundaresan, L., Harly, C., Nedellec, S., Olive, D., Scotet, E., Bonneville, M. & Adams, E. J. The Molecular Basis for Modulation of Human Vγ9Vδ2 T Cell Responses by CD277/Butyrophilin-3 (BTN3A)-specific Antibodies. *J Biol Chem* **287,** 32780–32790 (2012).

106. Yang, Y., Li, L., Yuan, L., Zhou, X., Duan, J., Xiao, H., Cai, N., Han, S., Ma, X., Liu, W., Chen, C.-C., Wang, L., Li, X., Chen, J., Kang, N., Chen, J., Shen, Z., Malwal, S. R., Liu, W., Shi,

Y., Oldfield, E., Guo, R.-T. & Zhang, Y. A Structural Change in Butyrophilin upon Phosphoantigen Binding Underlies Phosphoantigen-Mediated Vγ9Vδ2 T Cell Activation. *Immunity* **50,** 1043-1053.e5 (2019).

107. Wang, H., Nada, M. H., Tanaka, Y., Sakuraba, S. & Morita, C. T. Critical Roles for Coiled-Coil Dimers of Butyrophilin 3A1 in the Sensing of Prenyl Pyrophosphates by Human Vγ2Vδ2 T Cells. *J Immunol* **203,** 607–626 (2019).

108. Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M. & Sligar, S. G. Chapter Eleven Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods Enzymol* **464,** 211–231 (2009).

109. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* **157,** 38–46 (2007).

110. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative Stable Alignment and Clustering of 2D Transmission Electron Microscope Images. *Structure* **20,** 237–247 (2012).