uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Hoa P. Nguyen
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biochemistry)
GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Structural Studies of Polyspecific Anti-Carbohydrate Antibodies

TITRE DE LA THÈSE / TITLE OF THESIS

Stephen Evans
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Earl Brown

Danielle Carrier

Louis Delbaere

Martin Young

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Structural Studies of Polyspecific Anti-Carbohydrate Antibodies

Hoa P. Nguyen

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in Biochemistry

Department of Biochemistry, Microbiology & Immunology
Faculty of Medicine
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-15035-1
Our file Notre référence
ISBN: 978-0-494-15035-1

Abstract

The inheritance of germline immunoglobulin genes must ensure the recognition of prevalent pathogens and the ability to adapt to new pathogens. The current understanding of the generation of diversity of immunoglobulin binding sites involves the recombination of the B-lymphocyte germline genes during development. In general, the production of immunoglobulins proceeds with the help of T-lymphocytes through a process called "affinity maturation" involving mutations that increase antigen specificity. Carbohydrate epitopes, however, do not illicit such a response, and so are more dependent on the primary germline repertoire. Although anti-carbohydrate antibodies exhibit exquisite specifity, to date there have been few structural reports of complexes due to their low binding affinities. Here, high-resolution X-ray crystal structures of antigen-binding fragments from two highly homologous immunoglobulins S25-2 and S45-18 in their unliganded forms and bound to several carbohydrate epitopes of 3-deoxy-*D-manno*-oct-2-ulosonic acid (Kdo) found in Gram-negative bacteria including members of the family *Chlamydiaceae.* The combining sites of both immunoglobulins are similarly composed of a common germline-conserved "pocket" that binds a non-reducing terminal Kdo monosaccharide, and a "groove"-like remainder of the binding site that uses the same residues multiply in interactions with different epitopes in the case of the germline-conserved S25-2, or that has undergone mutations that enhance the specificity for a single epitope in the case of S45-18. The structural basis of this adaptability is accomplished firstly, by the combination of a specific set of germline V_H and V_κ genes that encodes a binding pocket that recognizes the non-reducing terminal Kdo; secondly, by a difference in the nature of the D and J_H mini-genes between S25-2 and S45-18; and thirdly, by the contributions of conserved germline

residues compared to regions of amino acid mutation. Furthermore, the strategy of coupling a strict binding pocket with a more adaptive region demonstrates how anti-carbohydrate immunoglobulins can bind a limited range of related epitopes without necessarily incurring the entropic penalty of immobilisation of flexible regions.

To my parents and to my family,

for their generosity, patience and sacrifice.

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Bragg

Acknowledgements

This work would not be possible without the support, encouragement and wisdom of my supervisor, Dr. Stephen V. Evans. I also thank the members of the Evans lab: Sonia and Roxanne; the members of my advisory committee Dr. Denis Williamson, Dr. Carol P. Huber, Dr. Svetlana Borisova, and Dr. John E. Baenziger for their insight and guidance; the members and support staff of the department of Biochemistry, Microbiology & Immunlogy at the University of Ottawa (thank you Carol Ann); and to my collaborators at the laboratories of Dr. Nina Seto and Dr. C. Roger Mackenzie at NRC, Dr. Helmut Brade at the Research Center Borstel, Dr. Paul Kosma at the University of Agricultural Sciences Vienna; the staff at Brookhaven National Labs NSLS beamline X8C, and other help from the crystallography laboratories of Queen's University, and of the University of Toronto. Finally, I acknowledge other important people in my life for their unending support and faith: Natalie, Steve, Kien, Tim, Ling, Irene, Jonathan and Neil.

Table of Contents

List of Abbreviations used

Ab	antibody
APC	antigen-presenting cell
ATP	adenosine triphosphate
BcR	B-cell receptor
CDR	complementarity determining region
EB	elementary body
Fab	antigen-binding fragment
Fc	crystallizable fragment
FR	framework region
Gal	galactose
Glc	glucose
IEF	isoelectric focusing
Ig	immunoglobulin
Kdo	3-deoxy-*D-manno*-oct-2-ulosonic acid
LPS	lipopolysaccharide
mAb	monoclonal antibody
Man	mannose
MBL	mannose-binding lectin
MHC	major histocompatibility complex
NAc	N-acetyl
NaCacod	sodium cacodylate
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen-associated molecular pattern
pI	isoelectric point
PRR	pattern recognition receptor
RAG	recombination-activation gene
RB	reticulate body
Rha	rhamnose
SDS	sodium dodecylsulphate
TcR	T-cell receptor
TdT	Deoxynucleotidyl transferase
TLR	toll-like receptor

List of Figures

List of Tables

Preface

The limited amount of genomic information dedicated to the immune response must contain the capability to protect against common pathogens, while being able to adapt to new challenges.

For antibodies, the current understanding of the generation of structural diversity in binding sites relies on combinations of genetic rearrangements of several germline genes that form each site, followed by an "affinity maturation process" where clonal selection and hypermutation results in enhanced affinity towards an epitope. However, carbohydrates in general do not stimulate T-cell help and this affinity maturation process does not occur. This study probes the germline response against carbohydrates by determining the three-dimensional structures in liganded and unliganded forms of antibody-binding pockets formed largely from conserved germline sequences. We have shown that these antibodies combine hereditary and flexible binding subsites that allow them to adapt to several related but distinct ubiquitous carbohydrate epitopes found on bacterial surfaces.

Epitopes of 3-deoxy-*D-manno*-oct-2-ulosonic acid (Kdo), which is a sugar found in the lipopolysaccharide (LPS) of all Gram-negative bacteria, are used as models to explore the recognition of common bacterial carbohydrate epitopes. LPS has been the subject of vaccine and diagnostic investigations for decades. This thesis presents a range of clinically-relevant Kdo epitopes found on rough ("Re")-type LPS and from members of the bacterial family *Chlamydiacaea* in complex with two antibodies. The high-resolution structures of a remarkable antibody (S25-2) that has conserved germline sequences has been determined in two unliganded forms and in complex with several distinct but related

carbohydrate epitopes. The unliganded and complexed structures of the related antibody S45-18 further reveals how specific mutations can enhance carbohydrate specificity.

Section A: Introduction and Methods

1.0 Introduction

1.1 Immunoglobulin response

1.1.1 Carbohydrates in biological systems

1.1.1.1 Carbohydrate structure and function

Carbohydrates show both small and large variations in physical and chemical properties which can be altered by small modifications, and specific recognition in protein-carbohydrate interactions is critical to many biological processes: immune defense, recognition and effector pathways, blood coagulation (Gralnick *et al.,* 1976; Gralnick *et al.,* 1983); cell maturation (Kaushansky *et al.,* 1987; Muramatsu, 2000; Akama *et al.,* 2002; Muramatsu, 2002) and tumour metastasis (Singhal & Hakomori, 1990). Carbohydrates can also take part in unique microbial biological functions which are essential for survival and reproduction: membrane homeostasis (Belunis *et al.,* 1995), host defense evasion (Alexander & Elder, 1984), and host cell adhesion (Krivan *et al.,* 1988).

Carbohydrates range from monosaccharides to complex polysaccharides and occur in linear or branched configurations. This work focuses on derivatives of 6-membered ring structures with an etheric ring oxygen and one out-of-ring carbon atom. The substituent hydroxyl (-OH) groups can be arranged in axial or equatorial orientations with respect to the plane of the ring. These are stereoisomers, molecules with the same chemical formula and chemical attachments but different orientations of their functional groups.

Nomenclature exists for each sugar stereoisomer that reflects their different physical and chemical properties.

Three common naturally occurring hexoses are *D*-glucose, *D*-galactose and *D*-mannose, each with unique orientations of their hydroxyl groups. The monosaccharide unit Kdo (3-deoxy-*D-manno*-oct-2-ulosonic acid) is an 8-carbon molecule. It consists of a 3-deoxy hexose ring with a carboxylate group attached at C2 and a 1,2-ethanediol group at C6 (Fig. 1.1).

Generally the sugar ring is found mainly in "chair" conformations rather than "boat" conformations to minimize steric crowding of the hydroxyl groups. Glycosidic bonds are etheric covalent bonds formed by the condensation of the hydroxyl groups. More than one glycosidic bond can occur in a monosaccharide leading to linear or branched structures of oligo- and polysaccharides. Glycosidic bonds (Fig. 1.2A) have some degree of rotational freedom, generally limited by the steric and electronic forces between sugar residues.

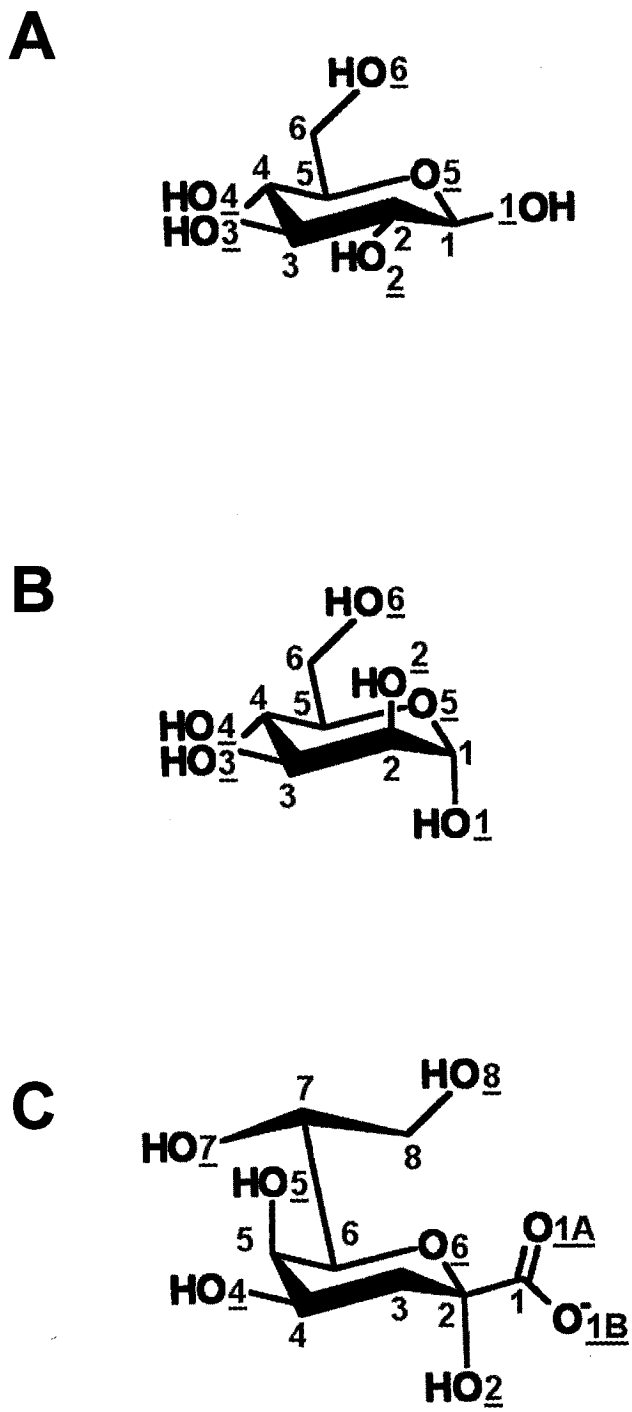


Figure 1.1 Chair conformation representations of **A)** β*D*-glucose **B)** α*D*-mannose **C)** α*D*-Kdo (3-deoxy-*D-manno*-oct-2-ulosonic acid). Hydroxyl group configurations attached to the first numbered ring carbon atom are designated by α (axial) and β (equatorial). Carbons are numbered in *green*. Oxygen atoms are numbered in *green* and *underlined*.

1.1.1.2 Protein-carbohydrate interactions

A condensation reaction between the amino and carbonyl groups of adjacent amino acids forms a peptide bond (Fig. 1.2B). The bond resonance between the atoms involved in the peptide bond gives it a planar character which reduces the freedom available during folding. Local secondary structures such as alpha helices and beta sheets are stabilized by non-covalent hydrogen bonding interactions of adjacent atoms. With the exception of sugars containing $\alpha(2\rightarrow8)$ glycosidic bonds, for example in sialic acid and Kdo, peptides have greater flexibility.

The recognition of carbohydrate structures by proteins in general utilizes dipole-dipole, hydrogen bonding, salt bridging, water-bridging, stacking (between non-polar atoms), and van der Waals forces (London dispersion forces) (Fig. 1.3).

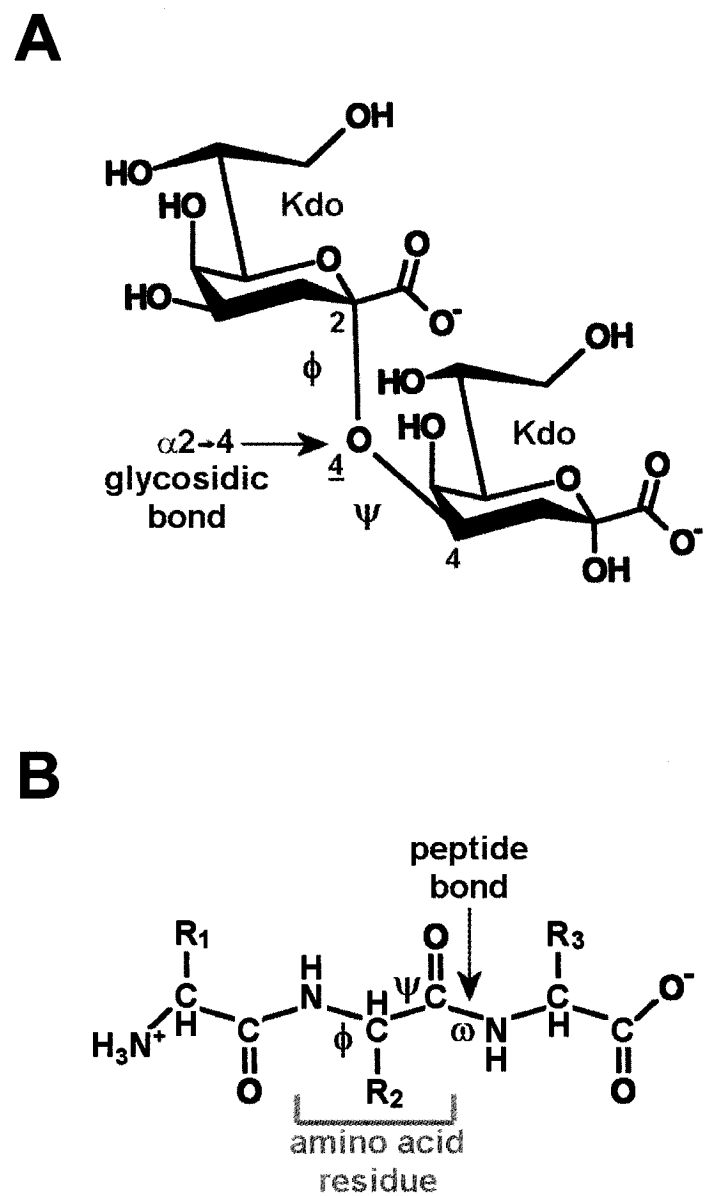


Figure 1.2 **A)** The glycosidic bond is an ether bond formed between two saccharide units. The carbon and oxygen atoms (*underlined*) are numbered. **B)** The peptide bond is formed between two amino acids. R*n* denotes the side chain. The dihedral angles are indicated.

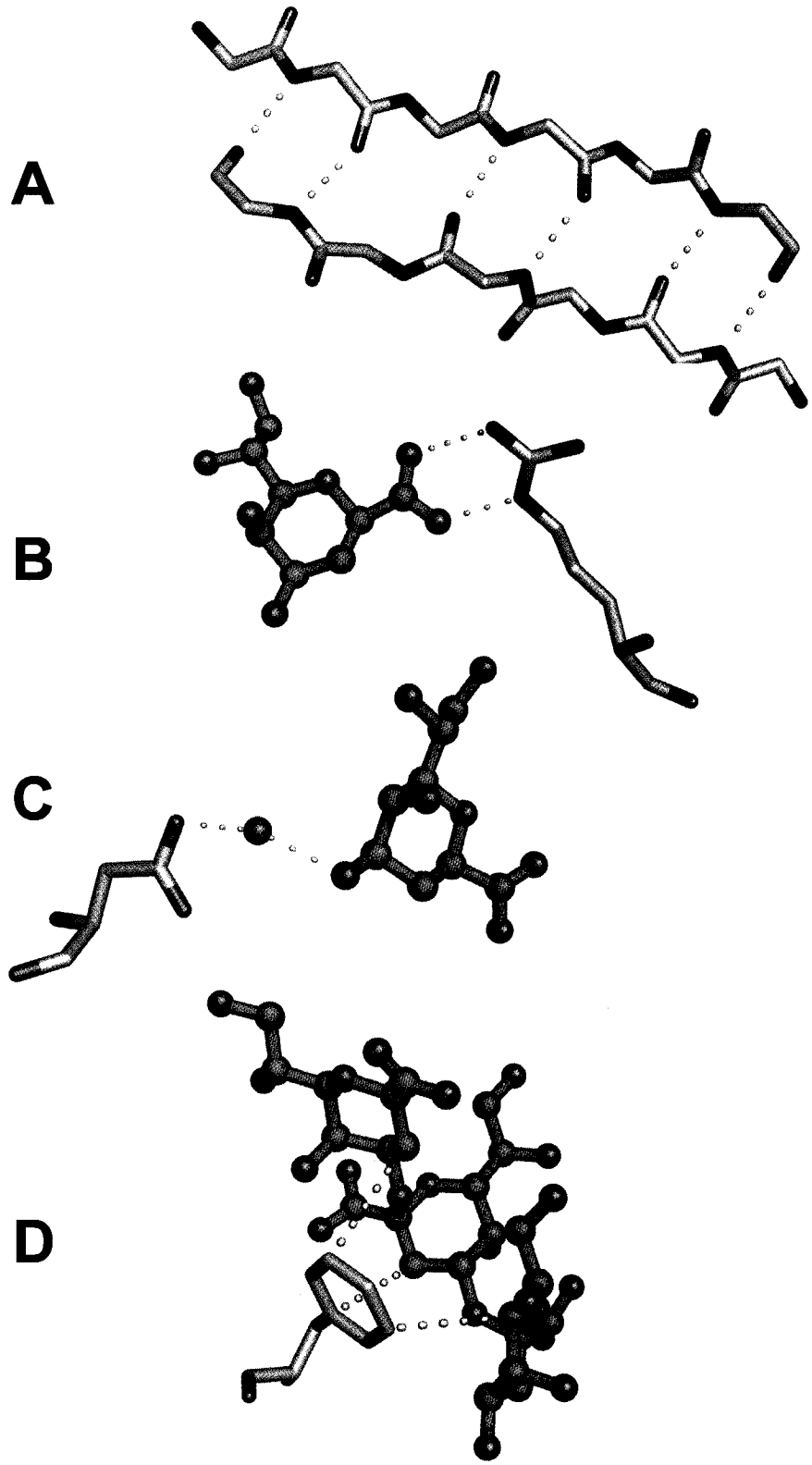


Figure 1.3 Examples of non-covalent interatomic interactions: **A)** hydrogen-bonding in the polypeptide backbone that forms a beta-sheet (amino acid side chains and hydrogen atoms have been removed for clarity), **B)** salt-bridging between charged groups, **C)** bridging water interaction with a monosaccharide (*green ball and stick*) hydroxyl group and **D)** van der Waals interactions. Hydrogen atoms have been removed for clarity. Protein carbon atoms are shown in *grey*, nitrogen in *blue*, and oxygen in *red.*

Several families of proteins can interact with carbohydrates, and have been extensively reviewed: enzymes (Singhal & Hakomori, 1990), immunoglobulins (Bundle & Young, 1992), lectins (Sharon & Lis, 1990; Lasky, 1992; Brandley *et al.,* 1993; Sharon, 1993), and other non-catalytic carbohydrate recognizing domains (Charnock *et al.,* 2002).

Two important aspects of the immune system are the abilities to specifically recognize a molecule, and to form a response against it. The term "antigen" refers to a molecule which the immune system recognizes. These can be of foreign origin to the host such as products of pathogens, or from the host itself as products of normal or abnormal physiology such as in the case of tumorigenesis. An "immunogen" refers to a molecule that can create an immune response. Another common term is a "hapten", which refers to a molecule that is recognized by the immune system, but by itself does not elicit an immune response. Antigens may be of peptide origin; however, there are also immune responses directed against carbohydrates, lipids and other molecules, using pathways that can be different than those elicited against peptide antigens.

The immune system shows specific recognition of many carbohydrates: alloimmunity such as ABO sugar blood group specificity (Landsteiner 1901), changes in surface carbohydrates on transformed tumor cells (Killion *et al.,* 1976; Fukuda 1996); and also chemotaxis in inflammatory responses (Phillips *et al.,* 1990; Lasky 1992), complement pathway activation (Nose & Wigzell 1983; Coloma *et al.,* 2000), and detection of carbohydrates from microorganisms (Feinberg *et al.,* 2000; Feinberg *et al.,* 2001).

Prior to these studies, there have been only four different carbohydrate antigen-antibody complexes reported. These include: antibodies with different O-polysaccharides of LPS (Cygler *et al.,* 1991; Villeneuve *et al.,* 2000; Vyas *et al.,* 2002), and with the tumour-specific Lewis Y epitope (Jeffrey *et al.,* 1995). The epitopes involved in these

previous reports are unrelated and affinity matured antibodies with little conservation of their respective germline sequences, and did not discuss whether binding to multiple epitopes was observed by germline immunoglobulins.

1.1.2 Antigen recognition by immunoglobulins: a historical perspective

The debate as to how the immune system recognizes, generates, maintains, and inherits molecular recognition dates back over 100 years when Behring and Kitasato (Behring & Kitasato, 1890) first used the term "antibody" to describe substances produced that circulated in the blood of animals that had become immune to diptheria and tetanus.

The concept of inheritance of specificity was introduced in the early 1900s by the "germline theory". Ehrlich proposed in the "side chain theory" (Ehrlich, 1900) that foreign particles in the body are recognized by the "side chain" of an immune cell. Interaction and binding with the side chain causes increased production of that side chain and the target is labelled for later destruction by other cells. However, Landsteiner questioned the genetic storage of "pre-encoded" immune recognition against all possible epitopes (Landsteiner *et al.,* 1945), and weakened Ehrlich's theory with immunization experiments using synthetic immunogens not found in biological systems. Landsteiner showed that the immune response was not based solely on pre-determined recognition, since it was unlikely that the organism had already possessed immunity against these synthetic molecules prior to immunization.

A mechanism of how the immune system adapts to molecules was introduced in the "instructive template theory" of Breinl and Haurowitz (Breinl & Haurowitz, 1930) and later

elaborated by Pauling (1940). Central to this theory was the binding of a novel antigen to a flexible antibody that would adopt a complementary structure. These theories accounted for binding a diverse range of novel antigens; however, they failed to explain how in subsequent immune challenges the antibody response was quicker to respond and the binding affinity was often increased. These theories also did not account for the ability to distinguish between self and non-self antigens.

To account for the increase in response by the immune system upon subsequent challenge with a similar antigen, Jerne (1955) proposed that this was the result of proliferating immune cell lines. Jerne won a Nobel prize in 1984 in recognition for his work in immunology. Burnet (1959) further elaborated this mechanism as the "clonal selection theory", where each antibody-producing cell produces a unique antibody that is generated randomly, and displays the antibody on its surface. About 15 years after this proposal, the B-lymphocyte was discovered to be this antibody-producing cell and the process of affinity maturation in activated proliferating B-lymphocytes resulted in the increase of antigen specificity (French *et al.,* 1989). The process of rearrangement in the inherited (germline) genes responsible for the structural characteristics and specificities of an immunoglobulin was revealed by Tonegawa (1983). The effect of combinatorial rearrangement of germline genes helped to account for the ability of a finite number of genes being able to recognize an almost infinite amount of antigens. Tonegawa received a Nobel prize in 1987 for this work.

1.1.3 Development of immune specificity

1.1.3.1 The usage of germline genes

The immune system surveils physiology and defends against foreign invaders. This is accomplished through the inheritance of antigen recognition encoded in the germline genes. These recognition molecules include immunoglobulins and other receptors, on the surface of immune cells or as soluble molecules. The "innate" immune pathways represent the first line of defense and surveillance against pathophysiology, and are often followed by activation of the antigen-dependent "adaptive" immune pathways. The genes encoding the recognition molecules in the innate pathways are typically unchanged from their germline sequences and offer a quick response, whereas the adaptive pathways are often altered during the immune response and associated generally with higher specificity and "memory" of the antigen for subsequent exposures (Fig. 1.4).

Pattern recognition receptors (PRRs) are molecules that target common microbial structures known as pathogen-associated molecular patterns (PAMPs), which are typically unique products of bacterial metabolism and include lipopolysaccharide in Gram-negative, peptidoglycan in Gram-positive bacteria, and nucleotide structures (Medzhitov & Janeway, 2002). These functions are often critical to microbial survival and so are common and have remained unchanged in microorganisms. Some PAMPs feature carbohydrates unique to the microorganism, which allows them to be recognized as "non-self" antigen by the host immune defenses (Janeway, 1992). PRRs that bind carbohydrate include the mannose binding lectin, macrophage mannose receptor, toll-like receptors (TLRs) as reviewed in Holmskov *et al.* (2003) as well as immunoglobulins. These receptors are typically on the surfaces of immune cells that circulate throughout the body such as dendritic cells,

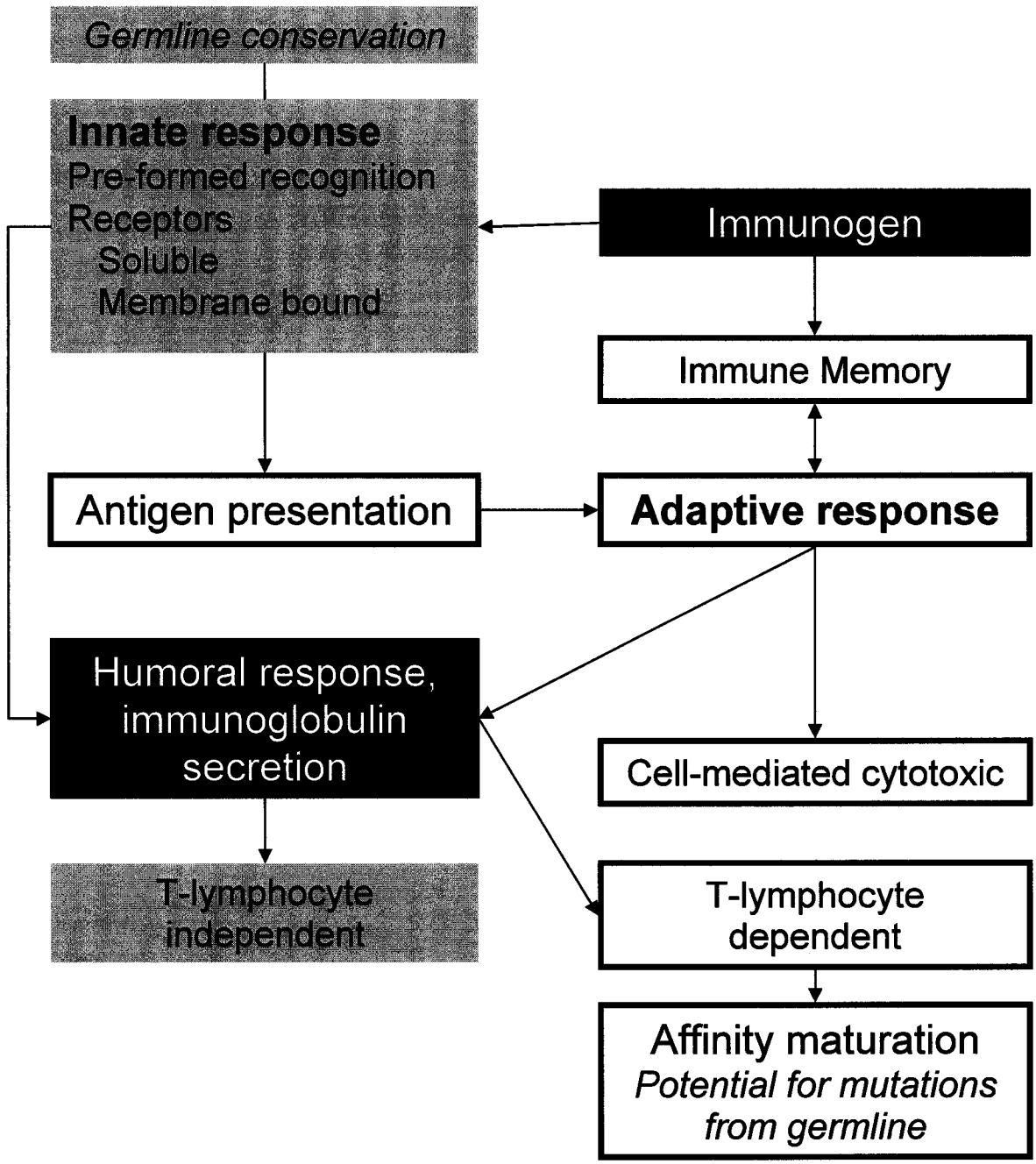


Figure 1.4 Simplified schematic of the immune reponses that lead to the production and secretion of immunoglobulin, employing generally conserved germline identity (*grey boxes*) and the T-lymphocyte dependent pathways (*outlined boxes*) that typically lead to mutations of the germline sequences (adapted from Janeway, 2001).

macrophages and B-lymphocytes, surveying for abnormal and pathophysiology. Soluble PRRs such as complement and immunoglobulin play a role in phagocytosis, proteolytic digestion, initiation of apoptotic pathways or kill by forming pores in the microorganism membranes.

The family of Toll-like receptors (TLRs) in mammalian systems is believed to be a phylogenetically primitive form of immunity, and share some genetic similarities with Toll receptors from *Drosophila* (Rock *et al.,* 1998). Ten TLRs have been discovered and bind different types of non-self PAMPs including LPS and peptidoglycans; however, their three-dimensional structures have not yet been reported. TLR-mediated binding of LPS involves TLR4 and other molecules such as LPS-binding protein (LBP) and CD14.

Members of a protein grouping called lectins each have similar carbohydrate recognition domains (CRDs), although they differ in sequence (Vijayan & Chandra, 1999; Holmskov *et al.,* 2003). Lectins were first discovered as agglutinating agents of erythrocytes and were later found to interact with the carbohydrates in other molecules involved in transport, signal transduction, and immune pathways. Lectins play a critical role in immunity (Ikeda *et al.,* 1987; Kawasaki, 1999; Vijayan & Chandra, 1999; Lawson & Reid, 2000; Holmskov *et al.,* 2003) where they are found as membrane-bound receptors on immune cells such as macrophages and dendritic cells, or as a soluble form in circulation. A well known immune pathway is the lectin-dependent activation of the complement system (Janeway 2001) initiated by mannose-binding protein lectin (MBL) (Holmskov *et al.,* 2003). The complement pathway can opsonise or coat pathogen surfaces so that they can be later phagocytized by immune cells (for elimination and antigen presentation to activate the adaptive response), by modulation of cytokine and oxidant responses, and by formation of the membrane attack complex.

The cellular components of the "adaptive" immune response include the T- and B-lymphocytes. During their development, the germline genes of these lymphocytes undergo genetic combinatorial rearrangements that results in a myriad of antigenic specificities. Effective recognition of antigen is achieved by few or many mutations to the germline genes that occur during successive rounds of proliferation. In general, during an adaptive response, T-lymphocytes differentiate and carry out separate functions (Janeway, 2001). The "helper" T-lymphocytes are involved in the proliferation of B-lymphocyte clones and the transformation to plasma cells that secrete soluble immunoglobulins. Immunoglobulin mediated responses are called humoral immunity and are particularly effective against extracellular pathophysiology. Helper T-lymphocytes also activate "cytotoxic" T-lymphocytes. This response, called cell-mediated immunity is particularly effective against intracellar pathophysiology. Helper T-lymphocytes also take part in the transformation of T- and B-lymphocytes into memory cells that offer a quicker response in subsequent immune challenges.

Recombinations of different germline genes and mutations can lead to enhancement of antigen specificity, however they may also develop recognition of self (Wucherpfennig, 2001). T-lymphocyte "regulator" cells participate in immune functions such as peripheral tolerance by deleting or inactivating immune cell lines that develop recognition of self-antigens.

1.1.3.2 Immune tolerance of self-antigens

In order to effectively distinguish between normal and abnormal physiology, the immune system must be able to distinguish between self and non-self antigens. The self antigens must be "tolerated" as to not produce an immune response. T-lymphocytes are important mediators of recognizing self and non-self, particularly in the case of T-lymphocyte dependent activation of B-lymphocytes. T- and B-lymphocytes originate from the bone marrow, however T-lymphocytes further migrate to the thymus to complete their maturation (Fig. 1.5). Deletion of self-reactive T-lymphocytes occurs in the thymus minimizing the occurrence of activation of self-reactive B-cell lymphocytes. Deletion of B-lymphocyte self-reactive clones occurs in the bone marrow in a process called "negative selection".

1.1.3.3 T-lymphocyte dependent mechanisms

Although the antibody structures discussed in this thesis derive from B-lymphocytes, it is useful to briefly review the role of T-lymphocytes in affinity maturation of antibodies. Antigen-presenting-cells (APCs) such as dendritic cells, macrophages and B-lymphocytes, and particles from the sites of infection are circulated between the peripheral blood and lymphatic fluid. Extracellular or intracellular molecules that have been phagocytized are enzymatically broken down and their polypeptide fragments are presented on the surface of APCs bound to molecules called major-histocompatibility complexes (MHCs). MHC type I and MHC type II present fragments of molecules of intracellular and extracellular origin respectively.

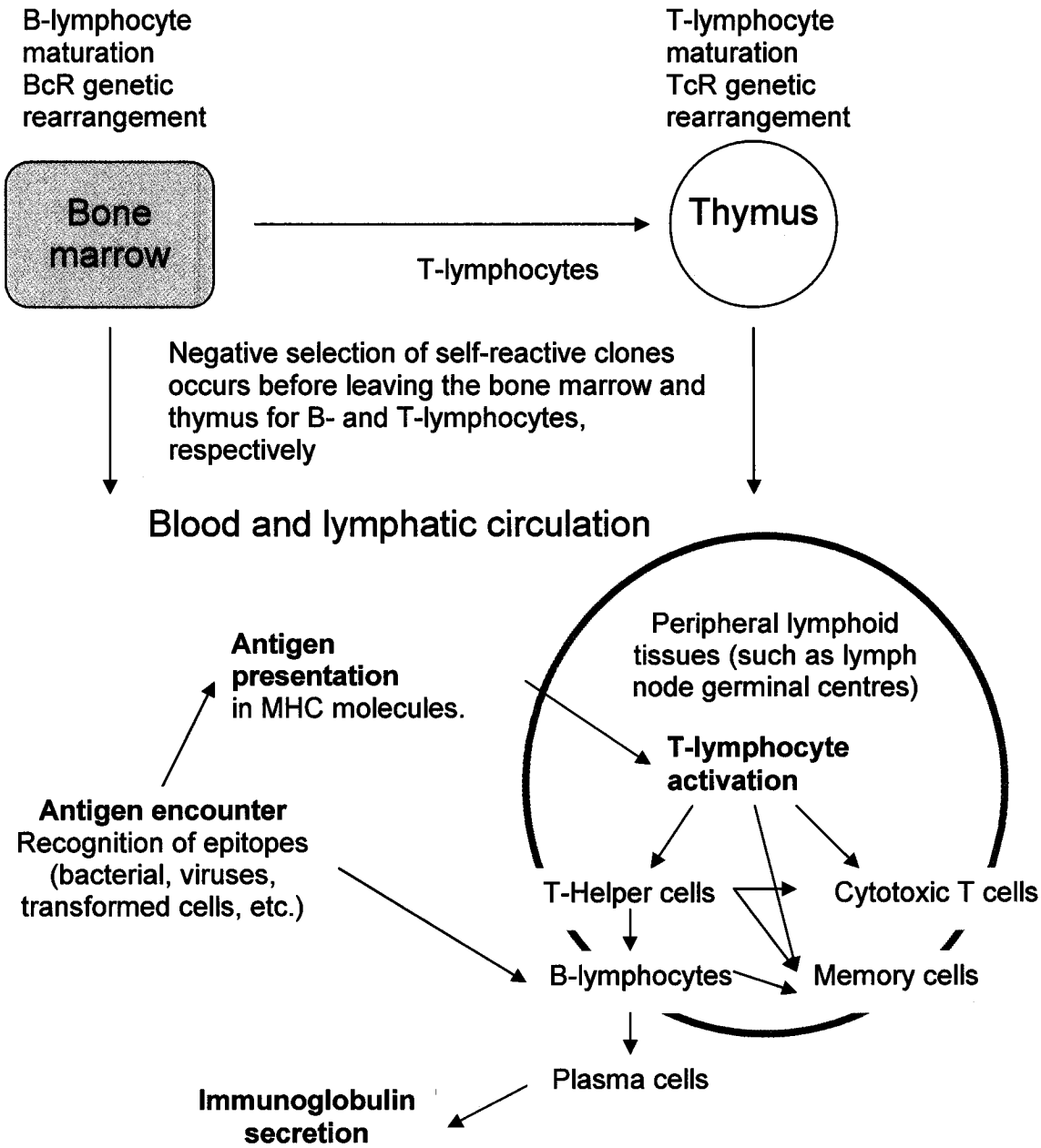


Figure 1.5 The origins and functions of B- and T-lymphocytes. After maturation in the bone narrow and thymus, lymphocytes can alternate between lymphatic and blood circulation. Antigen presenting cells (dendritic cells, macrophages and B-lymphocytes) process and present peptide antigens (obtained through phagocytosis or through receptor-mediated endocytosis) bound to MHC molecules. Recognition of these complexes by TcR activates the T-lymphocyte to perform specialised roles including the activation of B-lymphocytes to become immunoglobulin-secreting plasma cells. Some antigens can activate the secretion of immunoglobulins without T-lymphocyte help.

In the germinal centres (for example in lymph nodes), T-lymphocytes bind to MHC and develop into helper T-lymphocytes. Due to prior negative selection, generally these T-lymphocytes do not recognize the presentation of self-antigens by APCs. The helper T-lymphocytes can proceed to activate lines of B-lymphocytes that also display the same MHC complex. Proliferation of activated B-lymphocyte lines occurs. During this process, the immunoglobulin genes undergo several point mutations to localized "hypervariable" regions that in the polypeptide form part of the antigen-binding site. These immunoglobulins are then displayed on the B-lymphocyte surface as B-cell receptors (BcR). B-lymphocytes undergo successive rounds of proliferation, hypermutation and selection. This process selects for increasingly greater antigen specificity and is known as "affinity maturation". Mutations that result in poor antigen recognition or are self-reactive undergo apoptosis. Activated B-lymphocytes become plasma cells that secrete antibodies, or become memory B-cells. Immunoglobulin class-switching is an event that requires helper T-lymphocytes. This event is observed during recognition of a previous antigen, typically resulting in the production of IgG over IgM.

The immune response varies according to the nature of the carbohydrate antigen. MHC molecules present linear oligopeptides. While LPS, for example, is not typically presented by MHC molecules, fragments of carbohydrates that are attached to peptides can be presented and result in a T-lymphocyte dependent response against the carbohydrate moiety, enhancing the immune response in some cases (Alexander & Elder, 1984). A clinical example that takes advantage of this is the polysaccharide pneumococcal vaccine which employs a capsular carbohydrate antigen bound to a protein.

Additionally, some pre-formed PRRs of the innate immune response such as TLRs bind common PAMPs (*i.e.* LPS and peptidoglycan) and signal activation of T-lymphocytes cells to ultimately activate humoral and cell-mediated responses.

1.1.3.4 The T-lymphocyte independent responses

There are immune pathways that are not linked to the adaptive response. In a T-lymphocyte independent response, affinity maturation does not occur and activation of the B-lymphocyte is accomplished through binding of BcRs that have retained mostly germline identity. These BcR and secreted immunoglobulins called "natural antibodies" have evolved from selective environmental pressure and recognize PAMPs such as LPS. Natural antibodies offer the host protection without having first encountered the antigen. The resulting immune response without T-lymphocyte help results in relatively lower immunoglobulin production.

1.1.3.5 Immunoglobulin structure and the rearrangement of germline genes

During maturation, successful pairing of heavy and light chains results in survival of the B-lymphocyte and continued expression of the rearranged genes as immunoglobulin surface receptors (BcR). Otherwise, in the event of unsuccessful heavy and light chain pairing, the B-lymphocyte undergoes apoptosis.

Tonegawa discovered that lymphocyte receptors such as the BcR, and the T-cell receptor (TcR) were formed from the rearrangement of different genes. It is relevant to this work to elaborate on the B-lymphocyte development in the human and murine genomes. There are three separate locations on chromosomes for these gene clusters that form an immunoglobulin, one for the genes that form a heavy molecular weight chain: variable (V), diversity (D), joining (J), and constant (C); and separate locations for the genes encoding light molecular weight chains *kappa* and *lambda*, (which do not contain D genes) where only one type (either κ or λ) will associate with a heavy chain to form an immunoglobulin (Fig. 1.6).

Table 1.1 Reported number of germline genes (including pseudogenes) in the human and mouse genomes.

Gene	Human	Mouse
V_H	129	216
V_κ	76	158
V_λ	74	9
D_H (heavy chain only)	27	14
J_H	9	7
$J_{\kappa+\lambda}$	16	11
C_H	11	9
$C_{\kappa+\lambda}$	12	6

(Lefranc *et al.*, 1999; Ruiz *et al.*, 2000; Lefranc, 2001).

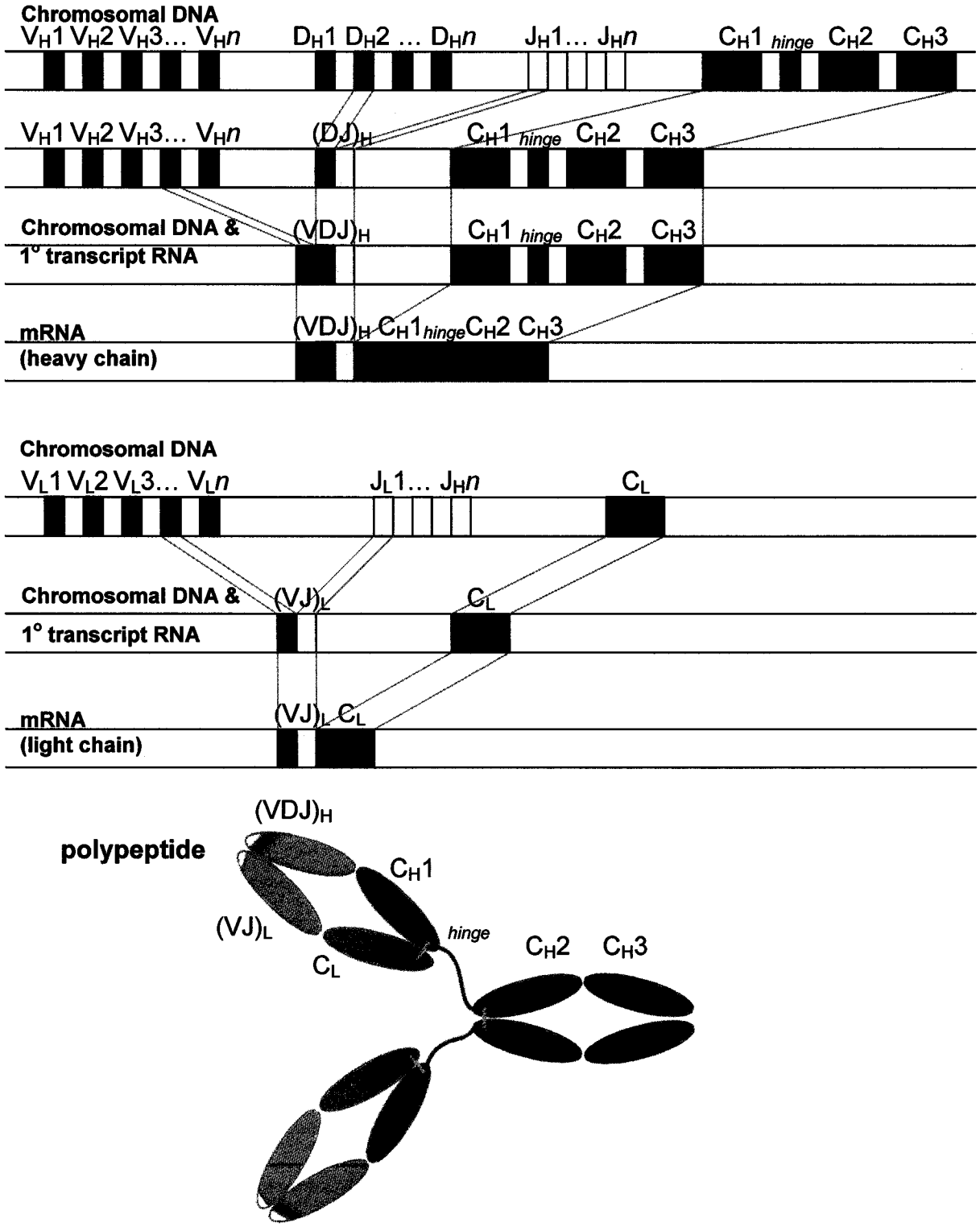


Figure 1.6 Representation of the genetic rearrangements that encode the heavy (*top*) and light (*middle*) chains that are translated and then associated with one another to form an immunoglobulin (*bottom*) (adapted from Janeway, 2001).

The family of glycoproteins known as immunoglobulins share a common set of germline genes that encode for surface bound form BcR as well as the secreted form. Briefly, an immunoglobulin has a "Y" shaped structure that is a dimer composed of two light chains and two heavy chains (Fig. 1.7A). The antigen-binding fragment (Fab) portion of the immunoglobulin is composed of a variable domain where most of the sequence variability occurs and a single constant domain from each light and heavy chain. The immunoglobulin fold characteristic of this glycoprotein family describes a highly conserved secondary structure pattern. This Greek-key type connectivity (Fig. 1.7B) that forms a two-layer sandwich is composed of 9 anti-parallel β-strands in the variable domain, and of 7 anti-parallel β-strands in the constant domain.

A hinge region connects the Fab fragment to the Fc fragment (named for being "crystallizable") that is composed of the heavy chain constant (C) domains. The constant domain genes (μ, γ, α, ε and δ respectively) determine the five classes of immunoglobulin: IgM, IgG, IgA, IgE and IgD respectively. IgM is found as a monomer on the surface of naïve B-lymphocytes. Early immune response secreted immunglobulins contain a high proportion of monomeric and pentameric IgM. In the presence of chemical signals from helper T-cells, the immunoglobulin genes coding for the constant domains are substituted for other constant domains. IgG is a monomer and the main immunoglobulin present in secondary immune challenges. Many subclasses of IgG are possible IgG1, IgG2, IgG3 and IgG4 for human (and IgG1, IgG2a, IgG2b, and IgG3 for mice). IgG is particularly important in the clearance of bacteria and viruses. It is also able to cross the placenta and confer immunity to a developing fetus. IgA can be found as a monomer and secreted as a dimer from mucosal surfaces serving critical functions in the lung and gut. IgE is

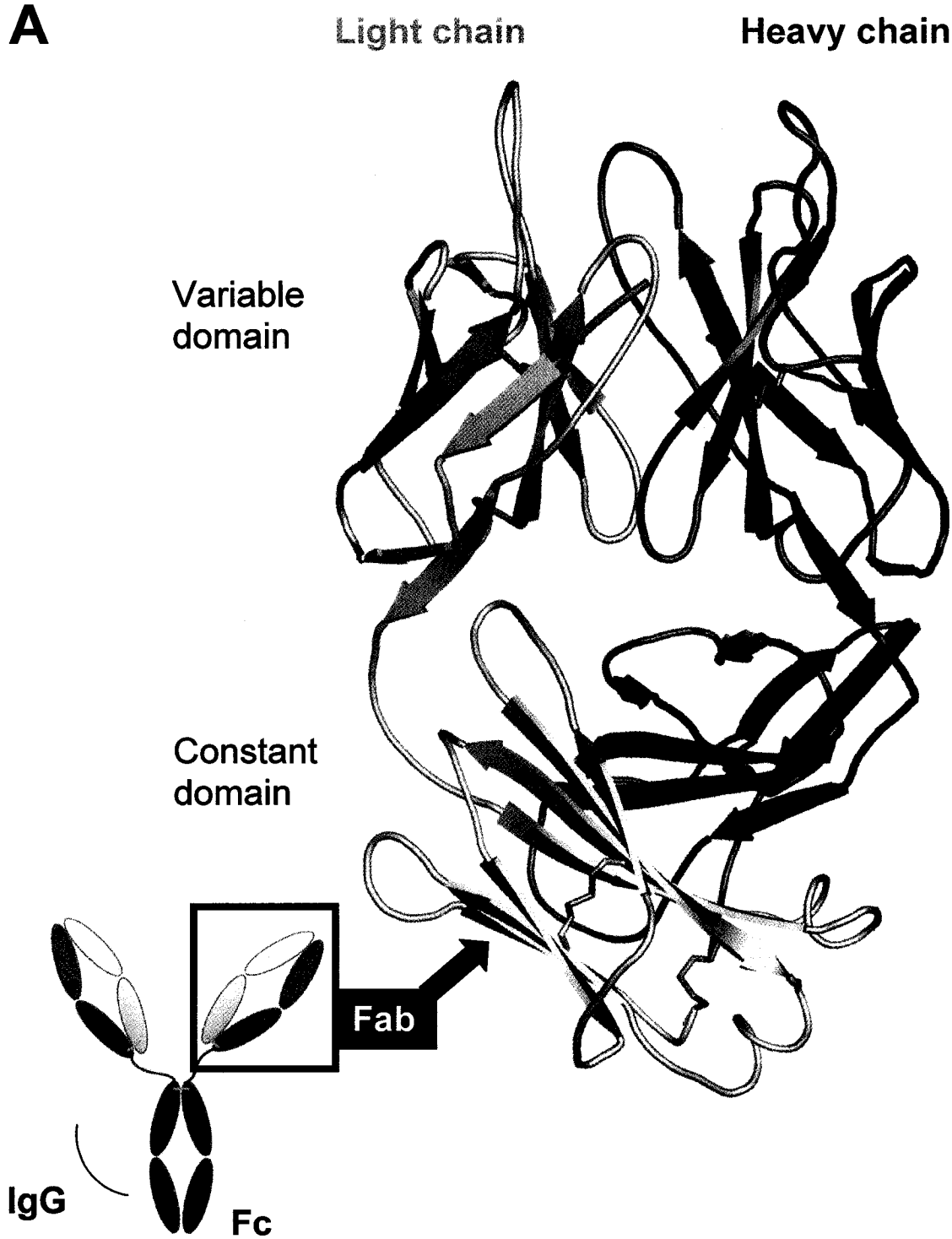


Figure 1.7 A) Two light (*gold*) and heavy (*blue*) chains form the "Y" shape of the IgG (*inset*) with two Fab regions each joined by a flexible "hinge region" loop to the Fc constant domains. *Enlarged* is the the structure of one Fab, composed of a variable and a constant domain from the light and heavy chains. Disulphide bridges are indicated in *orange*. A two-layer "sandwich" immunoglobulin-fold is composed of antiparallel β-strands.

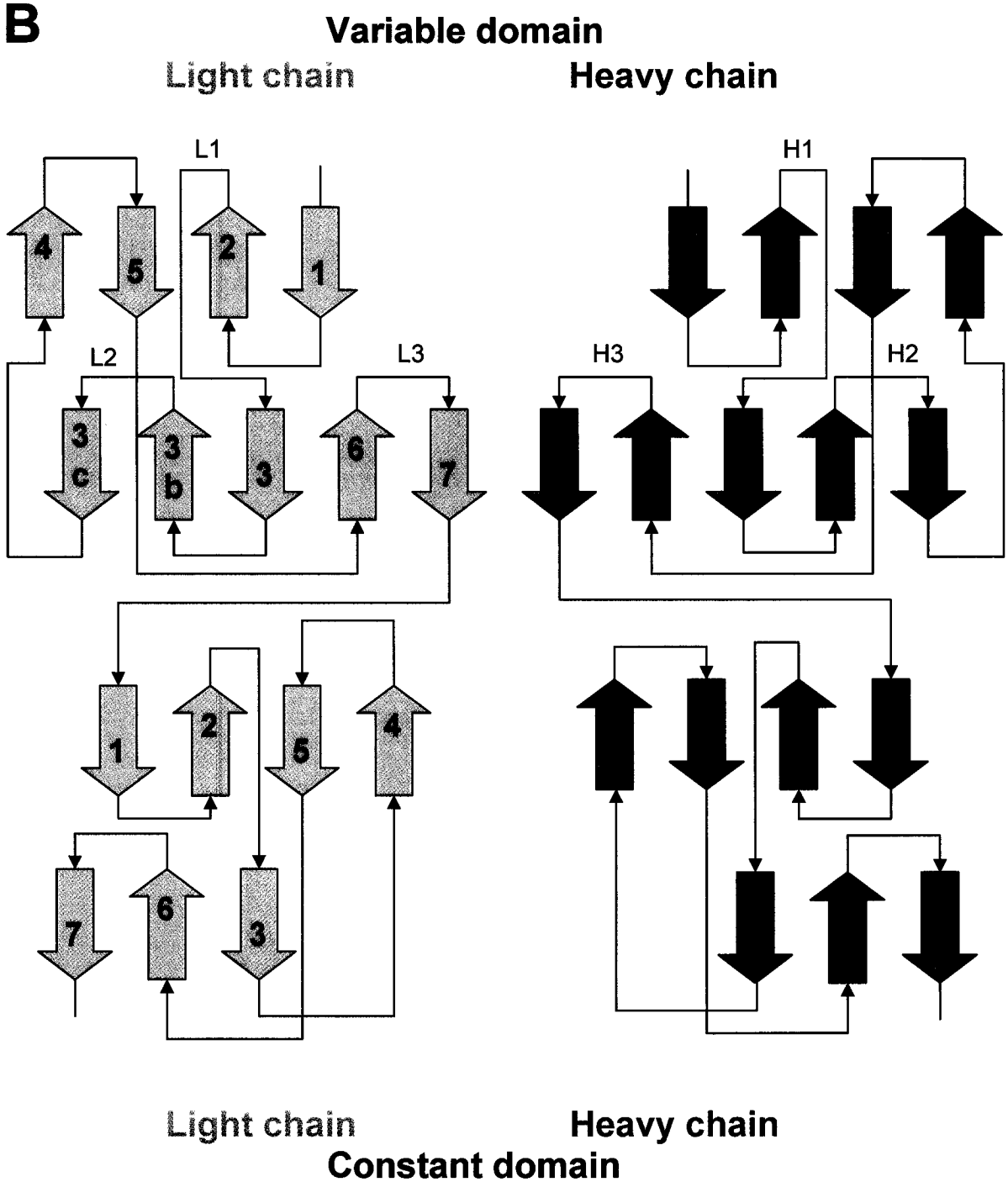


Figure 1.7 B) The connectivity of the β-strands for each domain and chain with the CDR loops highlighted in *red.*

implicated in hypersensitivity (*i.e.* allergy) responses, and also important during parasitic attacks. IgD is also a surface receptor in naïve B cells and less is known about its effector immune functions. The Fc domain is also recognized by receptors on the surfaces of immune effector cells. For example, macrophages have Fc receptors that initiate phagocytosis of the immunoglobulin-antigen complex.

The variable domains of the heavy and light immunoglobulin chains each exhibit three regions of localized sequence hypervariability. These are called complementarity determining regions (CDRs), and are numbered accordingly on the light chain: CDR L1, L2, L3; and on the heavy chain: CDR H1, H2, H3. The remaining amino acid sequences in between the CDRs form the framework regions (FR). Up to three CDRs from each chain can form the "combining" site, which interacts with the antigen.

The first two CDRs of each light and heavy chain, located at around 25-33 and 50-55 residues from the N-terminus, respectively, are encoded entirely by the corresponding light or heavy V genes. However, the third CDRs of each chain, L3 and H3, typically have more sequence variability as they include the location where the V and J genes join in the light chain, and where the V, D, and J genes join in the heavy chain. Based on sequence and structure comparisons, it has been reported that the conformations of the CDRs excluding CDR H3 fall into specific "canonical" groups based on the possible hydrogen bonding patterns in the backbone and also in the side chain. In previous reports, CDR H3 often provides many critical contacts with antigen, but due to its variability in sequence and length, there are limited rules of conformational prediction (Shirai *et al.,* 1996).

1.1.3.6 Non-template sequence diversity

In addition to the rearrangement of the different germline genes, and the random pairing of light and heavy chains – the antibody combining site gains further variability from the rearrangement process itself. The imprecise joinings of V and J genes in the light chain and V, D and J genes in the heavy chain result in changes in polypeptide sequence identity and length at the regions of joining. These changes within the CDRs can alter their structures and canonical classification (Tomlinson *et al.,* 1995).

In between each copy of an immunoglobulin gene are two conserved sequences: a heptamer and a nonamer separated by a spacer sequence that has a length of either 12 or 23 base pairs (Fig. 1.8). Recombination-activating gene (RAG)-1 and RAG-2 proteins bind to these heptamer-spacer-nonamer sequences and pair up with one another bringing two distant genes close together. As a RAG protein that recognizes a 12 bp spacer can only pair with one that recognizes a 23 bp spacer, this allows the spacer flanking the immunoglobulin gene to determine what pairing can occur. In the light chain, a V_κ gene with a flanking 12 bp spacer only pairs with the J_κ gene that has a 23 bp spacer. In the heavy chain, the V_H gene has a 23 bp spacer, and can pair with the D_H gene with a 12 bp spacer that has been paired with a J_H gene flanked by a 23 bp spacer. There are exceptions to this pairing rule, in some cases, more than one D_H gene (with flanking 12 bp spacers) have been found to have joined together (Janeway 2001).

In a typical rearrangement, the immunoglobulin genes found in between the paired RAG proteins are excised as DNA loop fragments. In this process, the DNA is broken (by a hairpin formation and by the random cleavage by the RAG protein on one strand) at different positions flanking the immunoglobulin gene, which accounts for P- (palindromic)

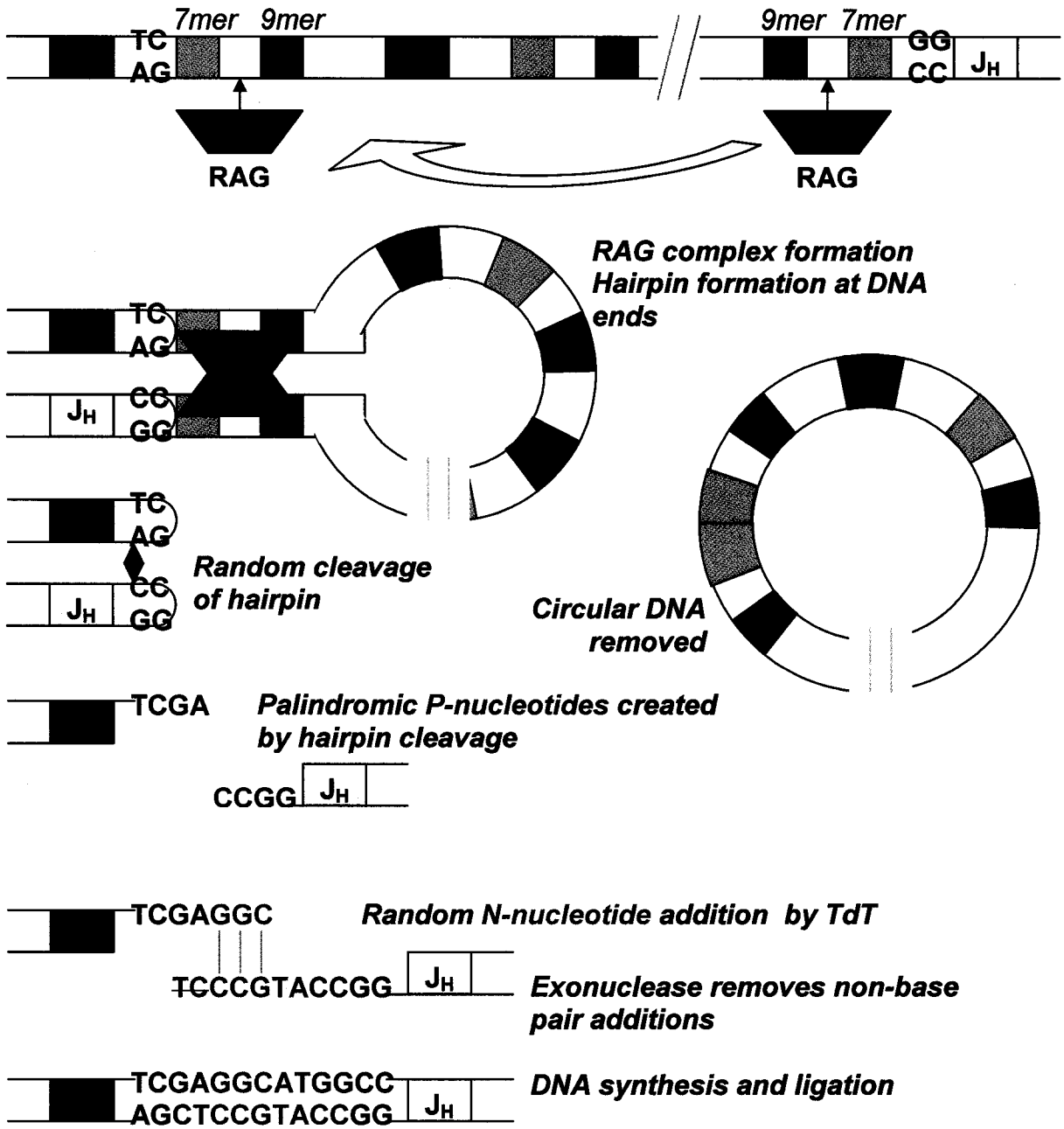


Figure 1.8 Excision of immunoglobulin genes, and the origins of non-template sequences near the genetic flanking regions. After binding the heptamer-spacer-nonamer sequences (*cyan-white-pink boxes*), the RAG protein (*green*) complex brings the two immunoglobulin genes together. Random, imprecise cleaving (*red diamond*) by RAG proteins that removes the sequences in between also accounts for uneven ends that form palindromic P-nucleotides (in *red*). N-nucleotides (in *blue*) are added by deoxynucleotidyl transferase (TdT) until base-pairing can occur. Exonuclease activity during DNA synthesis and ligation also removes base-pairs (adapted from Janeway, 2001).

nucleotide additions. Deoxynucleotidyl transferase (TdT) then randomly adds several N-(non-template coded) nucleotides after the P-nucleotide addition, usually until both DNA ends can pair with one another. An exonuclease is involved in removing non-base pairing additions.

As such, these mechanisms contribute to the longer length and higher sequence variability observed in the third CDR of the light and heavy chains (Hofle *et al.,* 2000; Janeway, 2001) where the genes join. Because of the additions, mutations and deletions that may occur during the genetic rearrangement process, it may be difficult to recognize the original individual germline genes, particularly at the flanking regions of the genes, and in the smaller genes such as the D and J genes. Some rearrangements result in functional frameshift mutations that may result in increased diversity, or in a stop codon forming a "pseudogene". As well, functional light and heavy chains may not readily combine to form an intact immunoglobulin. Genetic rearrangement can continue until a functional light and heavy chain combination occurs which allows the cell to express these genes as functional immunoglobulin surface receptors (BcR), otherwise in the case of a non-functional heavy and light chain pairing, the B-lymphocyte undergoes apoptosis (Janeway, 2001).

The template and non-template genetic information determine the effectiveness of the immune response to pathophysiology. The survival of the population ultimately determines the inheritance and persistence of these germline genes.

1.1.4 Polyspecificity and cross-reactivity

The concept of antibodies binding more than one antigen has been investigated for the past 50 years and manifests as a clinical condition in the case of autoimmune disease, and also presents problems in laboratory and diagnostic procedures. Well documented autoimmune conditions include the first reported case of gastric and thyroid autoimmune response in type 1 diabetes patients (Moore & Neilson, 1963), systemic lupus erythematous (Radic & Weigert, 1994), infection-related diseases such as multiple sclerosis (Wucherpfennig, 2001), Streptococcus induced rheumatic fever and associated heart and joint disease (Cunningham, 2003) and nephritis (Cunningham, 2000), autoimmune responses from lipopolysaccharide of different microorganisms (Moran & Prendergast, 2001; Boffey *et al.,* 2004), and including several conditions of autoimmune disease associated with chlamydial infection including atherosclerosis (Lamb *et al.,* 2003), heart disease (Bachmaier *et al.,* 1999), and reactive arthritis (Benjamin *et al.,* 1991; Bachmaier *et al.,* 1999).

In many reports the term "cross-reactive" has been used to illustrate the binding of ligands other than the antibody's cognate antigen. There is a lack of structural evidence in understanding many diseases associated with cross-reactivity. For example, molecular mimicry of the cognate antigen has been proposed, where different ligands are recognized in the binding pocket using similar interactions (Wucherpfennig, 2001). Bacteria have developed mechanisms that exploit such mimicry and so complicate their clearance by host. To evade recognition, bacteria employ molecular mimicry of host antigens and can horizontally transfer genes encoding host antigens (Medzhitov & Janeway, 2002). There are other factors that can lead to autoimmune disease, including genetic predisposition. For

example certain classes of MHC molecules with bound antigens (of bacterial or viral origin) could activate self-reactive populations of T-lymphocytes. Upon examination of binding to self-reactive T-lymphocyte clones using peptide screening experiments, it was revealed that the TcR also bound a range of antigens not related by sequence to the self antigen. This laxity in binding specificity has been referred to as "degeneracy" of the TcR. Wucherpfennig adds however that the possibility of molecular mimicry between the low sequence homology antigens is often overlooked in screening experiments.

To make a distinction from cross-reactivity, the binding of different antigens using different interactions in the same binding site has been previously defined as "polyspecificity" (Kramer *et al.,* 1997). Germline-encoded PRRs of the innate immune response can bind more than one antigen typically based on binding to an epitope that is common (PAMPs) that is part of the larger molecule. Many of these receptors are classified as lectins (Holmskov *et al.,* 2003) and have shown specificity based on the orientation of the functional groups attached to the sugar ring. There are few structural reports of antibody-carbohydrate complexes to illustrate the basis of such exquisite selectivity, thus limiting our understanding of how immunoglobulins can bind more than one antigen while restricting cross-reactivity. Previous binding studies have shown that immunoglobulins may bind a limited number of structures within a class of antigen (such as different molecules in the bacterial membrane) and that this polyspecificity is encoded by distinct subsets of V, D, or J germline genes (Perlmutter *et al.,* 1984; Tomlinson *et al.,* 1996; Seidl *et al.,* 1999). Recent structural reports have proposed possible mechanisms of polyspecific immunoglobulins that include multiple usage of the same residues within a fixed binding site and limited conformational flexibility in the CDR loops (Keitel *et al.,* 1997; Kramer *et al.,* 1997; James *et al.,* 2003), but there is yet no clinically revelant

examples. It is of interest to examine how the use of a set of germline genes can contribute to the specificity for a limited number of related antigens.

1.1.5 Conformation and entropy in carbohydrate-binding

Without the enhancement of affinity afforded by somatic hypermutation, immune carbohydrate recognition uses other strategies that enhance the likelihood of binding. In the Instructive Template theories, Pauling had proposed that antibody structure could adapt to complement the antigen. He proposed that structurally flexible portions of antibodies would come into contact with an antigen and adapt to its structure.

Antibody recognition of carbohydrates is often associated with low affinity binding, typically orders of magnitude below that found in protein-protein interactions. In the unliganded state, the polypeptide has more freedom of movement, and there is less order in both the protein and the surrounding solvent. However, in the liganded state decreased entropy results from the immobilisation of the bound portion of protein, resulting in an "entropic penalty". For favourable binding, this entropic loss is compensated by binding enthalpy (and possibly a net disorder in solvent molecules). In the case of antibodies binding carbohydrates, the bonds formed between protein and sugar may be few in number and of weak nature. One of the ways that carbohydrate-binding antibodies minimize entropic loss may be by sampling fewer conformations in the unliganded state. These considerations favour a binding pocket that minimizes conformational change but is yet able to recognize structurally diverse antigens.

1.2 Bacterial carbohydrates

1.2.1 Ubiquitous Lipopolysaccharide carbohydrate component

Lipopolysaccharide (LPS) is found in the outer layer of the asymmetric outer membrane of Gram negative bacteria (Fig. 1.9). The Lipid A LPS component is composed of a disaccharide of glucosamine phosphate acylated with fatty acid chains anchored in the lipid environment. The inner core carbohydrate moiety is composed of 3-deoxy-*D-manno*-oct-2-ulosonic acid (Kdo) residues linked to Lipid A at the glucosamine phosphate (Fig. 1.10). The α(2→4) Kdo disaccharide is a core structure found in the LPS of all Gram-negative bacteria. To this basic structure heptose and hexose additions form the outer core, to which the polysaccharide O-antigen is successively attached. In "rough" (Re) strains of Gram negative bacteria, the LPS is truncated at the inner core, exposing this Kdo α(2→4) disaccharide epitope.

Present in all species of the family *Chlamydiaceae* is a unique Kdo trisaccharide epitope αKdo-(2→8)-αKdo-(2→4)-αKdo that has not been observed in any other bacterium, and is thus an important taxonomic and diagnostic marker. One species, *Chlamydophila psittaci* also displays a trisaccharide epitope αKdo-(2→4)-αKdo-(2→4)-αKdo in addition to the family-specific epitope (Brade *et al.*, 2000).

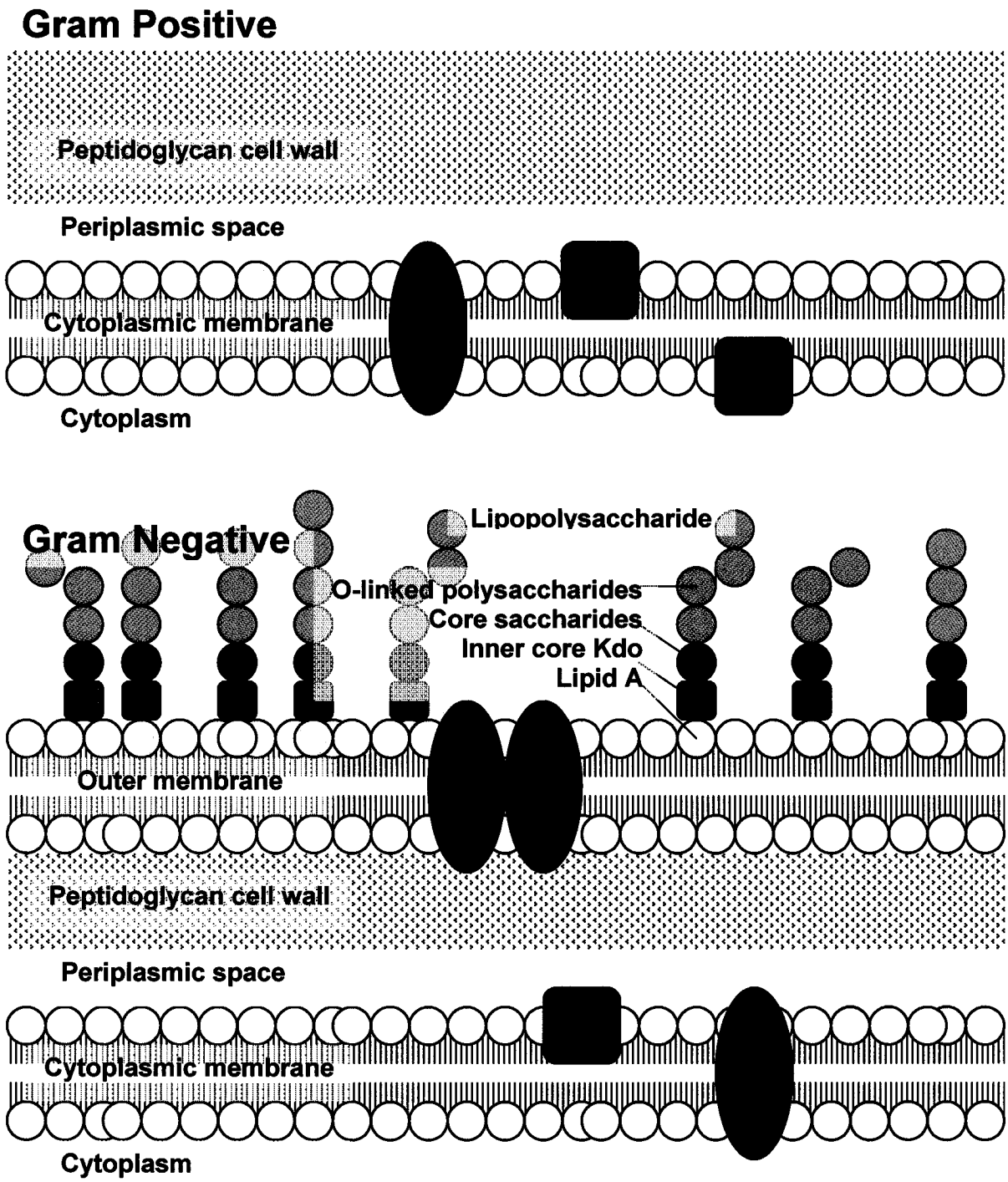


Figure 1.9 Gram-positive (*upper*) and Gram-negative (*lower*) bacterial surfaces differ in that Gram-positive has a thicker peptidoglycan cell wall but lacks the lipopolysaccharide (LPS) and outer membrane. Essential components of LPS are the Lipid A molecule (*yellow*) and the inner core Kdo saccharides (*orange*) to which the core (*brown*) and O-antigen (*pink*) saccharides are linked. Membrane proteins are represented in *green*.

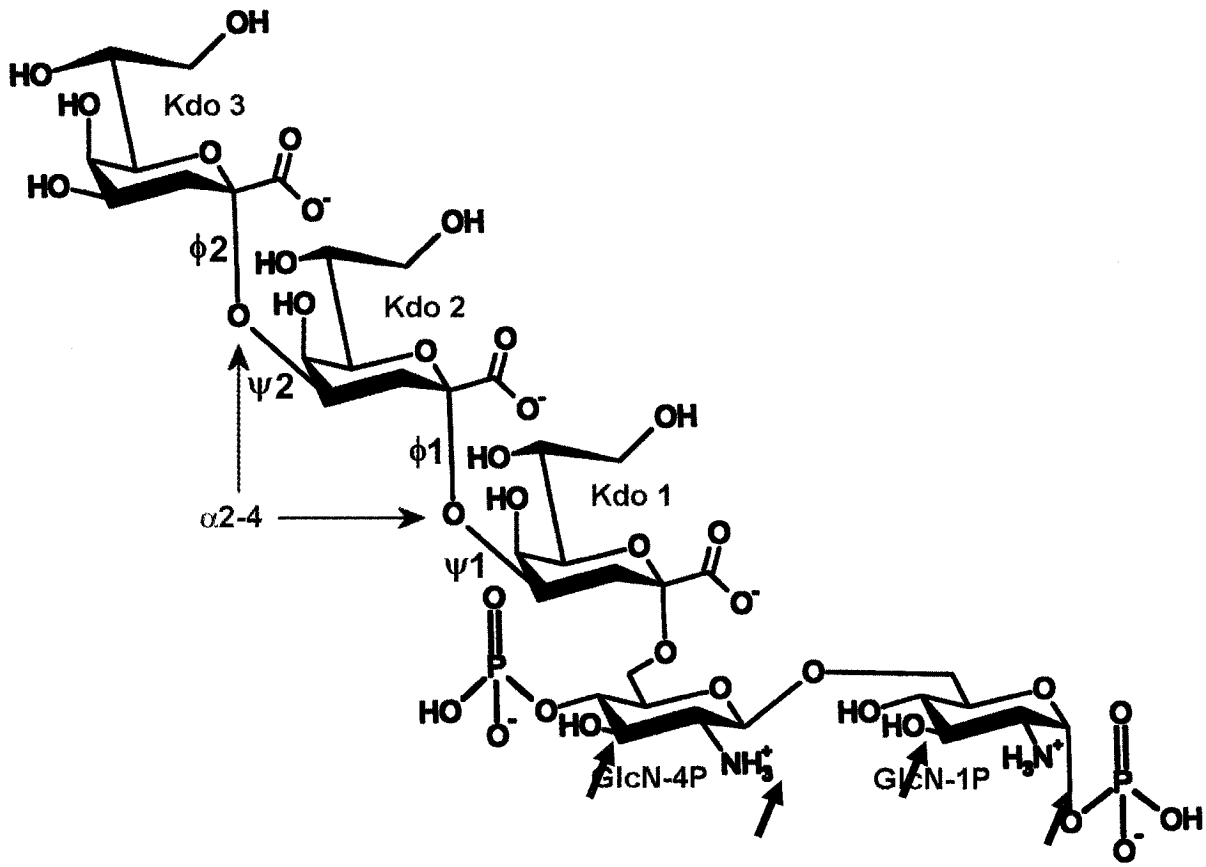


Figure 1.10 A portion of the purified LPS, αKdo-(2→4)- αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate from *C. psittaci*. This structure represents the LPS deacylated at the O- and N- groups indicated by *blue arrows*. The Kdo and Lipid A glucosamine phosphate residues are indicated and the glycosidic bonds and corresponding dihedral $(\phi,\psi)n$ angles indicated in *red*.

1.2.2 Clinical aspects

The LPS has been associated with toxic shock and disruption of many immune responses (Schaechter *et al.,* 1993). The O polysaccharide is generally associated with many toxic systemic effects (Schaechter *et al.,* 1993; Varki, 1999). The $\alpha(2\rightarrow4)$ disaccharide epitope of the Re-type LPS is the smallest LPS structure required to maintain bacterial membrane barrier functions (Nurminen *et al.,* 1983; Strain *et al.,* 1983), aiding in the survival and potential pathogenicity of the bacteria. Similar to the Re-type LPS, the LPS found in members of the family *Chlamydiaceae* elicit a mild systemic response and can disrupt functions of the immune system (Ryan & Sherris, 1994).

Chlamydiaceae are obligatory intracellular pathogens. Their life cycle alternates between an infectious extracellular form called the elementary body (EB) and a metabolically active intracellular reproductive form called the reticulate body (RB) (Fig. 1.11). The role played by the EB cell surface in the mimicry of nutrients, growth factors and hormones in order to enter the host cell via receptor-mediated endocytosis is not completely understood (Schaechter *et al.,* 1993). In the RB form, chlamydiaceae use host cell ATP and other nutrients to drive their growth and replication. Unlike other antigenic membrane components in the intra- and extracellular morphologies such as major outer membrane proteins that undergo structural changes to adapt to the environment, however chlamydial LPS distribution and structure are largely conserved.

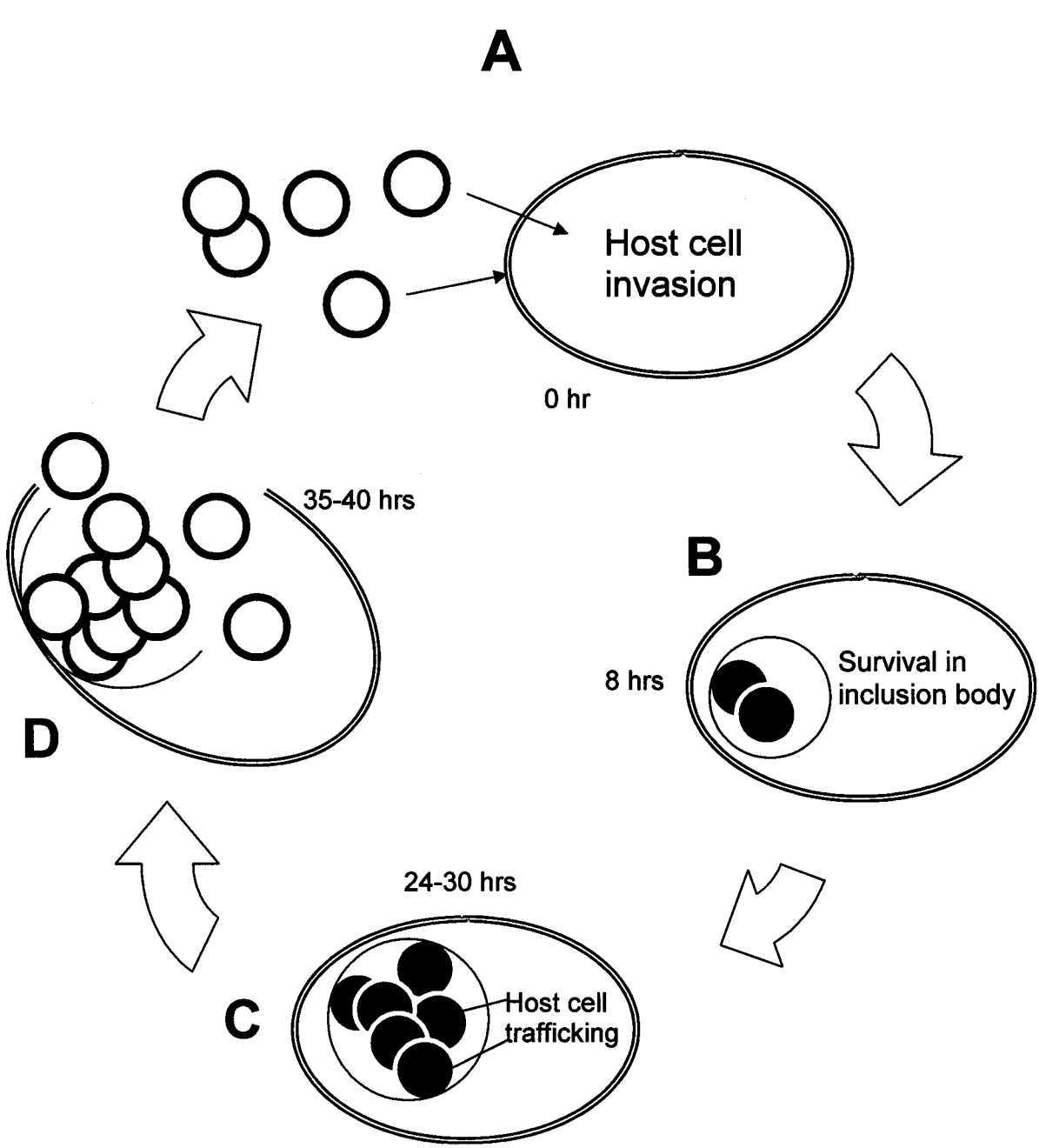


Figure 1.11 **A)** The chlamydial life cycle begins with the infection of a host cell by the elementary bodies (EB) (*black outline*). **B)** Protection from the host intracellular environment is afforded by survival in the inclusion body. Chlamydiaceae transform to the metabolically active reticulate body (RB) (*filled red*). **C)** The RB are able to traffic resources from the host for growth and replication **D)** The RB transform to the EB form in preparation for eventual host cell lysis and release (adapted from Ryan,1994).

Chlamydial infections disrupt many host physiological systems and are often insidious without obvious early symptoms. Lack of detection and surveillance contributes to the spread of disease and poor management particularly in less industrialized nations. *Chlamydia trachomatis* is a leading cause of secondary blindness (Schachter & Dawson, 1990), sexually transmitted infection of the urogenital tract including lymphogranuloma venereum (Schaechter *et al.,* 1993) and acquired infertility (Morell, 1995), and is also known to cause pneumonitis and urogenital tract infections in mice (de la Maza *et al.,* 1994). *Chlamydophila pneumoniae* causes infections of the respiratory tract (Kuo *et al.,* 1995), and has been linked to asthma bronchiale (Hahn *et al.,* 1998), atherosclerosis and heart disease (Bachmaier *et al.,* 1999; Grayston 2000; Kalayoglu *et al.,* 2000), multiple sclerosis and Alzheimer's disease (Yucesan & Sriram, 2001). *Chlamydophila psittaci* infection is observed in animals and is underestimated in humans, but manifests as a flu-like respiratory syndrome called psittacosis and also has been implicated in heart valve disease and abortion (Schaechter *et al.,* 1993). Furthermore, infection by many species of the family is also associated with reactive- and undifferentiated oligoarthritis (Schaechter *et al.,* 1993).

As a general strategy, vaccination in underdeveloped countries where infection is endemic is often more feasible than the distribution of antibiotics and monitoring of its effectiveness and compliance, and where the emergence of antibiotic-resistant bacteria is a concern. With antibiotics, reinfection by Chlamydia is common. The failure of vaccine attempts against Chlamydia is due to the targeting of major outer membrane proteins that tend to exhibit high rates of mutation. On the other hand, LPS carbohydrate moieties are ubiquitous and conserved at the family, genus and species levels and thus are potential targets for development of vaccines. The family specific epitope is believed to play an

important role in the neutralization of a strain of *C. pneumoniae* by antibody CP-33 (Peterson *et al.,* 1998).

1.2.3 Murine antibodies against synthetic antigens that mimic natural epitopes

S25-2 is a mAb that displays affinity for the family-specific terminal α(2→8)-linked Kdo residues, and binds the trisaccharide αKdo(2→8)-αKdo-(2→4)-αKdo (Fig. 1.12). Interestingly, S25-2 is also observed to bind the αKdo-(2→8)-αKdo and αKdo-(2→4)-αKdo disaccharides with significantly lower affinity (Muller-Loennies *et al.,* 2000). MAb S45-18 has a polypeptide sequence highly homologous to S25-2, but is specific for the trisaccharide αKdo-(2→4)-αKdo-(2→4)-αKdo and does not bind to the αKdo(2→8)-αKdo-(2→4)-αKdo trisaccharide epitope (Brade *et al.,* 2000).

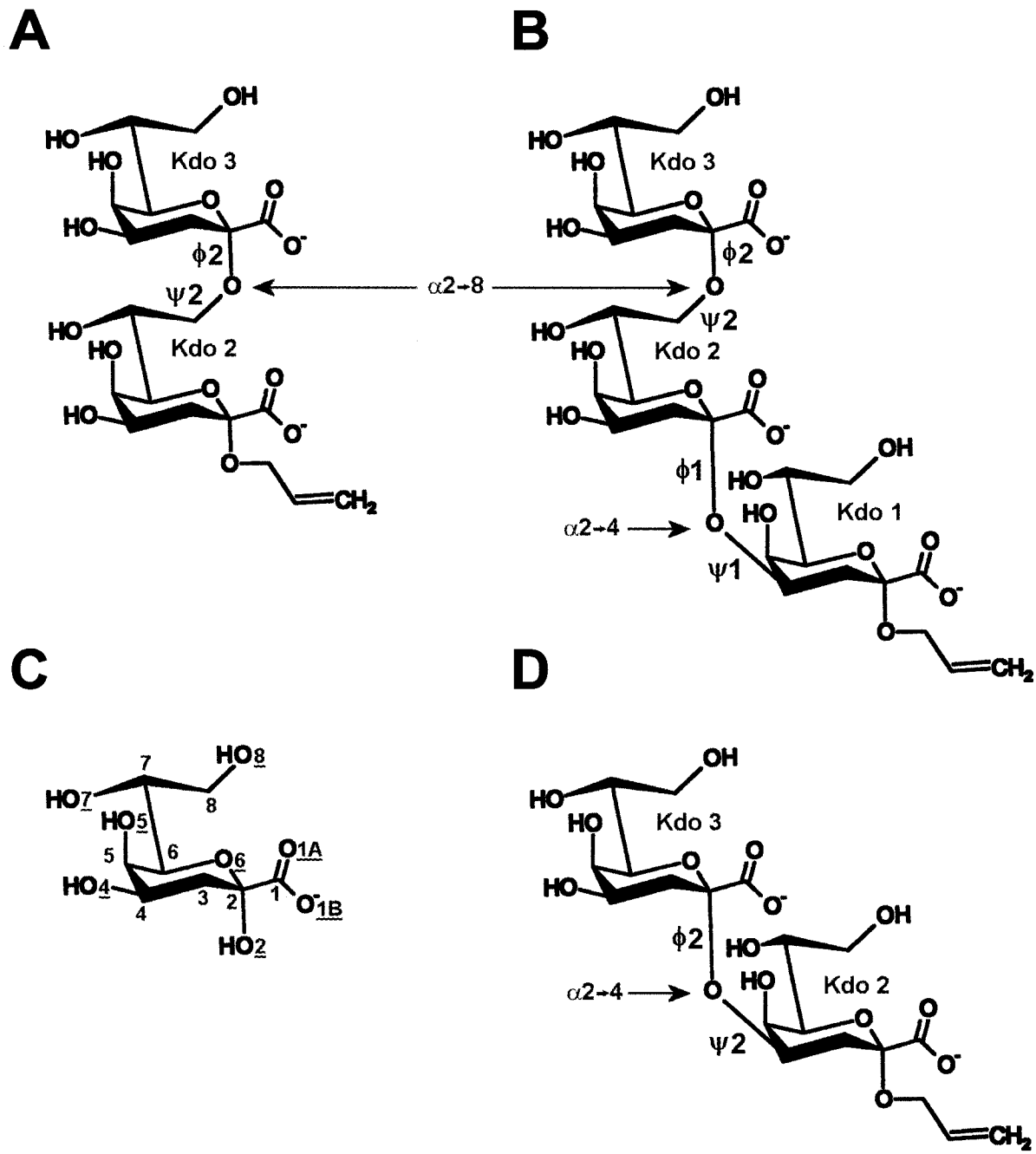


Figure 1.12 Kdo monosaccharide and (2→*O*)-allyl derivatives used in the co-crystallization of S25-2: **A)** αKdo-(2→8)- αKdo-(2→*O*)-allyl disaccharide **B)** αKdo-(2→8)- αKdo-(2→4)- αKdo-(2→*O*)-allyl trisaccharide, **C)** αKdo monosaccharide with numbering of carbon and oxygen (*underlined*) atoms and **D** αKdo-(2→4)- αKdo-(2→*O*)-allyl disaccharide. The Kdo residues are numbered in *green* and the glycosidic bonds and corresponding dihedral (ϕ,ψ)*n* angles labeled in *red.*

1.3 Determination of structure by X-ray Crystallography

1.3.1 X-ray crystallography

X-ray crystallography has allowed the determination of three-dimensional molecular structures of proteins beginning with myoglobin (Kendrew *et al.,* 1958) (Nobel prize 1962), DNA (Watson & Crick, 1953) from X-ray data from Maurice Wilkins and Rosalind Franklin (Nobel prize for Watson, Crick and Wilkins in 1962), of the first virus (Harrison, 1969), the first membrane protein from the photosynthetic bacterium *Rhodopseudomonas viridis* (Michel 1982) (Nobel prize 1988), and the complete ribosome structure (Cate *et al.,* 1999).

In this study of unliganded antibodies and antibodies in complex with carbohydrates, the underlying nature of their molecular recognition can be determined at high resolution using X-ray crystallography.

From the discovery of X-rays by Roentgen (Nobel prize 1901) to that of their diffraction of crystals by Laue (Nobel prize 1914), the concepts of this technique have been well documented (Sands, 1969; Glusker & Trueblood, 1985; McRee, 1993; Rhodes, 1993); however, a discussion of the relevant theory and methods of X-ray diffraction and structure solution used in this study will aid the reader in evaluation of the significance of the results.

1.3.2 Ordered Protein Crystal Growth

A crystal of the protein of interest is required for X-ray diffraction experiments. Many factors can influence the growth of a crystal, such as sample purity, solubility, pH, temperature, additives, and the presence of ligands (McPherson 1999). A common method used to grow a protein crystal is the hanging drop vapour diffusion method (Fig. 1.13), in which the protein sample mixed with precipitant and additive is suspended above a reservoir containing a higher precipitant concentration. Water then diffuses out of the drop towards equilibrium increasing the concentration of the protein in the drop, causing it to supersaturate. By fine-tuning the crystallization conditions, the rate of deposition can be controlled allowing molecules to pack in more ordered positions and orientations resulting in a well diffracting crystal.

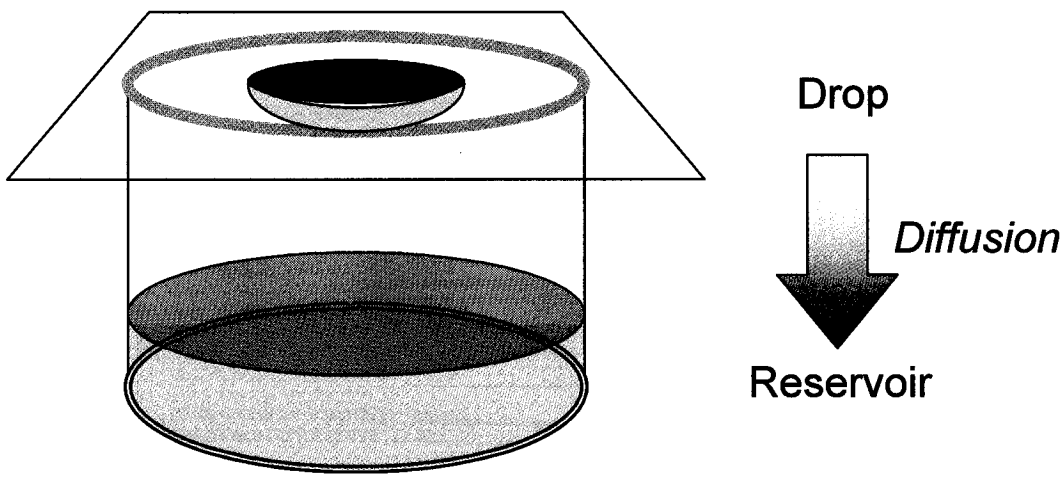


Figure 1.13 The hanging drop method of crystallization. A drop containing the protein (with or without the presence of ligand) is suspended above a reservoir solution in a sealed environment. The reservoir solution contains a higher concentration of solute and this creates a diffusion gradient. Equilibration of water occurs between the drop, the air in the closed vessel, and the reservoir. As water diffuses out of the drop, the protein and other solute concentrations increase. The slow rate at which the solution becomes saturated allows the protein molecules to arrange into a crystal.

In a well-ordered crystal, positions in which the surrounding environments are exactly identical, called equivalent, are defined as lattice points, and are related to one another in three-dimensional space by mathematical symmetry operations involving translations and rotations. One repeating volume-unit within the crystal containing these lattice points is called a unit cell. It is defined by axis lengths, ***a***, ***b***, and ***c***, with interaxis angles α, β, and γ. The restrictions on axes and angles in the unit cell are defined by 7 crystal systems (*i.e.* Fig. 1.14). However, considerations of crystal symmetry often make it advantageous to choose a unit cell that contains internal lattice points. These are described by the 14 Bravais lattices (*i.e.* Fig. 1.15) (see reference texts for representations of all Bravais lattices). Within each Bravais lattice there are several space groups – collections of operations that define the symmetry relationships and the locations of the equivalent points. Only 65 space groups are used to describe proteins made of left-handed chiral amino acids with a total of 230 space groups including mirror operations. The "asymmetric unit" contains atoms that are unique and not related by crystallographic symmetry operations. We need only to define the asymmetric unit to be able to obtain the entire unit cell through symmetry operations.

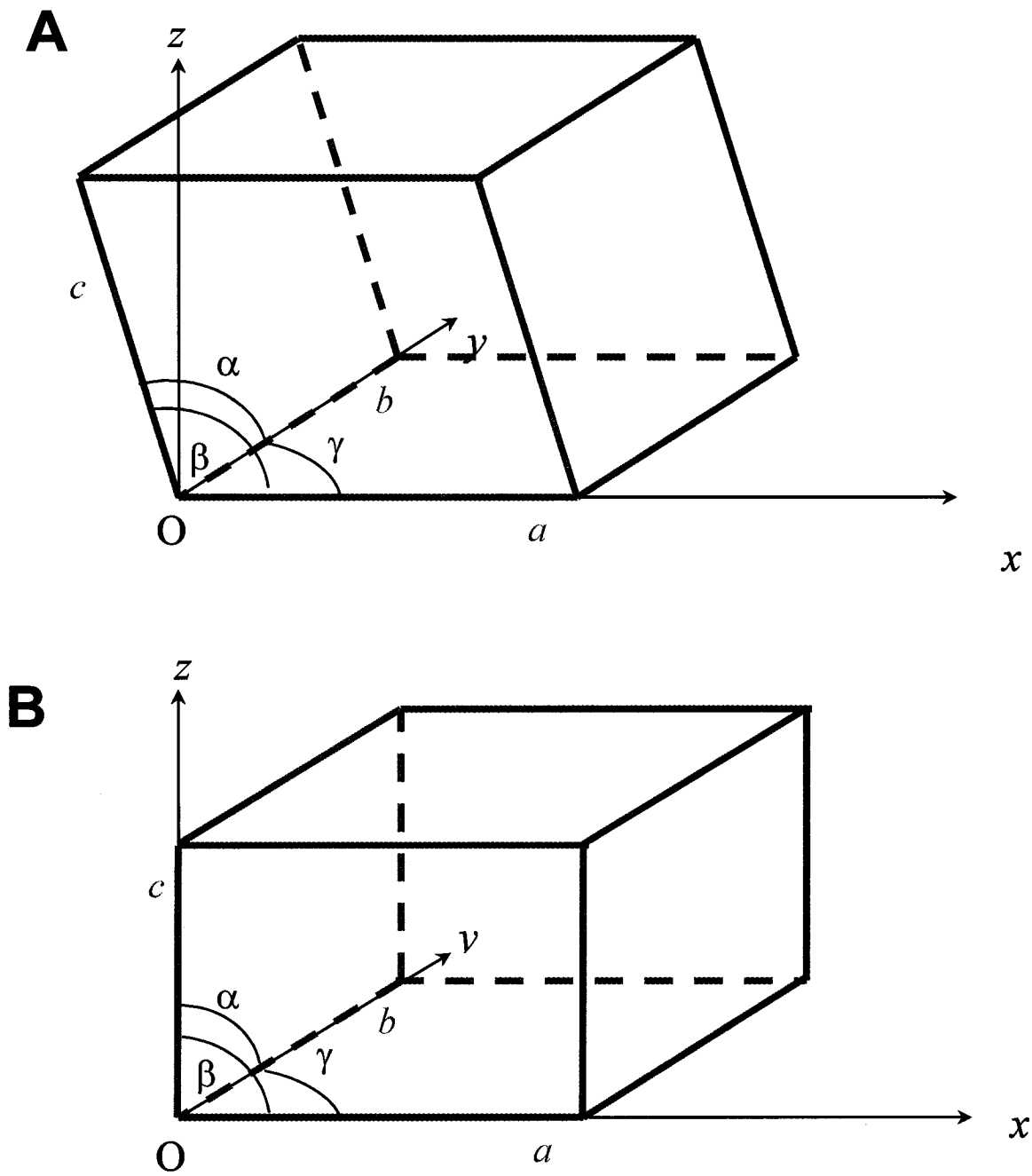


Figure 1.14 The unit cell is defined by axis lengths ***a***, ***b*** and ***c*** (*blue*) and interaxis angles α, β, γ (*red*). **A)** In a monoclinic crystal system such as observed in the crystal structure of unliganded S45-18, the introduction of symmetry along one axis has restricted the unit cell angles such that $\beta \neq 90°$ and $\alpha = \gamma = 90°$ and $a \neq b \neq c$. **B)** In an orthorhombic system with three axes of symmetry, $\alpha = \beta = \gamma = 90°$ and $a \neq b \neq c$.

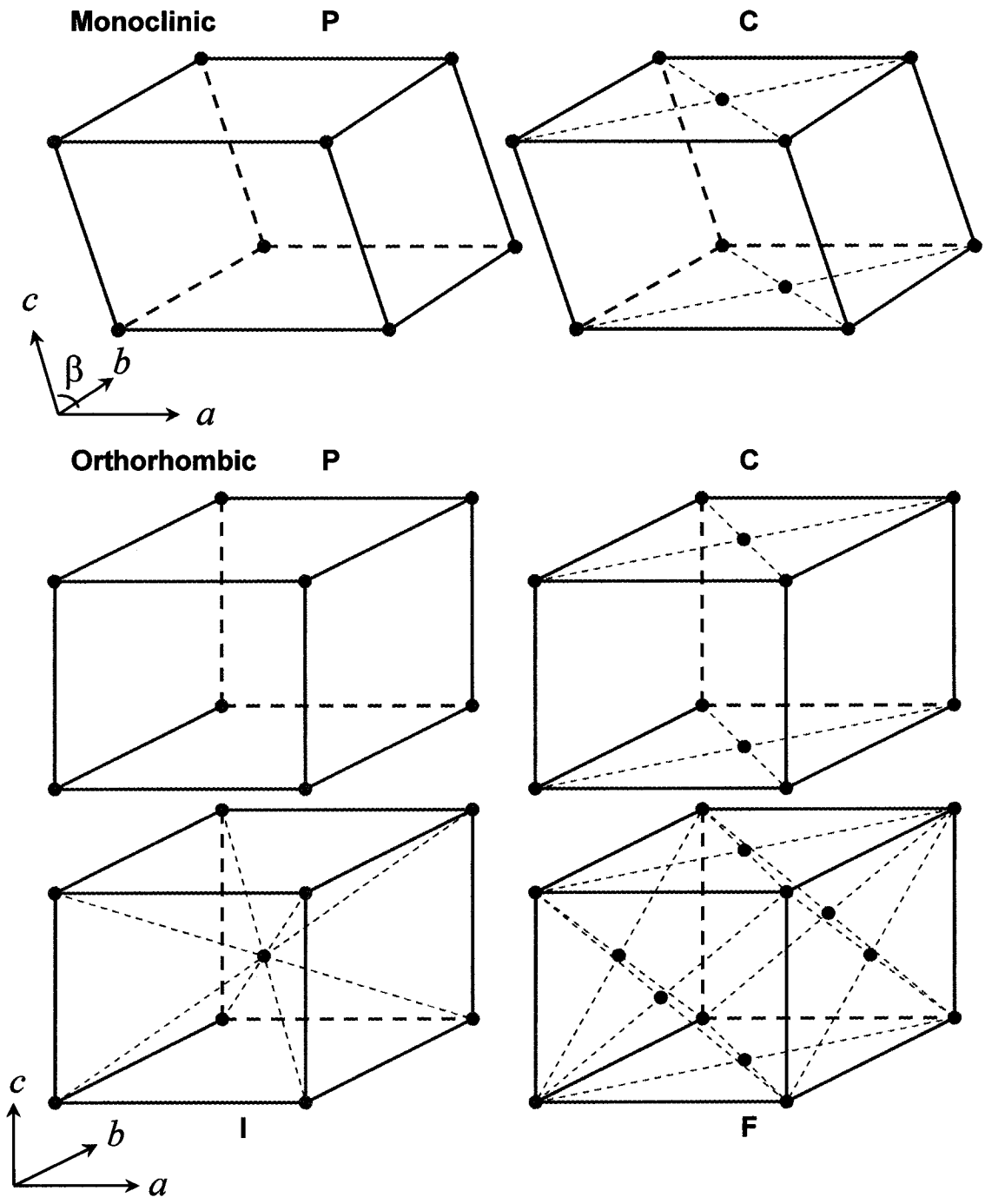


Figure 1.15 The unique arrangement of equivalent points (*solid dots*) can give rise to the Bravais lattices, for the monoclinic (*upper*) and orthorhombic (*lower*) crystal systems. **P** refers to an equivalent point only at the intersections of the unit cell edges. **C** "centering" refers to a point on the face defined as perpendicular to the **c** edge of the unit cell. Face-centering (**F**) refers to points at the diagonal intersection of each face. Body-centering **(I)** refers to a point at the diagonal intersection of the unit cell corners.

1.3.3 X-ray Diffraction

In the diffraction experiment (Fig. 1.16), a protein crystal of suitable size and quality is mounted in front of an X-ray beam. While most of the X-rays pass through the crystal unaltered, some are scattered in different directions by the electron clouds of the atoms. The various scattered X-rays from different positions in the crystal can combine constructively (and destructively) and appear on the diffraction pattern as spots of varying intensity called "reflections" because they can be thought of as being reflected from imaginary planes in the crystal defined by an index ***hkl*** (Fig. 1.17). Bragg's Law must be satisfied in order to observe (or predict) a reflection:

$$n\lambda = 2d_{hkl} \sin\theta$$

where **n** is an integer 1,2,3..., λ is the X-ray wavelength and is typically about 1.0 Å, and at the same scale as interatomic distances; d_{hkl} is the interplanar spacing between a set of parallel imaginary planes defined by an index ***hkl***, θ is the angle of incidence with and reflection from the plane.

The Miller index ***hkl*** separates the unit cell axis edge ***a*** into ***h*** sections, ***b*** into ***k*** sections, and ***c*** into ***l***. The planes perpendicular to the unit cell edges of the unit cell ***a, b,*** and ***c*** are then defined by Miller indices (1 0 0), (0 1 0) and (0 0 1) respectively. Each observed reflection on a diffraction pattern refers to the summation of diffraction from all the atoms in the crystal, weighted according to their distance from the planes defined by a index. The reflection (0 0 0) is taken to be the in-phase contribution of all atoms in the crystal.

From Bragg's law, the angle of diffraction θ is proportional to $1/d$. So for unit cells with longer axes and larger interplanar d_{hkl} spacings, the angle θ is smaller and the spots on

the diffraction pattern are closer together. As the ***hkl*** index increases, the smaller the interplanar d_{hkl} results in a higher resolution of data (able to distinguish a smaller distance in the unit cell), and at a larger value of θ.

The diffracted X-ray "reflections" actually form a three-dimensional spherical shape, but only a "slice" of this sphere is recorded electronically by an imaging detector as represented by Fig. 1.16. By rotating the crystal in discrete steps about an axis, all unique reflections of the sphere can be recorded, depending on the detector.

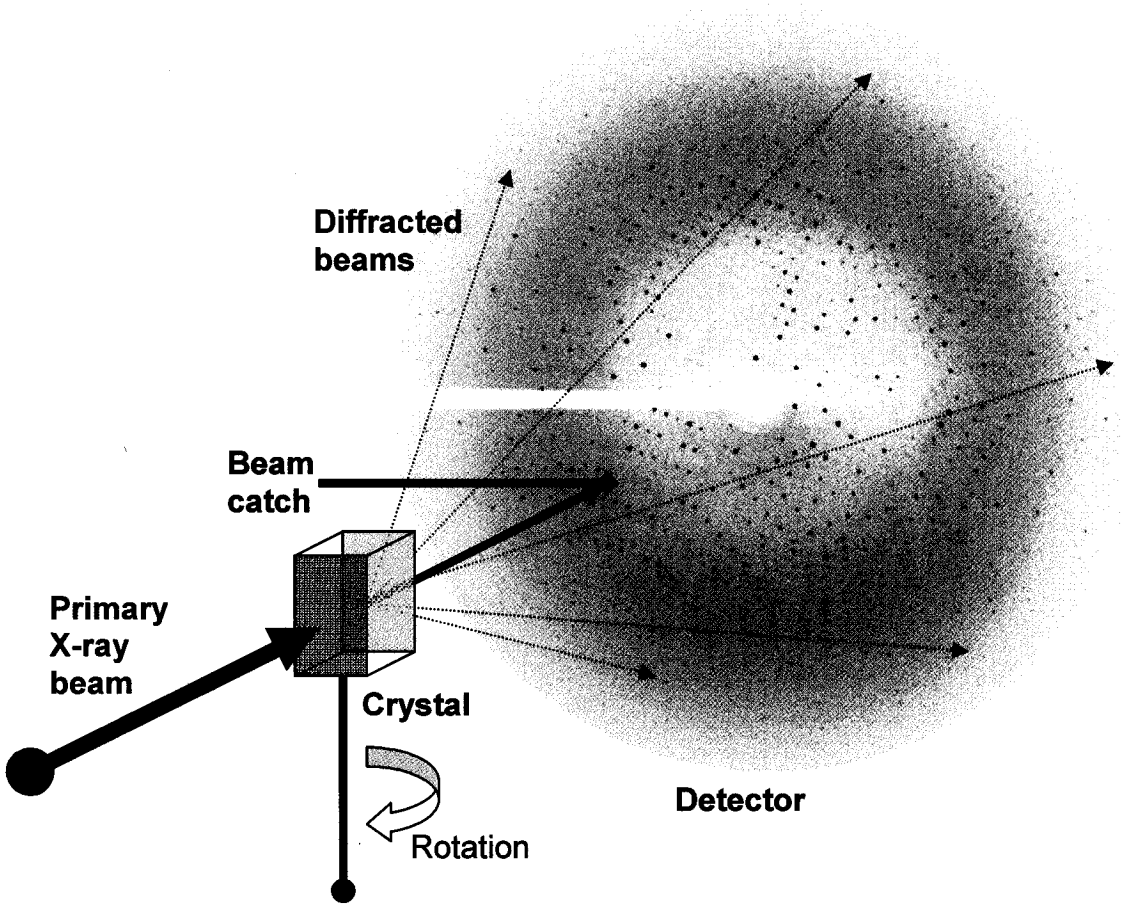


Figure 1.16 X-ray diffraction experiment. A primary X-ray beam *(red)* from a source is directed at the protein crystal. While most of the beam passes straight through, some are diffracted by the electron clouds at various angles. The diffracted beams from each position in the crystal interfere constructively and destructively with those from other positions to form a pattern that is recorded by the detector placed at a known distance. The crystal is rotated in steps through a range in order to obtain as much unique diffraction data as possible.

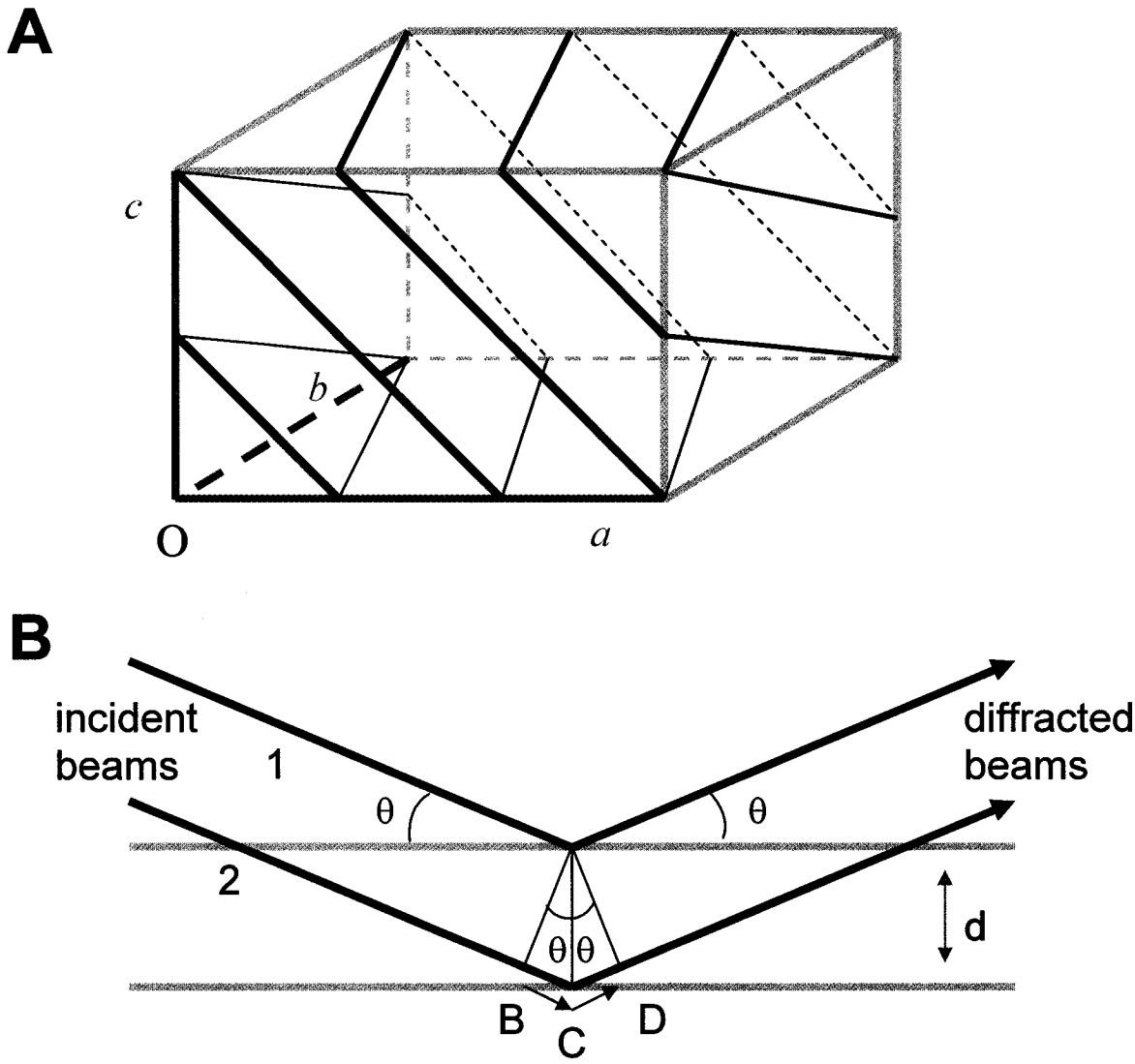


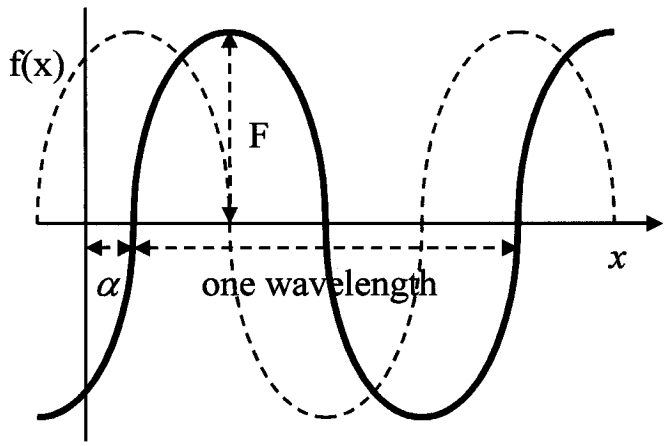
Figure 1.17 A) The ***hkl*** indices divide the unit cell into in a series of parallel planes. Here Miller index ***312*** divides unit cell axis ***a*** into $h = 3$ sections, ***b*** into $k = 1$ section, ***c*** into $l = 2$ sections. **B)** The condition to result in a "spot" for constructive interference of two parallel wavefronts is described by Bragg's law $n\lambda = 2d\sin\theta$. For example, two wavefronts (in *red*, 1 and 2) strike two planes spaced "d" apart and are diffracted at the same angle θ. The difference in the distance that beam 2 travels, BC+CD are equal to $2 \times d \sin\theta$ and constructive interference will occur if this distance is equal to one wavelength, λ, or an integral number of wavelengths, nλ.

1.3.4 Fourier Transform

In concept, we can imagine that all the atoms in a crystal can lie on various different sets of parallel planes defined by indices ***hkl***. In practice, to understand how all the atoms contribute to each reflection with respect to diffraction by a set of planes of index ***hkl***, we must first review the nature of each diffracted X-ray as a periodic wave function. For example, sine and cosine wave function equations can be written as:

$$f(x) = F\cos 2\pi(hx + \alpha),$$
$$f(x) = F\sin 2\pi(hx + \alpha)$$

and below showing the relationship between, for example, sine (*solid red*) and cosine (*dashed red*) functions:



where **f(x)** is the amplitude of the wave for any value of ***x*** where each repetition of a wavelength occurs at integer values of **x=1,2,3,…,n**; **F** is the maximum amplitude of the wave which is half the distance from peak to trough; **h** is the frequency expressed in wavelengths per cycle (and increasing **h** decreases the wavelength; the entire wavelength is 2π radians); and α is the "phase" shift which describes the position from a given origin.

Because of the periodic nature of the wave, a phase of 0 radians is the same as 2π, or 4π radians.

These simple wave functions can be summed to describe a more complicated wave function, in a Fourier series. Each Fourier term is a simple wave function as in below:

$$f(x)=F_1\cos 2\pi(1x+\alpha_1)+F_2\cos 2\pi(2x+\alpha_2)+F_3\cos 2\pi(3x+\alpha_3)+\ldots+F_n\cos 2\pi(nx+\alpha_n)$$

and re-written as:

$$f(x)=\sum_{h=1}^{n}F_h\cos 2\pi(hx+\alpha_h)$$

which is the summation of each term **h**, where **h=1,2,3,...n**. The Fourier series using sine functions can also be summed similarly.

The wave functions used to describe X-ray diffraction involve the use of complex numbers, which have the general form **a + *i*b**, where ***i*** is the imaginary number $(-1)^{1/2}$. The function **f(x) = cosθ + *i* sinθ** can be re-written as $\mathbf{f(x)} = e^{i\theta}$ where $\theta = 2\pi(hx)$ and for a one-dimensional wave (along the ***x*** direction), the Fourier function is:

$$f(x)=\sum_{h=1}^{n}F_h e^{2\pi i(hx)}$$

Then for a three-dimensional wave with (x,y,z),

$$f(x,y,z)=\sum_{h}\sum_{k}\sum_{l}F\,e^{2\pi i(hx+ky+lz)}$$

Applying this to X-ray diffraction, the contributions of every atom at **(x,y,z)** to each reflection with index ***hkl*** is described by a Fourier function called the structure factor $\boldsymbol{F_{hkl}}$ with two components: amplitude **F** and phase **α**:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

which describes a summation (over the total number of atoms **n**) of the individual atomic scattering factors f_j of atom **j**; and (x_j, y_j, z_j) are the coordinates of some atom **j** in the unit cell expressed as fractions of the unit-cell axis lengths; ***h, k, l*** are the frequencies as they appear in the wave function, indicated by the indices of that reflection (in the reciprocal lattice); and the phase is expressed in this form as the exponent term of ***e*** and it is important to note the contribution of the atomic coordinates **(x,y,z)** to the phase.

The atomic scattering function of atom **j** in turn is defined by:

$$f_j = f_o e^{-B(\sin^2\theta)/\lambda^2}$$

where scattering f_o is specific to the chemical nature of an atom (for example, varying with the different number of electrons) and assuming a spherical distribution of electron density. This equation also takes into account the thermal motion often referred to as the B-factor, and for scattering angle and wavelength. B can be related to absolute atomic motion by:

$$B = 8\pi^2 \bar{u}^2$$

where $\bar{u}^2$ is the mean square displacement of an atom. An Argand diagram can be used to represent the relationship between the real and imaginary components of the structure factor F_{hkl}, the contributions from individual atomic scattering f_j and the phase α (Fig. 1.18).

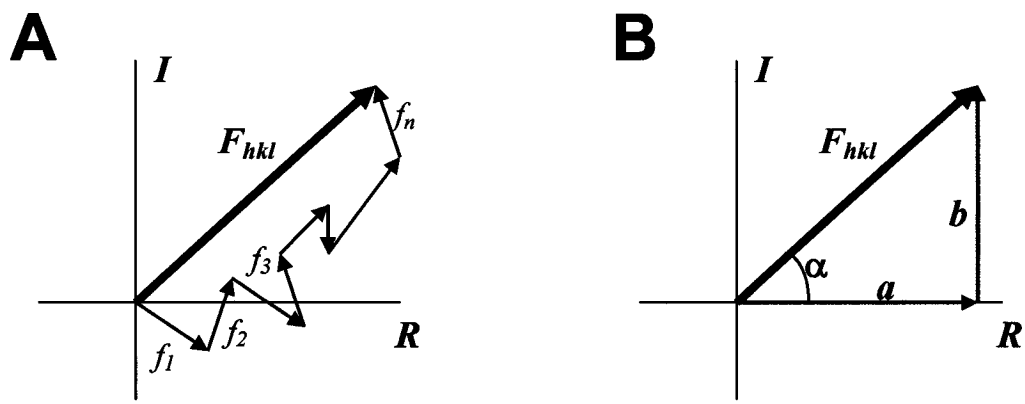


Figure 1.18 Argand diagrams representing the structure factor F_{hkl} in **A)** as a summation of the individual scattering f_j of **n** atoms with real (R) and imaginary (I) components, and in **B)** in the form **a** + *i***b** with the phase α.

1.3.5 The Phase Problem and Molecular Replacement

The information from the diffraction pattern can be used to create a model of the electron density ρ encountered by the X-rays. Again, the equations given here are to emphasize the importance of the contribution of every atom at **(x,y,z)** to each reflection ***hkl*** and to the electron density of the entire molecule. A Fourier transform of the structure factors gives of the electron density of the molecule:

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

where ρ is the electron density at spatial coordinates **(x,y,z)**, and **V** is the volume of the unit cell, and each F_{hkl} is a reflection derived from the diffraction pattern. However, F_{hkl} cannot be measured directly from the diffraction pattern. Each spot on a diffraction pattern only gives $\mathbf{I}_{hkl}$ which is proportional to $|F_{hkl}|^2$, from which only the amplitude **F** can be obtained but not the phase α_{hkl}, and thus no definitive solution for the atomic spatial coordinates **(x,y,z)**. This is referred to as the "phase problem" in X-ray crystallography.

In order to determine the phases, there are many approaches taken. These techniques generally begin with the Patterson function, which uses the measured intensities from the diffraction pattern, $|\mathbf{F}_{hkl}|^2$:

$$P(u,v,w) = \frac{1}{V}\sum_h\sum_k\sum_l |F_{hkl}^2| e^{-2\pi i(hu+kv+lw)}$$

at where the vector peak ***P(u,v,w)*** reflects the vectors between atoms in a molecule. Recall that the phase information is lost in the diffraction pattern, and it is not possible to directly interpret the atomic **(x,y,z)** positions. This information is not required in a Patterson function, and for two atoms **A** and **B**, their interatomic distances and relative positions are

expressed as a set of two vectors P_{AB} and P_{BA} each of equal magnitude but opposite directions. Each vector is brought to a common origin where there is also a peak for an atom's position with respect to itself, forming the Patterson map with a centre of symmetry.

In this study, a technique called molecular replacement (MR) is used to solve the phase problem. The initial phases are approximated from the existing three-dimensional **(x,y,z)** coordinates of a solved structure called a phasing model that is thought to share structural homology. Unliganded structures can also be used as the phasing model for their liganded counterparts. The phasing model must first be located in a similar orientation and position as the experimental data. This is accomplished by respectively performing rotation and translation searches in Patterson space.

1.3.6 Structure Refinement

After manually verifying and editing the amino acid sequence of the phasing model to reflect the experimental model, initial refinement treating the protein as several rigid-body groups is executed. This is proceeded by successive energy-minimization and simulated-annealing of the structure and generation of electron density maps. Manual inspection and intervention are required to ensure that the model fits appropriately into the electron density. Defining water and other molecules present in the environment is also required.

Besides visual inspection, a way to monitor the correctness of the model is with the calculation of the R factor:

$$R = \frac{\sum || F_{obs} | - | F_{calc} ||}{\sum | F_{obs} |}$$

where $|F_{obs}|$ is the amplitude of the structure-factor from the measured experiment data, and $|F_{calc}|$ is the amplitude of the structure-factor calculated from the current working model. The smaller the R value, the better the agreement of the working model to the experimental data. An iterative approach to solving a structure proceeds with cycles of computer calculation, electron density examination and manual adjustment.

2.0 Materials and Methods

The methods executed in this study involve steps of antibody purification from mouse ascites fluid, papain proteolysis to produce the Fab, HPLC column purification to purify the Fab, hanging-drop vapour diffusion method for protein crystallization and co-crystallization with antigen, and finally X-ray diffraction data collection and three-dimensional model building (Fig. 2.1).

2.1 Production of monoclonal immunoglobulins

Generation of specific antibodies against chlamydial epitopes has been previously described (Fu *et al.,* 1992). A panel of antibodies against chlamydial epitopes was provided by Dr. Helmut Brade in Borstel, Germany. Antibodies were raised against synthetic carbohydrates that mimicked their natural counterparts (supplied by Dr. Paul Kosma, University of Agriculture, Vienna, Austria) conjugated to BSA (Fu *et al.,* 1992). Briefly, BALB/c mice were immunized during a 60-day period by several injections of a mixture of Kdo-BSA glycoconjugates and Freund complete adjuvant. Their spleens were removed and the cells were fused 1:1 with myeloma cells using polyethylene glycol to form hybridomas and screened against LPS conjugated to BSA. Two such hybridoma lines were used to produce ascites containing monoclonal immunoglobulins $IgG_{1\kappa}$ S25-2 and S45-18 by the lab of Dr. C. Roger MacKenzie (IBS/NRC, Canada). The cloning and sequencing of the S45-18 variable domain genes was performed by the lab of Dr. Nina Seto (IBS/NRC, Canada) and has been described previously (Nguyen *et al.,* 2001).

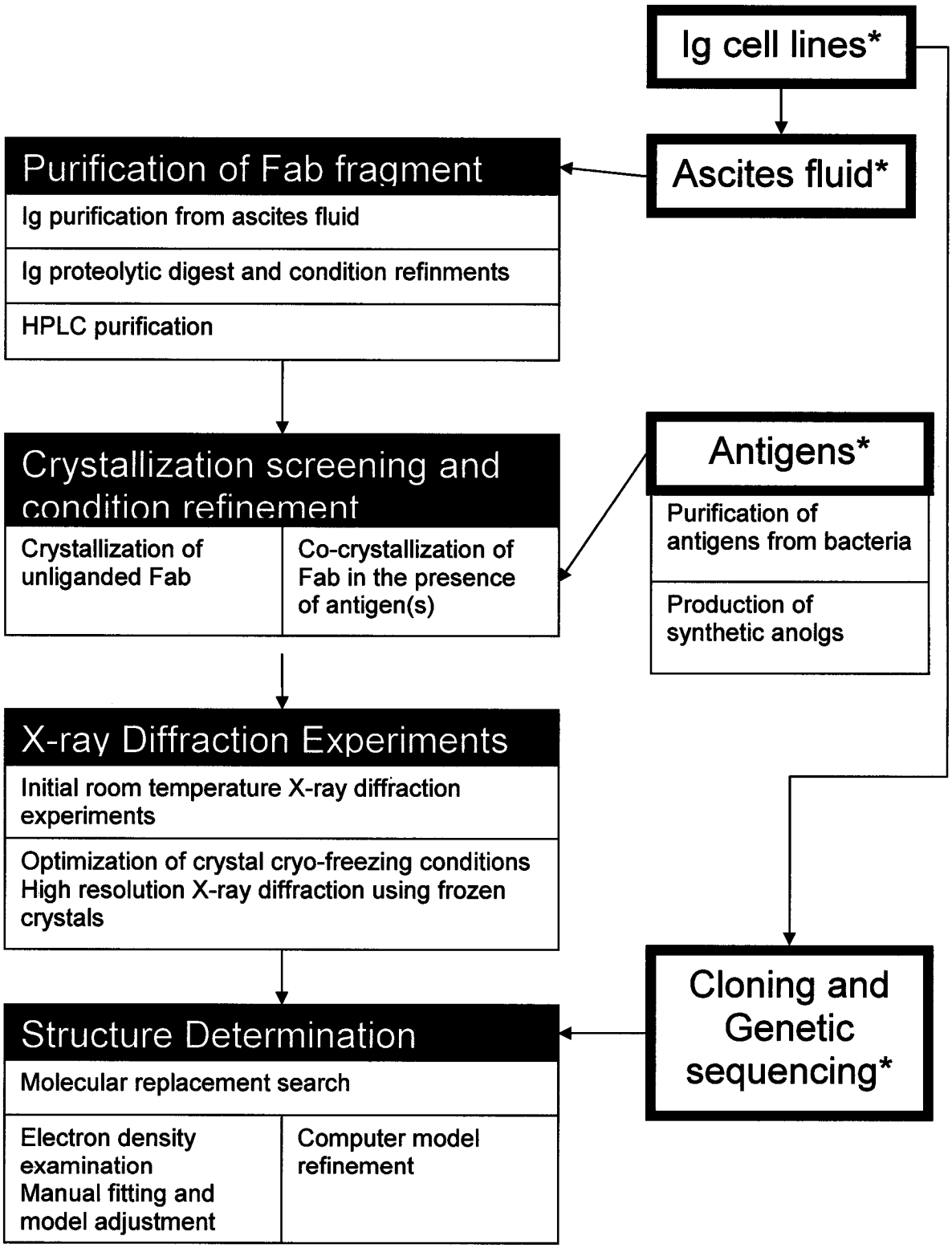


Figure 2.1 Outline of experimental design. *Material from collaboration.

2.2 *Preparation of Fab*

2.2.1 Caprylic acid precipitation of serum albumins

Two volumes of 60 mM sodium acetate buffer pH 4.3 were added per volume of ascites, which was brought to approximately pH 4.8. For each 10 mL of original ascites volume, 0.4 mL of caprylic (octanoic) acid (>99% purity, from Sigma) was added dropwise and the mixture was stirred end-over-end for 30 minutes. The effect of the caprylic acid was to precipitate serum albumins. After centrifugation at 5000 × *g* for 15 minutes, the supernatant portion (containing the immunoglobulins) was collected, and dialyzed overnight against 20 mM Tris-HCl pH 8.0 and 100 mM NaCl.

2.2.2 Immunoglobulin purification by affinity column

S45-18 was previously purified from ascites by Dr. Sonia Patenaude at the University of Ottawa using a protein A affinity column (Affi-Gel Protein A MAPS II Kit from BioRad with improved binding of immunoglobulin subclass G1). This purification is based on purifying the Ig by binding the Fc portion, which generally has an acidic pI. The process involves binding the antibody in a basic buffer environment, and then eluting it using a buffer of pH below the pI of the Fc portion, typically about pH 3.

During this period, no new S45-18 ascites was received. For S25-2, affinity column purification was not necessary due to the relatively low amounts of contaminating serum proteins after caprylic acid precipitation.

2.2.3 Papain proteolysis of S25-2 and S45-18 IgG

The immunoglobulins were subjected to papain proteolysis which cleaves at the hinge region of the heavy chain to produce two moles antigen-binding fragments (Fabs) and one mole Fc portion per mole of IgG (Fig. 2.2). Digest experiments contained papain (Papaya Latex mercuripapain from Sigma) in a ratio of 1:100 (w/w) papain:IgG, with 2 mM EDTA, 5 mM dithiothreitol and 50 mM Tris HCl pH 8.5, and were brought to a final volume with 20 mM Tris HCl pH 8.0. The concentration of IgG in the reaction mixture was just less than 1 mg mL^{-1}.

SDS-PAGE was used to optimize digestion times (Fig. 2.3). The reaction was quenched with 10 mM iodoacetamide after 4 hours for the S25-2 digest and after 2.5 hours for the S45-18 digest, where no further decrease was observed in the intensity of the respective protein band corresponding to the reported weight of an IgG heavy-chain (approximately 50 kDa). The solution was dialyzed overnight at 4°C with two changes of 20 mM Tris HCl buffer pH 8.0.

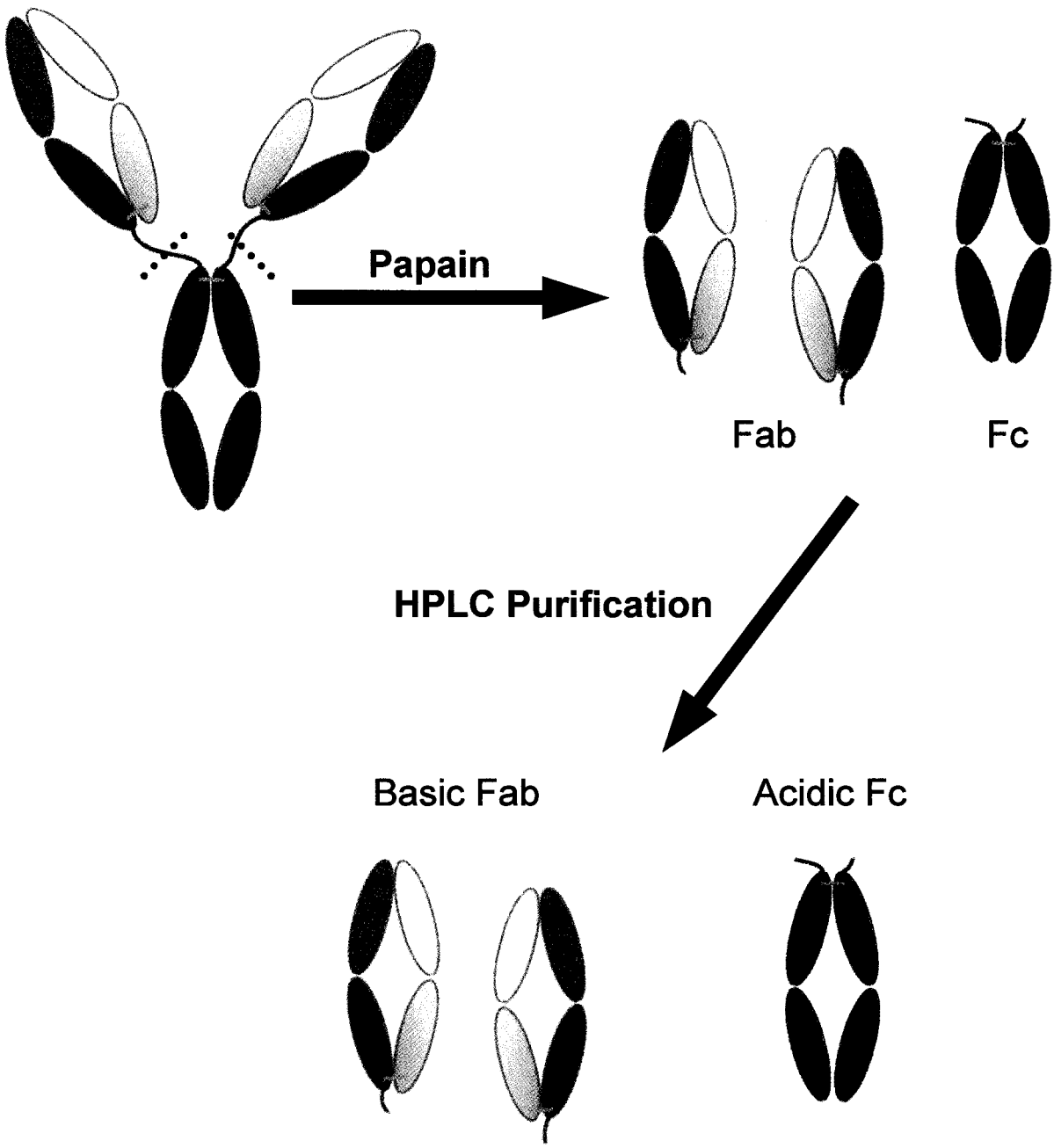


Figure 2.2 IgG papain proteolytic digest of the hinge region, followed by HPLC separation of the Fab based on the charge (pI) difference of the fragments.

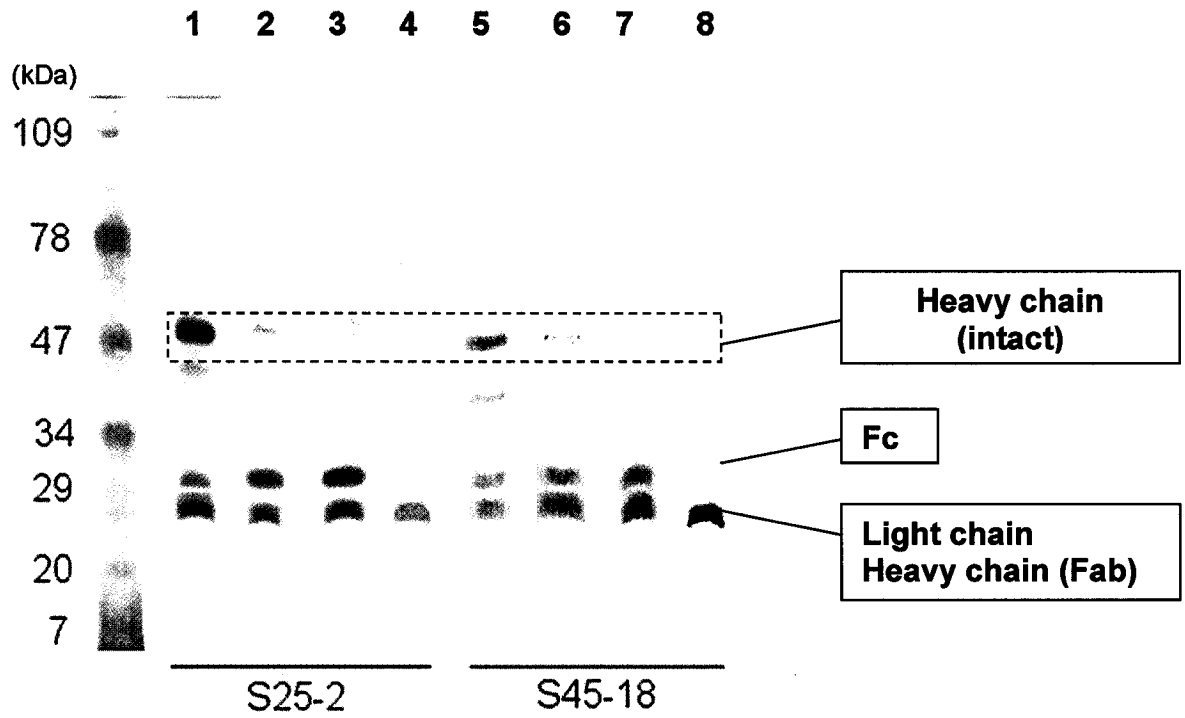


Figure 2.3 Examination of papain proteolytic digestion of IgG using sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). The total weight of the IgG (light and heavy chain dimer) is approximately 150 kDa. However, denaturing conditions disrupt the association between the light and heavy chains, appearing as approximately 50 and 25 kDa bands at time 0 (lane 1 and lane 5). As the cleavage at the hinge region proceeds releasing Fab and Fc, samples are taken at certain time points and prepared for SDS-PAGE. The 50 kDa heavy chain disappears (*dotted line*), and the Fab light and heavy chains appear as a 25 kDa band, and the Fc appear as a 30 kDa band. At 1 hour (lane 2 and lane 6), most of the heavy chain has disappeared. By 4 hours (lane 3) very little remains of the heavy chain of S25-2. The disappearance of the heavy chain occurs quicker for S45-18 in lane 7 at 2.5 hours. Samples from HPLC-purified basic (Fab) fractions are shown in lanes 4 and 8.

2.2.4 HPLC purification

The protein sample was filtered through a low protein binding 0.22 μm filter (Millipore) and degassed under vacuum before proceeding with HPLC cation-exchange purification (Gilson Series 800 HPLC system, Gilson Scientific Canada, Inc.; Cation exchange Shodex CM-825 column from Phenomenex). Digest volumes 0.8 - 1 mL were injected with a mobile phase of degassed 20 mM Tris pH 8.0 flowing at 1 mL min^{-1}.

Both the S25-2 Fab and the S45-18 Fab eluted in a 20-minute linear gradient of 0 - 1.2 M KCl, with corresponding Fab peaks appearing at about 1.0 M KCl (Fig. 2.4, 2.5). Samples were concentrated to 30 mg mL^{-1} using Centricon YM-10 centrifugal filter devices (Millipore) with a wash buffer of 20 mM Tris pH 8.0.

2.2.5 Sequencing of Antibodies

Genetic and N-terminal sequencing were provided by collaborators L. Brade, and H. Brade (Research Center Institute, Borstel, Germany), and C.R. Mackenzie and N.O.L. Seto (Institute for Biological Sciences, National Research Council, Ottawa, Canada) and has been described previously (Nguyen *et al.,* 2001).

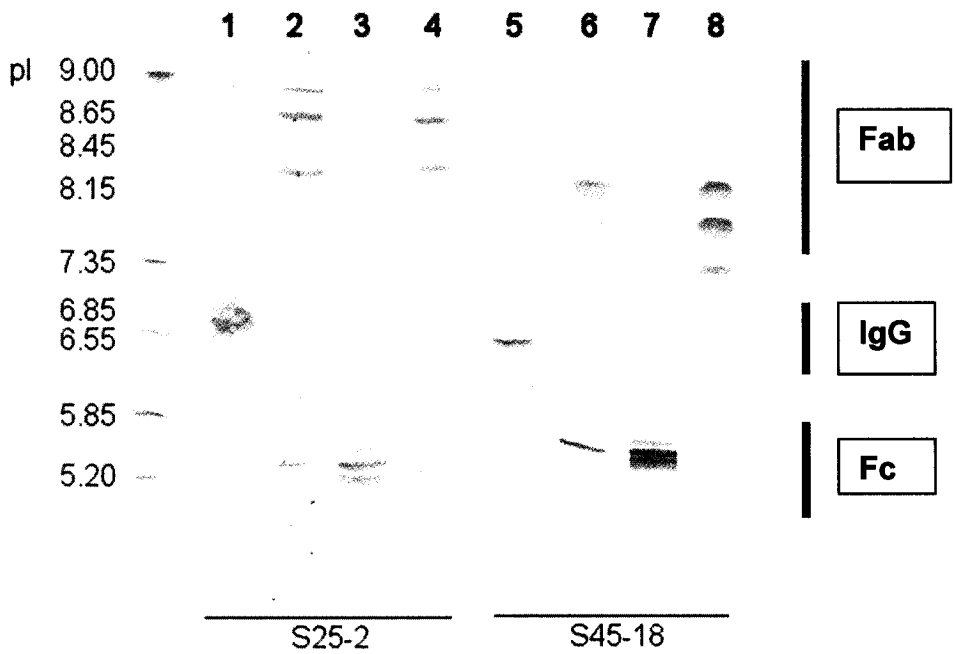


Figure 2.4 Examination of samples using isoelectric focusing (IEF) (diluted in a non-denaturing solution with HEPES pH 7.4 buffer). Lane 1 and lane 5 shows the intact IgG before digest for comparison with lanes 2 and 6 which represent samples immediately after the respective preparative digest experiments (4 hours for S25-2, and 2.5 hours for S45-18). The HPLC-eluted acidic and basic fractions are concentrated and shown on IEF. Lane 3 and lane 7 show the acidic fragments corresponding to the Fc. Lane 4 and lane 8 show the basic Fab fragments. The different bands present in each sample suggest different extent of proteolytic cleavage, and deamidation of Asn and/or Gln residues.

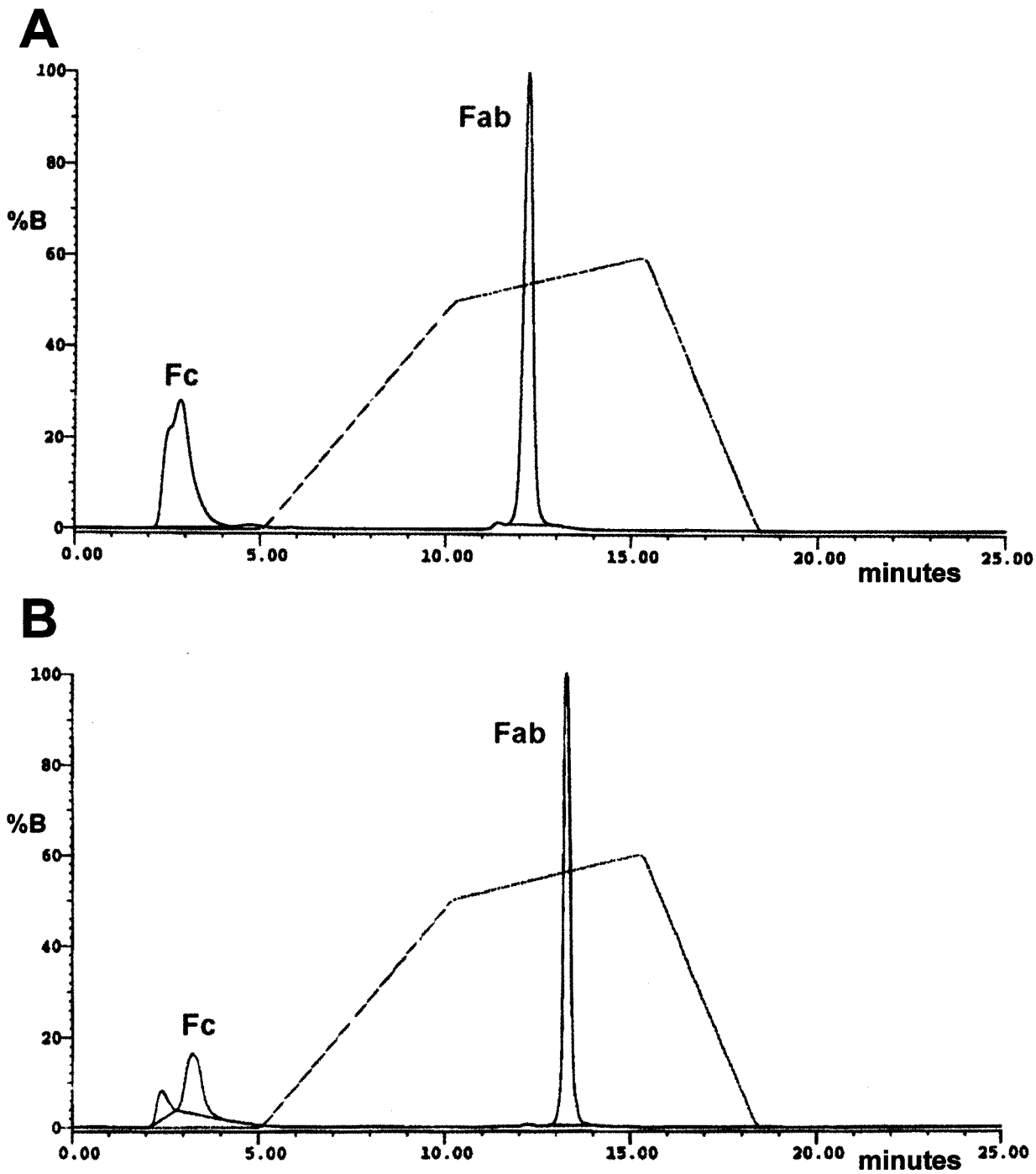


Figure 2.5 Elution profile of high performance liquid chromatography (HPLC) purification of the papain proteolytic digest of **A)** S25-2 and **B)** S45-18, using a cation-exchange column. The concentration of KCl is given as %B where 100% is 2.0 M KCl (Tris buffer pH 8.0). The salt concentration was steadily increased (*dotted line*) after the elution of the non-binding acidic Fc fraction (*red*) within 5 minutes after injection. The basic Fab (*blue*) was retained in the column until the concentration of KCl reaches 1.0 M and is eluted before 1.2 M.

2.3 *Crystallization of Fab*

2.3.1 Initial Screening

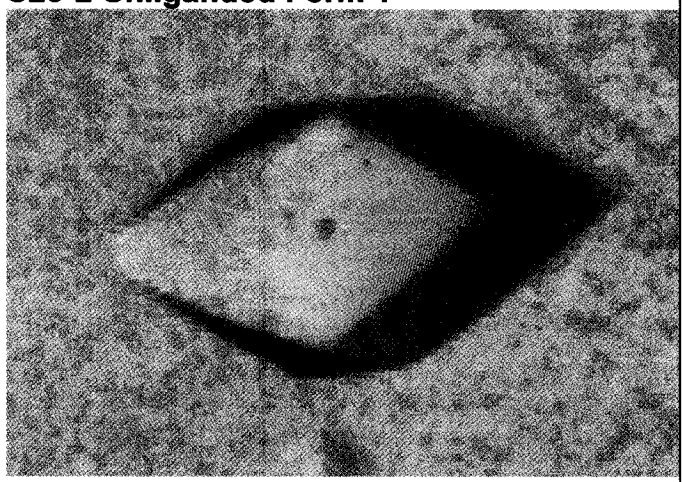
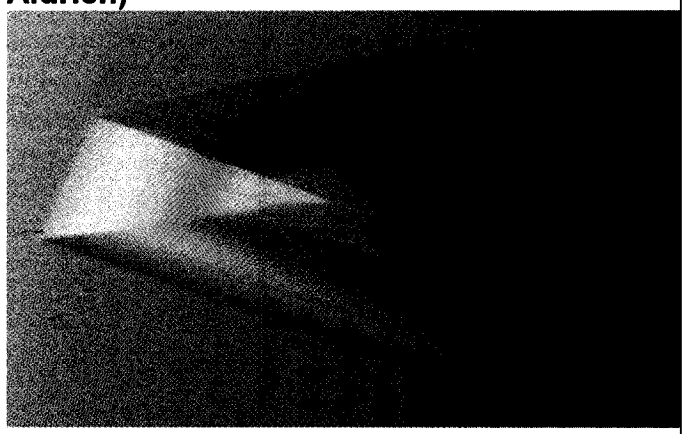
The various unliganded and liganded forms of Fab were crystallized via the hanging-drop vapor diffusion method. Forty-nine crystallization mixtures based on the Crystal Screen I (CS) conditions (Hampton Research) were used to test the crystal growth of the various protein samples. For co-crystallization in the presence of the corresponding antigens, the antigens were dissolved with the Fab solutions in the specific ratios and left overnight prior to setting up crystallization.

Drop sizes of 1-2 μL mixed in a 1:1 ratio of protein sample to precipitant were suspended over 1 mL of reservoir (precipitant) in 24-well Linbro plates (Hampton Research). Fragile crystal clusters of unliganded S25-2 Fab grew in CS condition #45. Rod-like clusters of S25-2 in the presence of antigens grew in CS #6. Unliganded S45-18 crystals were grown readily in several CS I conditions. In the presence of an 80-fold molar excess of antigen, S45-18 Fab initial screens suffered from phase separation. Well-formed crystals were produced upon minimizing phase separation by decreasing the precipitant concentration of CS #6.

2.3.2 Refinement of crystallization and preparation for diffraction studies

The various unliganded and liganded forms of Fab were crystallized typically in drops of 6 to 10 μL, suspended from covers of culture dishes (Corning) above a reservoir (1 mL), and then sealed with parafilm. The final drop conditions are presented in Table 2.1.

Table 2.1 Drop and reservoir conditions for crystallization experiments.

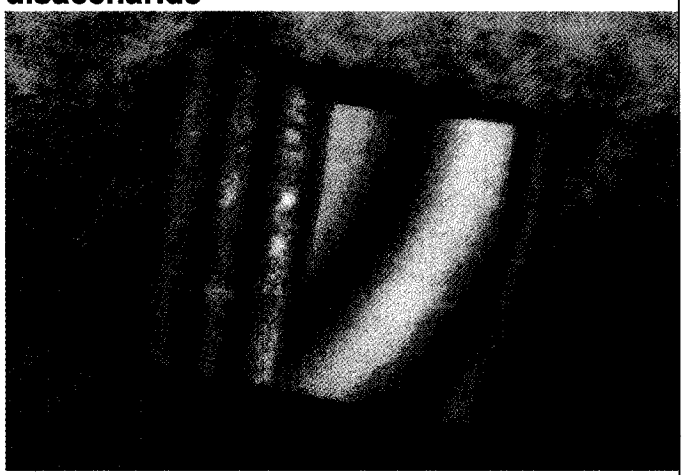
Crystal	Conditions
S25-2 Unliganded Form 1 	**Drop** 12 mg/mL Fab, 20 mM NaCacod pH 6.5, 25 mM Tris pH ~ 8.0, 45 mM ZnOAc, 45 mM $MgCl_2$, 4 % (w/v) PEG 8000, 4 % (w/v) PEG 4000 **Reservoir** 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 17% (w/v) PEG 4000
S25-2 Unliganded Form 2 	
S25-2 with Kdo monosaccharide (Sigma-Aldrich) 	**Drop** 15 mg/mL Fab, 150 × sugar (mol:mol), 20 mM Tris pH ~ 8, 35 mM $MgCl_2$, 22 mM $ZnCl_2$, 7 % (w/v) PEG 4000, 9 % (w/v) Ethylene Glycol **Reservoir** 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 25% (w/v) PEG 4000

S25-2 with αKdo-(2→8)- αKdo-(2→4)- αKdo-(2→*O*)-allyl trisaccharide[1]

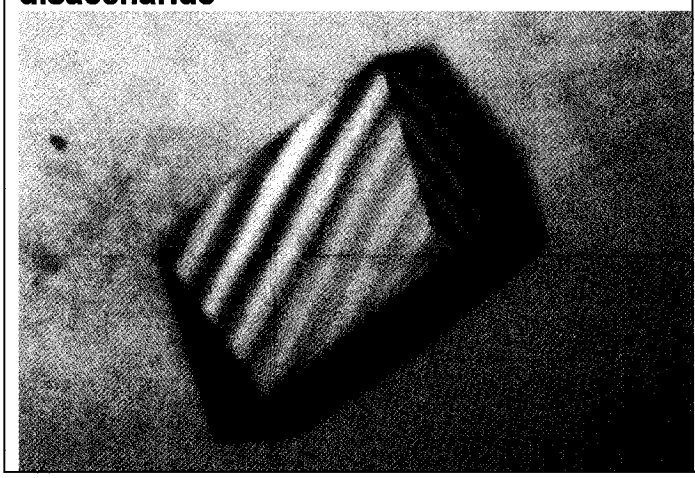
Drop
15 mg/mL Fab,
80 × sugar (mol:mol),
20 mM Tris pH ~ 8,
35 mM $MgCl_2$,
22 mM $ZnCl_2$,
7 % (w/v) PEG 4000,
9 % (w/v) Ethylene Glycol
3% (w/v) Glycerol

Reservoir
100 mM Tris pH 8.5,
200 mM $MgCl_2$,
25% (w/v) PEG 4000

S25-2 with αKdo-(2→8)- αKdo-(2→*O*)-allyl disaccharide

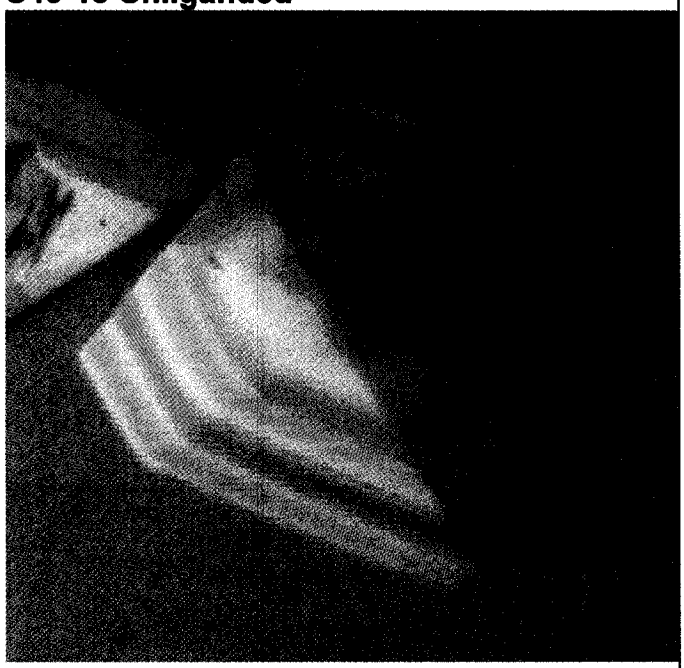
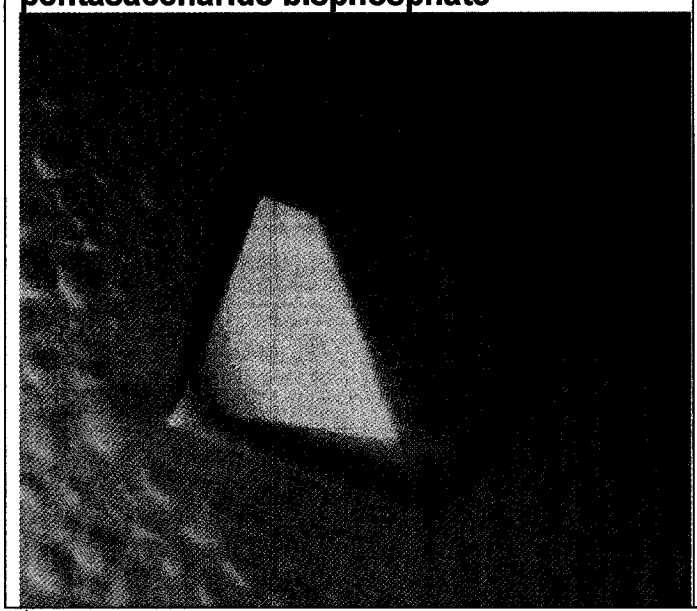


25-2 with αKdo-(2→4)- αKdo-(2-*O*)-allyl disaccharide



Drop
18 mg/mL Fab,
100 × sugar (mol:mol),
20 mM Tris pH ~ 8,
43 mM $MgCl_2$,
28 mM $ZnCl_2$,
8 % (w/v) PEG 4000,
11 % (w/v) Ethylene Glycol

Reservoir
100 mM Tris pH 8.5,
200 mM $MgCl_2$,
25% (w/v) PEG 4000

S45-18 Unliganded 	**Drop** 6-8 mg/mL Fab, 25 mM K_3PO_4 pH 7.8, 10% (w/v) PEG 8000 **Reservoir** 25 mM K_3PO_4 pH 7.8, 20% (w/v) PEG 8000
S45-18 with αKdo-(2→4)- αKdo-(2→4)- αKdo-(2→6)-βGlcN-4P-(1→6)- αGlcN-1P pentasaccharide bisphosphate[2] 	**Drop** 9 mg/mL Fab, 80 × sugar (mol:mol), 65 mM Tris pH 8.5, 130 mM $MgCl_2$, 13% (w/v) PEG 4000 **Reservoir** 100 mM Tris pH 8.5, 200 mM $MgCl_2$, 20% (w/v) PEG 4000

[1] (2→*O*)-allyl analogs were provided by P. Kosma (University of Agricultural Sciences, Vienna, Austria)

[2] pentasaccharide antigen provided by H. Brade (Research Center Borstel, Borstel, Germany)

There were different crystal morphologies observed in the drops containing unliganded S25-2 crystals. Attempts were made to select the different forms for diffraction experiments.

Crystals of various sizes (0.15 to 0.85 mm) were either mounted in glass capillaries (Charles Supper Co.) for room temperature diffraction, or washed with a drop-condition solution but without the presence of additional protein or sugar and with 20% of the water replaced by ethylene glycol or glycerol, after which crystals were then mounted on a rayon loop (Hampton), frozen in liquid propane (Praxair), and stored in cryo-vials (Hampton Research).

2.4 Diffraction and Data Collection

The first crystal of diffraction quality collected was S45-18 Fab grown in presence of the pentasaccharide antigen. Preliminary and room-temperature data sets of liganded S45-18 Fab were collected at the University of Toronto (RU-200 rotating anode, Rigaku, Tokyo, Japan; MAR Research 345 image plate detector, Evanston, IL). Diffraction data from initial crystals of S25-2 Fab grown in the presence of $\alpha(2\rightarrow8)$ Kdo disaccharide (Appendix Table A.1) were collected at Queen's University (RU-300 rotating anode, Rigaku, Tokyo, Japan; MAR Research 300 image plate detector & single axis goniostat, Evanston, IL).

High resolution data for unliganded and liganded S25-2 (Table 2.2) and S45-18 (Table 2.3) were collected at Brookhaven National Labs (Upton, New York) at the National Synchrotron Light Source (NSLS) beamline X8C (Area Detector Systems Corporation Quantum 4Q CCD detector, Poway, CA; MAR single-axis goniostat used from June '99 to March '01, MAR Research, Evanston, IL; Single-axis ω/ 2θ arm used from April '01 to November '01, Cystal Logic, Los Angeles, CA; Oxford Cryosystems model 600 low temperature system, Oxford, UK). DENZO and SCALEPACK were used as part of the HKL 2000 package (Otwinowski & Minor 1996) for data integration and scaling.

Table 2.2 Cryo data refinement and model statistics for unliganded and liganded forms of S25-2 Fab.

	S25-2 Fab					
	Unliganded form 1	Unliganded form 2	α(2→8)-α(2→4) trisaccharide	α(2→8) disaccharide	α(2→4) disaccharide	Mono-saccharide
Space group	$P2_12_12$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	*a* = 96.8 Å, *b* = 112.2 Å, *c* = 41.5 Å	*a* = 84.7 Å, *b* = 96.0 Å, *c* = 114.2 Å	*a* = 45.6 Å, *b* = 80.9 Å, *c* = 130.6 Å	*a* = 45.9 Å, *b* = 81.6 Å, *c* = 131.5 Å	*a* = 45.9 Å, *b* = 81.8 Å, *c* = 130.3 Å	*a* = 45.9 Å, *b* = 81.8 Å, *c* = 130.3 Å
No./ASU	1	2	1	1	1	1
No. Obs	203,789	230,670	279,457	870,735	363,226	183,080
Unique reflections[a]	31,851 (2,827)	38,076 (3,282)	77,711 (6,994)	84,235 (8,455)	49,410 (4,918)	44,748 (4,791)
Resolution Range (Å)	20 - 1.96	20 - 2.28	20 - 1.49	20 - 1.45	20 - 1.74	20 - 1.73
Completeness (%)[a]	96.1 (87.5)	88.4 (77.5)	97.3 (88.9)	95.2 (96.9)	97.9 (96.4)	92.3 (84.8)
R_{sym}[a,b]	0.054 (0.328)	0.053 (0.380)	0.042 (0.398)	0.051 (0.247)	0.043 (0.381)	0.043 (0.353)
$\langle I/\sigma \rangle$[a]	28.3 (3.2)	29.7 (4.2)	29.9 (2.3)	38.1 (9.4)	40.5 (3.9)	22.7 (3.2)
Polypeptide[c]	3,399	6,798	3,399	3,399	3,413	3,399
Sugar atoms[c]	0	0	49	34	34	16
Solvent atoms[c]	273	409	473	516	339	397
R_{work}[d]	0.223	0.216	0.200	0.203	0.215	0.202
R_{free}[d]	0.275	0.287	0.223	0.225	0.252	0.236

[a] values in parenthesis refer to the highest resolution shell which corresponds to 2.03 – 1.96 Å for S25-2 unliganded form 1, 2.36 - 2.28 Å for S25-2 unliganded form 2, 1.54 – 1.49 Å for S25-2 - α(2→8)-α(2→4) trisaccharide, 1.50 – 1.45 Å for S25-2 - α(2→8) disaccharide, 1.80 – 1.74 Å for the S25-2 - α(2→4) disaccharide, 1.79 – 1.73 Å for S25-2 – monosaccharide

[b] $R_{sym} = \Sigma|(I_{hkl} - I_{<hkl>})| / (\Sigma I_{hkl})$ where $I_{<hkl>}$ is the mean intensity of all reflections equivalent to reflection *hkl* by symmetry.

[c] non-hydrogen atoms.

[d] R_{work}, $R_{free} = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, 10% of the data was used for the calculation of R_{free}.

Table 2.3 Cryo data refinement and model statistics for unliganded and liganded forms of S45-18 Fab.

	S45-18 Fab	
	Unliganded	Pentasaccharide
Space group	C2	$P2_12_12_1$
Unit Cell	*a* = 76.4 Å, *b* = 170.3 Å, *c* = 77.1 Å β = 114.4°	*a* = 71.2 Å, *b* = 113.9 Å, *c* = 133.9 Å
No./ASU	2	2
No. Obs	234,850	406,104
Unique reflections[a]	75,932 (6,553)	106,111 (10,237)
Resolution Range (Å)	20 - 1.79	20 - 1.75
Completeness (%)[a]	90.4 (78.5)	95.8 (93.1)
R_{sym}[a,b]	0.030 (0.133)	0.063 (0.396)
$\langle I/\sigma \rangle$[a]	41.6 (6.9)	18.2 (2.4)
Polypeptide[c]	6,850	6,850
Sugar atoms[c]	0	136
Solvent atoms[c]	529	770
R_{work}[d]	0.220	0.212
R_{free}[d]	0.247	0.247

[a] values in parenthesis refer to the highest resolution shell which corresponds to , 1.85 – 1.79 Å for unliganded S45-18 and 1.81 – 1.75 Å for liganded S45-18.

[b] $R_{sym} = \Sigma|(I_{hkl} - I_{<hkl>})| / (\Sigma I_{hkl})$ where $I_{<hkl>}$ is the mean intensity of all reflections equivalent to reflection *hkl* by symmetry.

[c] non-hydrogen atoms.

[d] R_{work}, $R_{free} = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, 10% of the data was used for the calculation of R_{free}.

2.5 Molecular Replacement, Structure Refinement and Manual Fitting

X-PLOR (Brunger 1992) and CNS (Brunger *et al.,* 1998) were used to find the appropriate molecular replacement solutions, and to perform computer refinements. The initial search probe was unliganded YsT 9.1 Fab produced in response to *Brucella abortus* A antigen (Rose *et al.,* 1993), used for the room temperature unliganded S45-18 Fab diffraction data. The rotation and translation searches in Patterson space used reflections in the 20 – 4 Å range. Rigid body refinement was performed where the domains variable light, variable heavy, constant light, and constant heavy were each treated as a solid body and readjusted with respect to one another. After computer positional refinement, electron density maps were calculated and the model was inspected and manually adjusted where necessary using FRODO (Jones 1985), SETOR (Evans 1993) or PyMol (DeLano 2004). Typically about 30 – 50 % additional reflections are added to the number of reflections used in each cycle of computer and manual refinements until all the data have been included. At around 2.8 Å resolution, a CNS search was made for electron density (using a *Fo-Fc* map) was carried out to find peaks that could represent water molecules. Each water molecule added by the computer search was then inspected and readjusted manually if required. At this resolution, electron density for the Kdo sugar was clearly seen from crystals that were co-crystallized with ligand, and the respective sugar structure was added to the model. Cycles of simulated annealing ("slow-cooling"), and individual atomic B-factor refinements were also performed.

The unliganded room temperature S45-18 Fab model was used as a probe for the initial remaining Fab data, including those for the room temperature liganded S45-18 Fab and liganded S25-2 Fab.

Section B: Results and Discussion

3.0 Conservation of germline in S25-2 and S45-18

The unliganded and liganded structures of the antibodies S25-2 and S45-18 examined here have many important implications to our understanding of the immune system. This is the first report of liganded antibody structures specific against a panel of clinically relevant related LPS epitopes (Nguyen *et al.,* 2003). A binding site that recognizes only a limited number of carbohydrates sheds light upon how the immune system can maximize the usage of encoded genes by the inheritance of polyspecificity while minimizing cross-reactivity with self antigens.

The ability of the immune system to adapt to a diverse amount of structures has been discussed for decades. It is understood that two branches of the immune system, namely the innate and adaptive, make use of the germline information in different ways. The innate immune system offers an early response to foreign antigens and is often the initiator of the adaptive response, and thus is critical for survival of the host. It employs many preformed receptors (in the absence of antigen) including immunoglobulins that retain germline identity. The past 70 years has seen much progress in understanding the adaptive immune system including clonal selection, affinity maturation, genetic and non-template mechanisms to create paratope diversity. However the large structural and chemical variation in carbohydrates is often not recognized by the adaptive immune system with the exception of those bound to carriers presented by MHC. However germline

conservation in affinity-matured immunoglobulins presents an opportunity to understand the contributions of germline identity to antigen recognition.

The effectiveness of the response through natural selection by prevalent pathogens shapes the inherited repertoire. To adapt to antigen diversity, the genetic code must be maximized to recognize as many epitopes as possible. The concept of immunoglobulin multi-epitope binding has significant advantages to adapting to pathogenic diversity; however, this may also result in undesirable recognition of self-antigens. Exploring the structural basis of binding of closely related immunoglobulins S25-2 and S45-18 allows us to understand how the inherited germline sequences are used to adapt to multiple epitopes while minimizing self-antigen recognition.

Like many PRRs, S25-2 and S45-18 target carbohydrates that are not found in the host, such as Kdo which forms a critical part of LPS in all Gram-negative bacteria. Carbohydrate recognition by the innate immune system has often been discussed in terms of recognition of large epitopes of repeating sugar units. In host immune systems, preformed receptors in the absence of antigen, including immunoglobulins, perform duties of surveillance and defense. Their germline genes are encoded with some redundancy in the targets (that are common in bacteria such as PAMPs) which is necessary if a pathogen is able to compromise a specific response pathway.

In this study, specific interactions between receptor molecules of the innate immune system with individual sugars have been revealed by X-ray crystallography reports. These studies allow us to compare the atomic details of specificity, the mode of binding and the structural features of the carbohydrates that are recognized.

3.1 T-lymphocyte dependence and germline conservation

As mentioned is section 1.1.3.5, combinatorial genetic rearrangements of V, D and J genes in the production of light and heavy chains greatly add to the diversity of binding sites. Further non-template diversity is created by the joining mechanisms. The joining of a light and heavy chain completes the combining site. The structural information to allow selection between several epitopes is inherited in the germline repertoire by not only each individual gene but as well by the combination of genes. This is unique to the development of BcR and TcR and is not found in other PRRs previously mentioned.

The conservation of germline identity such as in S25-2 can be compared to the development of immunoglobulin production in the absence of T-lymphocyte help. Some sub-populations of B-lymphocytes do not require T-lymphocyte activation (and do not undergo affinity maturation) to be effective against pathogens. The characterisation of BcR immunoglobulins from CD5+ B1 lymphocytes and marginal zone B-lymphocytes in the spleen has shown germline conservation. Also common to T-lymphocyte independent production of immunoglobulin, is the presence of junctional diversity observed as deletions and non-template N and P insertions (Boffey *et al.,* 2004). A common target of these natural immunoglobulins is LPS. By cross-linking many BcR, LPS can induce the production of soluble immunoglobulins. Given that this process is T-lymphocyte independent, smaller titres of immunoglobulins are produced, and there is no development of immune memory for subsequent challenges without helper T-lymphocytes. As well, immunoglobulin class-switching is restricted and mainly IgM are secreted.

The ability of germline immunoglobulins to recognize more than just the cognate antigen has been documented, such as in the anti-LPS response to *Campylobacter jejuni*

(Boffey *et al.,* 2004). T-lymphocyte independent activation was indicated by mainly germline IgM and a few IgG3 with the absence of other sub-classes of IgG. The immunoglobulins were shown to bind actin, thyroglobulin, tubulin, ssDNA and dsDNA. This increased binding to multiple epitopes was observed more by IgM compared to IgG, due to an avidity effect of the IgM pentamer.

Although T-lymphocyte dependent pathways can lead to increased sequence diversity through somatic hypermutation in the proliferation B-lymphocyte clones, at a phenotypic level, not all mutations away from germline result in a more effective response towards antigen and some mutations can increase the range of epitopes bound while others limit the number of bound epitopes. Clones that effectively show increased antigen affinity are selected for. Within these affinity matured immunoglobulins, the combination of point mutations with the conservation of key germline residues form the basis of antigen recognition (Tomlinson *et al.,* 1996).

Considering the production of IgG S25-2 using synthetic Kdo-BSA glycoconjugates likely through a T-lymphocyte dependent pathway with class switching, it is remarkable that the germline identities are significantly conserved. There have been other reports of affinity matured immunoglobulins that have retained germline identity and are effective during an immune challenge (Roost *et al.,* 1995). Roost *et al.* revealed an early quick neutralizing response by immunoglobulins to a glycoprotein epitope from vesicular stomatitis virus. Initially, early immunoglobulins that retained germline identity were produced. Interestingly, those produced later during the immunization also included some that had retained germline identity and others with some mutations. All immunoglobulins recognized the same epitope and with no improvements in affinity over time.

There have been many complexes reported of immunoglobulins bound to ligands such as proteins and lipids, but to date there have only been four structural reports of antibody-carbohydrate complexes whose structures have been determined in three-dimensions (Table 3.1). These reports examine largely unrelated epitopes with affinity-matured antibodies with low conservation of germline identity. The lack of comparison to their corresponding germline sequences limits the understanding of binding adaptability encoded by inherited germline sequences, and also how affinity maturation has altered specificity.

Table 3.1 Previously determined antibody-carbohydrate complexes.

Antibody	Bacterial source and ligand	Resolution	Reference
Se-155-4	***Salmonella* serogroup B O-antigen,** Fab with αD-Gal[αD-Abe]-αD-Man	**2.05Å**	**(Rose *et al.*, 1993)**
BR96	**Lewis Y antigen,** MBR96 Fab and cBR96 Fab with nonoate methyl ester derivative of Lewis Y αFuc- (1→2)- βGal-(1→4)-(αFuc-[1→3])-GlcNAc	**2.8, 2.5Å**	**(Jeffrey *et al.*, 1995; Sheriff *et al.*, 1996)**
S-20-4	***Vibrio cholerae* O1 serotype Ogawa O-polysaccharide,** Unliganded, Fab with monosaccharide D-perosamine [4-amino-(4→6)-dideoxy-D-mannose], Fab with disaccharide αD-perosamine-(1→2)-αD-perosamine	**2.3Å** **2.3Å** **2.8Å**	**(Villeneuve *et al.*, 2000)**
SYA/J6	***Shigella flexneri* O-polysaccharide** Unliganded, Fab with BC*D αL-Rha*p*-(1→3)-αL-2-deoxy-Rha*p*-(1→3)-βD-Glc*p*NAc, Fab with ABCDA' →2)-αL-Rha*p*-(1→2)-αL-Rha*p*-(1→3)-αL-Rha*p*-(1→3)-βD-Glc*p*NAc-(1→2)-αL-Rha*p*	**2.8Å** **2.3Å** **2.5Å**	**(Vyas *et al.*, 2002)**

3.2 Germline identity in S25-2 and S45-18

A comparison of the polypeptide sequences (Fig. 3.1) of S25-2 with known murine germline sequences shows that S25-2 corresponds to the germline genes Ig V_κ8-21 (GenBank nucleotide sequence accession code Y15982), J_κ2 (V00777) which form the variable light chain, and Ig V_H7S3 (J00525), J_H3 (V00777), and D_H SP2.9 (13199) which form the variable heavy chain (IMGT, the international ImMunoGeneTics database http://imgt.cines.fr, initiator and coordinator: Marie-Paule Lefranc, Montpellier, France) with only three mutations.

S45-18 uses the same V_κ8-21 and J_κ2, as well as V_H7S3 germlines as S25-2. However, the identity of the S45-18 D_H region corresponds to either of two related germline genes. S45-18 also differs from S25-2 by using J_H4. As a result the CDR H3 between the two immunoglobulins share neither length nor sequence homology.

In general, mutations occurred only in the residues flanking the germline genes. The mutation in CDR L3 in both S25-2 and S45-18 at Tyr to Arg L95 is the first residue encoded by the J_κ2 gene. In S25-2, there is a mutation (TAT to CAT) corresponding to germline Tyr to His H96. Also flanking the D_H region are Asp H95 and Glu H100a that are likely non-template insertions. The other two mutations in CDR H3, Trp to Arg H100b (TGG to CGG), and Ala to Ser H100d (GCT to TCT), are at the beginning of the J_H3 mini-gene. In S45-18, there are three insertions Asp H95 (common with S25-2), Ile H95a and Tyr H95b before the D_H region, and one insertion Asp H100a after it. There are two deletions at the beginning of J_H4 region, and a mutation Ala to Gly H100b (GCT to GGT).

S45-18 contains seven mutations from the germline in the light chain, including five mutations in the kappa light variable (V_κ) gene with three mutations in the framework

regions at Ser to Phe L5, Ser to Phe L43, and Thr to Ser L76, and two in the V_κ CDR regions at Asn to Ser L31, and Ser to Ala L52. In the S45-18 J_κ region, there are two mutations. The mutation Tyr to Arg L95 (similar to S25-2, at the beginning of $J_\kappa 2$) is found in CDR L3, and another, Lys to Ile L106, at the last residue of the J_κ gene in the framework region.

There are ten mutations from germline in the S45-18 heavy chain (and four non-template residue insertions flanking the D_H region). There are four mutations in the framework regions (all in the V_H region) at Lys to Ile H3, Ser to Pro H14, Ala to Ser H23, and Ala to Val H93. The remaining six mutations and four insertions are in the H CDR regions: Ala to Pro H52c, Asn to Lys H52d, and Pro to Ala H59 in CDR H2 contained in the V_H region; in the D_H region, there are the insertions Asp H95, Ile H95a and Tyr H95b at the beginning of the D_H region and Asp H100a at the end, with the mutations Tyr to Phe H97 and Tyr/His to Gly H98; and one in the $J_H 4$ gene at Ala to Gly H100b.

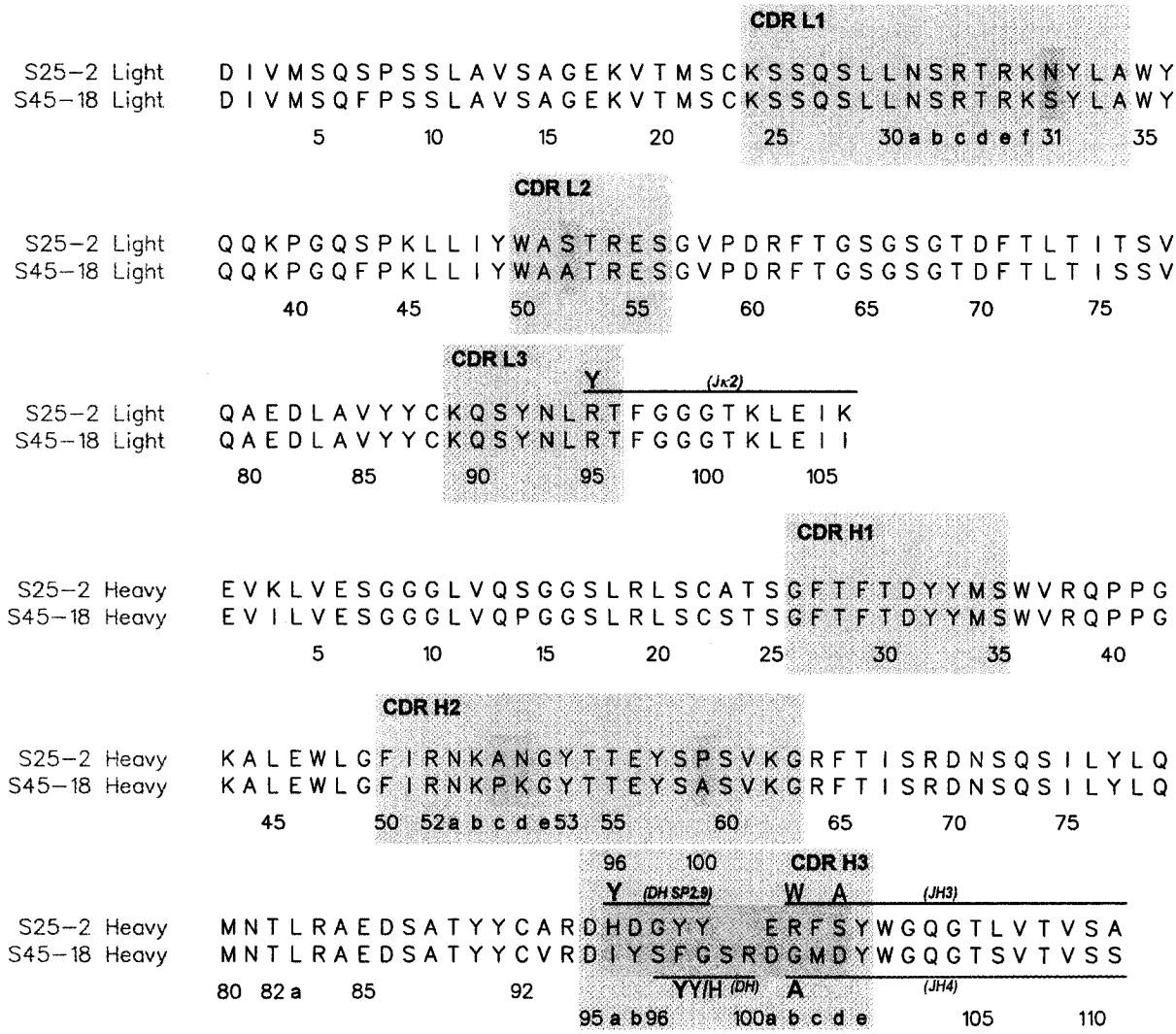


Figure 3.1 Amino acid sequence alignment of the variable domains of S25-2 and S45-18, numbered according to Chothia (Chothia & Lesk 1987; Chothia *et al.*, 1992; Al-Lazikani *et al.*, 1997) and the predicted CDR regions (*grey box*) based on sequence. S45-18 shares V_κ and V_H genes with S25-2 with differences highlighted in *yellow*, and mutations from germline in S25-2 are indicated in *red*. The names of the D and J genes are indicated and their regions marked in S25-2 (*red bar*) and in S45-18 (*blue bar*) and residues indicated above (in *red* for S25-2) and below (in *blue* for S45-18) are the germline residues.

3.3 *Crystal morphology and space group differ in the presence of ligand*

The Fab from mAb S25-2 has been crystallized liganded to the α(2→8)-α(2→4) Kdo trisaccharide, the α(2→8) Kdo disaccharide, α(2→4) Kdo disaccharide, Kdo monosaccharide (Fig. 1.12), and in two unliganded forms. The Fab from mAb S45-18 has been crystallized liganded to a pentasaccharide that contains the α(2→4)-α(2→4) Kdo trisaccharide (Fig. 1.10) as well as in the unliganded form.

For both S25-2 and S45-18 Fab, changes were observed in the morphology and space group of the crystals of the unliganded crystals and of those grown in the presence of antigen, as previously mentioned (Section 2, Table 2.1, 2.2, 2.3).

S25-2 Fab grown in the absence of antigen exhibited different forms. Although various crystals of unliganded S25-2 were grown in many forms, crystals of diffractable quality were typically improved with the presence of Zn^{2+} and Mg^{2+} ions. These crystals are described by the space group $P2_12_12$. Crystals of S25-2 grown in the presence of antigen were observed to be of similar form and with Mg^{2+} ions grew in the space group $P2_12_12_1$.

Crystals of S45-18 in absence of antigen grew in various different screening conditions. One condition was chosen based on visual inspection under a polarizing light microscope (Leica). The form collected exhibited C2 space group symmetry with two molecules of Fab per asymmetric unit. S45-18 in presence of its carbohydrate antigen αKdo-(2→4)- αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate grew only in one condition among those found in the unliganded conditions. This form had a $P2_12_12_1$ space group with two molecules of Fab per asymmetric unit.

3.4 Conservation of overall structure

The use of the same V_K and V_H germline genes accounts for the same overall conformation, association of light and heavy chains, and positioning of the CDR loops in the two immunoglobulins. Both antibodies possess the characteristic immunoglobulin fold resembling a "sandwich" of two β-sheets of anti-parallel strands in a Greek-key connectivity. Each light and heavy chain is composed of a 9-stranded beta-sheet in the variable domain, and a 7-stranded beta-sheet in the first constant domain. There is no difference between the hydrogen bonding pattern of the β-sheets in the liganded and unliganded structures of S25-2 and S45-18 (Appendix Table A.2).

Superposition of the light chain (variable and constant domains) reveals no significant differences in the position of the heavy chains (Fig 3.2A) of the two antibodies. Superposition of the light constant domain found a good fit between the light and heavy constant domains of all structures. Interestingly, while the S25-2 structures did not exhibit any significant differences in the variable domain upon ligation, the S45-18 structures differed in the orientation of the variable domain in the presence of ligand, across the elbow region (Fig 3.2B). However, superposition of the framework residues in the variablfe domain shows that both light and heavy variable domains maintain their association upon ligation in both S25-2 and S45-18. Additionally, the relative arrangement of the complementarity determining loops (CDRs) that form the combining site (CDRs L1, L3, H1, H2, and H3) is conserved (Fig 3.2C, 3.2D).

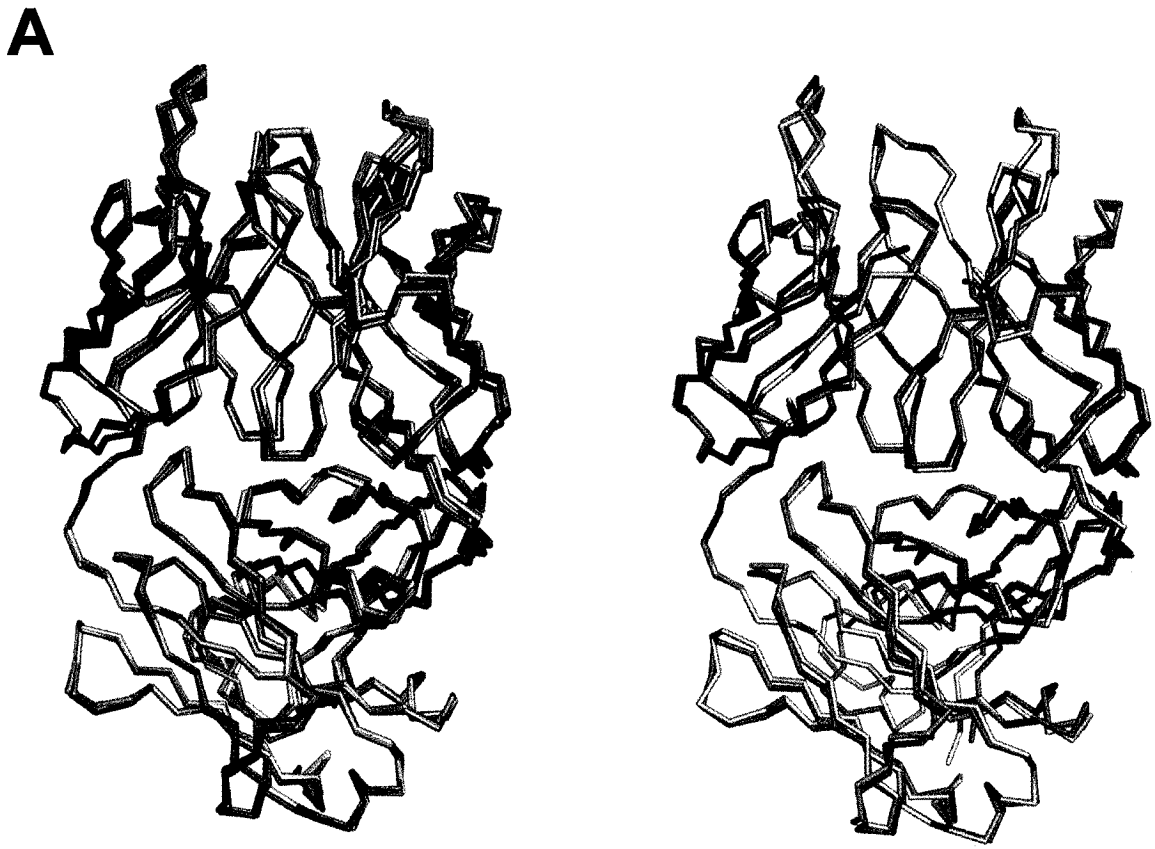


Figure 3.2 A) Superposition of the light chain variable and constant domains. The Fab backbone Cα trace (antigens removed for clarity) of S25-2 (*left*) complexed with αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*), and unliganded form 1 (*light blue*), unliganded form 2 (*dark blue*), and S45-18 (*right*) complexed with the pentasaccharide (*green*) and unliganded (*white*).

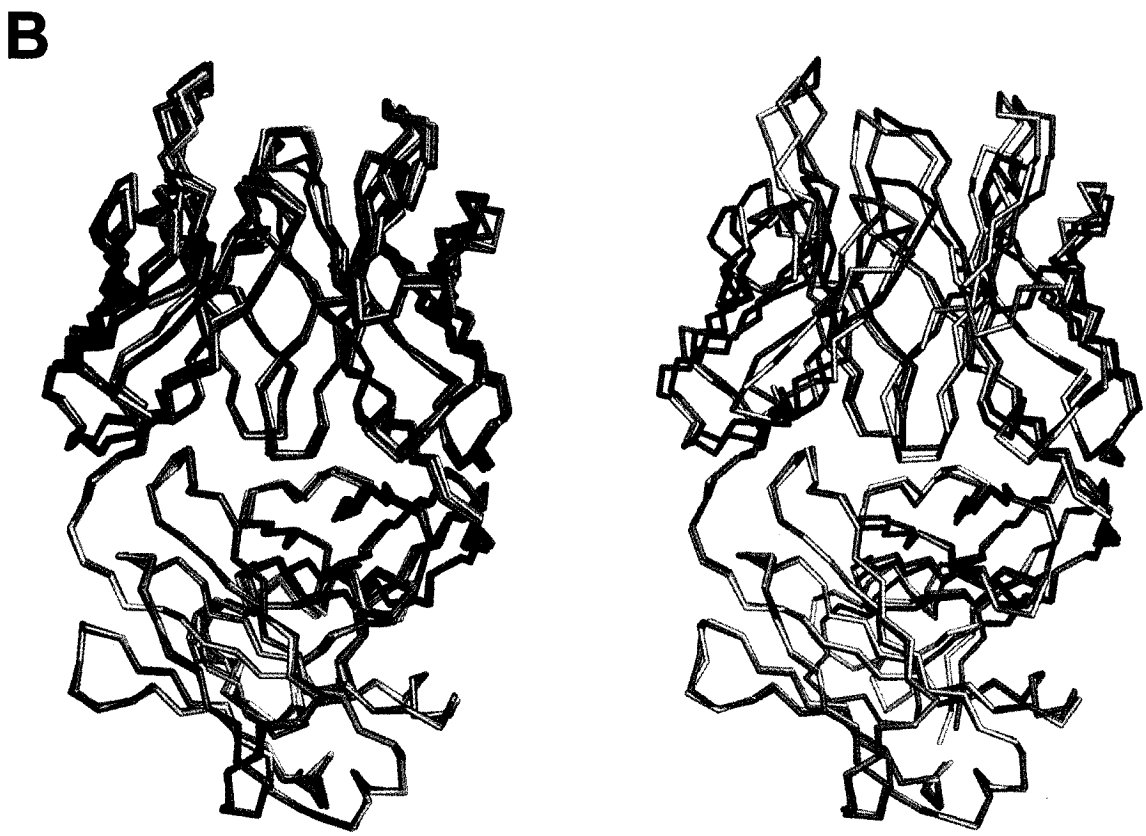


Figure 3.2 **B)** The superposition of the light constant domains reveals differences in the orientation of the entire S45-18 variable domain upon binding. The Fab backbone Cα trace (antigens removed for clarity) of S25-2 (*left*) complexed with αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*), and unliganded form 1 (*light blue*), unliganded form 2 (*dark blue*), and S45-18 (*right*) complexed with the pentasaccharide (*green*) and unliganded (*white*).

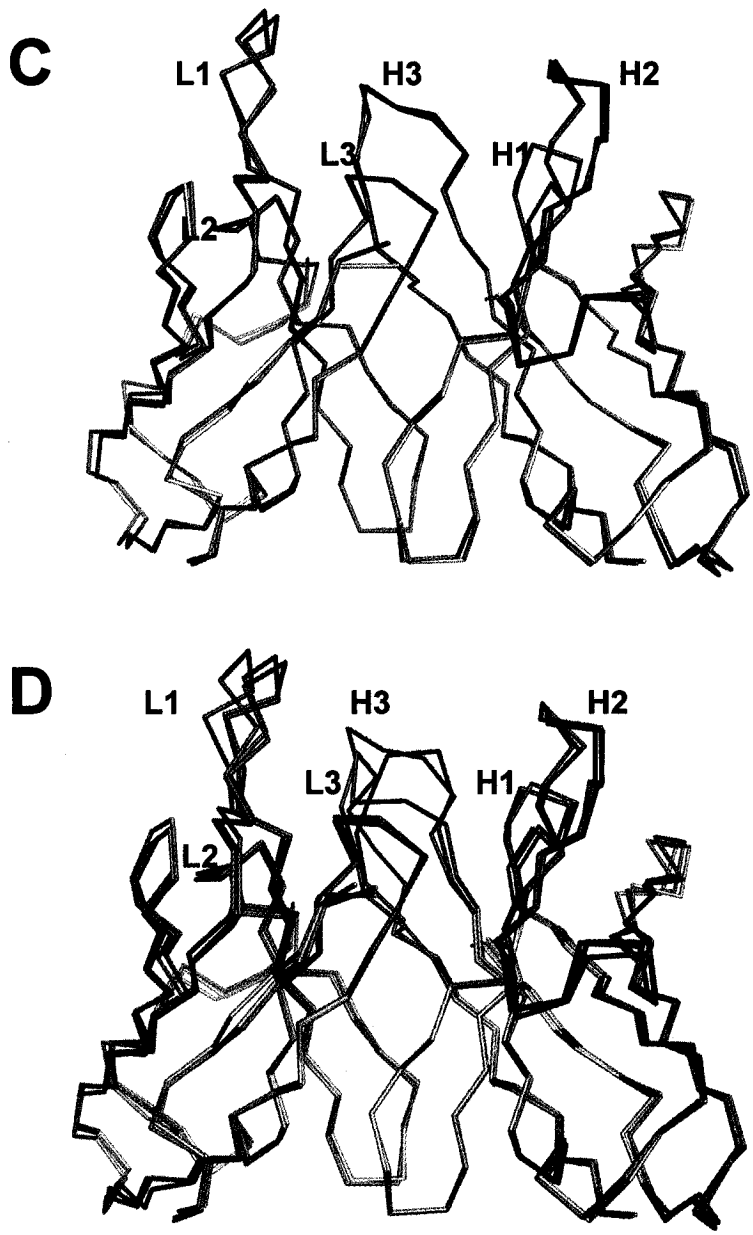


Figure 3.2 **C)** Superposition of the framework regions in the variable domain. Highlighted are the light *(pink)* and heavy (blue) CDR loops. The shift in the variable region of S45-18 does not alter the association of the light and heavy chain CDRs shown here in the unliganded and liganded forms. **D)** Comparison of the liganded S45-18 with the unliganded and liganded S25-2 structures with the framework regions superpositioned with good agreement. The Fab backbone Cα trace (antigens removed for clarity) of S25-2 (*bottom*) complexed with αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*), and unliganded form 1 (*light blue*), unliganded form 2 (*dark blue*), and S45-18 (*top*) complexed with the pentasaccharide (*green*) and unliganded (*white*).

3.5 Canonical conservation of CDRs L1, L2, L3, and H1, H2

Six of the loops connecting the β-strands form the CDRs, and investigations have analyzed the sequence-structure relationships of the CDRs by different approaches. Before many structural reports appeared, Kabat's group (Potter 1983) identified the CDRs based on residues that exhibited the highest variability. Later, Chothia (Chothia & Lesk 1987; Chothia *et al.,* 1992; Tomlinson *et al.,* 1995; Al-Lazikani *et al.,* 1997) analyzed the structures of the CDRs of the available structurally-characterized Fabs and classified them based on the torsion angles and hydrogen bonding patterns that resulted in distinct classes of structures named "canonical" conformations. The identity of a residue was found to effect the canonical conformation particularly where there was an interaction involving the side chain. As sequence identity has been the basis of conformation prediction, similar sequences often share the same canonical structure. Specifically, it has been observed that members of the same canonical form for each CDR often have germline residues conserved during affinity maturation (Tomlinson *et al.,* 1995). Nakamura's group made contributions to predicting the structure of CDRs as well (Shirai *et al.,* 1996), in particular the CDR H3 which has shown to possess the most variable sequence and length and to date has the least amount of predictability.

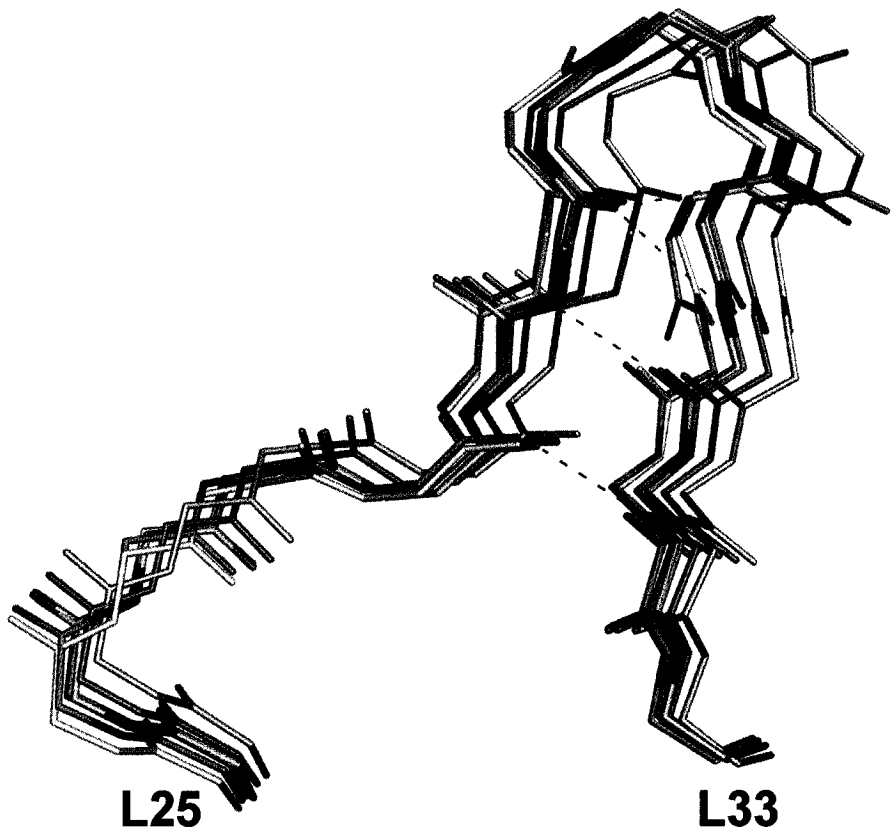
Chothia's and Nakamura's guidelines based on the variable domain sequences of S25-2 and S45-18 (Fig. 3.1) generally apply, although an overall change in the canonical conformation occurs in the S25-2 CDR H3 region, but not S45-18.

The first CDR of the light chain (CDR L1) from S25-2 (and similarly from S45-18) begins at residue Lys L24, located after the disulphide bridge between CDR L1 Cys L23 and Cys L88 of an adjacent β-strand. This disulphide bridge occurs at the base of CDR L1,

while some flexibility in position is observed at the tip (Fig. 3.3), particularly in the unliganded S25-2 structures. Chothia predicts that CDR L1 typically ends before a Trp, at Ala L34 in the case of S25-2 and S45-18. S25-2 identifies with the germline sequence in CDR L1; however, S45-18 has a Asn L31 to Ser mutation. For both S25-2 and S45-18 in the absence and presence of ligand, CDR L1 belongs to the κ class 3 of CDR L1 canonical structures.

CDR L2 has been predicted to begin typically 16 residues after L1 (but 15 residues in the case of both S25-2 and S45-18, (Fig. 3.4), in general preceded by Ile-Tyr, but also has been shown in other antibodies with Val-Tyr, Ile-Lys or Ile-Phe. It has been previously reported as a shorter CDR, comprising generally 7 residues, and in S25-2 and S45-18 ending with Ser L56. CDR L2 in S25-2 and S45-18 are classified as class 1 canonical conformations.

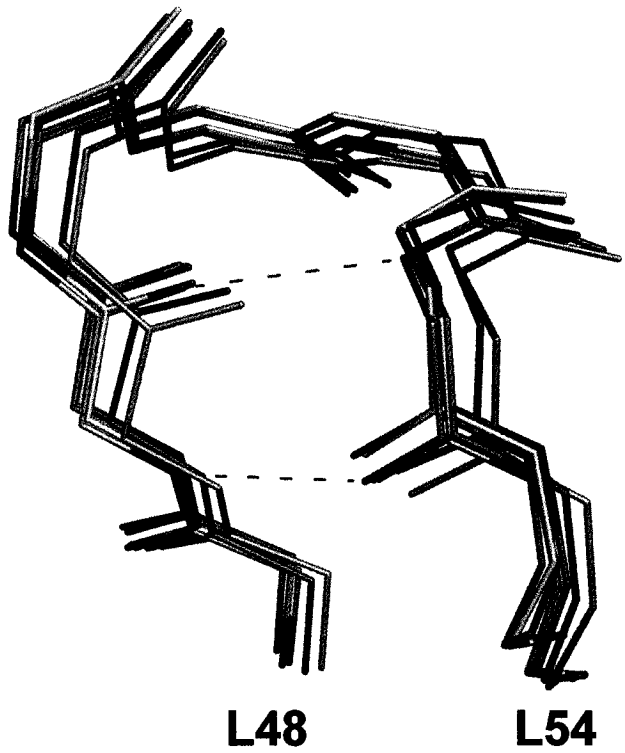
CDR L3 has been reported to begin after a Cys residue, that is Cys L88 in S25-2 and S45-18 (in a disulphide bridge interaction with Cys L23) (Fig. 3.5) and ends before Phe-Gly-*XXX*-Gly (where *XXX* represents any amino acid residue), at Thr L96. In both S25-2 and S45-18, there is a mutation from germline Tyr L95 to Arg as the first residue contributed by the J_H2 gene. The CDRs L3 in both immunoglobulins most closely resemble a class 3 canonical conformation.



Hydrogen bond pairs:

29	O ---	32 N
30a	N ---	30f O
30a	O ---	30e N
30a	O ---	30f N

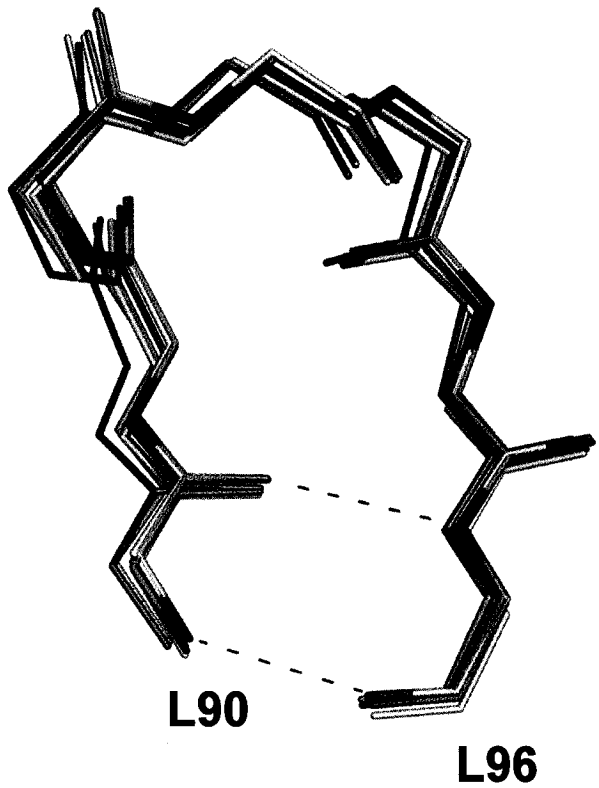
Figure 3.3 Comparison of the positions of CDR L1, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well. However at the tip of CDR L1, the unliganded form 1 (*light blue*), and unliganded form 2 (*dark blue*) do not overlap well. The forms of S45-18 liganded (*green*) and unliganded (*white*) overlap with each other, and only differ slightly with those of liganded S25-2. The CDR L1 canonical κ class 3 and hydrogen bonding is conserved in all structures.



Hydrogen bond pairs:

49	N	---	53 O
49	O	---	53 N

Figure 3.4 Comparison of the positions of CDR L2, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with in αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well with the two forms of S45-18 liganded (*green*) and unliganded (*white*). The differences in the unliganded S25-2 form 1 (*light blue*), and unliganded form 2 (*dark blue*) are not as pronounced as they are in CDR L1. CDR L2 shows a canonical class 1 in all structures.

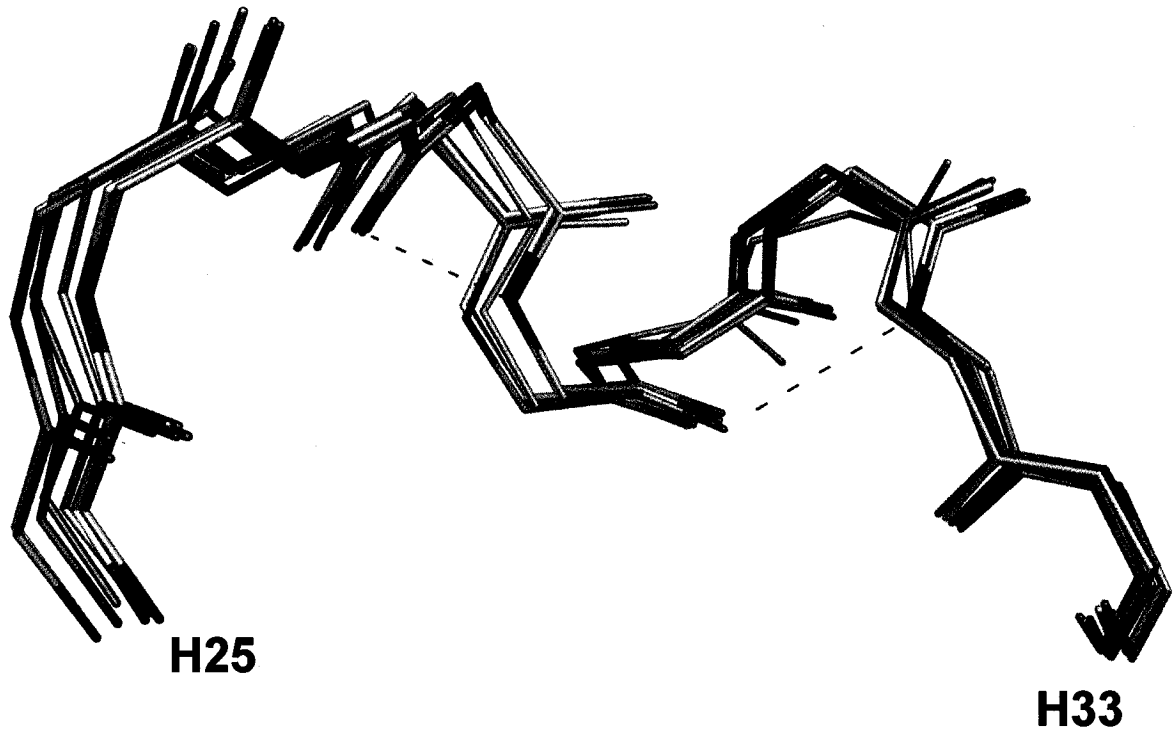


Hydrogen bond pairs:

90	N	---	96 O
90	O	---	96 N

Figure 3.5 Comparison of the positions of CDR L3, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with in αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well with the two forms of S45-18 liganded (*green*) and unliganded (*white*), and with the S25-2 unliganded forms 1 (*light blue*) and unliganded form 2 (*dark blue*). CDR L3 is a class 3 canonical structure, and like CDR H1 is identical in sequence in both S25-2 and S45-18, and has the greatest structural similarity of all the CDRs.

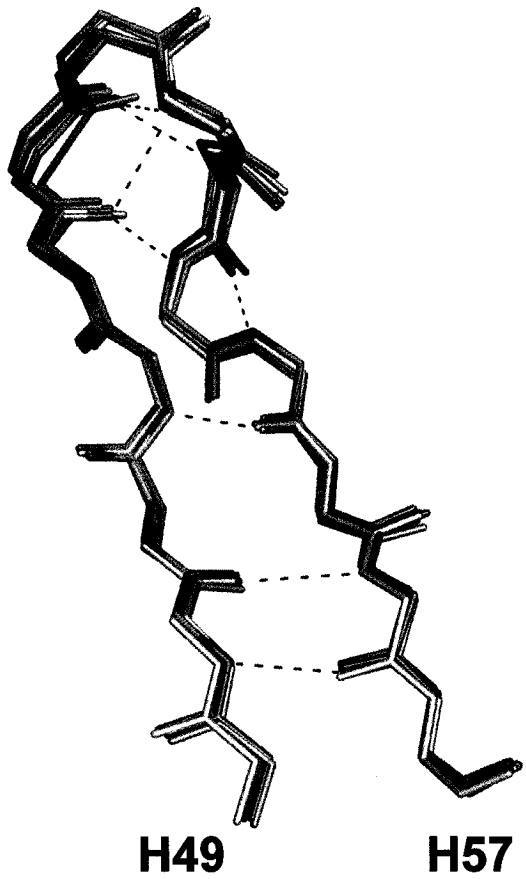
CDR H1 from both antibodies begins after Cys-*XXX-XXX-XXX* at Gly H25. It typically ends before Trp-Val, Trp-Ile or Trp-Ala, and in both S25-2 and S45-18 the last residue of CDR H1 is Ser H35. CDR H1 is a class 1 canonical conformation (Fig. 3.6). A hydrogen bond between Phe H29 and Tyr H32 maintains a kinked local structure, which maintains the adjacent Tyr H33 residue in a position that can allow for different interactions with the respective antigens of S25-2 and S45-18. CDR H2 has been reported to begin after Leu-Glu-Trp-Leu-Gly (with some other variations in sequence observed in other reports). CDR H2 is classified as a class 4 canonical conformation (Fig. 3.7). The conservation of the backbone position here is critical as there are germline mutations in S45-18. The mutation from S25-2 (and germline) Ala to Pro H52c in S45-18 occurs at the tip of this loop and does not affect the conformation. The mutation S25-2 (and germline) Asn to Lys H52d in S45-18 allows for increased specificity of its ligand.



Hydrogen bond pairs:

27	N	---	29 N
29	O	---	32 N

Figure 3.6 Comparison of the positions of CDR H1, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with in αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well with the two forms of S45-18 liganded (*green*) and unliganded (*white*), and with only slight differences with the S25-2 unliganded forms 1 (*light blue*) and unliganded form 2 (*dark blue*). This CDR H1 structure is class 1, and like CDR L3 is identical in sequence in both S25-2 and S45-18.



Hydrogen bond pairs:

50	N ---	56 O
50	O ---	56 N
52	N ---	54 O
52a	O ---	53 N
52a	O ---	52d N
52b	O ---	52e N
52b	O ---	52d N

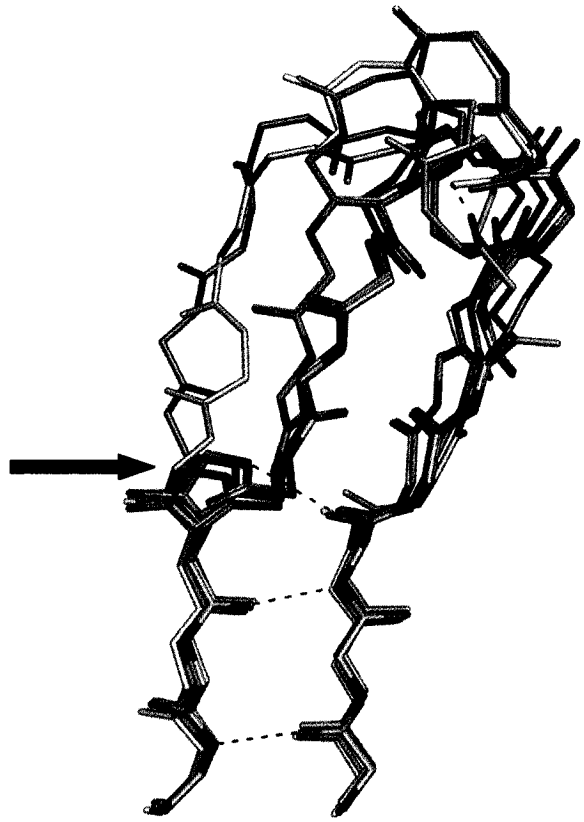
Figure 3.7 Comparison of the positions of CDR H2, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with in αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well with the two forms of S45-18 liganded (*green*) and unliganded (*white*), and the S25-2 unliganded form 1 (*light blue*) and unliganded form 2 (*dark blue*). This CDR H2 canonical class 4 structure is conserved in both S25-2 and S45-18 and contains differences in sequence at residues that interact with antigen located at the tip of the loop.

3.6 Different CDR H3 conformations in S25-2 and in S45-18

Generally, CDR H3 begins after about 100 residues from the N-terminus after Cys-*XXX*-*XXX*. In the germline and S25-2, this sequence is Cys-Ala-Arg whereas in S45-18 the sequence is Cys-Val-Arg. CDR H3 shows the most variability as a result of a rearrangement of three minigenes VDJ compared to CDR L3 only using the VJ genes. The main sequence contributors are the D and J genes, which are different for S25-2 and S45-18. The S25-2 CDR-H3 resembles its germline sequence (with only 3 mutations and 2 insertions at the V-D and D-J junctions, out of a total number of 11 residues) more than S45-18 to its germline sequence (3 mutations and 4 insertions at the junctions, out of 13 CDR H3 residues). The residues following CDR H3 are Trp-Gly-*XXX*-Gly, where residue *XXX* is Gln in the germline, S25-2 and S45-18. Nakamura's group proposed guidelines to predict the structure of CDR H3. The conformation of S45-18 CDR H3 does not change in the presence of antigen (Fig. 3.8), adopting a conformation that is predicted by Nakamura. However, Nakamura's predictions are only correct for the liganded forms of S25-2.

The mAb S25-2 displays a limited flexibility in the conformation of CDR H3 (and to a lesser extent L1) in the unliganded compared to liganded states. Electron density is more continuous in the liganded structures of S25-2, which defines a backbone conformation that is different than in the unliganded forms. The unliganded structures of S25-2 each crystallized in unique space groups, and showed only slightly different polypeptide backbone conformations for CDR H3; however, these were significantly different than that found in the liganded S25-2 structures.

The liganded CDR H3 structures in S25-2 and S45-18 are similar both with "kinked" polypeptide backbones at residue H100d and a conserved hydrogen bond (H94 O



H102 H92

Hydrogen bond pairs:

92 O	---	102 N
94 N	---	100e N
94 O	---	100d N (liganded S25-2 and both forms of S45-18)
97 O	---	100 N (unliganded S25-2 form 1)

Figure 3.8 Comparison of the positions of CDR H3, after superposition of the framework regions of the variable domain. The liganded forms of S25-2 with αKdo-(2→8)- αKdo-(2→4)-αKdo trisaccharide (*orange*), α(2→8) disaccharide (*yellow*), α(2→4) disaccharide (*magenta*), monosaccharide (*cyan*) overlap well with the two forms of S45-18 liganded (*green*) and unliganded (*white*) until the tip of CDR H3 where there is a two-residue insert in S45-18. There is conserved "kink" in the polypeptide backbone of the bound S25-2 and both forms of S45-18, with a conserved hydrogen bond (*arrow*). This kink is not present in the S25-2 unliganded form 1 (*light blue*) and unliganded form 2 (*dark blue*). This CDR H2 canonical class 4 structure is conserved in both S25-2 and S45-18 and contains differences in sequence at residues that interact with antigen located at the tip of the loop.

to H100d N). According to Nakamura (Shirai *et al.,* 1996), a kinked conformation at the base of CDR H3 in S45-18 arises from interactions between the residue preceding the first (0^{th}) in the CDR, specifically an Arg (H94), and one residue preceding the last (*n-1*) an Asp (H100d), whose side chains interact with each other in a salt bridge. In S25-2, the corresponding (*n-1*) residue is Ser H100d instead of Asp, and as such a salt bridge is not possible. In this case, Nakamura predicts a kinked conformation if there is a Trp at the (*n+1*) position (H101) that forms an interaction with the backbone carbonyl of the (*n-2*) residue which can be any residue. In the case of S25-2, the residue *(n-2)* is Phe H100c. Although these residues are present, CDR H3 is only canonical in the liganded form.

Between the canonical form and the unliganded S25-2, the polypeptide backbones show structural differences beginning at residue H96 and then converging at residue H100d. The side chain of the (*n-1*) residue Ser H100d is oriented away from that of the (0^{th}) residue Arg H94, and the (*n+1*) Trp H101 is too distant to interact with the (*n-2*) Phe H100c. In the binding of the Kdo monosaccharide to S25-2, the interaction of a single CDR H3 residue (*n-4*) Glu H100a with Kdo3 O4 is associated with reordering the remaining portion of CDR H3 (that exhibits high temperature factors), which includes the first residue Asp H95 to the (*n-1*) residue Ser H100d such that a canonical conformation is stabilized by intra-CDR interactions. This canonical form is identical to the remaining mono-, di- and trisaccharide S25-2 forms.

The polypeptide backbone between S45-18 and the bound conformation of S25-2 diverge after residue S25-2 Gly H98 and S45-18 H96 and become superimposable again at residue H100a in both antibodies. The tip of CDR H3 does not have a regular hydrogen-bonding pattern as suggested in Nakamura's predictions. Due to genetic rearrangement at the VDJ junctions, there are insertions in this region. The structural consequence of a two-

residue increase in S45-18 CDR H3 length allows Phe H97 to protrude into the binding pocket and plays a role in selecting the α(2→4)- α(2→4) Kdo trisaccharide. Other differences in specificity include a shift of the position of the terminal Kdo3 by 0.8 Å, related to the CDR H3 residue Glu H100a in S45-18 and Asp H100a in S25-2, and which is an insertion between the D and J genes in both immunoglobulins.

3.7 Germline conservation of a binding site for a common Kdo residue

The identical pairing of the V_κ and V_H genes leads to similar combining sites in both S25-2 and S45-18 where at the centre is located a common binding subsite for the terminal non-reducing Kdo3 residue. Kabat first proposed that the binding of a linear carbohydrate antigen involves binding sites that are formed from "grooves" and "cavities" (Kabat 1988). The conserved shape of the combining site in both antibodies features an "open" groove to accommodate the Kdo1 and Kdo2 residues with a cavity for the Kdo3 residue (Fig. 3.9).

Hydrogen bonding occurs between the polypeptide backbone and side chains either directly or through bridging water interactions with the -OH groups on C4, C5, ring-O6, C7, C8, and the glycosidic oxygen (Table 3.2). This pattern differs between different parts of the different antigens, with the exception of generally conserved H-bonding around the terminal Kdo3 residue (Fig. 3.10) compared to the remainder of the binding site (Fig. 3.11). It is remarkable that intreractions in the Kdo3 binding pocket are largely formed with germline residues from V_κ and V_H. As well, conserved germline residues of $V\kappa$ and V_H that do not contact the antigen include aromatic residues (discussed later) that aid in maintaining the position of the antigen-contacting residues.

Other reports have shown that it is not only the conservation of single genes, but the combinations of genes have been evolutionarily retained in the repertoire against specific types of antigen. For example, phosphatidylcholine, a common epitope on pathogens, elicits responses from particular VDJ germline genes with little non-encoded alterations. While both S25-2 and S45-18 use V_κ8-21 (Y15982) for the light chain and V_H7 S3 (J00525) for the heavy chain, a previous study of responses in mice against

phosphatidylcholine (Seidl *et al.,* 1999) revealed a predominance of the usage of germlines genes $V_\kappa 9$ and $V_\kappa 4$ in the light chain with $V_H 11$ and $V_H 12$ in the heavy chain. Another example is found in mAbs against GD3 epitopes, where there is predominant usage of the V_H 7183.3b gene (Boffey *et al.,* 2004).

Table 3.2 Interactions in the binding site, and with the non-reducing terminal Kdo3 indicated in bold.

	S25-2				S45-18		
Unique interactions	α(2→8)-α(2→4)[1]	α(2→8)[2]	α(2→4)[3]	Mono[4]	Penta[5]	Antigen Atom	Type of Interaction
Tyr L32 OH - WAT11	x	x	x	x		Kdo3 O7	H bond bridging water
Ser L91 O	x	x	x	x	x	Kdo3 O5	H bond
Tyr L92 O	x	x	x	x	x	Kdo3 O7	H bond
Leu L94 N – WAT 2	x	x	x	x	x	Kdo3 O5, O6	H bond bridging water
Arg L95 NH	x	x	x	x	x	Kdo3 O4	H bond
Arg H52 NH, Nε	x	x	x	x	x	Kdo3 O1	Salt bridge
Asn H52d (S25-2) Nδ – WAT24	x	x				Kdo2 O1, O6, O7	H bond
Asn H52d (S25-2) Oδ – WAT28	x	x				Kdo2 O7	Bridging water
Lys H52d (S45-18) NZ					x	Kdo2 O5, O6	H bond
His H96 (S25-2) Nε	x	x				Kdo2 O4, O5	H bond
Phe H97 (S45-18) Ring					x	Kdo3, Kdo2, Kdo1 rings	van der Waals
Gly H98 (S25-2) O	x	x				Kdo2 O4	H bond
Glu H100a (S25-2) Oε	x	x	x	x		Kdo3 O4	H bond
Glu H100a (S25-2) Oε – WAT22	x	x				Kdo2 O4	H bond bridging water
Asp H100a (S45-18) Oδ - WAT1					x	Kdo3 O4	H bond bridging water

[1] αKdo-(2→8)- αKdo-(2→4)- αKdo-(2→*O*)-allyl trisaccharide,
[2] αKdo-(2→8)- αKdo-(2→*O*)-allyl disaccharide,
[3] αKdo-(2→4)- αKdo-(2→*O*)-allyl disaccharide.
[4] αKdo monosaccharide,
[5] αKdo-(2→4)- αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate

	S25-2				S45-18		
Multiple usage interactions	α(2→8)-α(2→4)[1]	α(2→8)[2]	α(2→4)[3]	Mono[4]	Penta[5]	Antigen Atom	Type of Interaction
Asn L30a Nδ					x	Kdo1 O1	H bond
Nδ - WAT3	x	x	x	x	x	Kdo3 O7	H bond via bridging water
Nδ – WAT20	x	x		x		Kdo3 O8	H bond via bridging water
Arg L30c NH	x					Kdo1 O5	H bond
			x			Kdo2 O1	Salt bridge
NH2 - WAT11	x	x	x	x		Kdo3 O7	H bond via bridging water
						Kdo1 O1	Salt bridge
Nε					x	Kdo1 O5, O6	H bond
Tyr H33 OH	x	x				Kdo3 O1, Kdo2 O7	H bond
			x		x	Kdo3 O1, Kdo2 O5	H bond
				x		Kdo3 O1	H bond
Asn H52d (S25-2) Oδ	x	x				Kdo2 O7	H bond
			x			Kdo2 O7 (different position)	H bond
Asn H52d (S25-2) Nδ – WAT23	x	x				Kdo2 O7	H bond via bridging water
			x			Kdo2 O5	

[1] αKdo-(2→8)- αKdo-(2→4)- αKdo-(2→*O*)-allyl trisaccharide,
[2] αKdo-(2→8)- αKdo-(2→*O*)-allyl disaccharide,
[3] αKdo-(2→4)- αKdo-(2→*O*)-allyl disaccharide.
[4] αKdo monosaccharide,
[5] αKdo-(2→4)- αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate

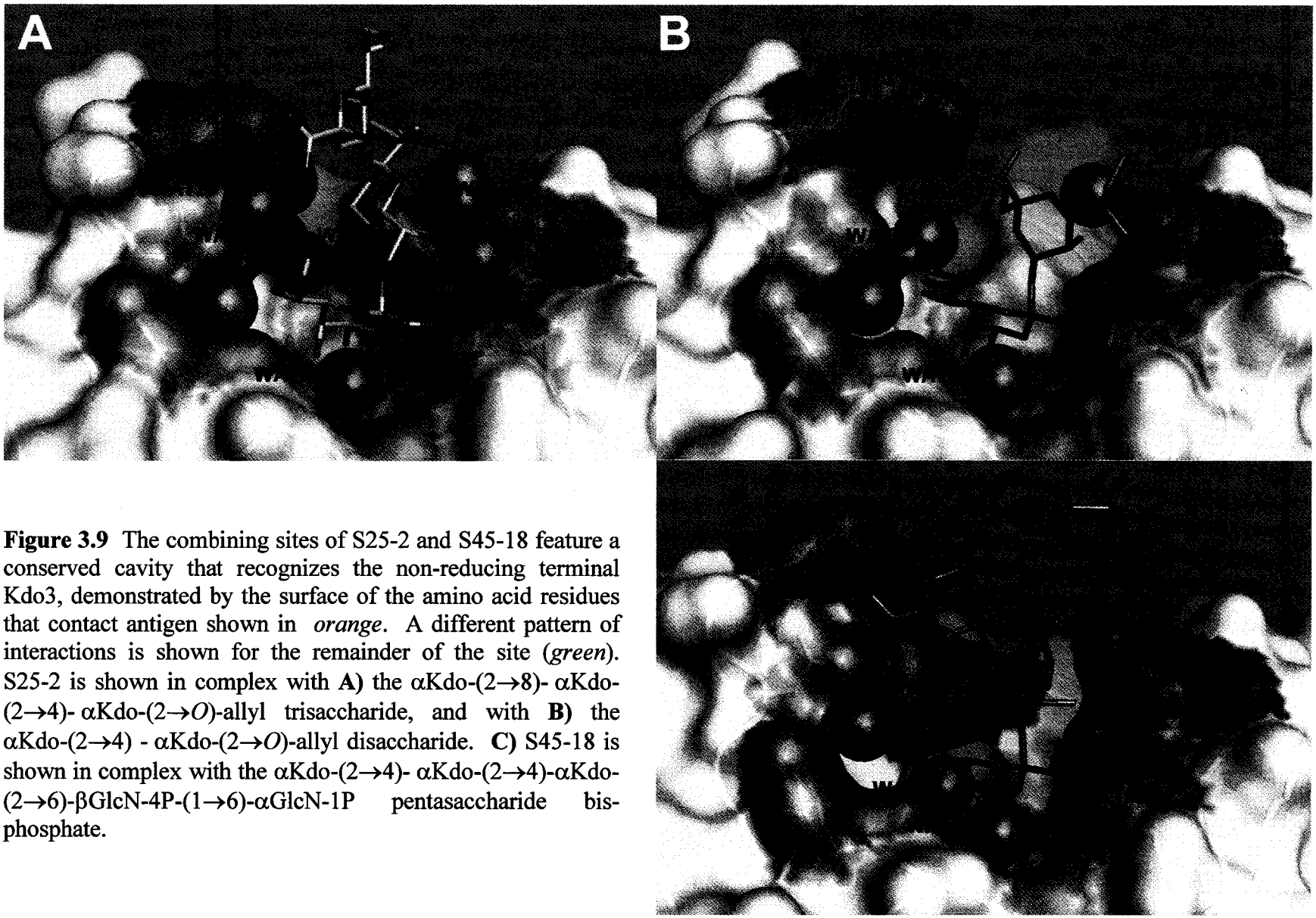
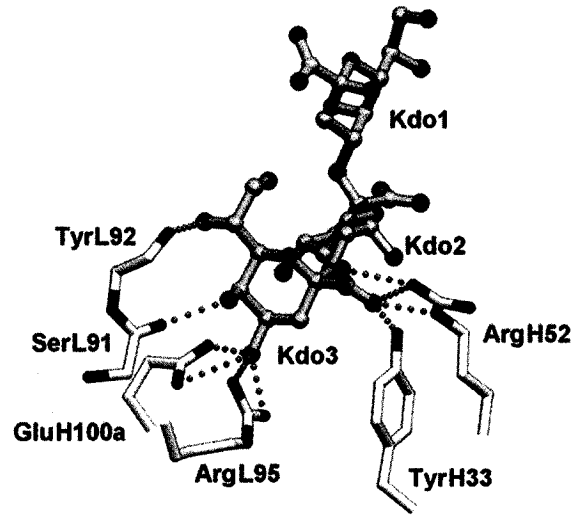
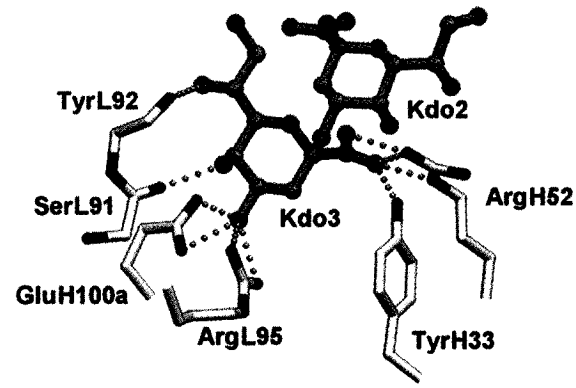


Figure 3.9 The combining sites of S25-2 and S45-18 feature a conserved cavity that recognizes the non-reducing terminal Kdo3, demonstrated by the surface of the amino acid residues that contact antigen shown in *orange*. A different pattern of interactions is shown for the remainder of the site (*green*). S25-2 is shown in complex with **A)** the αKdo-(2→8)- αKdo-(2→4)- αKdo-(2→*O*)-allyl trisaccharide, and with **B)** the αKdo-(2→4) - αKdo-(2→*O*)-allyl disaccharide. **C)** S45-18 is shown in complex with the αKdo-(2→4)- αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bis-phosphate.



S25-2



S45-18

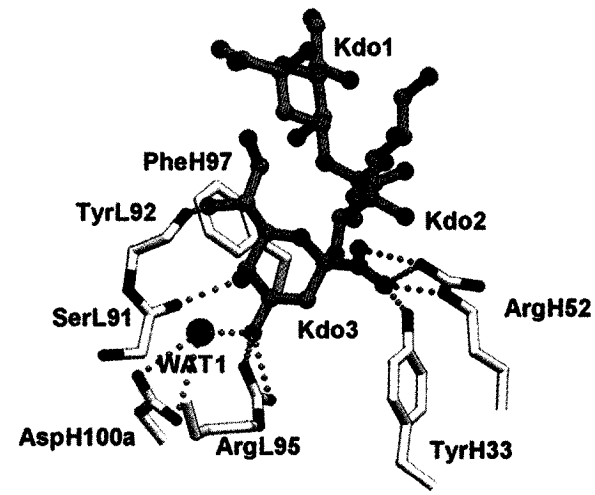


Figure 3.10 The binding environment surrounding the respective non-reducing terminal Kdo3 of the different antigens is conserved in both immunoglobulins. Shown are S25-2 bound to the α(2→8)-α(2→4) trisaccharide (*orange*), α(2→4) disaccharide (*magenta*), and S45-18 bound to the α(2→4)-α(2→4) trisaccharide portion of the pentasaccharide antigen (*green*).

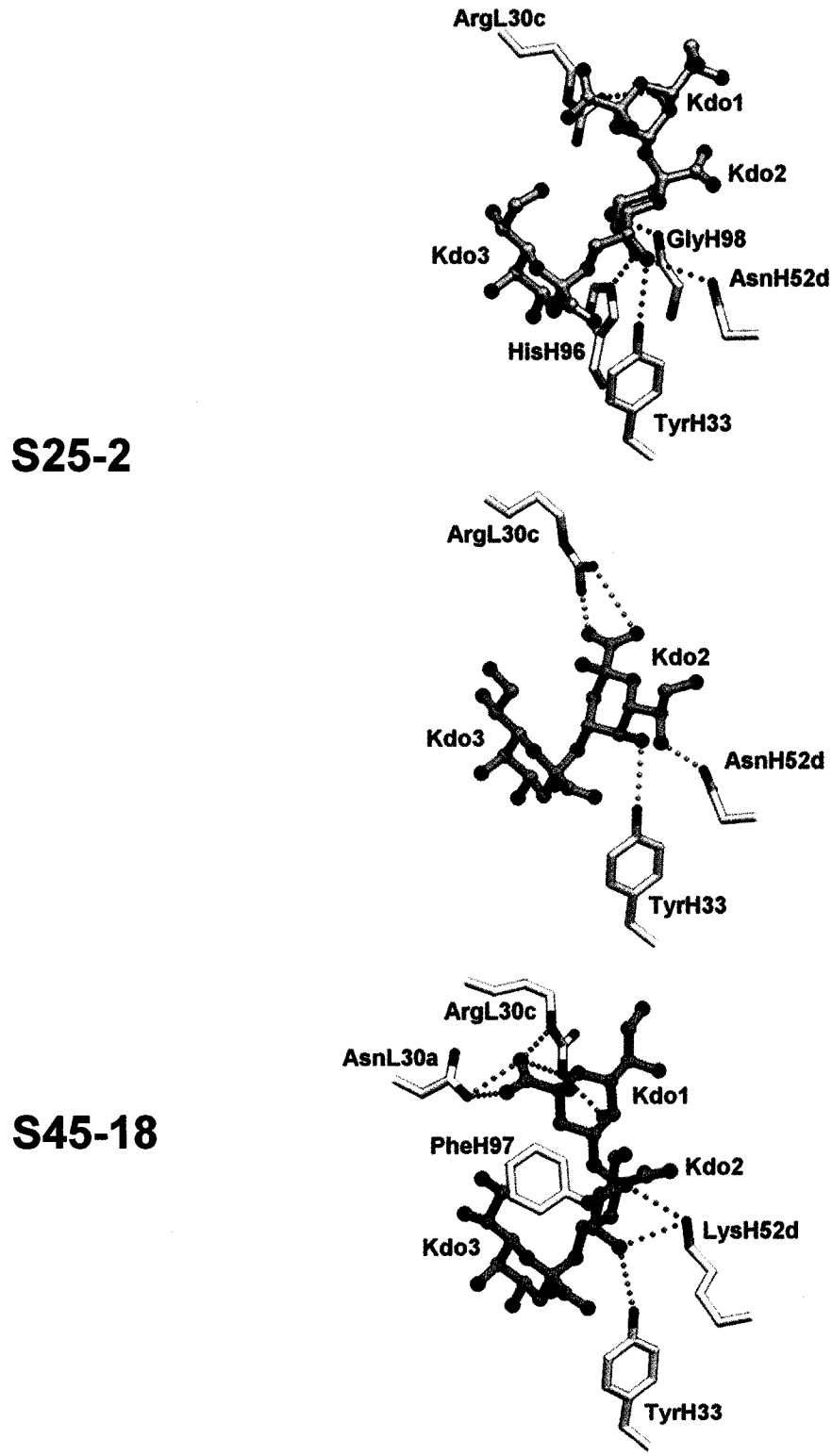


Figure 3.11 The remainder of the binding site is able to accommodate epitopes that exhibit different structures. While the environment of Kdo3 uses largely the same interactions, Kdo2 and Kdo1 are recognized by a combination of unique residues and the multiple usage of the same residues. *The antigens are coloured as in Figure 3.10.*

4.0 Differences in germline usage and mutations

S25-2 and S45-18 serve as examples of the conservation of germline identity and mutations that play roles in modulating the specificity of antigen. Aspects of their adaptability and selectivity are revealed through comparison of germline sequences, point mutations, conformational change and multiple usage of residues.

4.1 Distribution of conservation and mutation in the combining site

Examination of the residues that line the combining site of S25-2 and S45-18 reveals that germline conserved residues are found in the centre of the binding site, largely accounting for Kdo3 recognition, while point mutations (particularly in S45-18) are found more toward the periphery (Fig. 4.1).

Tomlinson *et al.* (1996) have previously reported on the distribution of germline and mutated residues. Their analysis involved immunoglobulin sequence diversity from two origins: firstly, from different inherited germline genes, and secondly, from the mutations occurring during somatic replication and affinity maturation. The variation in germline identity across *different* genes is greatest in residues that make contact with antigen lining the centre of the binding site. Germline residues that were found to be variable but not involved in interactions with antigen may have roles in induced fit mechanisms, which has been previously observed (Wedemayer *et al.,* 1997). Tomlinson also reported that variation by somatic mutation is found in residues that are at the periphery of the binding site. As well from Tomlinson, non-template additions were found at the periphery of the binding site.

Previously, somatic mutations have also been shown to be important in both lock-and-key and induced fit binding. Wedemayer found that somatic mutations in affinity-matured antibodies also helped organize a distant binding site, whereas Foote & Milstein (1994) through the use of kinetic binding experiments revealed that certain somatic mutations (as well as conserved germline residues) occurred more frequently than others in immunoglobulins that sample several conformations in solution.

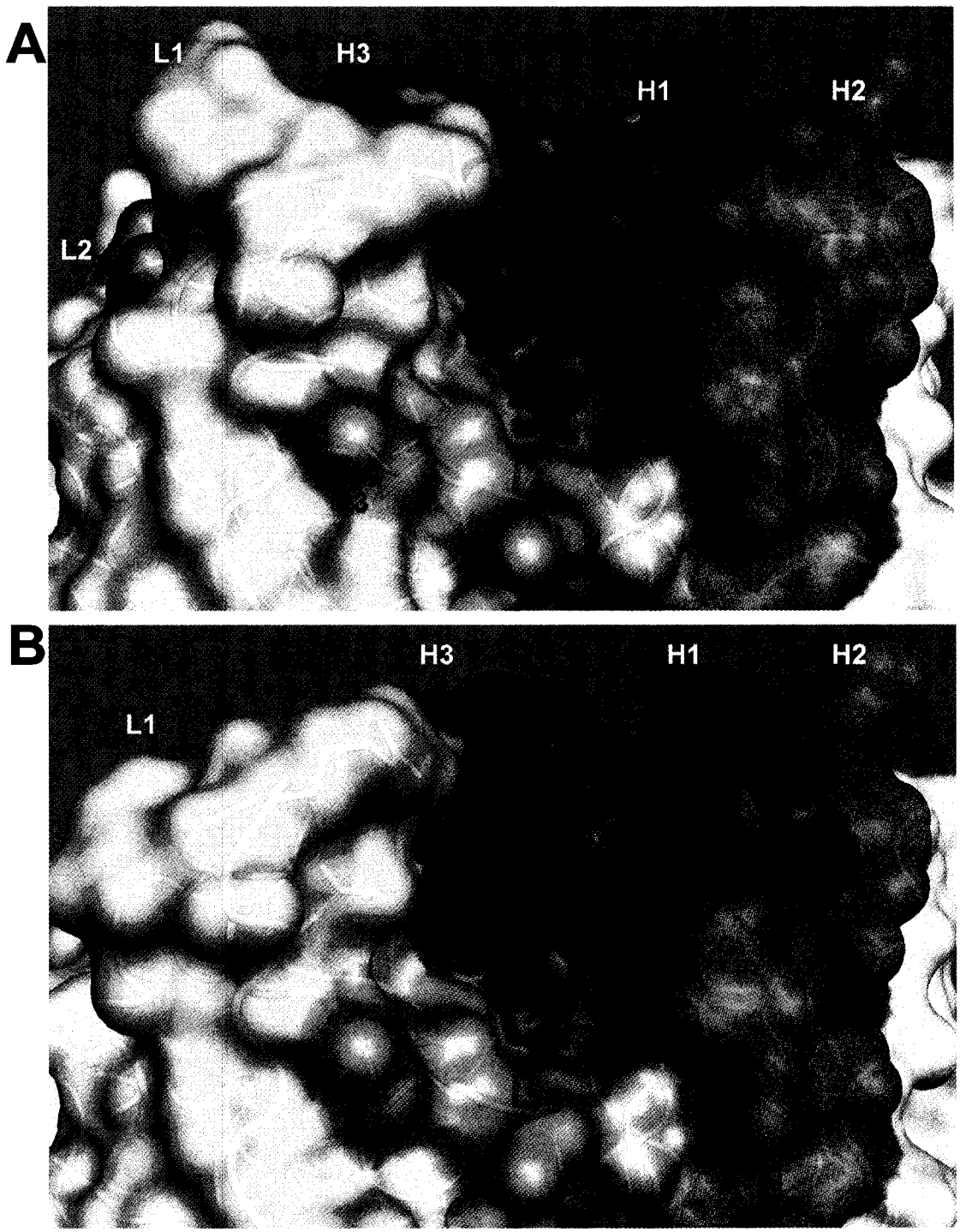


Figure 4.1 Surface representations featuring point mutations (*red*) and nontemplate additions (*magenta*) in the binding pockets of **A)** S25-2 (bound to its trisaccharide antigen) and **B)** liganded S45-18. In comparison the surfaces of the conserved germlines sequences are highlighted for the CDRs L1 in *yellow*, L2 in *orange*, L3 in *pink*, H1 in *green*, H2 in *cyan*, and H3 in *blue*. There is a larger conservation of the CDR germlines in S25-2, compared to S45-18 which features more mutations at the periphery of the binding pocket.

4.2 D and J genes complement conservation of V genes

The common usage of V_κ and V_H germline genes in S25-2 and S45-18 is complemented by different D and J genes in each antibody accompanied by point mutations to enhance specificity, while not disrupting the binding of the Kdo3 residue. Following the V_κ gene in both antibodies, the first residue of a common J_κ gene is the mutation of germline Tyr to Arg L95. Simulation of Arg L95 back to germline Tyr followed by rounds of computer energy minimization indicate that the tyrosine would fulfill the same role in forming a hydrogen bond to Kdo3 O4 (Fig. 4.2). The mutation to Arg contributes to the local positive charge in the binding pocket, as will be discussed later.

In S45-18, a different set of D_H and J_H genes confers specificity for the α(2→4)-α(2→4) Kdo trisaccharide epitope, while excluding the binding of the α(2→4) disaccharide or Kdo monosaccharide. The D_H and J_H genes of S45-18 contribute a different interaction with Kdo3 that increases specificity without disrupting Kdo3 recognition: Asp H100a (that is Glu in S25-2) created by non-template additions occurring at the flanking regions between the respective D_H and J_H genes. The role of this mutation is related to the observed 0.8Å shift of Kdo3 (with their framework regions superpositioned) (Fig. 4.3), which is accomplished using largely the same set of hydrogen bonds. S45-18 prevents this shift (and so decreases the ability to recognize the α(2→8) disaccharide) by Asp instead of Glu H100a and the insertion of the buried bridging WAT1 molecule (also present in the unliganded structure, although open to the solvent).

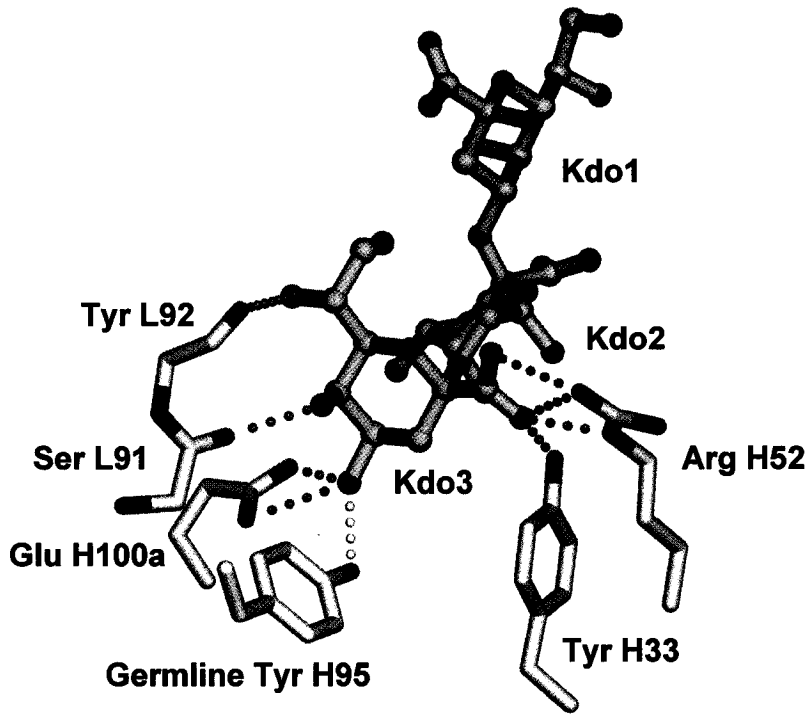


Figure 4.2 In a computer simulation, Arg L95 is changed back to the germline Tyr (*yellow*) in the binding pocket of liganded S25-2 shown here with the α(2→8)-α(2→4) trisaccharide (*orange*). Like Arg L95 at this position, Tyr L95 can form a hydrogen bond interaction (*dotted yellow*) with Kdo3 O4.

The most important contributions by the D_H and J_H minigene are to CDR H3, which only makes 1 contact with Kdo3, and several contacts with the remaining Kdo residues, and as discussed later plays an important role in conformational rearrangement and ligand specificity. This role is also achieved while complementing the conserved identity of the germline V genes, where CDR H3 residues critical for antigen recognition form interactions with conserved germline V residues. An example of this is the Tyr to Phe H97 mutation (in S45-18 CDR H3) that recognizes all three Kdo hexose rings. The orientation of the Phe H97 side chain is maintained even in the absence of ligand through aromatic stacking with the conserved V_κ germline residue Tyr L32.

The complementary roles of D and J have also been suggested in previous reports, where the conservation of V_κ and V_H genes is more critical in determining responses against classes of epitopes such as anti-phosphatidylcholine immunoglobulins that have conserved V genes (Seidl *et al.,* 1999), similar to the bound antigens of S25-2 and S45-18 having a non-reducing terminal Kdo residue.

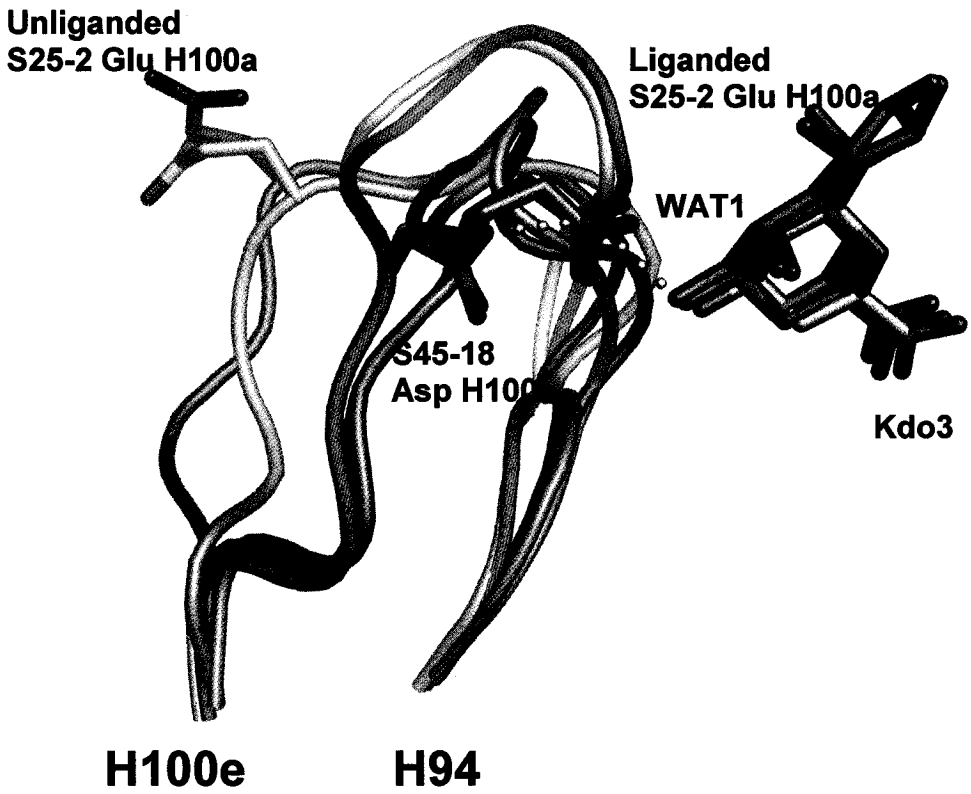


Figure 4.3 With superposition of the framework regions in the variable domains, small differences can be observed in the Kdo3 binding pocket. Shown above is a cartoon represention of CDR H3. The side chain of Glu H100a is displayed for S25-2 complexed with α(2→8)-α(2→4) trisaccharide (*orange*), α(2→4) disaccharide (*magenta*), and of unliganded form 1 (*light blue*) and form 2 (*blue*). The difference in this side chain in S25-2 is 12 Å measured from the Cδ atom. A shift of up to 0.8 Å is found in the positions of Kdo3 (Kdo1&2 residues not shown for clarity). S45-18 prevents this shift using the water molecule WAT1 (*green sphere*) bridging with the side chain Asp H100a, that is in the same position in the liganded (*green*) and unliganded forms (*white*). WAT1 is also present in the same position in the unliganded structure.

4.2 Mutations of germline genes enhance specificity

In the remainder of the binding pocket, mutations away from germline sequence and differences between S25-2 and S45-18 can be examined in terms of their contributions to specificity through steric and electronic factors.

The differential binding by S25-2 and S45-18 of their respective Kdo2 and Kdo1 residues is reflected in the shape of the remainder of their binding site. In this area, while S25-2 remains largely unchanged from the germline, S45-18 features non-template additions in combination with germline mutations that complement the shape of its ligand. The additions Asp H95, Ile H95a and Tyr H95b during genetic rearrangement (between the V and D genes) in S45-18 contribute a two-residue increase in length over S25-2, that allows the tip of CDR H3 to insert into the binding pocket (Fig. 4.4). At the tip of this loop are two germline mutations Tyr to Phe H97 and Tyr/His to Gly H98. Gly H98 allows for a kink in the polypeptide backbone that is maintained by hydrogen bonding between the adjacent backbone Ser H99N with the side chain Ser H96 Oγ. This positions Phe H97 parallel to Kdo1 and perpendicular to Kdo2 and Kdo3 of the α(2→4)-α(2→4) trisaccharide portion of the pentasaccharide antigen. Phe H97 is exposed to solvent in the unliganded S45-18 structure. This example of van der Waals recognition in S45-18 is not observed in S25-2.

The corresponding region of the S25-2 combining site is observed to be as more "open" than in S45-18 and is able to accommodate the unique Kdo2 positions of the different antigens bound by S25-2 (Fig 3.9).

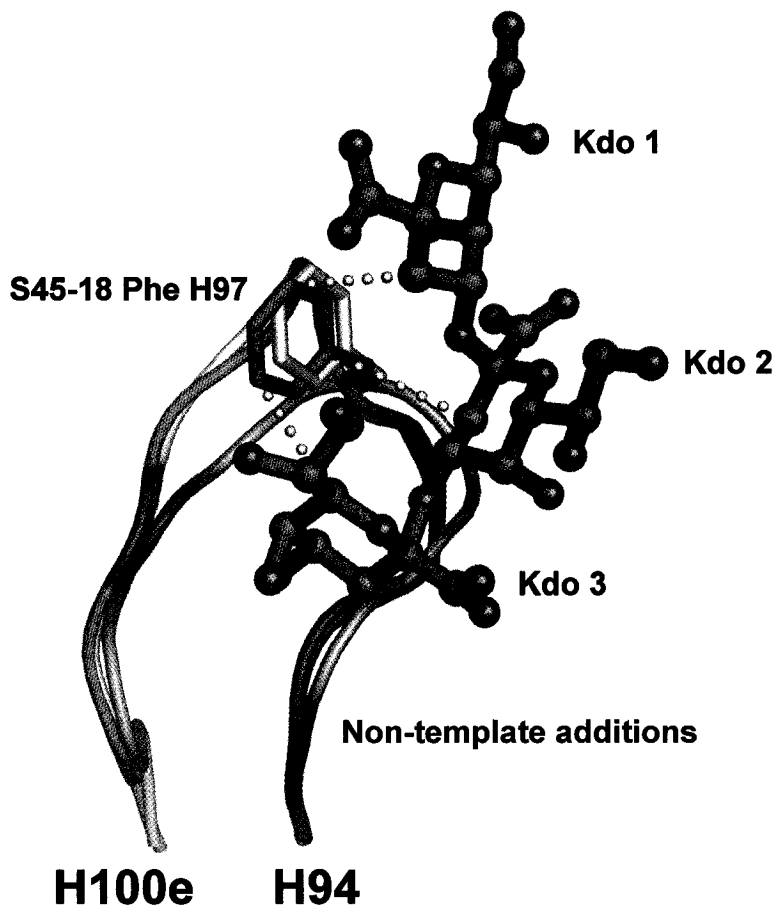


Figure 4.4 The genetic rearrangement of VDJ genes results in a different number of non-template additions (*magenta*) at positions between V-D and D-J – a total of four in S45-18 (unliganded, *white*; liganded, *green*), and two in S25-2 (complexed, *orange*). The two-residue length difference allows formation of a kink in the tip of CDR H3 that positions the mutation Phe H97 (*side chain shown*) closer into the binding pocket where it can interact with all three Kdo residues of the S45-18 antigen. The orientation of the Phe H97 is similar in the unliganded structure.

In addition to steric factors in S25-2, charge distribution lining the binding pocket (Fig. 4.5) contributes to selectivity. S25-2, in comparison to germline, has two mutations in CDR H3: Trp to Arg H100b and Ala to Ser H100d that are part of the flanking DJ regions (which are prone to non-template mutations from imprecise joining mechanisms), however do not make contact with antigen. In general the mutations that create Arg could serve to provide a positively charged environment for the binding of the anionic Kdo epitope. (The corresponding function is fulfilled in S45-18 by Arg H100 within the D region that is conserved in its germline sequence.)

The increase in the proportion of cationic to anionic amino acid residues has been previously reported in antibodies that bind negatively-charged sialic acid (Boffey *et al.*, 2004). However there are no direct charged interactions between the residues mutated from germline and antigen, and the germline Tyr L95 can also fulfill the hydrogen bonding role of Arg L95 observed in S25-2 and S45-18.

The Asn to Lys H52d mutation (Fig 3.11) in S45-18 is an example of a germline and S25-2 conserved residue that is responsible for altering antigen specificity. In S25-2, Asn H52d is a multi-usage residue that provides direct contact to the corresponding Kdo O7 that is in different positions in the $\alpha(2\rightarrow8)$ di- and trisaccharide and $\alpha(2\rightarrow4)$ disaccharide antigens. Asn H52d also incorporates the usage of a water molecule that is not present with the Lys H52d mutation in S45-18. Making use of its longer side chain in S45-18, the Lys H52d side chain confers specificity for the different position of Kdo2. This mutation transforms a multiple-usage residue in S25-2 to a specific unique interaction in S45-18.

Figure 4.5 A) Distribution of charges (*blue* for positive, *red* for negative surface charge) and water molecules in the binding pocket. The S25-2 α(2→8)-α(2→4)-(2)-allyl trisccharide complex features the most water molecules distributed around the Kdo3 and Kdo2 residues.

Figure 4.5 B) Distribution of charges (*blue* for positive, *red* for negative surface charge) and water molecules in the binding pocket of S25-2 with the α(2→4) disaccharide bound. There are significantly less water molecules around Kdo2. This residue does not lie in the same position as the water molecules surrounding the Kdo2 of the α(2→8) epitope.

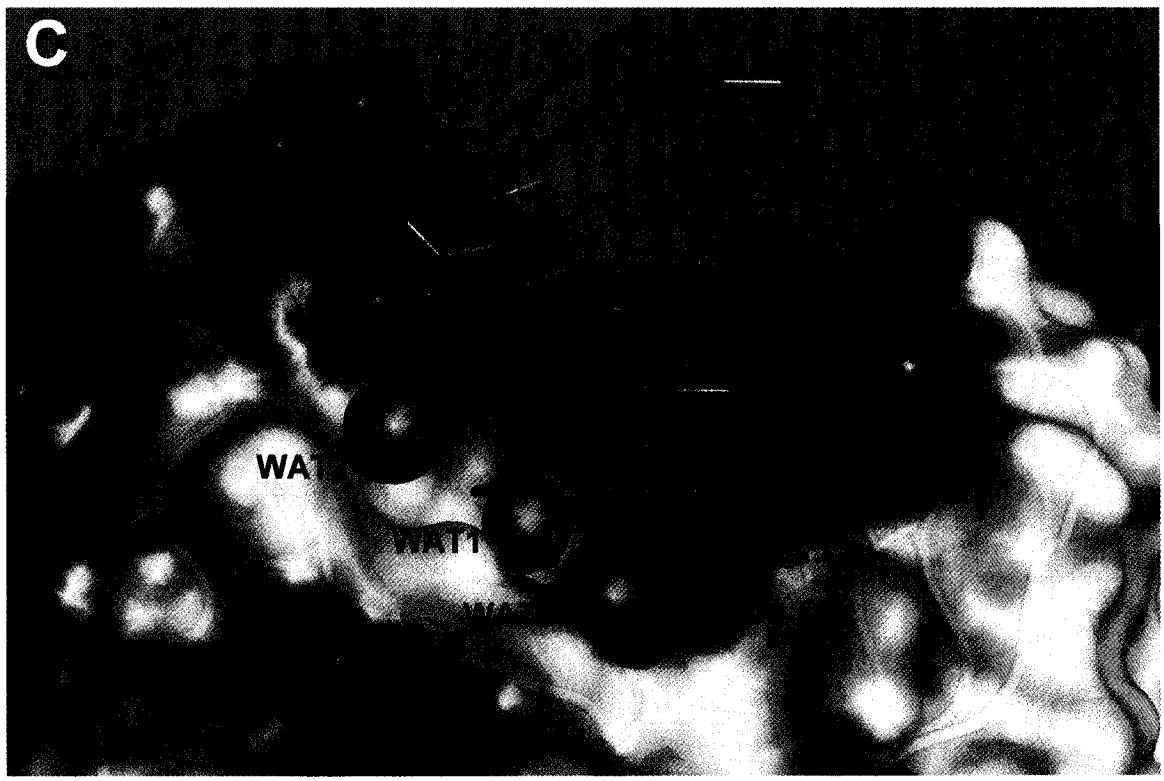


Figure 4.5 **C)** Distribution of charges (*blue* for positive, *red* for negative surface charge) and water molecules in the binding pocket of S45-18. The binding sites of S25-2 and S45-18 feature a conserved Kdo3 binding pocket that share two common water molecules, WAT2 and WAT3. Bridging similar interactions, WAT2 is found at the base of the binding pocket, while WAT3 varies in position in the bound S45-18 complex. The more basic (*blue*) binding pocket of S45-18 is achieved through gain of basic residues such as CDR H2 Lys H52d and CDR H3 Arg H100, and by removing the acidic (*red*) Glu H100a lining the Kdo3 pocket in S25-2 by mutation to Asp H100d and usage of the water molecule WAT1.

The binding consequences of mutations away from germline have been explored previously (Wedemayer *et al.,* 1997; Yin *et al.,* 2001; Li *et al.,* 2003). In the Wedemayer (1997) report, nine amino acid changes in antibody 48G7 (six in the heavy chain, three in the light chain) from the germline resulted in a 30,000 fold affinity increase for the ligand, a nitrophenyl phosphonate transition state analog. In 45G7, mutations occurred not only at the binding site but at distant sites that allowed the stabilization of a conformation that was better able to bind the ligand. A significant difference in binding is observed in the mature 48G7 versus the germline. While 48G7 bound the ligand using a lock-and-key mechanism, the germline bound through an induced fit such as in the case of S25-2. Moreover, the mutations result in the structure of the mature antibody approximating the bound state conformation of the germline, such that the mature antibody binding site is "preorganized" for ligand. The mutations contribute to more hydrogen bonding to the ligand. This is also observed in S45-18, however with accompanying substitution of DJ genes that contribute to CDR H3. Besides enthalpic contributions, the mutations immobilize the flexible regions of the germline antibody, and thus reduce the entropic penalty of ligand binding.

A similar finding was made when comparing the germline antibody and an affinity-matured form 28B4 that differ by 9 amino acid substitutions (Yin *et al.,* 2001), where both bound the same hapten as in Wedemayer et al., with the phosphonate groups in the same position but the p-nitrophenyl groups in different positions. The germline antibody bound with accompanying conformational changes to CDR L1 and H3, and the affinity-matured antibody bound in lock-and-key fashion. 28B4 shares identical germline genes Ig $V_{\kappa}1$ and Ig $J_{\kappa}1$ with 20-1 and NZA6 (Yin *et al.,* 2001), but is observed to bind different epitopes

illustrating that conserved germline genes and combinations of genes undergo few single mutations to adapt to different antigens.

4.3 Aromatic residues in anti-carbohydrate recognition

In the germline, S25-2 and S45-18 sequences, there is a total of up to 33 aromatic residues (Appendix Table A.4); however, only a few of these residues actually have their side chains lining the binding pocket: Tyr L32, Tyr L92, Tyr H33, Phe H50 and Phe H97 (in S45-18), with only S45-18 Phe H97 forming van der Waals interactions with antigen using its aromatic side chain, and Tyr H33 forming unique hydrogen bond interactions with different antigens using its –OH group. To compare, Phe H97 is a mutation from germline Tyr, while Tyr H33 is a conserved germline residue.

Aromatic residues in S25-2 and S45-18 that are created by somatic mutation or conserved from germline, play different roles in specificity. The mutations of germline residues that result in aromatic residues are observed to confer specificity by direct interactions with antigen (and excluding other antigens), whereas conserved aromatic germline residues complement somatic mutations that interact with antigen, and play roles in multiple usage interactions.

The mutation of germline Tyr to Phe H97 in S45-18 changes the role played by each aromatic residue in antigen specificity. This is accomplished by eliminating the possibility of forming interactions between the hydroxyl group of Tyr and surrounding residues and with hydroxyl groups on the αKdo-(2→4)-αKdo-(2→4)-αKdo trisaccharide which would re-orient the ligand. Instead with Phe H97, hydrophobic van der Waals stacking interactions are formed between the side chain against carbons in all three rings of

the Kdo antigen which are oriented in such a way that the hydroxyl groups face away from the phenyl side chain.

The conservation of the germline residue Tyr L32 complements Phe H97 through aromatic stacking interactions that maintain the orientation of the Phe H97 side chain. Since Phe H97 is not found in S25-2, the ring of Tyr L32 in S25-2 is rotated 180 degrees with respect to that in S45-18. Another role of Tyr L32 in S25-2 is associated with the conformational flexibility of S25-2 CDR H3. In one form of unliganded S25-2, the ring of Tyr L32 interacts with the germline Gly H98 Cα.

The conservation of germline aromatic residues in both S25-2 and S45-18 also demonstrates multiple usage of the same amino acid. Tyr H33 binds the terminal non-reducing Kdo3 residue, while making unique interactions with the Kdo2 residues of different antigens. An aromatic T-stacking interaction between Tyr H33 and Phe H50 maintains Tyr H33 in a critical position to be used multiply.

Finally, there are other mutations that create aromatic residues that do not play a significant role in antigen binding, such as in the V_κ gene of S45-18: Ser to Phe L7 and Ser to Phe L43, both in the framework regions. As well there are mutations that remove germline aromatic residues such as Tyr to Arg L95 in both S25-2 and S45-18, Tyr to His H95a in S25-2, and Tyr/His to Gly H98 in S45-18. Of these, the mutation to Arg L95 has been previously discussed serving a role possibly to maintain a hydrogen bond with the Kdo3 residue, while providing a positive charge for the binding pocket.

The use of aromatic side chains in the direct binding of antigen has been previously reported. Padlan in the examination of lysozyme-immunoglobulin complexes (1990) had noted that aromatic residues are exposed to solvent in immunoglobulin binding sites, in

particular tyrosine providing non-polar faces for ligand binding, instead of aliphatic chains. Similarly, the first report of an antibody-carbohydrate complex (Cygler *et al.,* 1991) revealed a binding pocket lined with aromatic side chains and coordinated water molecules. The IgG1λ Se155-4 was crystallized in the presence of a dodecasaccharide composed of three repeating units of *Salmonellae* serogroup B oligosaccharide {→3)-αD-Gal-(1→2)-[αD-Abe-(1→3)]-αD-Man-(1→4)-αL-Rha-(1→}, with electron density visible for the first three carbohydrate residues. The antigen was bound with one face involved in hydrogen bonding, and the other exposed to a hydrophobic environment allowing van der Waals stacking of aromatic residues against the sugar ring carbon atoms. The authors reason that this serves to select for abequose (3,6-dideoxy-D-galactose), and sterically exclude antigens that contain its C4 epimer, 3,6-dideoxy-D-glucose. The binding site of the anti-tumour antibody BR96 is also lined with aromatic residues. A complex of this antibody with a nonoate ester of Lewis Y tetrasaccharide antigen αFuc-(1→2)- βGal-(1→4)-(αFuc-[1→3])-GlcNAc reveals a mixture of van der Waals interactions and hydrogen bonding interactions (Jeffrey *et al.,* 1995). It was reasoned that while surface complementarity was achieved with burial of the antigen, the desolvation of all four residues likely accounts for the low observed binding affinity.

4.4 Antigens show different conformations

Clear electron density was found surrounding the antigens (Fig. 4.6), with no surrounding unaccounted electron density – suggesting a single predominant conformation in each complex.

A superposition of the non-reducing terminal Kdo3 residues (Fig. 4.7) reveals good overlap of the corresponding residues in the di- and trisaccharide antigens containing the α(2→8) linkage bound by S25-2. This is also reflected in the similar observed measurement of the dihedral (ϕ2, ψ2) angles about the glycosidic bond between Kdo2 and Kdo3 (listed in Appendix Table A.3). In constrast, comparison of the α(2→4) disaccharide antigen bound by S25-2 and the terminal α(2→4) disaccharide portion of the pentasaccharide antigen bound by S45-18 revealed marked differences in the position of Kdo2, as well as differences in the dihedral (ϕ2, ψ2) angles.

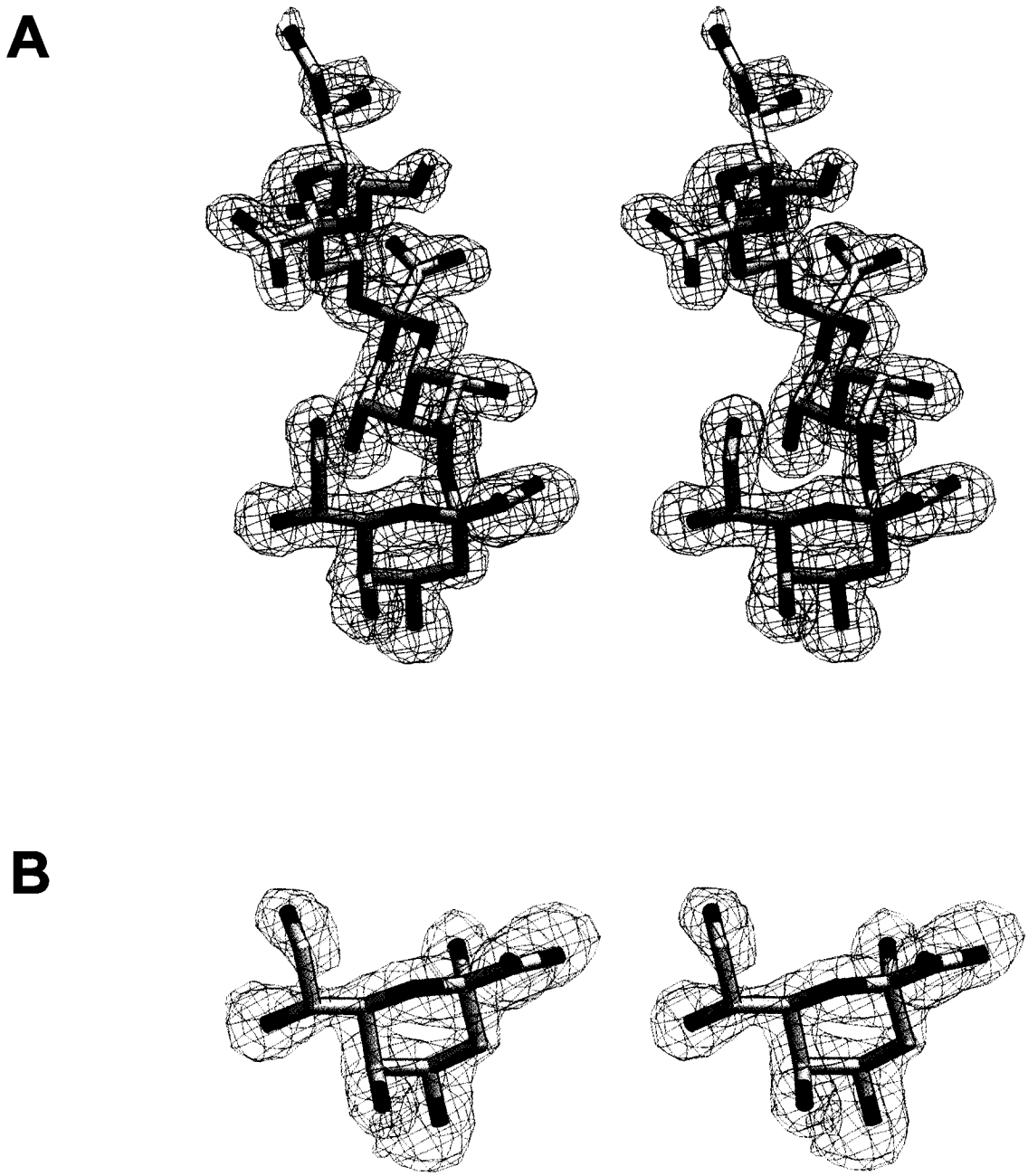


Figure 4.6 Stereo diagrams of *2Fo-Fc* σA-weighted electron density maps contoured at 1.0σ, for the **A)** αKdo-(2→8)- αKdo-(2→4)-αKdo-(2→*O*)-allyl trisaccharide (electron density of the αKdo-(2→8)-αKdo-(2→*O*)-allyl disaccharide is similar, *not shown*), **B)** αKdo monosaccharide.

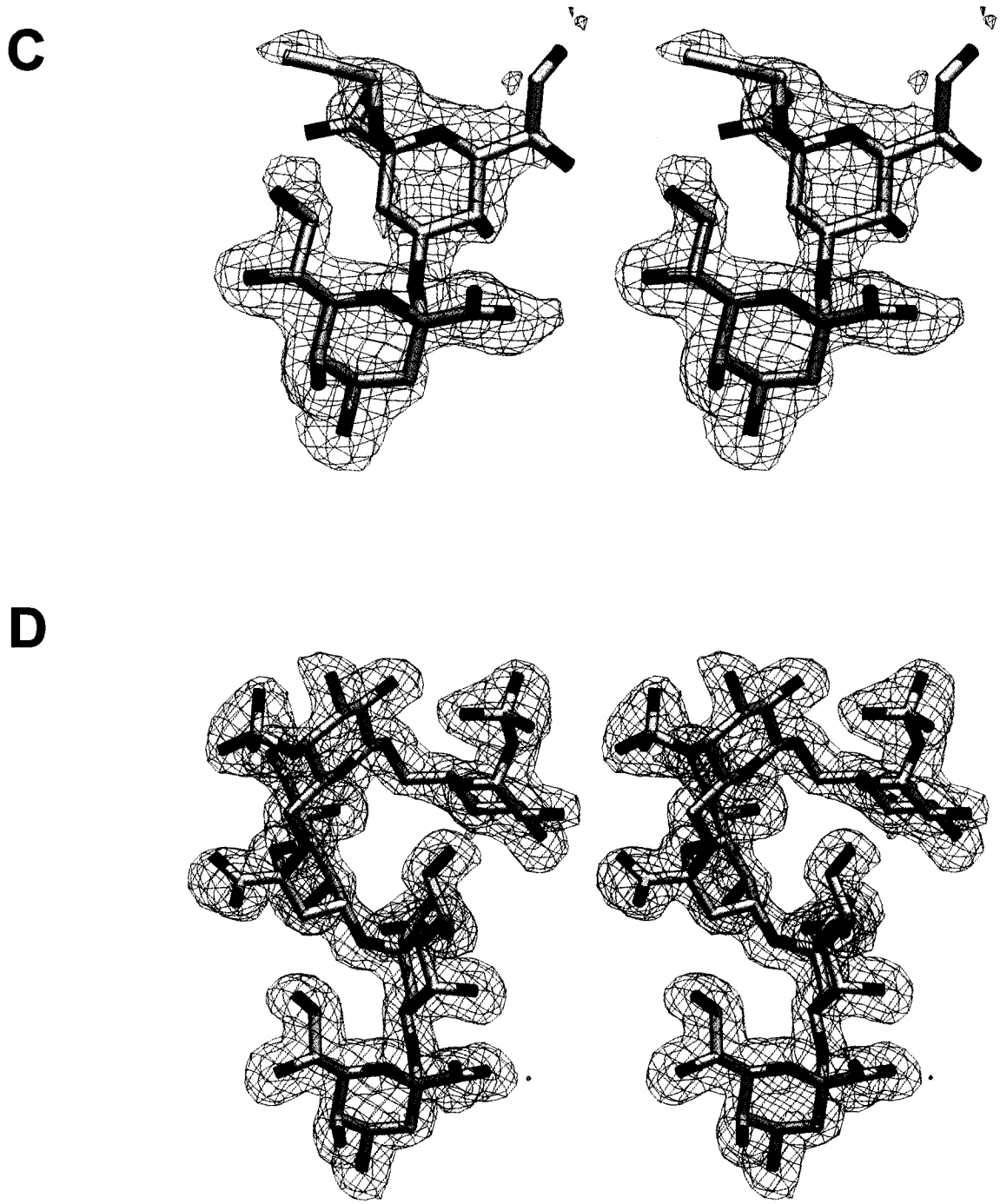


Figure 4.6 **C)** αKdo-(2→4)- αKdo-(2→*O*)-allyl disaccharide and **D)** αKdo-(2→4)– αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate. Atoms colored are carbon (*white*), oxygen (*red*), nitrogen (*blue*), and phophorus (*yellow*).

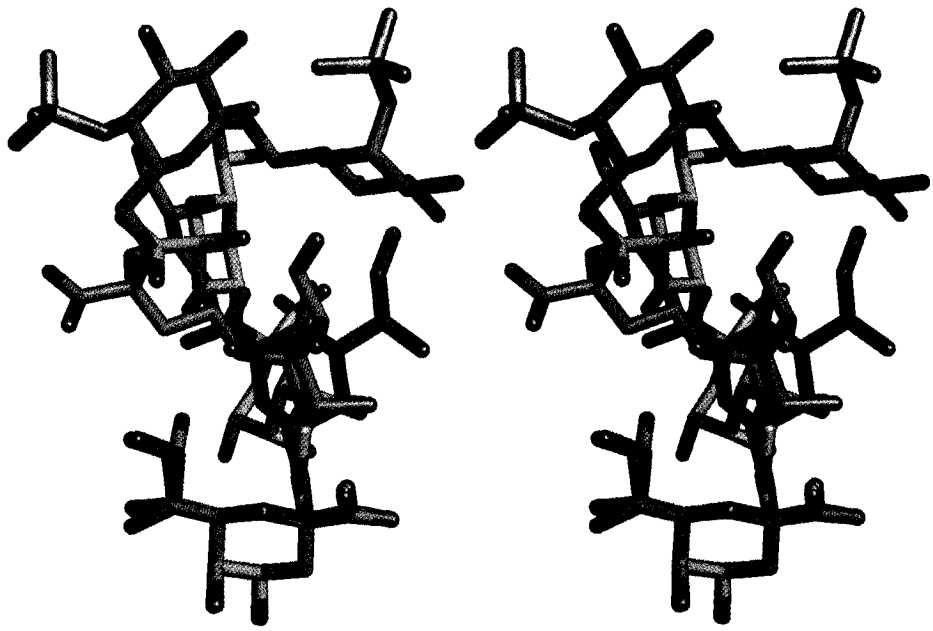


Figure 4.7 Stereo diagram of the antigens with superposition of their non-reducing terminal Kdo3 residue to illustrate that the remaining residues of the epitopes are not structurally mimicking one another: αKdo-(2→8)- αKdo-(2→4)-αKdo-(2→*O*)-allyl trisaccharide (*orange*), αKdo-(2→4)- αKdo-(2→*O*)-allyl disaccharide (magenta) and αKdo-(2→4)– αKdo-(2→4)-αKdo-(2→6)-βGlcN-4P-(1→6)-αGlcN-1P pentasaccharide bisphosphate (*green*). The αKdo-(2→8)-αKdo-(2→*O*)-allyl disaccharide position corresponds well with the trisaccharide and is not shown. The (2→*O*)-allyl groups have also been removed for clarity.

One aspect of binding recognition is the conformational flexibility exhibited by the carbohydrate molecules. In solution, a carbohydrate can adopt a strict range of conformations that are energetically most favourable called global minima, compared to other ranges in conformation referred to as local minima. These energy levels and conformations are associated with electronic and steric forces between the atoms, and interactions with solvent molecules (Kirschner & Woods 2001).

Previous studies of the α(2→8) and α(2→4) Kdo disaccharides in solution and bound to antibodies have been performed using NMR (Sokolowski *et al.,* 1998; Haselhorst *et al.,* 1999; Maaheimo *et al.,* 2000). In particular they revealed that the α(2→8) disaccharide is bound to S25-2 in a similar conformation compared to the global minimum observed in solution (Sokolowski *et al.,* 1998). An X-ray structure of the crystallized α(2→8) disaccharide revealed a hydrogen bond between the non-reducing Kdo3 carboxylate group and the proton donor of Kdo2 O7 that is also present in the solution conformation determined by NMR. Two further immunoglobulins S25-39 and S23-24 both unique in polypeptide sequence were also shown to bind the α(2→8) disaccharide in a similar conformation to that bound by S25-2 (Haselhorst *et al.,* 1999; Maaheimo *et al.,* 2000). However, upon closer inspection of the crystal structure of the α(2→8) di- and trisaccharides bound to S25-2, the intramolecular hydrogen bond observed in the free antigen crystal structure is not present (at a distance of 3.2 Å) in the antigen-immunoglobulin complex.

From these studies, different immunoglobulin-bound conformations of α(2→4) were revealed as bound by S25-2, S25-39 and S23-24. In comparison to S25-2's affinity for the two epitopes, S23-24 has greater affinity for the α(2→4) disaccharide, whereas S25-

39 has greater affinity for the α(2→8) disaccharide (Maaheimo *et al.*, 2000). Maaheimo reported the conformation of an unbound α(2→4) Kdo disaccharide that was similar to those bound by S23-24. However, S25-2 and S25-39 bound this disaccharide epitope in a different higher relative energy conformation that was not highly populated in solution.

In comparison, the α(2→4) Kdo disaccharide in the immunoglobulin-25-2 complex reveals a conformation that is different than previously proposed in NMR studies, however is still within the possible conformations observed bound to other immunoglobulins. The corresponding terminal two residues of the pentasaccharide epitope bound by S45-18 also adopts a different conformation than those reported previously bound or in solution. The α(2→4) disaccharide bound by S25-2 and the αKdo2-(2→4)- αKdo3 disaccharide portion of the S45-18 trisaccharide does not overlap. S45-18 forms a "bulge" on the CDR H3 face of the binding site with an aromatic residue that interacts with the hexose rings of the α(2→4)-α(2→4) trisaccharide (Fig. 4.5C), and sterically occludes the larger S25-2 α(2→8)-α(2→4) trisaccharide. In comparison, the binding site of S25-2 has a more "open" shape, enough to accommodate the S45-18 α(2→4)-α(2→4) trisaccharide. It is possible that the α(2→4)-α(2→4) trisaccharide may be in a different conformation when bound by S45-18. S25-2 co-crystallized in the presence of a α(2→4)- α(2→4) trisaccharide and the S45-18 pentasaccharide produced multiple morphologies of crystals within each hanging-drop experiment, none of which were of diffractable quality.

A bacterial antigen adopting different conformations that resemble self-antigens can evade immune surveillance. However, it is possible that this molecular mimicry can activate a self-reactive clone, and therefore has consequences relating to autoimmune disease. For example, the LPS of *Helicobacter pylori* is believed to mimic Lewis blood

group antigens (Appelmelk *et al.,* 1996; Moran & Prendergast 2001). LPS from *Campylobacter jejuni* contains sialic acid and can structurally mimic gangliosides found on peripheral nerve tissue (Moran & Prendergast, 2001; Boffey *et al.,* 2004).

4.5 *Differences in mode of binding*

The observation of Fab S25-2 liganded to two epitopes with different glycosidic linkages emphasizes the flexibility of this germline antibody. While all complexes show Kdo3 to occupy similar positions in the binding site, the corresponding remaining Kdo residues lie in significantly different conformations, and several corresponding amino acid residues are observed to display multiple functionalities. The α(2→4) linked Kdo2 is recognized by 3 interactions, including a bifurcated salt bridge between the negatively charged Kdo carboxylate group and positively charged side chain of Arg L30c (Fig. 3.11). The α(2→8) linked Kdo2 in the di- and trisaccharide complexes are recognized by the same identical interactions: three residues Arg L30c, Tyr H33, Asn H52d. S25-2 also recognizes the α(2→4) linked Kdo2 using these residues as well, and using two additional interactions from His H96 and Gly H98. There are also bridging water residues found here that are not found surrounding the bound Kdo2 of the α(2→4) antigen.

4.5.1 Direct multiple usage of the remainder of the binding pocket

In terms of the shape of the region that recognizes Kdo1 and Kdo2, a "groove" is formed by the opposition of CDR L1 by CDRs H1 and H2 comprised of side chains that shift positions to adapt to the different Kdo2 and Kdo1 antigens. Both S25-2 and S45-18 share an identical V_κ gene that encodes CDR L1. However, there is one mutation Ser to Asn L31 in S45-18 that does not contact antigen. Bifurcated salt bridges with Arg L30c are observed with the C1-carboxylate Kdo2 of the α(2→4) linked disaccharide, and with that of Kdo1 of the S45-18 trisaccharide. (This study is the first structural description of a salt bridge interaction between an immunoglobulin to a bacterial carbohydrate. As well, Arg

H52 also forms a bifurcated salt bridge with the Kdo3 carboxylate group.) With further reorientation of its side chain in the S25-2 α(2→8)-α(2→4) Kdo trisaccharide complex, Arg L30c forms a hydrogen bond to Kdo1 O5.

Tyr H33 forms a hydrogen bond to O5 of Kdo2 in the α(2→4) disaccharide, while forming a hydrogen bond to O7 in the α(2→8)-α(2→4) trisaccharide complex, however the positions of these two atoms seem to be similar. However in S45-18, Tyr H33 forms an interaction with the S45-18 α(2→4)-α(2→4) trisaccharide at Kdo2 O5 that is in a different position by 1.4Å. Asn H52d forms a hydrogen bond with O7 of Kdo2 in both complexes, even though Kdo2 lies in significantly different conformations.

4.5.2 Contribution of water molecules to polyspecificity

Besides WAT1 found in the S45-18 Kdo3 pocket, other water molecules are found distributed differently among the complexes previously shown (Fig. 4.5). Comparing the S25-2 liganded and unliganded structures, only 1 water molecule around the binding site is common (WAT2). In the absence of ligand in S25-2, this water molecule forms a hydrogen bond network that exists around Leu L94 and Arg L95. Interestingly, electron density for this water molecule is not clear for the unliganded S45-18, indicating that it is not always present in the absence of antigen. Conserved in all the liganded S45-18 and S25-2 structures, WAT2 serves to bridge the interaction of Leu L94 backbone N to Kdo3 O5 and O6.

The S25-2 binding of the α(2→8)-α(2→4) trisaccharide and α(2→8) disaccharide utilizes 8 water molecules in total, with 4 around Kdo3 (WAT 2, 3, 11 & 20), and 4 around Kdo2 (WAT 22, 23, 24 & 28) (Fig. 4.5A). WAT3 bridges Asn L30a with Kdo3 O7.

WAT11 forms hydrogen bonds bridging both Arg L30c and Tyr H32 in the S25-2 complexes to Kdo3 O7. While all of the Kdo3 waters are conserved in the monosaccharide complex, only the three WAT2, WAT3 and WAT11 are present in the α(2→4) complex (Fig. 4.5B). WAT20 uniquely interacts with Asn L30a by bridging an interaction with Kdo3 O8 in only the α(2→8) di- and trisaccharide, and interestingly in the monosaccharide complexes.

WAT23 is the only common non-Kdo3 water to all the S25-2 complexes. It is in a slightly different position in the α(2→4) complex, and bridges a different interaction. In the α(2→8) linked antigen complexes, it is close to WAT28 and bridges an interaction between Asn H52d Oδ and Kdo2 O5 and O7. In the α(2→4) linked antigen complexes, it interacts with Kdo2 O5, which is close to the position of the Kdo O7 of the α(2→8) linked ligands. Notably, the water molecules surrounding the α(2→8) linked Kdo2 do not lie in the same position as the α(2→4) linked Kdo2.

Between the liganded forms of S25-2 and S45-18, only WAT2 and WAT3 bridge the same interactions from Asn L30a Nδ to Kdo3 O7 (Fig. 4.5C). However, the position of WAT3 in S45-18 is shifted by 1.8 Å. In the S45-18 pentasaccharide complex, there is a buried water (WAT1) molecule in the Kdo3 recognition subsite, but no water molecules found in the remaining subsites. S45-18 WAT1 is also present in the absence of ligand as well.

Quiocho *et al.* (1989) suggest that water molecules play a role in modulating protein-carbohydrate recognition. S25-2 and S45-18 both use water molecules to bind antigens, however via different modes. In S25-2, the use of water molecules is to enchance complementarity with different ligands. S25-2 binds the Kdo2 residues of the α(2→8) di-

and trisaccharides using several bridging water molecules that are not present when binding Kdo2 of the α(2→4) disaccharide. Furthermore, S25-2 binds the common Kdo3 residue using the same water molecules except for a water molecule bridging an interaction to Kdo3 O8. Kdo3 O8 of the α(2→4) disaccharide is in a different position than in the bound α(2→8) Kdo antigens (but similar to the Kdo3 O8 of the S45-18 α(2→4) trisaccharide). The different usage of water molecules, and the fact that the antigens did not occupy the same positions as the water molecules of another complex suggest that the antigen conformations themselves are not structurally mimicking each other.

The role of water molecules in S45-18 contrasts that in S25-2. Specifically S45-18 uses a buried water molecule to narrow the range of specificity by restricting the movement of the terminal Kdo3 residue. The remainder of the binding pocket of S45-18 upon ligand binding does not allow space for any water molecules.

4.5.3 Polyspecificity: conformational change and multiple usage of residues

A significant entropic penalty generally occurs in the immobilization of large flexible loops that limits antigen binding. Inspection of the differences in electron density surrounding the S25-2 and S45-18 structures along with observation of high thermal (B) values can reveal regions of conformational flexibility. In the two forms of unliganded S25-2, the CDRs H3 show decreased electron density and elevated B values compared to the other CDRs, whereas such differences are not observed when comparing the CDRs in the liganded structures of S25-2, or in the unliganded and liganded forms of S45-18 (binding through a lock-and-key mode).

There is a paucity of reports of antibodies that can bind different classes of antigens while displaying conformational changes in their CDRs. Conformational flexibility may not be necessary for the selection of small haptens since the enthalpic gain may not be sufficient to compensate for the entropic penalties associated with the immobilization of large backbone changes. An immunoglobulin that is able to bind both large and small antigens would likely use a different surface to bind each type of antigen, for example, using a pocket with only a limited number of interactions and little conformational change for binding different haptens, versus the use of a larger complementary surface to bind larger peptide molecules.

Although the concept of antibodies adapting their conformations to antigens has been around for decades (Breinl & Haurowitz 1930; Pauling 1940), little structural evidence has been gathered. In terms of general binding by proteins, the lock-and-key theory and the induced fit theory by Koshland (Koshland 1976) can be applied to antibody binding.

Immunoglobulin conformational change in binding a single ligand was examined by Rini *et al.* (1992) describing antibody Fab 17/9 unliganded and in complex with a nonapeptide. This study revealed an induced fit mechanism in the conformation of CDR H3 upon ligation. Another example is found in a study (Sheriff *et al.*, 1996) that compared the X-ray structures of immunoglobulin BR96 in the unliganded and liganded forms, showing a change in conformation of a CDR upon binding of the Lewis Y antigen (and similarly observed in S25-2 reorganizing its CDR H3 to bind its antigens). The authors in these papers raised the idea of this structural flexibility being the basis for the potential to bind other related epitopes as well. In the case of BR96, since the *in vivo* antigen displayed on tumours is believed to be larger than the Lewis Y tetrasaccharide epitope bound in the

study, the conformational flexibility of BR96 may be the basis of polyspecificity. However, no structural evidence has been reported as yet to support this proposal.

Rearrangements of side chains upon binding different antigens, such as with antibody D1.3 which binds lysozyme and an anti-idiotype antibody differently (Bhat *et al.,* 1990) would seem energetically more favourable than involving conformational changes to the polypeptide backbone.

Recently an antibody, SPE7 was reported to show conformational flexibility in its CDRs that was associated with binding different ligands (James *et al.,* 2003). SPE7 was presented in two unliganded forms, and in complexed forms with different ligands, including three haptens and a larger protein antigen. Using kinetic studies previously employed by Foote and Milstein (1994), James *et al.* (2003) interpreted their results to suggest the existence of different subpopulations of unliganded antibody. However comparing the structures of the unliganded forms to the complexed forms, and applying kinetic studies, the authors observed that while one unliganded form bound the haptens, the other unliganded form did not resemble the protein-bound form.

4.5.4 Antigen binding converts S25-2 CDR H3 to a canonical conformation

S25-2 CDR H3 converts to the same canonical conformation upon binding a range of antigens (Fig. 3.8), despite using different interactions to recognize each antigen. The observation that antigen binding is able to stabilize a conformation, and even convert a structure to a canonical form has been shown previously in the anti-carbohydrate BR96 (Sheriff *et al.,* 1996). Nakamura's group (Shirai *et al.,* 1996) examined 55 structures of CDR-H3 that have been deposited in the Protein Data Bank, where previously tertiary

structure prediction was difficult due to the variations in amino acid sequence and length (ranging from 5 to 17 residues). Of the S25-2 structures, only the unliganded CDR H3 was not predicted by rules developed by Nakamura.

The change in conformation of S25-2 CDR H3 is mediated by only one interaction that is observed to be common in the interactions with different antigens, Glu H100a with Kdo3 O4. This Cδ in the side chain moves approximately 12 Å between the unliganded and liganded states. Consequently the binding of the terminal Kdo3 reorients CDR H3 such that the remainder of the loop can form intra-CDR H3 hydrogen bonding and converts to a canonical structure. The change alters the role of CDR H3 to one similar to other CDRs of the V_κ and V_H genes with conserved germline sequences, in that this canonical form is critical to allowing further unique interactions with the α(2→4) and α(2→8) linked Kdo2 in different orientations. Furthermore, the difference between CDR H3 S25-2 Glu H100a and S45-18 Asp H100a is associated with, respectively, the induced-fit binding in S25-2 versus lock-and-key binding in S45-18, and the exclusion of other epitopes.

4.5.5 Multiple usage in other immunoglobulins

This is the first report of the recognition of clinically relevant epitopes of distinct but closely related carbohydrates by a conserved germline gene combination where the specifity is associated with differences in the small genetic D and J components, point mutations, and multiple usage of residues.

Advances in peptide sequence screening have improved our understanding of polyspecificity at the amino acid residue level (Marchalonis *et al.,* 2001). There have been previous reports of polyspecific recognition by multiple usage of amino acid residues to

peptide antigens. By scanning combinatorial libraries, the anti-p24 (HIV) monoclonal antibody CB4-1 was found to bind several unrelated peptides (Kramer *et al.,* 1997). In an effort to understand the basis for this molecular cross-reactivity, the structures of CB4-1 unliganded and complexed with four of these peptides were later determined (Keitel *et al.,* 1997). Superposition of the binding sites in the presence of ligand revealed that the ligands occupied different regions of the binding site and made different amino acid contacts due to their markedly different structures. Interestingly, the CB4-1 polypeptide backbone does not undergo conformational changes between the unliganded or bound structures. The authors note that to bind ligands that are relatively the same size of the binding pocket, it is often not necessary for an induced fit mechanism of binding which may have entropic penalties.

Multiple usage of the same binding site residues was also demonstrated in another Fab of IgG3k SYA/J6 that recognized the repeating tetrasaccharide unit (ABCD) from *Shigella flexneri* O-polysaccharide (Vyas *et al.,* 2002). The report compared structures of SYA/J6 unliganded and bound with a pentasaccharide ABCDA' with this nomenclature representing [→2)-αL-Rha*p*-(1→2)-αL-Rha*p*-(1→3)-αL-Rha*p*-(1→3)-βD-Glc*p*NAc-(1→2)-αL-Rha*p*-(1→], and a partial epitope trisaccharide BC*D, where C* is a modified synthetic 2-deoxy-L-Rha analog of the C residue. Based on the structure of the pentasaccharide ABCDA' complex, Vyas *et al.* modeled the antibody in complex with the O-polysaccharide.

Interestingly, the partial trisaccharide BC*D with the synthetic 2-deoxy analog binds differently, where the corresponding sugar rings are rotated along the glycosidic bonds with respect to BCD observed in the ABCDA' complex structure. The authors show that some differences in hydrogen bonding patterns are observed and are due to the absence

of the A (and A') residues that in the ABCDA' complex impose steric constraints on the adjacent B and D residues, and also due to the recognition of the 2-OH group in the native C residue which is not present in the 2-deoxy C* analog.

Although this partial epitope binding is not biologically relevant, it demonstrates the flexibility of the residues in forming interactions with different portions of the sugar. In the S25-2 binding of the αKdo-(2→4)- αKdo and αKdo-(2→8)- αKdo disaccharides which are exposed on chlamydial LPS, with each Kdo2 residue bound in a markedly different location by unique interactions, illustrating the multifunctional nature of the binding pocket. S25-2 and S45-18 show multiple-usage of residues is possible by the conservation of the germline V_κ and V_H gene residues and the D_H and J_H genes of S25-2, while those of S45-18 confer a more limited ligand recognition.

4.6 *Balancing polyspecificity and cross-reactivity*

From a physiological perspective, although there is an advantage for antibodies such as S25-2 in binding multiple epitopes, limiting this binding to prevent self-antigen recognition is necessary. While the immunoglobulin response uses hundreds of genes that combine to encode antigen binding sites, non-immunoglobulin receptors in comparison are encoded from fewer genes. Particularly in the population of pre-formed receptors (in the absence of antigen stimulation) used in the innate immune response, we see families of related molecules each binding a limited range of ligand structures.

4.6.1 Comparison of polyspecificity in S25-2 and S45-18 with non-immunoglobulin receptors

Previous reports of non-immunoglobulin receptors can be used to compare against the specificity of S25-2 and S45-18. While the latter two antibodies bind a limited range of related antigens, non-immunoglobulin PRRs have a range of specificities that include being able to distinguish between small differences in antigen, to binding large antigens such as LPS and peptidoglycan using the same surface.

4.6.1.1 Terminal Kdo binding limits polyspecificity

S25-2 and S45-18 both have a binding pocket that recognizes the non-reducing terminal Kdo residue. This subsite restricts not only the nature of the antigen to include those possessing a non-reducing terminal Kdo residue, but also reduces the number of possible orientations in which the ligands are bound.

In comparison to the surfaces of larger PRRs, the receptor CD14 has been shown to bind different classes of ligands such as LPS and peptidoglycan. Recently, the unliganded CD14 structure has been reported (Kim *et al.,* 2005). CD14 binds LPS, and then transfers it to a TLR4 complex that functions to activate the immune response. The initial studies of specificity involved mutagenesis and binding experiments that suggest certain regions are necessary for LPS-binding and for LPS-transfer. Although crystals of CD14 were not grown in the presence of ligands, inspection of the structure allowed for a proposal of the recognition of the different molecules. For example the lipid portion of LPS is believed to bind a large hydrophobic surface at the NH2-terminal pocket. The helices around this area are mobile and proposed to undergo some local conformational changes. The carbohydrate portion of the LPS (namely the O-polysaccharide) is believed to bind further towards the COOH-terminal side of the binding pocket. This region is believed to bind peptidoglycan competitively. The authors suggest that the basis of this broad ligand recognition is due to the burial of the hydrophobic surface of the binding pocket by hydrophobic parts of the ligand. Given that the binding of LPS O-polysaccharide is competitive with that of peptidoglycan, a liganded structure may reveal the interations involved that accommodate both classes of antigen.

As previously discussed, polyspecific binding has been observed in the binding pocket of SYA/J6, where the partial epitope BC*D was bound in a different conformation than the corresponding BCD sugars of the ABCDA' saccharide. This "promiscuity" in ligand binding is prevented by S25-2 and S45-18 by first binding the non-reducing terminal Kdo residue. As previously mentioned in the case of S25-2, the polypeptide backbone of CDR H3 changes position upon binding the terminal Kdo monosaccharide, converting to a canonical structure that is able to form further interactions with different antigens.

4.6.1.2 S25-2 and S45-18 exclude the recognition of mannose

The critical component in antigens bound by S25-2 and S45-18 is Kdo (3-deoxy-*D*-*manno*-oct-2-ulosonic acid). The corresponding hydroxyl groups corresponding to mannose 3- and 4-OH on the Kdo ring are the 4- and 5-OH groups respectively due to the different numbering nomenclature for Kdo (Fig. 1.1C), and are also in the same axial configuration as found in mannose.

Recent published structures of immune lectins allow us to examine how they can bind a wide range of bacterial carbohydrates. For example, collectins (collagen-like lectins) participate in the innate immune functions in vertebrates, and contain C-type CRDs attached to collagen-like triple-helical regions (Feinberg *et al.,* 2000). Collectin members previously studied include mannose-binding lectin (MBL), surfactant (and also on mucosal surfaces) proteins SP-A and SP-D, bovine conglutinin and bovine collectin-43 (Kawasaki, 1999; Holmskov *et al.,* 2003). As previously mentioned MBL was found to activate the complement pathway, while SP-A and SP-D can only agglutinize and opsonize microorganisms.

Previous reports of lectins have demonstrated the ability to distinguish between different carbohydrate monosaccharides. The specificity varies amoung members of this family, ranging from the binding of different types of LPS to bacterial cell-wall lipoteichoic acids. Interestingly, SP-A was found to bind the Lipid A and rough LPS (truncated at the core Kdo), but not smooth LPS or peptidoglycan, however the structural basis of this binding has not yet been reported. The CRD of a mammalian MBL has been co-crystallized in presence of a high mannose oligomer (Weis *et al.,* 1992), revealing that a

terminal sugar with 3- and 4-OH groups in an equitorial orientation are required for binding, as found in mannose (Fig. 1.1B).

This specificity for positions of the hydroxyl groups has also been observed in many plant lectins previously solved in complex with their ligands (Bundle & Young, 1992). Interestingly, while the plant lectins showed the ability to also bind galactose (with 4-OH in the axial position), many of the immune receptors exclude its binding. This selectivity is an example of specific binding of a chemical feature that is shared by many terminal sugars found in microbial organisms but not in the host cells. Substrates including mannose, glucose, L-fucose, N-acetyl-mannosamine, and N-acetyl-glucosamine would be bound, but not galactose with its 4-OH groups in a non-binding orientation. Typically carbohydrates on host cells are terminated by sialic acid or galactose. Across different species there are only subtle preferences for the various sugars; however interestingly, a related MBL in the carp family (*Cyprinidae*) is believed to bind galactose (Vitved *et al.,* 2000).

The number of binding sites on a receptor molecule also effects binding. Typical BcR are monomeric forms of class IgM, but are later secreted as pentameric forms (for increased binding through an avidity effect). MBL employs trimers of polypeptides in a triple-helix each with a CRD binding in the millimolar range which then combine in 3, 4, 5 and 6 unit homooligomers to achieve a result of binding in the picomolar range. To compare, the bivalent IgG's S25-2 and S45-18 bind their antigens in the micromolar range (Haselhorst *et al.,* 1999; Brade *et al.,* 2000; Maaheimo *et al.,* 2000; Muller-Loennies *et al.,* 2000).

Attempts to crystallize S25-2 (or S45-18) in the presence of mannose have not been successful while the near germline S25-2 has been crystallized in the presence of the Kdo monosaccharide. In the common terminal Kdo3 subpocket in S25-2 and S45-18, there is a

preferred recognition of Kdo over mannose or glucose – the carboxylate group recognized by salt bridging interactions is a critical aspect of the Kdo3 recognition, as well as many conserved interactions with the 1,2-ethandiol substituent O7 and O8, and the orientations of the hydroxyl groups in the hexose ring which differ between mannose and Kdo (Fig 1.1).

Because these are conserved germline gene products, lectins and other first-response receptors illustrate the concept of natural selection based on the ability to adequately adapt quickly to pathogen. Overall, these binding sites each bind a limited range of ligands. Polyspecificity has been examined previously using blocking studies (using antibodies directed at and thus blocking a part of the binding site) and from mutagenesis techniques that have shown residues involved in multiple ligand recognition. In comparison, the structures of S25-2 and S45-18 exemplifies binding specificity at a finer level of the monosaccharide similar to previous reports of lectins, while using the same surface to bind antigens of different chemical and structural nature, and as well reveal that antigen recognition is limited by the common mode of binding of the non-reducing terminal Kdo residue.

4.6.2 Limiting polyspecificity by coupling subsites that show different modes of binding

The use of subsites in S25-2 and S45-18 serves to balance specificity and adaptability to antigen. The inheritance of the V genes that form largely the Kdo3 subsite confers a survival advantage to the host, while the combination of this subsite with the remainder of the binding pocket formed from smaller (DJ) germline genes (that exhibit more sequence variability) ensures adaptation to antigen through point mutation and non-template variation.

The previously reported structure of an immunoglobulin Fab S-20-4 in complex with the terminal monosaccharide and disaccharide perosamine units in the Ogawa O-specific polysaccharide from *Vibrio cholerae* (Villeneuve *et al.,* 2000) reveals a binding site that features two binding pockets. One pocket binds the terminal perosamine, and there is an adjacent groove to bind the remaining residue. Like the terminal Kdo pocket in S25-2 and S45-18 recognizing the C1 carboxylate group, S-20-4 has a pocket that is specific for a functional group, a 2-O-methyl group of the terminal perosamine residue of the antigen, and does not bind a similar antigen that only differs in the absence of the 2-O-methyl group.

In the combining sites of S25-2 and S45-18, a different number of interactions is found in each subsite. While there are many interactions with the terminal Kdo residue in one subsite, there are fewer interactions with Kdo2 in the remainder of the binding site (Fig. 3.11). In S-20-4, hydrogen bond interactions to the hydroxyl groups in both the monosaccharide and disaccharide perosamine antigens are present, with the exception of one hydroxyl group that is recognized on the monosaccharide that is not present on the corresponding terminal residue of the disaccharide, and only one interaction is formed to this second residue.

One subsite forming more interactions (such as the common terminal Kdo3 binding pocket) may reflect its function of excluding ligands that are not related, as previously discussed. While Villeneuve *et al.* presents a list of binding values of S-20-4 to other closely related synthetic analogs (of the monosaccharide), their corresponding bound structures have not yet been reported.

It may be that through loss of immune tolerance (during the development of the T- or B-lymphocytes) or mutation of residues in the binding subsites can lead to a loss of

antigen selectivity and ultimately self-reactivity. The details of the subsite binding of individual carbohydrate residues is similar to previous reports of peptide recognition mediated by subsites that recognize individual residues, such as in non-immunoglobulin receptors and enzymes (where mutations in their subsites have been shown to result in pathophysiology). For example, a genetic cause of autoimmune disease has been demonstrated to involve TcR receptors that have subsites that exhibit polyspecificity of the amino acid residues of the presented peptide bound to the MHC (Wucherpfennig, 2001).

While the Kdo3 binding subsite limits the binding of epitopes to those containing a non-reducing terminal Kdo, which is typically found in bacterial LPS, the remainder of the binding sites of S25-2 and S45-18 demonstrate specificity on the basis of glycosidic linkage and number of residues. This specificity of the remainder of the binding sites in S25-2 involves a near-germline sequence through an induced fit mechanism of binding, while in S45-18 involves specific point mutations and usage of a different set of $(DJ)_H$ genes through a lock-and-key mode of binding. Given this, mutations to the common Kdo3 subsite may affect selectivity of ligand type (such as epitopes containing a terminal non-reducing Kdo monosaccharide), whereas mutations to the remainder of the binding pocket may result in more subtle changes in specificity (such as for glycosidic linkage position or number of carbohydrate residues).

5.0 Conclusions and Future Work

5.1 Conclusions

The genome is limited by the space devoted to encoding immune functions that recognize abnormal pathophysiology. The breadth of the antigen response therefore is achieved by combining many different immune pathways and molecules. Although B-lymphocyte immunoglobulin- and T-lymphocyte receptors are unique in their maturation involving recombination of several individual genes, their functions as molecules of recognition can be compared to non-immunoglobulin germline receptors with respect to conservation of germline identity and its role in polyspecificity. S25-2 and S45-18 both share usage of residues from the same V_κ and V_H immunoglobulin germline genes. As revealed in the mode of binding of these immunoglobulins, the three roles of these germline segments are then firstly to conserve a binding pocket against a structural component that is common in all LPS inner core epitopes; secondly, to form a binding pocket that allows for the substitution of CDR H3 by contributions of different D and J genes (and to make use of non-template additions) that determine polypeptide conformational flexibility and specificity; and thirdly, to position germline-conserved residues in the remainder of the binding pocket so that they can be used multiply to adapt to different epitopes, and are complemented by affinity-matured residues to enhance specificity. Epitope recognition in this multi-usage portion of the binding pocket is also aided by incorporating different patterns of bridging water molecules into this portion of the binding pocket.

On a phenotypic level, the moderation of polyspecificity and cross-reactivity is achieved in S25-2 and S45-18 by using a binding subsite that is specific for Kdo. This

pocket binds the Kdo monosaccharide and makes interactions with the carboxylate and 1,2-ethanediol substituents, and the ring hydroxyl groups, allowing it to exclude other sugars such as *D*-mannose. The non-reducing terminal Kdo is a core LPS structural component that is exposed in pathogens including members of the family *Chlamydiaceae* and other Gram-negative bacteria. The inheritance and usage of the combination of genes that encode a specific monosaccharide binding pocket with the possibility to couple this site to other parts of the binding site that confer specificity illustrates the efficiency of the immunoglobulin germline to adapt to a range of antigens.

5.2 Future Work

Point mutations can be used to further investigate the effects of amino acid residue identity and position on antigen specificity, including the mutation of Arg to germline Tyr L95, as mentioned in section 4.2. Mutations at this region at the end of the V_κ gene and beginning of the J_κ mini-gene can affect CDR L3's contribution to ligand specificity. Another residue of interest is CDR H3 H100a, which is Glu in S25-2 and Asp in S45-18. In S25-2 in particular, this residue is associated with the induced-fit binding of ligand and the conversion from a non-canonical to canonical CDR conformation. Furthermore, reducing the conformational flexibility of CDR H3 may increase binding affinity by decreasing the entropic penalty, and as such, mutations within CDR H3 or in the surrounding region that stabilize the bound conformation can be investigated.

This study is the first structural report of the recognition of a saccharide carboxylate group by an immunoglobulin. This is of interest since the salt-bridging interactions to the Kdo carboxylate groups are associated with an increase in binding affinity. While Kdo is

present in all Gram-negative bacteria, another common saccharide that features a carboxylate group is sialic acid (N-acetylneuraminic acid). Polysaccharides containing sialic acid are immunological determinants in virulent pathogens such as streptococci and menigococci. This similarity between sialic acid and Kdo presents an opportunity to explore the structural basis of recognition of sialic acid by antibodies and with respect to affinity maturation and germline conservation.

6.0 Appendices

6.1 Appendix Tables

Table A.1 Room temperature preliminary X-ray diffraction and model statistics.

	S45-18 Fab		S25-2 Fab
	Unliganded	Pentasaccharide	α(2→8) disaccharide
Space group	C2	$P2_12_12_1$	$P2_12_12_1$
Unit Cell	*a* = 77.1, *b* = 173.2, *c* = 77.1 Å β = 115.0°	*a* = 71.2, *b* = 123.2, *c* = 137.8 Å	*A* = 80.8, *b* = 40.7, *c* = 133.5 Å
No./ASU	2	2	1
Observations	68,816	93,168	43,656
Unique reflections	28,693	33,578	11,983
Resolution Range	15 – 2.50 Å	15 - 2.50 Å	15 – 2.70 Å
Completeness[a]	92.0 (91.5)	78.4 (91.2)	94.8 (77.8)
Rsym[b]	0.098 (0.434)	0.111 (0.428)	0.104 (0.410)
$\langle I/\sigma \rangle$[a]	7.82 (2.1)	6.23 (1.2)	9.9 (2.5)
Polypeptide[c]	6,786	6,850	3,399
Sugar atoms[c]	0	136	0
Solvent atoms[c]	0	0	34
Rwork[d]	0.177	0.250	0.223

[a] values in parenthesis refer to the highest resolution shell, which corresponds to 2.80 – 2.70 Å for S25-2 - α(2→8) disaccharide, for unliganded S45-18 and for liganded S45-18.
[b] $R_{sym} = \Sigma|(I_{hkl} - I_{<hkl>})| / (\Sigma I_{hkl})$ where $I_{<hkl>}$ is the mean intensity of all reflections equivalent to reflection *hkl* by symmetry.
[c] non-hydrogen atoms.
[d] R_{work}, $\Sigma ||F_o| - |F_c|| / \Sigma |F_o|$,

Table A.2 Conservation of the β-sheets in the immunoglobulin folds.

Secondary Structure	S25-2 trisaccharide	S25-2 Unliganded	S45-18 Pentasaccharide
Light chain			
Variable domain			
βL1 [1]	L4-13	L4-13	L4-13
βL2	L19-25	L19-25	L19-25
βL3	L33-38	L33-38	L33-38
βL4	L45-49	L45-49	L45-49
βL5	L52-54	L52-54	L52-54
βL6	L62-65	L62-65	L62-65
βL7	L70-75	L70-75	L70-75
βL8	L85-90	L84-90	L84-90
βL9	L101-105	L101-106	L101-106
Constant domain			
βL10	L113-117	L113-117	L113-117
βL11	L128-138	L128-138	L128-138
βL12	L144-149	L144-149	L144-149
βL13	L158-162	L158-162	L158-162
βL14	L172-181	L174-181	L172-181
βL15	L190-196	L190-196	L191-196
βL16	L204-209	L204-209	L204-208
Heavy chain			
Variable domain			
βH1 [2]	H3-12	H3-7	H3-12
βH2	H18-25	H18-25	H18-25
βH3	H33-39	H33-39	H33-39
βH4	H46-51	H46-51	H46-51
βH5	H55-57	H55-57	H55-57
βH6	H65-70	H65-70	H65-70
βH7	H74-80	H74-80	H74-80
βH8	H88-96	H88-95	H88-96
βH9	H105-110	H105-107	H105-110
Constant domain			
βH10	H118-122	H118-122	H118-122
βH11	H133-143	H133-143	H133-143
βH12	H149-152	H149-152	H149-151
βH13 [3]	H161-169	H161-169	H161-169
βH14	H172-182	H172-182	H172-182
βH15	H192-197	H192-197	H192-197
βH16	H201-207	H201-207	H201-207

[1] break in hydrogen bonding pattern spanning residues L7-9 in βL1 due to Pro L8 in light chain
[2] break in hydrogen bonding pattern H8-10 in βH1 due to glycines at H8-10.
[3] break in hydrogen bonding pattern spanning residues H163-166 in βH13 due to Pro H165

Table A.3 Measured dihedral angles of the bound ligands in the crystal structures.

Antigen complex	Dihedral angle name	Atoms involved	Measurement (degrees)
S25-2 αKdo-(2→8)-αKdo disaccharide	ϕ2	$C1^{Kdo3}$-$C2^{Kdo3}$-$O8^{Kdo2}$-$C8^{Kdo2}$	-42
	ψ2	$C2^{3}$-$O2^{2}$-$C8^{2}$-$H8a^{2}$	-88
S25-2 αKdo-(2→8)-αKdo-(2→4)- αKdo trisaccharide	ϕ2	$C1^{3}$-$C2^{3}$-$O8^{2}$-$C8^{2}$	-38
	ψ2	$C2^{3}$-$O2^{2}$-$C8^{2}$-$H8a^{2}$	-88
	ϕ1	$C1^{2}$-$C2^{2}$-$O4^{1}$-$C4^{1}$	-43
	ψ1	$C2^{2}$-$O2^{2}$-$C4^{1}$-$H4^{1}$	-33
S25-2 αKdo-(2→4)-αKdo disaccharide	ϕ2	C^{3}-$C2^{3}$-$O4^{2}$-$C4^{2}$	-52
	ψ2	$C2^{3}$-$O2^{2}$-$C4^{2}$-$H4^{2}$	-27
S45-18 αKdo-(2→4)-αKdo-(2→4)- αKdo trisaccharide portion of the pentasaccharide	ϕ2	$C1^{3}$-$C2^{3}$-$O4^{2}$-$C4^{2}$	-66
	ψ2	$C2^{2}$-$O2^{2}$-$C4^{1}$-$H4^{1}$	-69
	ϕ1	$C1^{2}$-$C2^{2}$-$O4^{1}$-$C4^{1}$	-49
	ψ1	$C2^{2}$-$O2^{1}$-$C4^{1}$-$H4^{1}$	-51

Table A.4 Locations of aromatic residues in S25-2 and S45-18.

Residues	Region	Side chain in binding pocket?	Interacts with antigen?
Phe L7 (S45-18)	LFR1		
Tyr L32	L1	✓	
Trp L35 Tyr L36 Phe L43 (S45-18)	LFR2		
Tyr L49 TrpW L50	L2		
Phe L62 Phe L71	LFR3		
Tyr L92 Tyr L95 (germline)	L3	✓	
Phe L97	LFR4		
Phe H27 Phe H29 Tyr H32 **Tyr H33**	H1	 ✓	 ✓
Trp H36 Trp H47	HFR2		
Phe H50 Tyr H53 Tyr H57	H2	✓	
Phe H65 Tyr H77 Tyr H90 Tyr H91	HFR3		
Tyr H95b (S45-18) **Phe H97 (S45-18)** Tyr H99 (S25-2) Tyr H100 (S25-2) Trp H100b (germline) Phe H100c (S25-2) Tyr H100e Trp H101	H3	 ✓	 ✓

6.2 Relevant publications

1. Nguyen, H. P., N. O. Seto, L. Brade, P. Kosma, H. Brade *et al.,* (2001). "Crystallization and preliminary X-ray diffraction analysis of two homologous antigen-binding fragments in complex with different carbohydrate antigens." Acta Crystallogr D Biol Crystallogr **57**(Pt 12): 1872-6.

2. Nguyen, H. P., N. O. Seto, C. R. MacKenzie, L. Brade, P. Kosma *et al.,* (2003). "Germline antibody recognition of distinct carbohydrate epitopes." Nat Struct Biol **10**(12): 1019-25.

7.0 References

Akama, T. O., H. Nakagawa, K. Sugihara, S. Narisawa, C. Ohyama *et al.,* (2002). "Germ cell survival through carbohydrate-mediated interaction with Sertoli cells." Science **295**(5552): 124-7.

Alexander, S. & J. H. Elder (1984). "Carbohydrate dramatically influences immune reactivity of antisera to viral glycoprotein antigens." Science **226**(4680): 1328-30.

Al-Lazikani, B., A. M. Lesk & C. Chothia (1997). "Standard conformations for the canonical structures of immunoglobulins." J Mol Biol **273**(4): 927-48.

Appelmelk, B. J., I. Simoons-Smit, R. Negrini, A. P. Moran, G. O. Aspinall *et al.,* (1996). "Potential role of molecular mimicry between Helicobacter pylori lipopolysaccharide and host Lewis blood group antigens in autoimmunity." Infect Immun **64**(6): 2031-40.

Bachmaier, K., N. Neu, L. M. de la Maza, S. Pal, A. Hessel *et al.,* (1999). "Chlamydia infections and heart disease linked through antigenic mimicry." Science **283**(5406): 1335-9.

Behring, E. & S. Kitasato (1890). Deutsch. med. Wochenschr. **16**: 1113-1114.

Belunis, C. J., T. Clementz, S. M. Carty & C. R. Raetz (1995). "Inhibition of lipopolysaccharide biosynthesis and cell growth following inactivation of the kdtA gene in Escherichia coli." J Biol Chem **270**(46): 27646-52.

Benjamin, R. J., J. A. Madrigal & P. Parham (1991). "Peptide binding to empty HLA-B27 molecules of viable human cells." Nature **351**(6321): 74-7.

Bhat, T. N., G. A. Bentley, T. O. Fischmann, G. Boulot & R. J. Poljak (1990). "Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding." Nature **347**(6292): 483-5.

Boffey, J., D. Nicholl, E. R. Wagner, K. Townson, C. Goodyear *et al.,* (2004). "Innate murine B cells produce anti-disialosyl antibodies reactive with Campylobacter jejuni LPS and gangliosides that are polyreactive and encoded by a restricted set of unmutated V genes." J Neuroimmunol **152**(1-2): 98-111.

Brade, L., A. Rozalski, P. Kosma & H. Brade (2000). "A monoclonal antibody recognizing the 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) trisaccharide alphaKdo(2-->4)alphaKdo(2-->4)alphaKdo of Chlamydophila psittaci 6BC lipopolysaccharide." J Endotoxin Res **6**(5): 361-8.

Brandley, B. K., M. Kiso, S. Abbas, P. Nikrad, O. Srivasatava *et al.,* (1993). "Structure-function studies on selectin carbohydrate ligands. Modifications to fucose, sialic acid and sulphate as a sialic acid replacement." Glycobiology **3**(6): 633-41.

Breinl, F. & F. Haurowitz (1930). Z. Physiol. Chem **192**: 45.

Brunger, A. T. (1992). X-PLOR Manual, Version 3.0. Connecticut, USA, Yale University.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros *et al.,* (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Bundle, D. R. & N. M. Young (1992). "Carbohydrate-protein interactions in antibodies and lectins." Curr Opin Struct Biol **2**: 666-73.

Cate, J. H., M. M. Yusupov, G. Z. Yusupova, T. N. Earnest & H. F. Noller (1999). "X-ray crystal structures of 70S ribosome functional complexes." Science **285**(5436): 2095-104.

Charnock, S. J., D. N. Bolam, D. Nurizzo, L. Szabo, V. A. McKie *et al.,* (2002). "Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose." Proc Natl Acad Sci U S A **99**(22): 14077-82.

Chothia, C. & A. M. Lesk (1987). "Canonical structures for the hypervariable regions of immunoglobulins." J Mol Biol **196**(4): 901-17.

Chothia, C., A. M. Lesk, E. Gherardi, I. M. Tomlinson, G. Walter *et al.,* (1992). "Structural repertoire of the human VH segments." J Mol Biol **227**(3): 799-817.

Coloma, M. J., A. Clift, L. Wims & S. L. Morrison (2000). "The role of carbohydrate in the assembly and function of polymeric IgG." Mol Immunol **37**(18): 1081-90.

Cunningham, M. W. (2000). "Pathogenesis of group A streptococcal infections." Clin Microbiol Rev **13**(3): 470-511.

Cunningham, M. W. (2003). "Autoimmunity and molecular mimicry in the pathogenesis of post-streptococcal heart disease." Front Biosci **8**: s533-43.

Cygler, M., D. R. Rose & D. R. Bundle (1991). "Recognition of a cell-surface oligosaccharide of pathogenic Salmonella by an antibody Fab fragment." Science **253**(5018): 442-5.

de la Maza, L. M., S. Pal, A. Khamesipour & E. M. Peterson (1994). "Intravaginal inoculation of mice with the Chlamydia trachomatis mouse pneumonitis biovar results in infertility." Infect Immun **62**(5): 2094-7.

DeLano, W. L. (2004). The PyMOL Molecular Graphics System. San Carlos, CA, USA., DeLano Scientific LLC.

Ehrlich, P. (1900). Proc. R. Soc. Lond. **66**: 424-448.

Evans, S. V. (1993). "SETOR: hardware-lighted three-dimensional solid model representations of macromolecules." J Mol Graph **11**(2): 134-8, 127-8.

Feinberg, H., D. A. Mitchell, K. Drickamer & W. I. Weis (2001). "Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR." Science **294**(5549): 2163-6.

Feinberg, H., S. Park-Snyder, A. R. Kolatkar, C. T. Heise, M. E. Taylor *et al.,* (2000). "Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor." J Biol Chem **275**(28): 21539-48.

French, D. L., R. Laskov & M. D. Scharff (1989). "The role of somatic hypermutation in the generation of antibody diversity." Science **244**(4909): 1152-7.

Fu, Y., M. Baumann, P. Kosma, L. Brade & H. Brade (1992). "A synthetic glycoconjugate representing the genus-specific epitope of chlamydial lipopolysaccharide exhibits the same specificity as its natural counterpart." Infect Immun **60**(4): 1314-21.

Fukuda, M. (1996). "Possible roles of tumor-associated carbohydrate antigens." Cancer Res **56**(10): 2237-44.

Glusker, J. P. & K. N. Trueblood (1985). Crystal structure analysis : a primer. New York, Oxford University Press.

Gralnick, H. R., B. S. Coller & Y. Sultan (1976). "Carbohydrate deficiency of the factor VIII/von Willebrand factor Protein in von Willebrand's disease variants." Science **192**(4234): 56-9.

Gralnick, H. R., S. B. Williams & M. E. Rick (1983). "Role of carbohydrate in multimeric structure of factor VIII/von Willebrand factor protein." Proc Natl Acad Sci U S A **80**(9): 2771-4.

Grayston, J. T. (2000). "Background and current knowledge of Chlamydia pneumoniae and atherosclerosis." J Infect Dis **181 Suppl 3**: S402-10.

Hahn, D. L., D. Bukstein, A. Luskin & H. Zeitz (1998). "Evidence for Chlamydia pneumoniae infection in steroid-dependent asthma." Ann Allergy Asthma Immunol **80**(1): 45-9.

Harrison, S. C. (1969). "Structure of tomato bushy stunt virus. I. The spherically averaged electron density." J Mol Biol **42**(3): 457-83.

Haselhorst, T., J. F. Espinosa, J. Jimenez-Barbero, T. Sokolowski, P. Kosma *et al.,* (1999). "NMR experiments reveal distinct antibody-bound conformations of a synthetic disaccharide representing a general structural element of bacterial lipopolysaccharide epitopes." Biochemistry **38**(20): 6449-59.

Hofle, M., D. S. Linthicum & T. Ioerger (2000). "Analysis of diversity of nucleotide and amino acid distributions in the VD and DJ joining regions in Ig heavy chains." Mol Immunol **37**(14): 827-35.

Holmskov, U., S. Thiel & J. C. Jensenius (2003). "Collections and ficolins: humoral lectins of the innate immune defense." Annu Rev Immunol **21**: 547-78.

Ikeda, K., T. Sannoh, N. Kawasaki, T. Kawasaki & I. Yamashina (1987). "Serum lectin with known structure activates complement through the classical pathway." J Biol Chem **262**(16): 7451-4.

James, L. C., P. Roversi & D. S. Tawfik (2003). "Antibody multispecificity mediated by conformational diversity." Science **299**(5611): 1362-7.

Janeway, C. (2001). Immunobiology : the immune system in health and disease. New York, Garland Pub.

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

Jeffrey, P. D., J. Bajorath, C. Y. Chang, D. Yelton, I. Hellstrom *et al.*, (1995). "The x-ray structure of an anti-tumour antibody in complex with antigen." Nat Struct Biol **2**(6): 466-71.

Jones, T. A. (1985). "Diffraction methods for biological macromolecules. Interactive computer graphics: FRODO." Methods Enzymol **115**: 157-71.

Kabat, E. A. (1988). "Antibody complementarity and antibody structure." J Immunol **141**(7 Suppl): S25-36.

Kalayoglu, M. V., Indrawati, R. P. Morrison, S. G. Morrison, Y. Yuan *et al.*, (2000). "Chlamydial virulence determinants in atherogenesis: the role of chlamydial lipopolysaccharide and heat shock protein 60 in macrophage-lipoprotein interactions." J Infect Dis **181 Suppl 3**: S483-9.

Kaushansky, K., P. J. O'Hara, C. E. Hart, J. W. Forstrom & F. S. Hagen (1987). "Role of carbohydrate in the function of human granulocyte-macrophage colony-stimulating factor." Biochemistry **26**(15): 4861-7.

Kawasaki, T. (1999). "Structure and biology of mannan-binding protein, MBP, an important component of innate immunity." Biochim Biophys Acta **1473**(1): 186-95.

Keitel, T., A. Kramer, H. Wessner, C. Scholz, J. Schneider-Mergener *et al.*, (1997). "Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity." Cell **91**(6): 811-20.

Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff *et al.,* (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-6.

Killion, J. J., M. A. Wallenbrock, J. A. Rogers, 3rd, G. M. Kollmorgen, W. A. Sansing *et al.,* (1976). "Tumorigenicity and the expression of cell-surface carbohydrates." Nature **261**(5555): 54-6.

Kim, J. I., C. J. Lee, M. S. Jin, C. H. Lee, S. G. Paik *et al.,* (2005). "Crystal Structure of CD14 and Its Implications for Lipopolysaccharide Signaling." J Biol Chem **280**(12): 11347-11351.

Kirschner, K. N. & R. J. Woods (2001). "Solvent interactions determine carbohydrate conformation." Proc Natl Acad Sci U S A **98**(19): 10541-5.

Koshland, D. E., Jr. (1976). "The evolution of function in enzymes." Fed Proc **35**(10): 2104-11.

Kramer, A., T. Keitel, K. Winkler, W. Stocklein, W. Hohne *et al.,* (1997). "Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody." Cell **91**(6): 799-809.

Krivan, H. C., D. D. Roberts & V. Ginsburg (1988). "Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc beta 1-4Gal found in some glycolipids." Proc Natl Acad Sci U S A **85**(16): 6157-61.

Kuo, C. C., L. A. Jackson, L. A. Campbell & J. T. Grayston (1995). "Chlamydia pneumoniae (TWAR)." Clin Microbiol Rev **8**(4): 451-61.

Lamb, D. J., W. El-Sankary & G. A. Ferns (2003). "Molecular mimicry in atherosclerosis: a role for heat shock proteins in immunisation." Atherosclerosis **167**(2): 177-85.

Landsteiner, K. (1901). "Ueber Agglutinationserscheinungen normalen menschlichen Blutes." Wein. Klin. Wschr. **14**: 1132-1134.

Landsteiner, K., L. Pauling & E. K. Landsteiner (1945). The specificity of serological reactions. Cambridge, Mass., Harvard university press.

Lasky, L. A. (1992). "Selectins: interpreters of cell-specific carbohydrate information during inflammation." Science **258**(5084): 964-9.

Lawson, P. R. & K. B. Reid (2000). "The roles of surfactant proteins A and D in innate immunity." Immunol Rev **173**: 66-78.

Lederberg, J. (1959). "Genes and Antibodies." Science **129**: 1649-1653.

Lefranc, M. P. (2001). "IMGT, the international ImMunoGeneTics database." Nucleic Acids Res **29**(1): 207-9.

Lefranc, M. P., V. Giudicelli, C. Ginestoux, J. Bodmer, W. Muller *et al.,* (1999). "IMGT, the international ImMunoGeneTics database." Nucleic Acids Res **27**(1): 209-12.

Li, Y., H. Li, F. Yang, S. J. Smith-Gill & R. A. Mariuzza (2003). "X-ray snapshots of the maturation of an antibody response to a protein antigen." Nat Struct Biol **10**(6): 482-8.

Maaheimo, H., P. Kosma, L. Brade, H. Brade & T. Peters (2000). "Mapping the binding of synthetic disaccharides representing epitopes of chlamydial lipopolysaccharide to antibodies with NMR." Biochemistry **39**(42): 12778-88.

Marchalonis, J. J., M. K. Adelman, I. F. Robey, S. F. Schluter & A. B. Edmundson (2001). "Exquisite specificity and peptide epitope recognition promiscuity, properties shared by antibodies from sharks to humans." J Mol Recognit **14**(2): 110-21.

McPherson, A. (1999). Crystallization of biological macromolecules. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

McRee, D. E. (1993). Practical protein crystallography. San Diego, Academic Press.

Medzhitov, R. & C. A. Janeway, Jr. (2002). "Decoding the patterns of self and nonself by the innate immune system." Science **296**(5566): 298-300.

Michel, H. (1982). "Three-dimensional crystals of a membrane protein complex. The photosynthetic reaction centre from Rhodopseudomonas viridis." J Mol Biol **158**(3): 567-72.

Moore, J. M. & J. M. Neilson (1963). "Antibodies to Gastric Mucosa and Thyroid in Diabetes Mellitus." Lancet **186**: 645-7.

Moran, A. P. & M. M. Prendergast (2001). "Molecular mimicry in Campylobacter jejuni and Helicobacter pylori lipopolysaccharides: contribution of gastrointestinal infections to autoimmunity." J Autoimmun **16**(3): 241-56.

Morell, V. (1995). "Attacking the causes of "silent" infertility." Science **269**(5225): 775-7.

Muller-Loennies, S., C. R. MacKenzie, S. I. Patenaude, S. V. Evans, P. Kosma *et al.,* (2000). "Characterization of high affinity monoclonal antibodies specific for chlamydial lipopolysaccharide." Glycobiology **10**(2): 121-30.

Muramatsu, T. (2000). "Essential roles of carbohydrate signals in development, immune response and tissue functions, as revealed by gene targeting." J Biochem (Tokyo) **127**(2): 171-6.

Muramatsu, T. (2002). "Development. Carbohydrate recognition in spermatogenesis." Science **295**(5552): 53-4.

Nguyen, H. P., N. O. Seto, L. Brade, P. Kosma, H. Brade *et al.,* (2001). "Crystallization and preliminary X-ray diffraction analysis of two homologous antigen-binding fragments in complex with different carbohydrate antigens." Acta Crystallogr D Biol Crystallogr **57**(Pt 12): 1872-6.

Nguyen, H. P., N. O. Seto, C. R. MacKenzie, L. Brade, P. Kosma *et al.,* (2003). "Germline antibody recognition of distinct carbohydrate epitopes." Nat Struct Biol **10**(12): 1019-25.

Nose, M. & H. Wigzell (1983). "Biological significance of carbohydrate chains on monoclonal antibodies." Proc Natl Acad Sci U S A **80**(21): 6632-6.

Nurminen, M., M. Leinonen, P. Saikku & P. H. Makela (1983). "The genus-specific antigen of Chlamydia: resemblance to the lipopolysaccharide of enteric bacteria." Science **220**(4603): 1279-81.

Otwinowski, Z. & W. Minor (1996). "Processing of X-ray Diffraction Data Collected in Oscillation Mode." Methods in Enzymology **276**: 307-326.

Pauling, L. (1940). "A Theory of the Structure and Process of Formation of Antibodies." J. Am. Chem. Soc. **62**: 2643.

Perlmutter, R. M., J. L. Klotz, M. W. Bond, M. Nahm, J. M. Davie *et al.,* (1984). "Multiple VH gene segments encode murine antistreptococcal antibodies." J Exp Med **159**(1): 179-92.

Peterson, E. M., L. M. de la Maza, L. Brade & H. Brade (1998). "Characterization of a neutralizing monoclonal antibody directed at the lipopolysaccharide of Chlamydia pneumoniae." Infect Immun **66**(8): 3848-55.

Phillips, M. L., E. Nudelman, F. C. Gaeta, M. Perez, A. K. Singhal *et al.,* (1990). "ELAM-1 mediates cell adhesion by recognition of a carbohydrate ligand, sialyl-Lex." Science **250**(4984): 1130-2.

Potter, M. (1983). "Structural correlates of immunoglobulin diversity." Surv Immunol Res **2**(1): 27-42.

Radic, M. Z. & M. Weigert (1994). "Genetic and structural evidence for antigen selection of anti-DNA antibodies." Annu Rev Immunol **12**: 487-520.

Rhodes, G. (1993). Crystallography made crystal clear : a guide for users of macromolecular models. San Diego, Academic Press.

Rock, F. L., G. Hardiman, J. C. Timans, R. A. Kastelein & J. F. Bazan (1998). "A family of human receptors structurally related to Drosophila Toll." Proc Natl Acad Sci U S A **95**(2): 588-93.

Roost, H. P., M. F. Bachmann, A. Haag, U. Kalinke, V. Pliska *et al.,* (1995). "Early high-affinity neutralizing anti-viral IgG responses without further overall improvements of affinity." Proc Natl Acad Sci U S A **92**(5): 1257-61.

Rose, D. R., M. Przybylska, R. J. To, C. S. Kayden, R. P. Oomen *et al.,* (1993). "Crystal structure to 2.45 A resolution of a monoclonal Fab specific for the Brucella A cell wall polysaccharide antigen." Protein Sci **2**(7): 1106-13.

Ruiz, M., V. Giudicelli, C. Ginestoux, P. Stoehr, J. Robinson *et al.,* (2000). "IMGT, the international ImMunoGeneTics database." Nucleic Acids Res **28**(1): 219-21.

Ryan, K. J. & J. C. Sherris (1994). Sherris medical microbiology : an introduction to infectious diseases. Norwalk, Conn., Appleton & Lange.

Sands, D. (1969). Introduction to crystallography. New York, W. A. Benjamin.

Schachter, J. & C. R. Dawson (1990). "The epidemiology of trachoma predicts more blindness in the future." Scand J Infect Dis Suppl **69**: 55-62.

Schaechter, M., G. Medoff & B. I. Eisenstein (1993). Mechanisms of microbial disease. Baltimore, William & Wilkins.

Seidl, K. J., J. A. Wilshire, J. D. MacKenzie, A. B. Kantor & L. A. Herzenberg (1999). "Predominant VH genes expressed in innate antibodies are associated with distinctive antigen-binding sites." Proc Natl Acad Sci U S A **96**(5): 2262-7.

Sharon, N. (1993). "Lectin-carbohydrate complexes of plants and animals: an atomic view." Trends Biochem Sci **18**(6): 221-6.

Sharon, N. & H. Lis (1990). "Legume lectins--a large family of homologous proteins." Faseb J **4**(14): 3198-208.

Sheriff, S., C. Y. Chang, P. D. Jeffrey & J. Bajorath (1996). "X-ray structure of the uncomplexed anti-tumor antibody BR96 and comparison with its antigen-bound form." J Mol Biol **259**(5): 938-46.

Shirai, H., A. Kidera & H. Nakamura (1996). "Structural classification of CDR-H3 in antibodies." FEBS Lett **399**(1-2): 1-8.

Singhal, A. & S. Hakomori (1990). "Molecular changes in carbohydrate antigens associated with cancer." Bioessays **12**(5): 223-30.

Sokolowski, T., T. Haselhorst, K. Scheffler, R. Weisemann, P. Kosma *et al.,* (1998). "Conformational analysis of a Chlamydia-specific disaccharide alpha-Kdo-(2-->8)-alpha-Kdo-(2-->O)-allyl in aqueous solution and bound to a monoclonal antibody: observation of intermolecular transfer NOEs." J Biomol NMR **12**(1): 123-33.

Strain, S. M., S. W. Fesik & I. M. Armitage (1983). "Characterization of lipopolysaccharide from a heptoseless mutant of Escherichia coli by carbon 13 nuclear magnetic resonance." J Biol Chem **258**(5): 2906-10.

Tomlinson, I. M., J. P. Cox, E. Gherardi, A. M. Lesk & C. Chothia (1995). "The structural repertoire of the human V kappa domain." Embo J **14**(18): 4628-38.

Tomlinson, I. M., G. Walter, P. T. Jones, P. H. Dear, E. L. Sonnhammer *et al.,* (1996). "The imprint of somatic hypermutation on the repertoire of human germline V genes." J Mol Biol **256**(5): 813-17.

Varki, A. (1999). Essentials of glycobiology. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Vijayan, M. & N. Chandra (1999). "Lectins." Curr Opin Struct Biol **9**(6): 707-14.

Villeneuve, S., H. Souchon, M. M. Riottot, J. C. Mazie, P. Lei *et al.,* (2000). "Crystal structure of an anti-carbohydrate antibody directed against Vibrio cholerae O1 in complex with antigen: molecular basis for serotype specificity." Proc Natl Acad Sci U S A **97**(15): 8433-8.

Vitved, L., U. Holmskov, C. Koch, B. Teisner, S. Hansen *et al.,* (2000). "The homologue of mannose-binding lectin in the carp family Cyprinidae is expressed at high level in spleen, and the deduced primary structure predicts affinity for galactose." Immunogenetics **51**(11): 955-64.

Vyas, N. K., M. N. Vyas, M. C. Chervenak, M. A. Johnson, B. M. Pinto *et al.,* (2002). "Molecular recognition of oligosaccharide epitopes by a monoclonal Fab specific for Shigella flexneri Y lipopolysaccharide: X-ray structures and thermodynamics." Biochemistry **41**(46): 13575-86.

Watson, J. D. & F. H. Crick (1953). "The structure of DNA." Cold Spring Harb Symp Quant Biol **18**: 123-31.

Wedemayer, G. J., P. A. Patten, L. H. Wang, P. G. Schultz & R. C. Stevens (1997). "Structural insights into the evolution of an antibody combining site." Science **276**(5319): 1665-9.

Weis, W. I., K. Drickamer & W. A. Hendrickson (1992). "Structure of a C-type mannose-binding protein complexed with an oligosaccharide." Nature **360**(6400): 127-34.

Wucherpfennig, K. W. (2001). "Structural basis of molecular mimicry." J Autoimmun **16**(3): 293-302.

Yin, J., E. C. Mundorff, P. L. Yang, K. U. Wendt, D. Hanway *et al.,* (2001). "A comparative analysis of the immunological evolution of antibody 28B4." Biochemistry **40**(36): 10764-73.

Yucesan, C. & S. Sriram (2001). "Chlamydia pneumoniae infection of the central nervous system." Curr Opin Neurol **14**(3): 355-9.