INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Université d'Ottawa · University of Ottawa

Structural Studies of the Biosynthesis and Recognition of the Human ABO(H) Blood Group Antigens

Sonia I. Patenaude

Thesis submitted to the Department of Biochemistry, Microbiology & Immunology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

University of Ottawa
Ottawa, Ontario, Canada
May, 2002

To the people I love, without whom this would not have been possible.

ABSTRACT

High-resolution crystal structures of the human ABO blood group glycosyltransferase enzymes, unliganded and in complex with substrate molecules, reveal the basis for their substrate specificity and suggest amino acid residues important in catalysis. The human ABO(H) blood group glycosyltransferases differ in only 4 amino acid residues and are therefore two of the most homologous, naturally occurring glycosyltransferases known that utilize different naturally occurring donor molecules. An N-acetylgalactosaminyltransferase (GTA) uses a UDP-GalNAc donor to convert the H-antigen to the A-antigen, while a galactosyltransferase (GTB) uses a UDP-Gal donor to convert the H-antigen to B antigen. The resulting A and B blood group antigens differ from each other only in the substitution on the terminal saccharide residue of an acetamido for an hydroxyl group yet the potentially catastrophic effects of a mismatched blood transfusion makes them a paradigm for specificity in biosynthetic and immune recognition.

The crystal structures of the catalytic domains of the cloned blood-group A and B enzymes have been determined to 1.80 and 1.65 Å resolution, respectively, and those of the catalytic domains of the A and B enzymes in complex with the H antigen disaccharide and UDP to 1.35 and 1.32 Å resolution, respectively. Glycosyltransferases that retain the stereochemistry of the donor sugar, such as GTA and GTB, have been postulated to function *via* a double-inversion mechanism, including a nucleophilic attack on the donor sugar anomeric carbon. The current structures support the double inversion mechanism and reveal, remarkably, that only two of the amino acids differing between GTA and GTB are positioned to select between the two donors and thus contribute to the stringent stereo- and regioselectivity in this biosynthesis. In addition, the structures of GTA and GTB in complex with H antigen alone have been determined to 1.58 and 1.46 Å resolution, respectively, and show that, contrary to expectations, the acceptor substrate can

bind in the absence of the donor substrate. The DxD motif in GTA and GTB coordinates a manganese ion, which in turn interacts with the UDP-sugar donor. Comparison of the GTA and GTB structures to the other two retaining glycosyltransferases (bovine α1-3galactosyltransferase, *N. meningitidis* LgtC) structures determined to date reveals that the coordination between the aspartate residues and manganese ion are the same. Interestingly, this coordination pattern is different from those observed in inverting glycosyltransferases, suggesting that these patterns may be characteristic of inverting *versus* retaining glycosyltranserases.

The structure of an antibody fragment specific for the human blood group A trisaccharide antigen has also been determined to near-atomic resolution. This structure shows a pronounced pocket at the antigen-binding site, which is formed by four of the six complementarity determining regions, and is of the appropriate size and shape to accommodate the terminal N-acetylgalactosamine residue of the A trisaccharide antigen. A model of the Fv-trisaccharide complex shows several protein-carbohydrate interactions that would allow specific recognition of the blood group A antigen acetamido group.

ACKNOWLEDGEMENTS

So many people contributed to this work in so many different ways. While it would be impossible to thank them all individually, I would like to acknowledge those who were closest to my work and my heart.

First and foremost I would like to thank my supervisor Dr. Stephen Evans not only for his guidance, but also for his friendship. Thank you for your encouragement and for challenging me when I needed it. The lessons I have learned, in science and in life, will not soon be forgotten. I would also like to thank Dr. Svetlana Borisova, who made this project possible. You have been a great help, and an incredible source of information and inspiration.

Life in the lab would not have been the same without the many people who have come and gone over the years. In particular I would like to thank Marcin, Roula, and Phuc Hoa for sharing many frustrations and even more laughs. You are all great friends, and my heart smiles when I think of you. I also thank Carol Ann and Joanne, who answered so many questions and listened to so many problems. I couldn't have made it without you, and will miss you both.

I would like to thank my parents. Without your encouragement and moral support, I could not have achieved this. There are no words to describe the debt that I owe you. I can now say with certainty: I've found something.

To John, who shares my life, I owe so much. For believing in me when I couldn't believe in myself, for seeing the light at the end of the tunnel when I couldn't, I thank you. This is as much your success as mine, and I am grateful to share it with you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	real component of the structure factor
Å	Angström
α_A	phase of the anomalously scattering atoms
ADA	N-(2-acetamido)-2-iminodiacetic acid
B	B-factor (thermal motion)
B	imaginary component of the structure factor
BGT	Bacteriophage T4 β-glucosyltransferase
BNL	Brookhaven National Laboratories
BSA	bovine serum albumin
CDR	complementarity determining regions
CHESS	Cornell High Energy Synchrotron Source
CNS	Crystallography & NMR system
Cα	alpha carbon
DTT	dithiothreitol
E. coli	*Escherichia coli*
f	corrected scattering factor
f'	dispersion component of the anomalous scattering factor
$\Delta f'$	real component of anomalous scattering factor
$\Delta f''$	imaginary (absorption) component of anomalous scattering factor
F	amplitude

F	structure factor
ΔF	Bijvoet difference
F(*hkl*)	structure factor of reflection *hkl*
F(hkl)	amplitude of reflection *hkl*
FA	anomalous scattering component of all atoms
f_a	anomalous scattering factor
Fab	antigen-binding antibody fragment
Fc	calculated amplitude
FH	structure factor of heavy atom
f_o	atomic scattering factor
Fo	observed amplitude
FP	structure factor of protein
FPH	structure factor of heavy atom-derivatized protein
FT	normal scattering component of all atoms
Fuc	L-fucose
Fv	antigen-binding variable domain
Gal	D-galactose
GalNAc	*N*-acetylgalactosamine
GalT	galactosyltransferase
α-GalT	retaining galactosyltransferase
α3GalT	Bovine α1,3-galactosyltransferase
β-GalT	inverting galactosyltransferases
β4Gal-T1	Bovine β1,4-galactosyltransferase T1

GlcAT-I	Human β1,3-glucuronyltransferase I
GlcNAc	*N*-acetylglucosamine
GnTI	Rabbit β1,2-*N*-acetylglucosaminyl- transferase I
GT	glycosyltransferase
GTA	human α3-N-acetyl-D-galactosaminyltransferase
GTB	human α3-D-galactosyltransferase
GtfB	*A. orientalis* UDP-glucosyltransferase
α_H	phase of heavy atom
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
α *(hkl)*	phase of reflection *hkl*
HL	Hendrickson-Lattman
I/σ	intensity to error ratio
IBS	Institute for Biological Sciences
IgG	immunoglobulin G
IgM	immunoglobulin M
K	degrees Kelvin
K_A	Michaelis constant for the acceptor substrate
K_B	Michaelis constant for the donor substrate
k_{cat}	catalytic constant
keV	kiloelectron-volt
LgtC	*N. meningitidis* α1,4-galactosyltransferase
MAD	multiwavelength anomalous dispersion
MIR	multiple isomorphous replacement

MMA	methylmercuryacetate
MPD	2-methyl-2,4-pentanediol
MurG	*E. coli* β1,2-*N*-acetylglucosaminyl- transferase
NaAc	sodium acetate
NRC	National Research Council of Canada
NSLS	National Synchrotron Light Source
α_P	phase of protein
PDB	Protein Data Bank
PEG	polyethyleneglycol
R_{free}	free R-factor
RMS	root mean square
RMSD	root mean square deviations
R_{work}	working R-factor
scFv	single chain antigen-binding variable domain
SeMet	selenomethionine
SnB	Shake-and-Bake
SpsA	*B. subtilis glycosyltransferase*
α_T	phase of normally scattering atoms
u^2	mean square displacement
UDP	uridinediphosphate
UDP-Gal	uridinediphospho-α-D-galactose
UDP-Gal-Hg	uridinediphospho-α-D-galactose-mercury
V_H	variable heavy chain

V_L	variable light chain
α	phase
$\Delta\alpha$	phase angle between normal and anomalously scattering atoms

GENERAL INTRODUCTION

In the century since Landsteiner established the existence of the ABO(H) histoblood groups (Landsteiner, 1900; Landsteiner, 1901), the blood group antigens have been shown to be carbohydrates synthesized by glycosyltransferases. Much is known of the substrate specificity and kinetic activity of the blood group glycosyltransferases, yet the mechanism by which these enzymes catalyze reactions or recognize substrates is not completely understood. Similarly, the mechanism by which the anti-blood group antibodies are able to distinguish between the carbohydrate antigens remains unclear.

The purpose of this work is to explore using X-ray crystallographic methods the structure-function relationship between glycosyltransferases, antibodies and their respective substrates of the ABO(H) blood group system. We will attempt to determine the structural aspects of substrate recognition and binding, as well as the basis for catalysis by the blood group glycosyltransferases. Finally we will determine whether these two classes of macromolecules utilize common recognition mechanisms.

The introductory chapter will provide a description of glycosyltransferases and carbohydrate-binding antibodies in general, followed by an abbreviated history of the ABO(H) histoblood groups. Finally, an overview of the crystallographic theory and methods used in this work will be presented.

The Methods section will be presented in two parts, the first addressing the human A and B blood group glycosyltransferases and the second an anti-blood group A antibody fragment. Likewise, Results and Discussion chapter will be divided into similar sections. A general discussion of the two classes of molecules, including a comparison of their specificity mechanisms, will then follow.

INTRODUCTION

1. PROTEIN-CARBOHYDRATE INTERACTIONS

Living cells require the presence of carbohydrate-binding proteins for a number of biological processes, including transportation, degradation, synthesis, signalling, antigen-binding and cell-cell adhesion (Rademacher *et al.*, 1988; Feizi, 1991). Numerous types of proteins are involved in these recognition events, including enzymes, antibodies and lectins; reviews on several classes are readily available (Vyas *et al.*, 1991; Bundle & Young, 1992; Sharon, 1993; Barondes *et al.*, 1994; Davies & Henrissat, 1995; Drickamer, 1995; Taylor, 1996; Drickamer, 1997; Henrissat & Davies, 1997; Weis, 1997). Several reviews of protein-carbohydrate interactions based on structural data have also been published (Quiocho, 1986; Quiocho, 1988; Quiocho, 1989; Sharon & Lis, 1990; Quiocho *et al.*, 1991; Vyas, 1991; Spurlino *et al.*, 1992; Toone, 1994; Meyer & Schulz, 1997) and reveal common features regarding the molecular basis of protein-carbohydrate interactions, which include: van der Waals forces and stacking interactions, direct hydrogen bonding, as well as the participation of ordered water molecules in hydrogen-bonding networks (Figure 1). Though such general observations regarding the nature of protein-carbohydrate interactions are known, each system possesses particular details that can only be determined experimentally. In addition, the features imparting specificity and affinity are likely to vary slightly depending on the protein-carbohydrate complex under study. Determining the molecular basis for recognition and specificity in protein-carbohydrate reactions is central to this study.

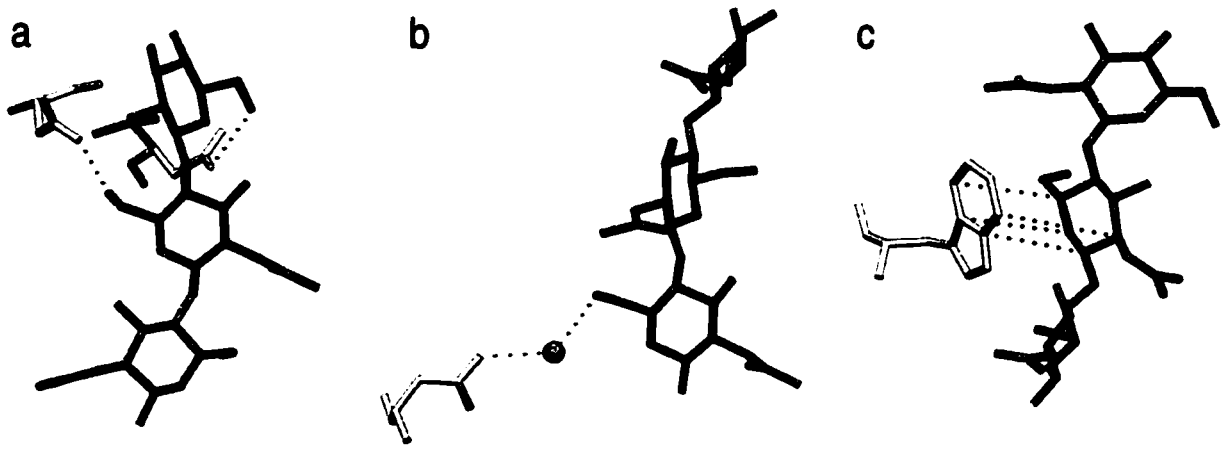

Figure 1. Common protein-carbohydrate interactions. In addition to surface complementarity, proteins and carbohydrates can interact in various ways. a) Hydrogen bonds (yellow dots) between the hydroxyl groups of the carbohydrate (green) and polar protein side chains (white) are often observed, as well as water (blue) -mediated hydrogen bonds to the protein (b) and stacking through van der Waals interactions (red dots) (c). The interactions depicted here were observed in the structure of hen egg-white lysozyme in complex with tetra-*N*-acetyl-chitotetraose (Maenaka *et al.*, 1995), PDB accession code 1LZC.

2. GLYCOSYLTRANSFERASES

The carbohydrate moieties of glycoproteins and glycolipids on the cell surface are widely varying structures synthesized by glycosyltransferases (GTs). Most glycosylation reactions are specific with respect to the anomeric configuration of the sugar and the site of addition; because of this, there exist many glycosyltransferases, each with its own donor, acceptor and linkage specificity (Paulson & Colley, 1989). The GTs transfer the sugar moiety from an activated nucleotide-sugar to an acceptor structure, which can be an oligosaccharide, lipid, protein or nucleic acid. In eukaryotic cells, most glycosyltransferases are localized to the endoplasmic reticulum and Golgi apparatus, generally adopting a type II membrane protein topology: in this case, a short N-terminal cytoplasmic segment, a single transmembrane domain, a stem region of variable length and a C-terminal catalytic domain (Paulson & Colley, 1989).

Glycosyltransferases can be classified according to several parameters, including: the stereochemistries of the substrates and products (Sinnott, 1990), the substrate specificity and reaction catalyzed (International Union of Biochemistry and Molecular Biology, 1992), and also sequence similarities. Recently, 27 families of GTs were identified based on both the amino acid similarities and substrate/product stereochemistries (Campbell *et al.*, 1997; Campbell *et al.*, 1998), and this number has since been updated to 56 (Coutinho & Henrissat, 1999).

I- GALACTOSYLTRANSFERASES

Glycosylation events and the resulting oligosaccharides can be crucial for the development, growth and function of an organism (Varki, 1993). It follows that glycosyltransferases that add the terminal sugar of glycoconjugates, such as sialyltransferases, fucosyltransferases and galactosyltransferases, can be important in the mediation of specific

recognition events or biological processes. Of particular interest in this work are galactosyltransferases.

Galactosyltransferases (GalTs) transfer the galactose from uridinediphospho-α-D-galactose (UDP-Gal) to the acceptor with either retention or inversion of the anomeric configuration of the C1 carbon of the donor sugar (Sinnott, 1990). The GalTs synthesize many oligosaccharides of biological importance, including lactose and the histoblood group antigens in eukaryotes as well as prokaryotic cell wall polysaccharides. GalTs have been divided into 5 families (Breton *et al.*, 1998a), based largely on the nature of the reaction, sequence identity and conserved structural features.

II- GLYCOSYLTRANSFERASE MECHANISM

The GT catalyzes the transfer of a monosaccharide from an activated nucleotide sugar to an acceptor molecule, the majority of these reactions being dependant upon the presence of a divalent cation such as manganese (Mn^{2+}) or magnesium (Mg^{2+}) ions. The chemical mechanism for glycosidic bond formation is one of the most important issues concerning GTs.

The GT enzymes can be classified as inverting or retaining based on the stereochemistry of the anomeric carbon (C1) of the donor sugar (Sinnott, 1990; Davies *et al.*, 1998; Zechel & Withers, 2000). The reaction mechanism of the glycosyltransferases is thought to be analogous to that of the glycosidases. The inverting reaction (Figure 2a) proceeds through a direct displacement mechanism: a general base deprotonates an acceptor hydroxyl group, which then attacks the anomeric (C1) carbon of the donor sugar in an $S_{N}2$-like reaction (Sinnott, 1990; Davies & Henrissat, 1995; Davies *et al.*, 1998). The reaction has been proposed to proceed through an oxocarbenium-ion-like transition state (Kim *et al.*, 1988; Sinnott, 1990; Murray *et al.*, 1997; Davies

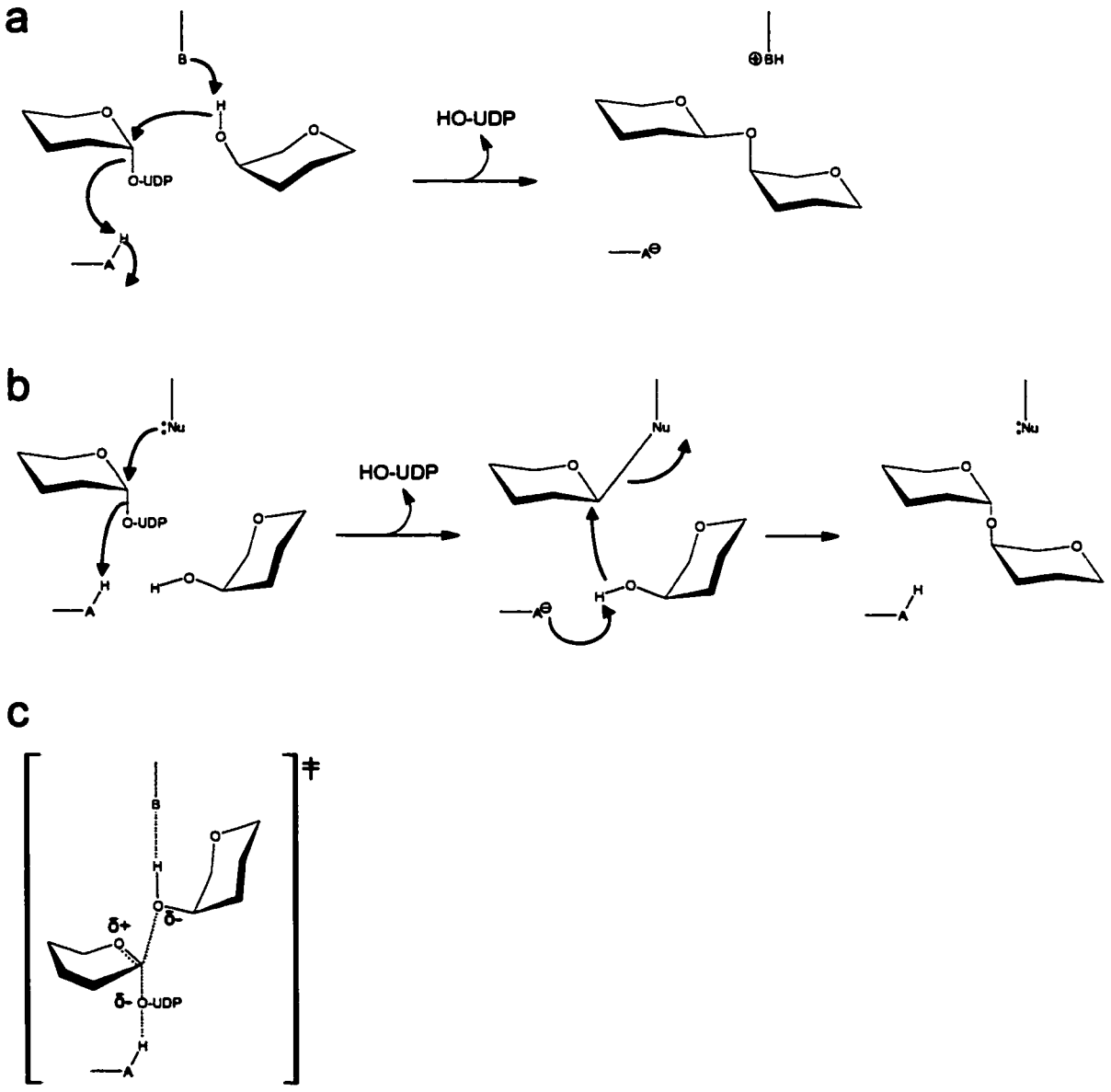

Figure 2. Proposed mechanisms of inverting and retaining glycosyltransferases. The inverting glycosyltransferases have been proposed to proceed through a nucleophilic attack on the anomeric carbon by a general base (a). Similarly, the retaining enzymes are thought to proceed by a double displacement mechanism (b), the intermediate consisting of a glycosyl-enzyme complex. Both types of reaction are believed to involve an oxocarbenium ion-like transition state (‡), similar to that shown in c (Zechel & Withers, 2000). This figure was adapted from Persson *et al.* (2001).

et al., 1998), and results in an inverted configuration of the C1 atom of the donor sugar. The retaining glycosyltransferases are thought to proceed through a double-displacement mechanism (Figure 2b) involving two such inversions. Here, a nucleophile provides the first attack on the anomeric carbon, forming an inverted glycosyl-enzyme intermediate, though no structural evidence has been reported for the existence of this species (Davies, 2001). Deprotonation of the acceptor hydroxyl allows for a second attack on carbon C1 and leads to a second inversion of configuration. The final product retains the original stereochemistry of the donor sugar.

III- CONSERVED MOTIFS

Though very few regions of sequence homology exist between or even within classes of glycosyltransferases, a number of conserved short sequence motifs have been identified. For example, conserved sequences were found in the β1,4galactosyltransferase family (Van Die *et al.*, 1997; Breton *et al.*, 1998a), including FNRA and WGGEDDD. Similarly, conserved motifs have been observed in the sialyltransferases (Datta & Paulson, 1997; Geremia *et al.*, 1997) and the fucosyltransferases (Breton *et al.*, 1998b; Oriol *et al.*, 1999).

Of all the short sequence motifs in GTs, the DXD motif is probably the most widespread in both eukaryotes and prokaryotes (Breton *et al.*, 1998a; Wiggins & Munro, 1998; Breton & Imberty, 1999). This motif consists of two Asp residues separated by any amino acid (x) and is present in almost all galactosyltransferases and in many other GT families (Breton *et al.*, 1998a). It always occurs after a stretch of hydrophobic residues (usually a β-sheet in GTs of known structure) and has been shown to bind the divalent cation necessary for catalysis (Busch *et al.*, 1998). The mutagenesis of the aspartate residues of the DXD motif (Hagen *et al.*, 1999) demonstrated that they are essential for the activity of the enzyme (Busch *et al.*, 1998; Persson *et al.*, 2001). In

certain structures, the second aspartate provides the only interaction between Mn^{2+} and the protein (Charnock & Davies, 1999; Unligil *et al.*, 2000). In some structures, the position of the coordinated manganese ion is appropriate for its positive charge to counter the negative charge on the β-phosphate resulting form the cleavage of the C1-phosphate bond (Unligil & Rini, 2000).

IV- PREVIOUSLY DETERMINED GLYCOSYLTRANSFERASE STRUCTURES

Until recently, few structures of glycosyltransferases had been determined. Since 1999, however, several structures have been reported. In the current section, structures published at the time of writing (Table 1) will be briefly reviewed.

Although little sequence homology exists between the nine GT structures solved to date, common elements regarding overall structure, substrate binding and catalysis have become apparent.

Each structure can be divided into N- and C-terminal domains, consisting of one or more β-sheets flanked on either side by α-helices. The N-terminal donor binding domain is an α/β polypeptide arranged in a distorted Rossmann fold nucleotide-binding motif. The active site is situated in a relatively solvent exposed cleft across this domain and the C-terminal acceptor binding domain. An exception is LgtC, where the binding groove is solvent-excluded. Recently, the GTs outlined in Table 1 have been classified into two folds, A and B, based on structural elements (Unligil & Rini, 2000). BGT, MurG and GtfB (Vrielink *et al.*, 1994; Ha *et al.*, 2000; Mulichak *et al.*, 2001), which form the group adopting fold A, are characterized by two similar domains and by the absence of a DXD motif. In these enzymes, the diphosphate of the donor is stabilized by basic residues (Vrielink *et al.*, 1994; Moréra *et al.*, 1999; Ha *et al.*, 2000; Moréra *et al.*, 2001) or a helix dipole (Mulichak *et al.*, 2001). The remaining 6 structures comprise the second

Table 1. Glycosyltransferase structures solved by X-ray crystallography

Glycosyltransferase	Source	Type[1]	Fold[4]	Ligands	Resolution	Reference
β-glucosyltransferase (BGT)	Bacteriophage T4	β[2]	A	unliganded	2.2 Å	(Vrielink *et al.*, 1994)
				UDP-Glc	2.2 Å	
				UDP-Glc	2.3 Å	(Moréra *et al.*, 1999)
				UDP	1.88 Å	
				UDP, Mg^{2+}	2.07 Å	(Moréra *et al.*, 2001)
				UDP, Mn^{2+}	1.65 Å	
				UDP, Ca^{2+}	1.65 Å	
β1,4-galactosyltransferase T1 (β4Gal-T1)	Bovine	β[3]	B	Unliganded	2.4 Å	(Gastinel *et al.*, 1999)
				UDP-Gal	2.4 Å	
SpsA	*B. subtilis*	β	B	Unliganded	1.5 Å	(Charnock & Davies, 1999)
				UDP, Mn^{2+}	1.7 Å	
				UDP, Mg^{2+}	2.0 Å	
				Mn^{2+}, dTDP	2.0 Å	(Tarbouriech *et al.*, 2001)
				Mg^{2+}, dTDP	1.98 Å	
β1,2-*N*-acetylglucosaminyl-transferase MurG	*E. coli*	β	A	unliganded	1.9 Å	(Ha *et al.*, 2000)
β1,2-*N*-acetylglucosaminyl-transferase I (GnTI)	Rabbit	β	B	Unliganded	1.5 Å	(Unligil *et al.*, 2000)
				UDP-GlcNAc	1.8 Å	
β1,3-glucuronyltransferase I (GlcAT-I)	Human	β	B	UDP, Mn^{2+}, Galβ1-3Galβ1-4Xyl	2.3 Å	(Pedersen *et al.*, 2000)
α1,4-galactosyltransferase (LgtC)	*N. meningitidis*	α	B	UDP-2F-Gal	2.0 Å	(Persson *et al.*, 2001)
				UDP-2F-Gal & 4'-deoxylactose	2.0 Å	
α1,3-galactosyltransferase (α3GalT)	Bovine	α	B	UMP, Mn^{2+}	2.3 Å	(Gastinel *et al.*, 2001)
				UDP-Gal, Mn	2.5 Å	
				UDP, Mn^{2+}	1.53 Å	(Boix *et al.*, 2001)
UDP-glucosyltransferase (GtfB)	*A. orientalis*	β	A	Unliganded	1.8 Å	(Mulichak *et al.*, 2001)

[1]Type of mechanism
[2]Inverting
[3]Retaining
[4]Type of fold

glycosyltransferase fold, B (Charnock & Davies, 1999; Gastinel *et al.*, 1999; Pedersen *et al.*, 2000; Unligil *et al.*, 2000; Boix *et al.*, 2001; Gastinel *et al.*, 2001; Persson *et al.*, 2001). The spatial

arrangement of structural elements related to uridinediphosphate (UDP) binding and the location of the DXD motif is similar in all members. A representative of each type of fold is shown in Figure 3.

Many of the native structures exhibit a disordered loop adjacent to the active site cleft (Vrielink *et al.*, 1994; Charnock & Davies, 1999; Gastinel *et al.*, 1999; Gastinel *et al.*, 2001; Mulichak *et al.*, 2001). In some structures (Vrielink *et al.*, 1994; Unligil *et al.*, 2000), this loop becomes ordered upon binding of the donor substrate. It was postulated that for LgtC, the absence of the UDP substrate would cause two loops "covering" the active site to become disordered (Persson *et al.*, 2001). The presence of flexible loops in proximity to the active site has been proposed to limit access to the active site, thereby limiting the premature hydrolysis of the donor (Charnock & Davies, 1999). Alternatively, such structures could play a role in the release mechanism of the enzyme products (Unligil *et al.*, 2000).

Of the structures solved in the presence of the nucleotide-sugar (Vrielink *et al.*, 1994; Charnock & Davies, 1999; Gastinel *et al.*, 1999; Moréra *et al.*, 1999; Unligil *et al.*, 2000; Gastinel *et al.*, 2001; Persson *et al.*, 2001), most show clear electron density only for the UDP moiety of the donor. Generally, the uracil and ribose rings as well as both phosphate groups interact with the protein. The manganese ion is coordinated by the DXD motif and interacts with both phosphate groups of UDP (Charnock & Davies, 1999; Pedersen *et al.*, 2000; Unligil *et al.*, 2000; Boix *et al.*, 2001; Gastinel *et al.*, 2001; Persson *et al.*, 2001). LgtC and GnTI (Unligil *et al.*, 2000; Persson *et al.*, 2001) are the only structures with highly ordered donor sugar moieties; in both structures, the sugar forms hydrogen bonds with the protein. In certain conditions, some GTs (Vrielink *et al.*, 1994; Unligil *et al.*, 2000) show spontaneous hydrolysis of the UDP-sugar donor can occur, thus rendering visualization of the sugar moiety difficult.

Only two structures have shown the presence of acceptor in the active site, GlcAT-I and LgtC (Pedersen *et al.*, 2000; Persson *et al.*, 2001). In the former, the acceptor binding site is

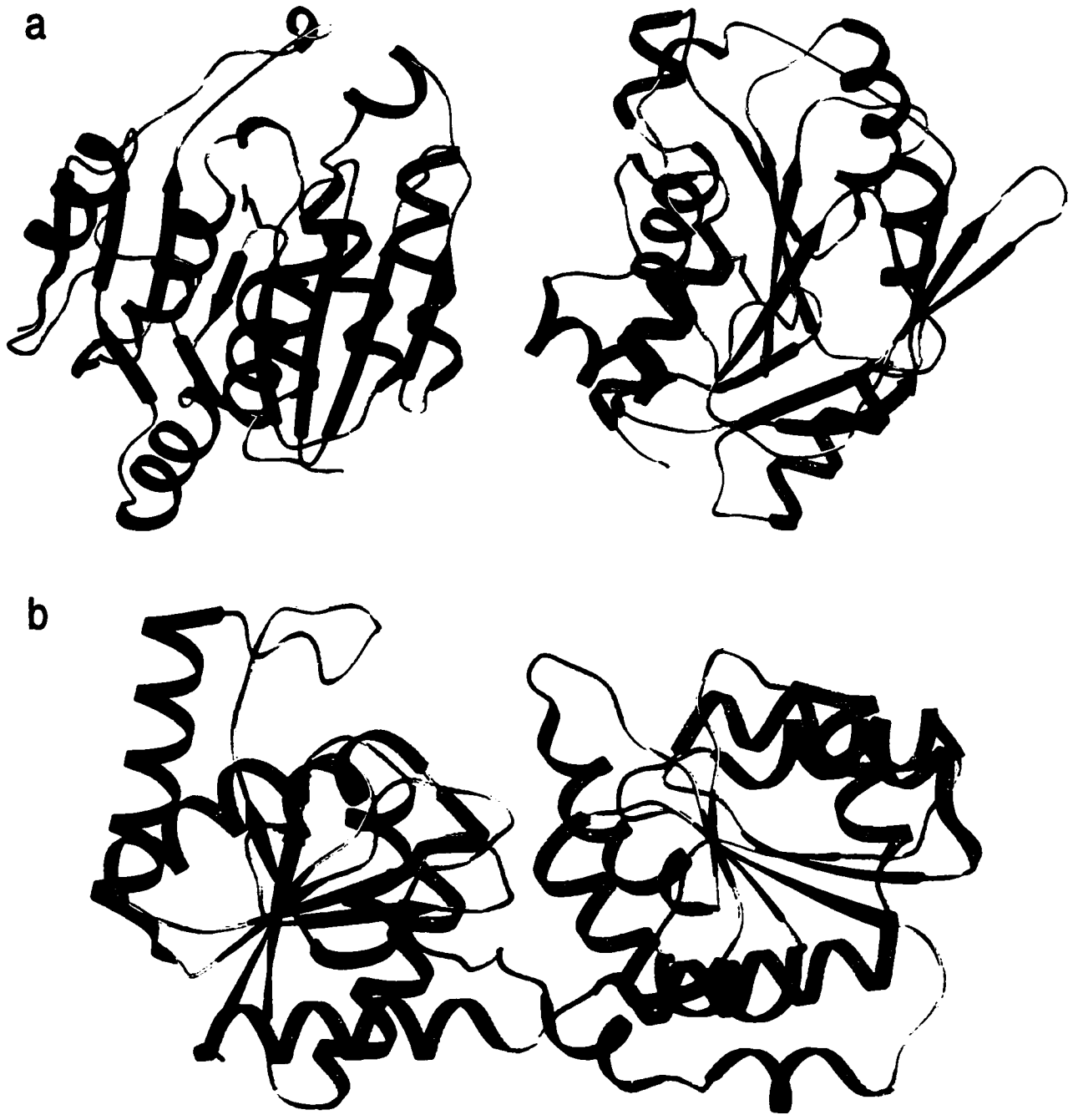

Figure 3. Two types of glycosyltransferase folds. The glycosyltransferase structures solved to date adopt one of two types of fold. Fold B is characterized by two dissimilar domains, one binding the donor substrate and the other binding the acceptor substrate; this fold is represented by SpsA (b; Charnock & Davies, 1999), where the left panel shows the enzymes' overall fold while the right panel is at 90°. The DXD motif of SpsA if formed by Thr97 Asp98 Asp99, with Asp99 shown in red. Fold A is characterized by two similar Rossman fold domains and is represented in a) by MurG (Ha *et al.*, 2000).

defined primarily by three aspartic acid residues. In LgtC, the acceptor binding site is formed only after the binding of UDP-Gal, as the acceptor binds proximal to the galactose moiety. However, in the GnTI structure (Unligil *et al.*, 2000) the ordering of the flexible loop upon UDP binding results in the formation of the acceptor binding site. Conversely, the GtfB (Mulichak *et al.*, 2001) structure suggests that binding of the acceptor structure would order the flexible loops adjacent to the active site, thus forming the UDP binding site.

The determination of the catalytic base or nucleophile in the GT structures is key to the elucidation of the mechanism of action. In most structures, the catalytic base/nucleophile has been identified as either an aspartate or glutamate residue. While the inverting reaction can be easily visualized, the details of the retaining reaction remain unclear. For both α3GalT and LgtC (Boix *et al.*, 2001; Gastinel *et al.*, 2001; Persson *et al.*, 2001), nucleophiles were proposed, though a S_Ni-like reaction seems equally plausible. In this type of reaction, the reactive hydroxyl of the acceptor would provide a nucleophilic attack on the anomeric centre from the same side as the UDP leaving group would depart.

3. CARBOHYDRATE-BINDING ANTIBODIES

I- LOW AFFINITY INTERACTIONS

Generally, antibodies have been observed to have much higher affinity for protein antigens than for carbohydrate antigens, with typical dissociation constants of 10^{-9} M^{-1} for proteins antigens *versus* 10^{-3}-10^{-6} M^{-1} for carbohydrate antigens (Vyas, 1991). Some of the differences in behaviour of anti-carbohydrate and anti-protein antibodies stem from the fact that the antigens are processed differently by the immune system.

When protein antigens are processed, they are presented on the surface of antigen-presenting cells in complex with major histocompatibility complex class II molecules. When T-cells recognize these complexes, cytokines are released causing stimulation of B-cells; this in turn leads to somatic hypermutation to the coding sequences of the variable regions (Janeway & Travers, 1997). The selection of higher affinity antibodies among these new antibodies occurs through a process called affinity maturation, recently reviewed by Janeway & Travers (1997). Unlike protein antigens, carbohydrate antigens are generally T-cell independent, therefore effectively limiting the B-cell response to the production of immunoglobulin M (IgM) antibodies (Janeway & Travers, 1997). The absence of somatic mutation of the rearranged antibody genes leads to the expression of essentially the germline sequences of these proteins, consequently limiting their affinity (Tonegawa, 1983; French *et al.*, 1989).

This can be clearly seen in studies that compared the free and hapten-bound structures of an anti-peptide antibody 48G7 and its germline predecessor (Wedemayer *et al.*, 1997; Yang & Schultz, 1999). The germline antibody was shown to bind the antigen with low affinity by an induced-fit mechanism whereas the mature 48G7 antibody possessed a combining site highly complementary to the hapten, leading to a 30,000-fold increase in binding affinity. Many of the

amino acid residues mutated in the maturation process were not observed to directly contact the antigen, but induced structural changes in the binding domains that resulted in improved binding. This pattern has been seen in other systems (Hawkins *et al.*, 1993; England *et al.*, 1999) and is consistent with the relative contributions of germline and somatic factors to antibody affinity (Tomlinson *et al.*, 1996; Ignatovich *et al.*, 1997).

II- ANTIBODY-CARBOHYDRATE INTERACTIONS

The basis for lowered affinities of antibodies for carbohydrates is not completely clear, though it is likely linked to the types of interactions that typically occur. In the limited number of antibody-carbohydrate complexes determined (Cygler *et al.*, 1991; Zdanov *et al.*, 1994; Jeffrey *et al.*, 1995; Villeneuve *et al.*, 2000), the shape of the antigen and the antibody surfaces are highly complementary but the surfaces buried upon complex formation (250-300 $Å^2$) are much smaller than those involved in antibody-protein complexes (1200-2000 $Å^2$) (Janin & Chothia, 1990; Lo Conte *et al.*, 1999). Buried water molecules can help mediate hydrogen bonds between the carbohydrate and protein by enhancing the antibody-antigen surface complementarity, though this will introduce an entropic penalty.

Most interactions between the antigen-binding antibody fragment (Fab) and the carbohydrate antigens in the current structures involve aromatic residues, as these are frequently found in antibody combining sites (Padlan, 1990). Some structures show interaction of the carbohydrate with a charged carboxylate side chain (Jeffrey *et al.*, 1995; Villeneuve *et al.*, 2000); this type of charged residue has been shown to confer specificity in interactions between carbohydrates and other types of proteins (Vyas, 1991).

In the absence of the affinity maturation process, anti-carbohydrate antibodies are often expressed in humans as pentameric IgM antibodies, which generally show low monomeric affinity for their carbohydrate antigens. This response has likely evolved to recognize the highly multivalent presentations of carbohydrate antigens on many pathogen surfaces, where the low monomeric affinity of antibodies can be transformed into significantly higher avidity due to the presence of 10 antigen-binding sites. Multivalency can also help confer a certain degree of specificity, as all of the binding sites must recognize the same carbohydrate structure.

With any antigen, an entropic penalty may be required to partially or completely desolvate the contacting surfaces, or immobilize the antigen and the antibody complementarity determining regions (CDRs) to form a complex. For protein antigens, this is offset by the high binding affinities of the complex. In carbohydrate antigens, however, these losses become important and are a key factor in binding affinity (Jeffrey *et al.*, 1995).

4. ABO Blood Groups

I- History

Since its discovery in 1901, the ABO(H) histoblood group has become one of the most well-known and studied cell-surface carbohydrate systems. Though only a brief history is given here, more details are readily available (Morgan & Watkins, 2000; Olsson & Chester, 2001; Palcic *et al.*, 2001; Watkins, 2001; Yamamoto, 2001).

a- Early discoveries

In 1900, Landsteiner observed that the serum from certain individuals agglutinates the blood of others (Landsteiner, 1900). Further experiments revealed that the general population could be classified into three groups based on the reactions of their red blood cells with what were later determined to be antibodies (Landsteiner, 1901). It was established that A individuals have the A antigen and anti-B antibodies, B individuals have the B antigen and anti-A antibodies, and O individuals carried neither antigen but have both anti-A and anti-B antibodies. The AB group in which individuals possess both antigens on the red cells but whose serum contains neither anti-A or anti-B antibodies, was described later (von Decastello & Sturli, 1902).

Once the inheritance of the blood groups was established (Epstein & Ottenberg, 1908; von Dungern & Hirszfeld, 1910), Bernstein proposed the three-allele system (Bernstein, 1924), where the *A*, *B* and *O* alleles would account for the four phenotypes A, B, AB and O. Originally, Landsteiner had described group O as the absence of A and B antigens (Landsteiner, 1901) and Bernstein's model held that individuals of genotype AA, BB and AB could not carry the *O* gene. However, evidence surfaced suggesting that O-specific antibodies existed (Schiff, 1927; Eisler, 1930) and that these antibodies cross-reacted with certain A, B and AB red cells (Schiff, 1934;

Witebsky & Klendsshoj, 1941; Morgan & Watkins, 1948). At this time, it was proposed that the O substance be renamed H and that this was the precursor to both the A and B substances (Morgan & Watkins, 1948); the H substance on the surface of erythrocytes could then be attributed to unconverted precursor.

b- ABO(H) antigens

Initially, difficulties in purification prevented the identification of the chemical nature of the blood group antigens on red blood cells, though it was shown that the blood group A active material consists of amino acids and carbohydrate (Landsteiner & Harte, 1940). Walter T. J. Morgan and colleagues established protocols to purify the blood group antigens from cysts of single donors of known blood group (Aminoff *et al.*, 1950; Annison & Morgan, 1952; Gibbons & Morgan, 1954), observing that the composition of the active macromolecules was 85-90% carbohydrate (Morgan, 1960). Four sugars were found: *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose (Fuc) and galactose (Gal) (Landsteiner & Harte, 1940; Bray & Stacey, 1946; Aminoff & Morgan, 1948).

Eventually, the blood group antigens were confirmed to be carbohydrate (Renkonen, 1948; Boyd & Reguera, 1949; Watkins & Morgan, 1952). Hemagglutination, precipitation and degradation inhibition experiments suggested the roles of fucose in H specificity (Watkins & Morgan, 1952), *N*-acetylgalactosamine in A specificity (Morgan & Watkins, 1953; Kabat & Leskowitz, 1955; Watkins & Morgan, 1955) and galactose in B specificity (Kabat & Leskowitz, 1955; Watkins & Morgan, 1955). This was further reinforced upon the isolation of the serologically active terminal disaccharides by acid hydrolysis (Côté & Morgan, 1956; Painter *et al.*, 1962; Painter *et al.*, 1963). The α1,3 linkages of the A and B antigen terminal sugars were shown to be identical and isolation of the H, A and B determinants by alkaline degradation revealed the α1,2 linkage of

Fuc to the subterminal Gal (Painter *et al.*, 1965; Lloyd *et al.*, 1966). A series of enzymatic degradation experiments showed that the biosynthesis of A and B antigens from the H substance required the addition of only one sugar residue and that the sugar added was the immunodominant GalNAc or Gal sugar, respectively (Watkins, 1956; Iseki *et al.*, 1959; Watkins, 1962; Harrap & Watkins, 1964). In 1990, monoclonal antibodies were used to identify the minimal structures for the A, B and H specificities (Figure 4) (Oriol *et al.*, 1990).

Though it was suggested in 1953 that the blood group activity was associated with the glycolipids of red blood cells (Yamakawa & Iida, 1953), it was later shown that the ABH antigens were carried by glycosphingolipids (Koscielak, 1963; Hakomori & Strycharz, 1968) and glycoproteins (Finne *et al.*, 1980).

c- Glycosyltransferases A and B

Initially, the blood group antigens were thought to be the direct products of genes (Haldane, 1937). When the findings of Watson and Crick (Watson & Crick, 1953) showed that the carbohydrate structures could not be the primary products of the *ABO* genes, it was proposed that the gene products control the formation or functioning of the enzymes responsible for the attachment of the immunodominant sugar (Watkins, 1962).

Following Leloir's identification of glycosyltransferases (Caputto *et al.*, 1950; reviewed in Leloir, 1974), it was proposed that the *A*, *B* and *H* genes encoded α3-N-acetyl-D-galactosaminyltransferase, α3-D-galactosyltransferase and α2-D-fucosyltransferase, respectively (Watkins, 1965; Watkins, 1966). Thus, the α2-D-fucosyltransferase would transfer a Fuc to the terminal β-galactose of a precursor oligosaccharide chain, forming the H antigen (Fucα(1→2)Galβ-OR); this product would then be the acceptor structure for the transfer of GalNAc by α3-N-acetyl-D-galactosaminyltransferase (GTA) to form the A antigen (GalNAcα(1→3)[Fucα(1→2)]Galβ-OR), or

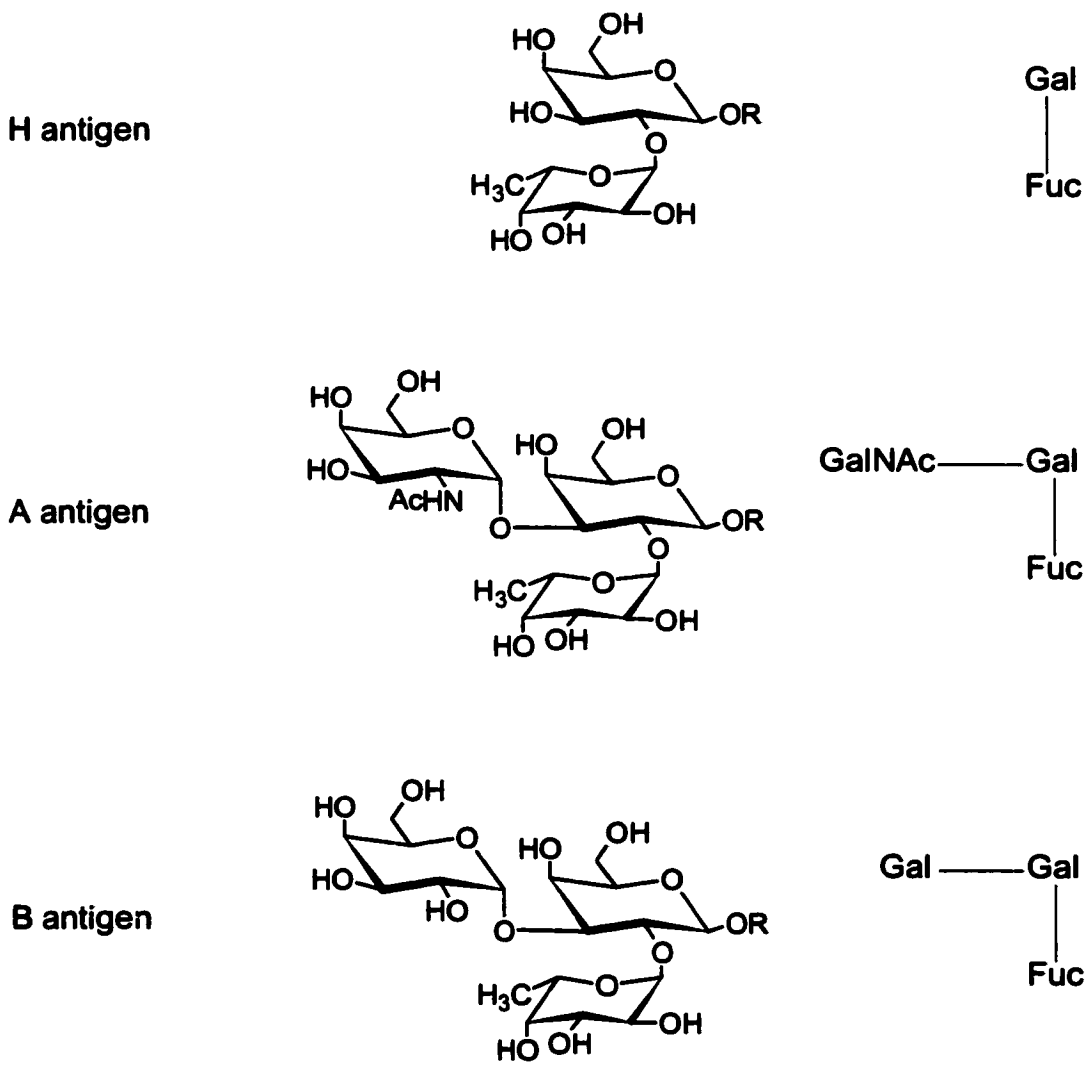

Figure 4. Minimal determinants of the ABH antigens. The minimal structures conferring A, B, or H specificity were determined using monoclonal antibodies. The H antigen, Fucα(1→2)Galβ-OR, is the precursor for the synthesis of GalNAcα(1→3)[Fucα(1→2)]Galβ-OR (A antigen) and Galα(1→3)[Fucα(1→2)]Galβ-OR (B antigen).

of Gal by α3-D-galactosyltransferase (GTB) to form the B antigen (Galα(1→3)[Fucα(1→2)]Galβ-OR) (Figure 5). A series of experiments in the late 1960s verified the proposed biosynthetic pathway (reviewed in Morgan & Watkins, 2000; Watkins, 2001).

Thus, the A and B glycosyltransferases catalyze reactions using almost identical substrates, which differ only in an acetamido *versus* a hydroxyl group on the donor sugars. In fact, the enzymes catalyzed such similar reactions that cross-reactivity could be observed (Yates & Watkins, 1982; Yates *et al.*, 1984; Greenwell *et al.*, 1986; Seto *et al.*, 1997), though significant differences in kinetics were measured (Greenwell *et al.*, 1986; Seto *et al.*, 1997).

In 1990, the application of molecular biology techniques allowed Hakomori's group to clone the *ABO* gene locus (Yamamoto *et al.*, 1990), based on the partial amino acid sequence of the A transferase (Clausen *et al.*, 1990). Their findings revealed that the predicted *A* and *B* genes differed in only seven nucleotide substitutions of the cDNA, resulting in 4 amino acid differences Arg176→Gly, Gly235→Ser, Leu266→Met and Gly268→Ala in GTA and GTB, respectively. These amino acids, often referred to as the "critical" residues, were postulated to account for the nucleotide-sugar specificity of each glycosyltransferase. The *O* gene showed a single nucleotide deletion near the N-terminus resulting in a frame-shift mutation (Yamamoto *et al.*, 1990). This introduces a premature stop codon, resulting in an inactive truncated protein.

Chimeric constructs of the A and B glycosyltransferase genes were produced to explore the importance of each amino acid in the A and B specificities (Yamamoto & Hakomori, 1990). The four differing amino acids between GTA and GTB were found to form the basis of the differential affinity of the enzymes for their donor substrates, UDP-GalNAc and UDP-Gal, respectively. It was determined that amino acid 176 did not affect the donor specificity of the enzymes, while amino acids 266 and 268 were found to be the most crucial for specificity. It was proposed that the loss of flexibility upon mutating residues 266 and 268 from A to B (where A or B shows the source of

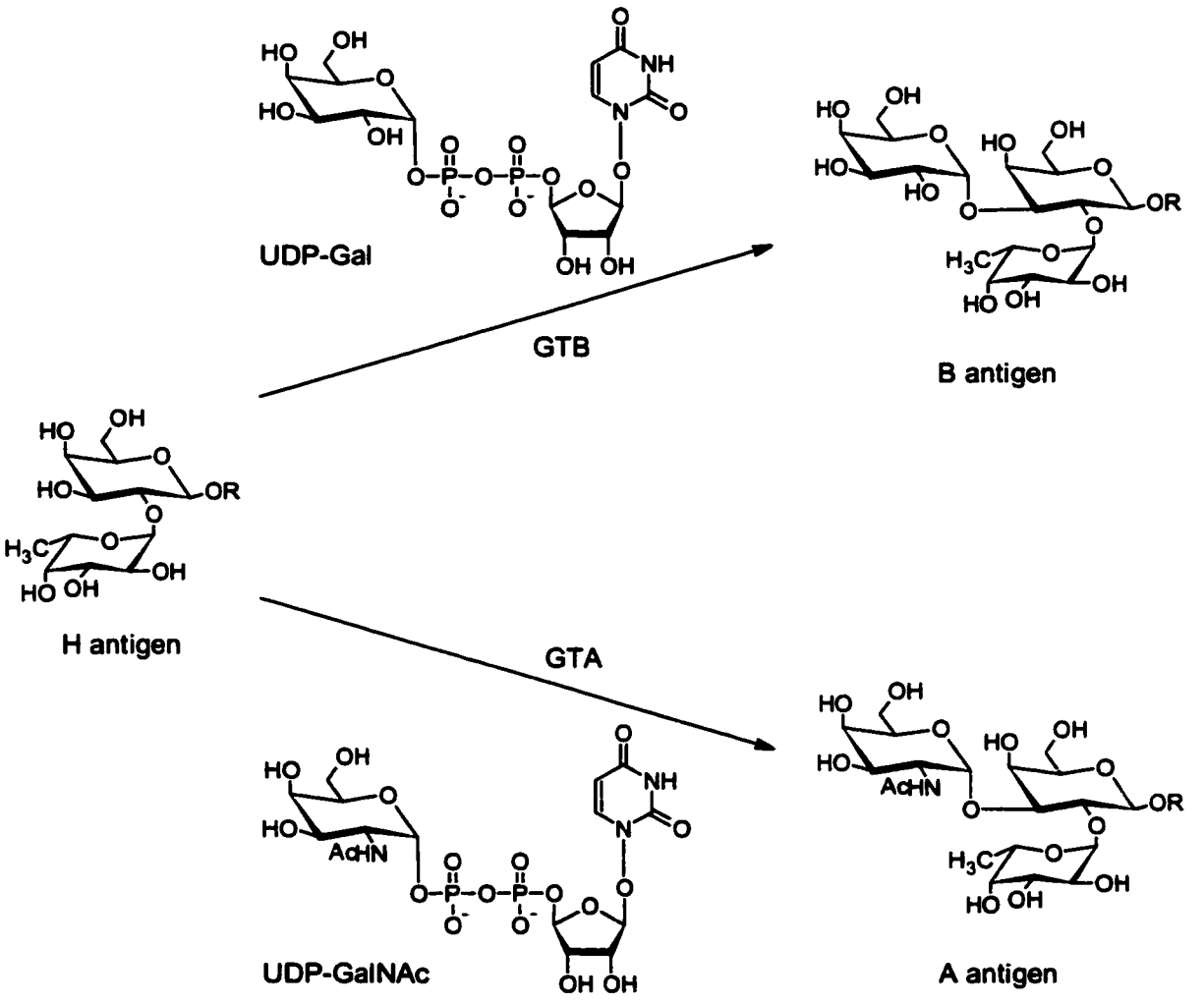

Figure 5. The biosynthesis of the A and B blood group antigens. Fucα(1→2)Galβ-OR (H antigen) is the acceptor substrate for both the A and B glycosyltranferases. GTB (α(1→3)galactosyltransferase, EC 2.4.1.37) transfers Gal from UDP-Gal to the H antigen, forming Galα(1→3)[Fucα(1→2)]Galβ-OR (B antigen). The transfer of GalNAc from UDP-GalNAc to the H antigen is catalyzed by GTA (α(1→3)N-acetylgalactosaminyltransferase, EC 2.4.1.40) and produces GalNAcα(1→3)[Fucα(1→2)]Galβ-OR (A antigen). The A and B antigens differ only in the substitution of an acetamido for a hydroxyl group, respectively, on the terminal sugar.

the critical amino acid) is crucial in determining substrate specificity based on the size difference between the Gal and GalNAc sugars. Further experiments at position 268 revealed that the size and charge of the amino acid side chain at this position was key in determining nucleotide-sugar specificity (Yamamoto & McNeill, 1996).

d- Naturally occurring ABO subgroups

Over the past century, a number of naturally occurring allelic variations of the four blood groups have been reported (reviewed in Olsson & Chester, 2001). Several subgroups of the *A* allele have been identified, the most common being A_2. These individuals express a GTA that displays different activity (Schachter *et al.*, 1973), different pI (Topping & Watkins, 1975) and different lengths (Yamamoto *et al.*, 1992) compared to the wild type enzyme. Numerous *O* alleles were identified in which the nucleotide deletion is retained, but which contain a number of mutants throughout the remainder of the gene (reviewed in Olsson & Chester, 2001). One *O* allele occurring at a significant frequency does not contain the nucleotide deletion, but shows the G268R mutation (Yamamoto *et al.*, 1993b), the side chain change being sufficient to cause total inactivation of the enzyme (Yamamoto & McNeill, 1996; Amado *et al.*, 2000). Most other naturally occurring *ABO* alleles are rare.

Naturally occurring chimeric glycosyltransferases having both A and B activity have also been found. *cis*-AB individuals have been identified that have the amino acids AAAB (where A or B shows the source of the critical amino acid) (Yamamoto *et al.*, 1993a; Fukumori *et al.*, 1996; Ogasawara *et al.*, 1996) or BBAB (Mifsud *et al.*, 2000). B(A) individuals have amino acids BABB and are basically of B phenotype, but their red blood cells react with anti-A antibodies (Yamamoto *et al.*, 1993c). These and other natural variations of the ABO alleles are well documented and have recently been reviewed by Olsson and Chester (2001).

e- Anti-blood group A and anti-blood group B antibodies

The anti-A and anti-B antibodies, first identified by Landsteiner, are present in persons lacking the corresponding antigen. The antibodies are highly specific and capable of unambiguously differentiating between trisaccharide antigens, which differ in only an acetamido *versus* a hydroxyl group on the terminal saccharide residue (Figure 4). They are expressed as IgM antibodies, which typically show low monovalent affinity for their antigens. This is overcome, however, by the IgM pentameric structure that allows for increased valency and clot formation.

II- BLOOD GROUP MISMATCH

The specificity of antibody recognition is demonstrated by the immune response to mismatched tissues. The ABO(H) system constitutes one of the major transplantation antigen systems and is reported to be the most important with regards to solid organ transplantation (Rydberg, 2001). Because the ABH antigens are expressed on almost all the cells in the body, transplantation of tissue containing foreign A/B antigens can result in graft rejection due to recognition of the tissue by the host's own antibodies. Though ABO-incompatible organ transplantation has been achieved (reviewed in Rydberg, 2001), these procedures remain tentative.

Haemolytic transfusion reactions resulting from the mismatch of blood during transfusion cause shortened survival of transfused red blood cells and are often severely toxic or fatal (Capon & Goldfinger, 1995; Seymour *et al.*, 1995). The recognition of foreign blood group antigens on the erythrocyte surface by the recipient's antibodies will cause the rapid agglutination of the red blood cells and complement activation. This can subsequently lead to massive intravascular haemolysis and, if left untreated, death.

III- STATISTICAL ASSOCIATIONS OF THE ABO(H) ANTIGENS

Changes in the blood group and related antigens constitute the major tumour-associated changes of glycosylation (Hakomori, 1999). Human gastric carcinoma, the majority of squamous cell and transitional cell carcinomas display the loss of A/B antigen expression (reviewed in Hakomori, 1999). The loss of A and B antigens has been correlated with the degree of malignancy and metastasis of a number of types of cancer (Davidsohn & Stejskal, 1972; Miyake *et al.*, 1992). Incompatible A expression has also been observed in cancer, where the A antigen is expressed in tumours of blood group B and O individuals. This is a grave problem due to the targeting of host immune surveillance in the early stages of cancer development. Though the mechanism by which these antigens are lost or gained in malignancy remains unknown, changes associated with the regulation of various GT gene expression, post-translational control or catalysis are likely causes (Singhal & Hakomori, 1990; David *et al.*, 1993; Meldgaard *et al.*, 1995; Hakomori, 1999).

In addition to changes of ABH antigen expression in malignancy, several other statistical associations with the blood groups have been reported (reviewed in Garratty, 1995). Among the more established associations is that between the A antigen, thrombosis and myocardial infarction. Among patients with thromboembolic disease, a raised proportion of blood type A individuals has been identified (Mourant *et al.*, 1971); the etiological pathways of cardiovascular disease were suggested to differ according to ABO(H) blood group due to characteristic variations in the cell surface/plasma proteins and blood rheology. The ABO(H) blood groups have also been proposed to contribute to the host defence mechanism and help confer resistance to bacterial or viral infections (Vogel *et al.*, 1960). For example, if a microbe displayed blood group B activity, the anti-B antibodies of type A and O individuals could contribute to the host defence against the infecting microbe.

5. CRYSTALLOGRAPHY

The determination of molecular structures by X-ray crystallography has become a powerful tool for studying structure-function relationships of macromolecules. The principles of crystallography are well documented (Sands, 1969; Glusker & Trueblood, 1972; *International Tables for X-ray Crystallography*, 1974; Blundell & Johnson, 1976; Stout & Jensen, 1989; Woolfson, 1997) and so the following discussion will briefly review basic theory and its application to structure solution methods used in this work.

i- CRYSTALLIZATION

The first step in macromolecular crystallography is the crystallization of the protein of interest. Factors that affect the ability of the protein to crystallize include pH, temperature, solubility and purity of the protein, as well as the effects of various additives (McPherson, 1999). The choice of crystallization method can also influence the crystallization process. It is therefore essential to understand the behaviour of the protein and to carefully observe its reaction to various environments. Finding the optimal conditions for obtaining diffraction-quality crystals may require numerous trials and significant amounts of highly pure protein.

By definition, a crystal consists of atoms arranged in a pattern that repeats periodically in three dimensions (Barrett, 1952). Ideally, as individual macromolecules precipitate, they are similarly influenced by the environment and are added to the growing crystal in an ordered manner. If the rate of deposition is slow enough to allow each molecule to orient itself favourably, a highly ordered crystal is possible.

In an ideal crystal, a set of reference points can be established where all points have exactly the same surroundings and are in identical positions with respect to the repeating pattern.

These are called lattice points. The lattice points can be connected in various ways to form a lattice that defines a three-dimensional series of identical building blocks related by translation. This building block is called the unit cell and represents a template for the entire crystal. The length of the edges (*a*, *b*, *c*) and the angles between the edges (α, β, γ) define the size and shape of the unit cell (Figure 6).

II- X-RAY EXPERIMENT

In order to visualize a molecule, the wavelength of the radiation must be comparable to the features to be resolved. X-rays are well suited to structural studies, as wavelengths of about 1.0 Å are used for X-ray crystallography. This is comparable to the ~1.54 Å length of a C-C bond and therefore allows the individual atoms in the molecular structure to be resolved.

Once crystals of suitable size and quality have been obtained, one can proceed to data collection. In a diffraction experiment (Figure 7), the primary X-ray beam is directed toward the crystal, causing acceleration of the molecules' electrons and diffraction of X-rays. The regular array within crystals is very important in this respect, as the repetition of the unit cells amplifies the diffraction to measurable values. If the conditions for diffraction are met, the scattered X-rays interfere constructively and the resulting diffracted beams are recorded on a detector. The captured image is termed the diffraction pattern. The mathematical Fourier transform of the pattern can lead to electron density maps, which in turn determines the location of the molecule's atoms.

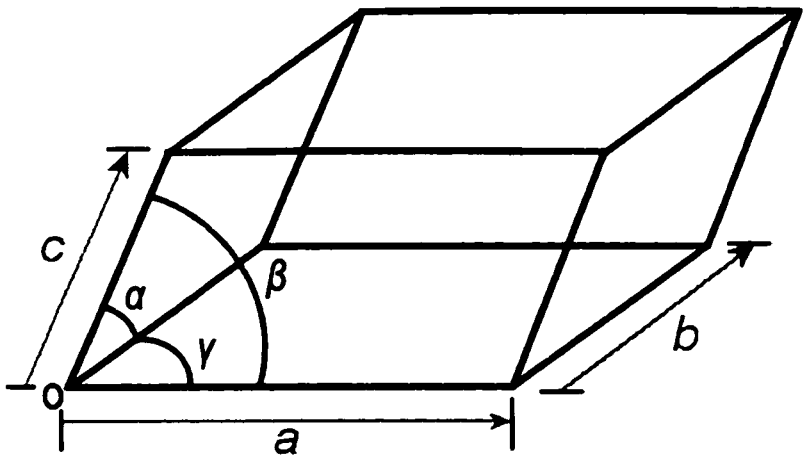

Figure 6. Schematic view of the unit cell. The size and shape of the unit cell are defined by the length of its three independent edges *a*, *b* and *c* and the angles α, β, γ between these edges. The crystal lattice is generated by repetition of the unit cell in three dimensions.

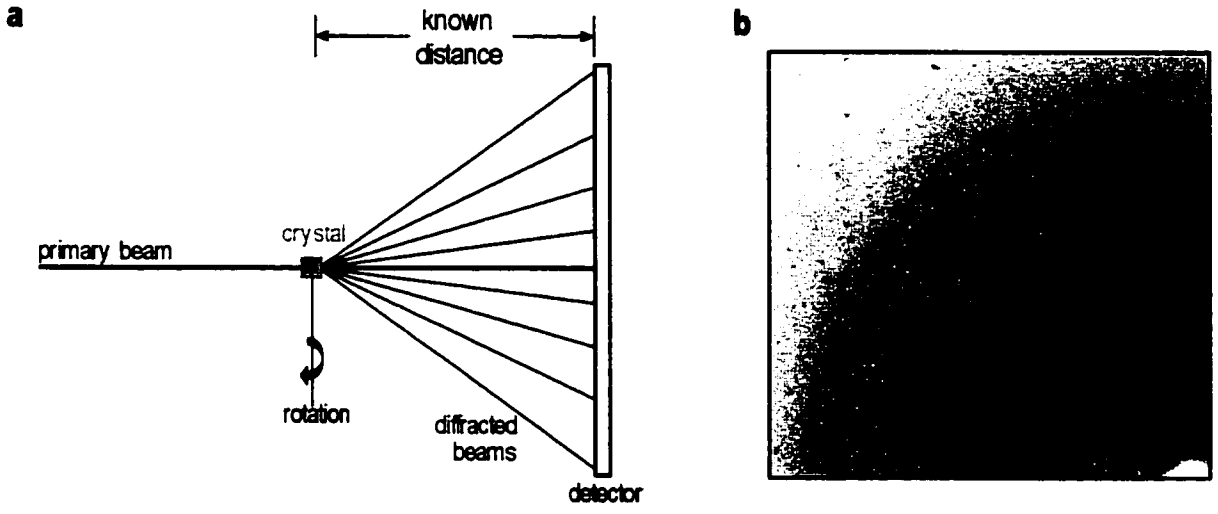

Figure 7. Schematic view of the diffraction experiment. a) A source produces X-rays (red), which are filtered through collimating mirror optics (beamline X8-C), producing a primary beam into which the protein crystal (blue) is centered. The interaction of the incident X-rays with the protein's electrons causes a fraction of the X-rays to be scattered in all directions. In some directions, destructive interference negates the scattering while it is reinforced in other directions. The resulting diffracted beams are recorded on a detector where they appear as spots on a diffraction pattern (b). Though the intensity of each reflection can be measured, the phase information is lost. During the collection of diffraction data, the crystal is methodically rotated to obtain the maximum number of reflections.

III- BRAGG'S LAW

It can be imagined that crystals are separated into sections by sets of parallel planes, where each plane represents the distribution of electron density in a single vector perpendicular to the plane (Figure 8a). These planes are a mathematical construction conceived as an aid in describing the diffraction of X-rays. Each set of planes is defined by Miller indices *hkl*, where *h* is the number of sections into which the *a* axis is cut, *k* is the number of sections into which the *b* axis is cut and *l* is the number of sections into which the *c* axis is cut. Each set of planes is responsible for one observation on the diffraction pattern; each observation is in turn given the corresponding Miller indices.

The Laue equations were derived to describe the diffraction of X-rays from a crystal, though Bragg more eloquently proposed that the diffraction of X-rays from crystals is equivalent to their reflection from various sets of planes in the crystal (Bragg, 1913) (Figure 8b). He established that the conditions for reflection are met when

$$\lambda = 2d\sin\theta \qquad (1)$$

where λ is the X-ray wavelength, d is the spacing between parallel planes in the crystal and θ is the angle of incidence and reflection of the X-rays. This equation is known as Bragg's law.

IV- STRUCTURE FACTORS

Each reflection can be described by a structure factor (**F**), which is composed of its amplitude (*F*) and phase (α). The structure factor of a given reflection (**F**(*hkl*)) is the summation of

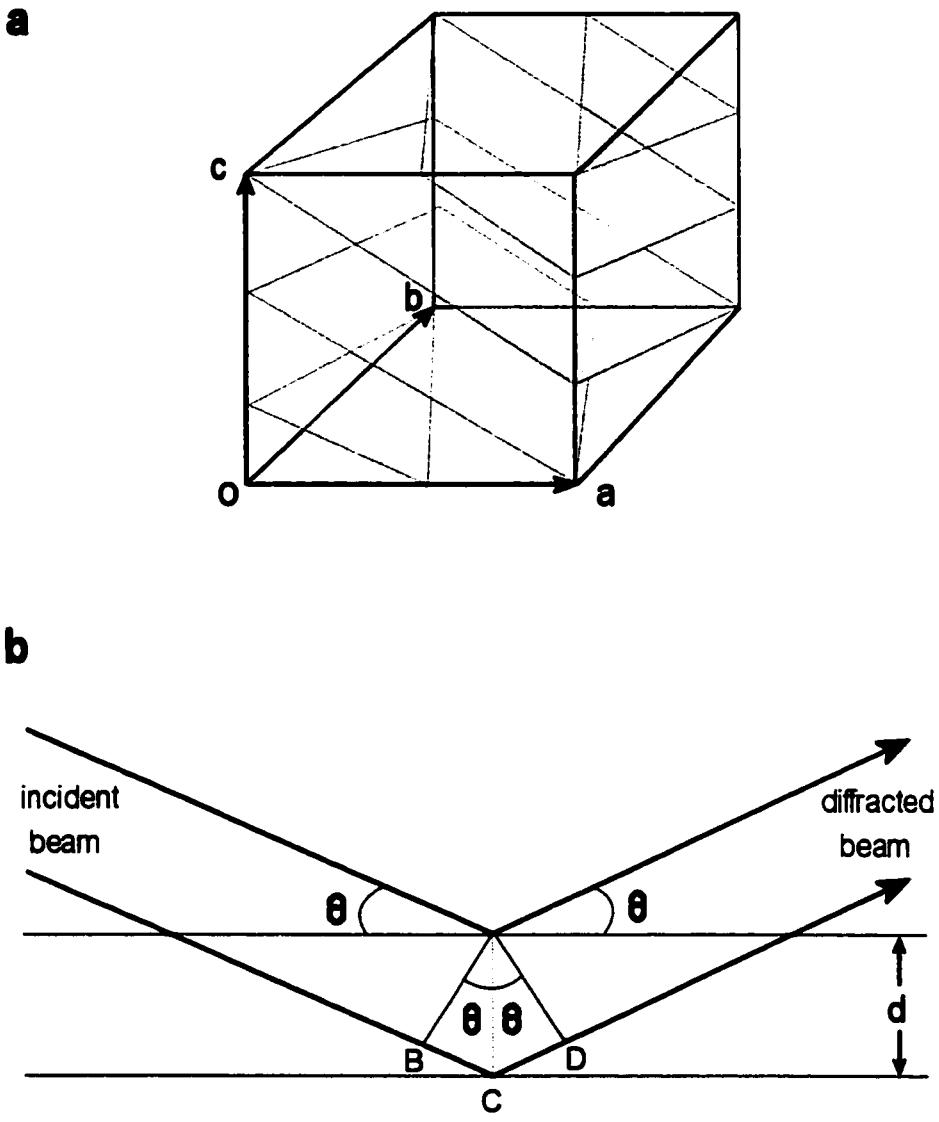

Figure 8. Mathematical construction of crystals and Bragg's law. The crystal can be constructed as sets of parallel planes from which the X-rays are diffracted. The 213 set of planes (blue) is shown in a). This mathematical construction led Bragg to establish the conditions for diffraction. According to Bragg's law the incident X-ray beam (red) strikes adjacent planes, separated by a distance d, at an angle θ. The beam is reflected by the same angle (θ); the difference between the distances traveled by X-rays reflected from adjacent planes is given by BC+CD. Constructive interference of diffracted beams will result when BC+CD is equivalent to the wavelength λ (or a multiple thereof) of the X-rays. This leads to the formal notation of Bragg's law, $n\lambda=2d\sin\theta$.

the contributions (phase and amplitude) of each atom in the unit cell, equation 2 and Figure 9a.

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j e^{2\pi i\left(hx_j+ky_j+lz_j\right)}$$

$$= F(hkl)e^{i\alpha(hkl)} \qquad (2)$$

where N is the total number of atoms and x, y, z represent the coordinates of atom j.

$$\mathbf{F}(hkl) = F(hkl)\cos[\alpha(hkl)] + iF(hkl)\sin[\alpha(hkl)]$$

$$= \mathrm{A} + i\mathrm{B} \qquad (3)$$

Equation 3 gives the structure factor in terms of its real (*A*) and imaginary (*iB*) components, which are indicated on the Argand diagram in Figure 9b.

The geometrical arrangement of planes in the crystal allows one to obtain equal yet opposite reflections. As the crystal is rotated in the X-ray beam, various Bragg planes come into diffracting position. The set of planes given in Figure 10a reflects the incident X-ray; a 180° rotation will allow the reflection of X-rays from the opposite side of the same set of planes. If the wavelength is distant from absorption edges of the atoms in the crystal, the resulting intensities of both reflections will be equal, but their phases will be opposite; this is known as Friedel's law (Figure 10b).

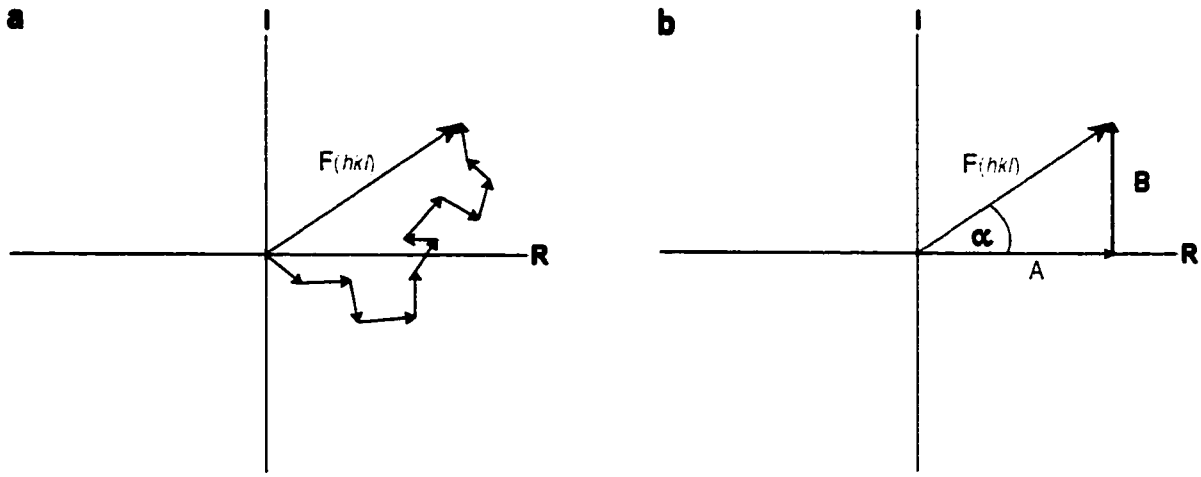

Figure 9. Defining the structure factor. The amplitude F and phase α of a reflection can be described by a structure factor. In a), the contribution of individual atoms (red) to a given reflection are summed to give the structure factor (blue) of that reflection. The Argand diagram in (b) gives the structure factor (blue) in terms of its real (A, green) and imaginary (B, red) components.

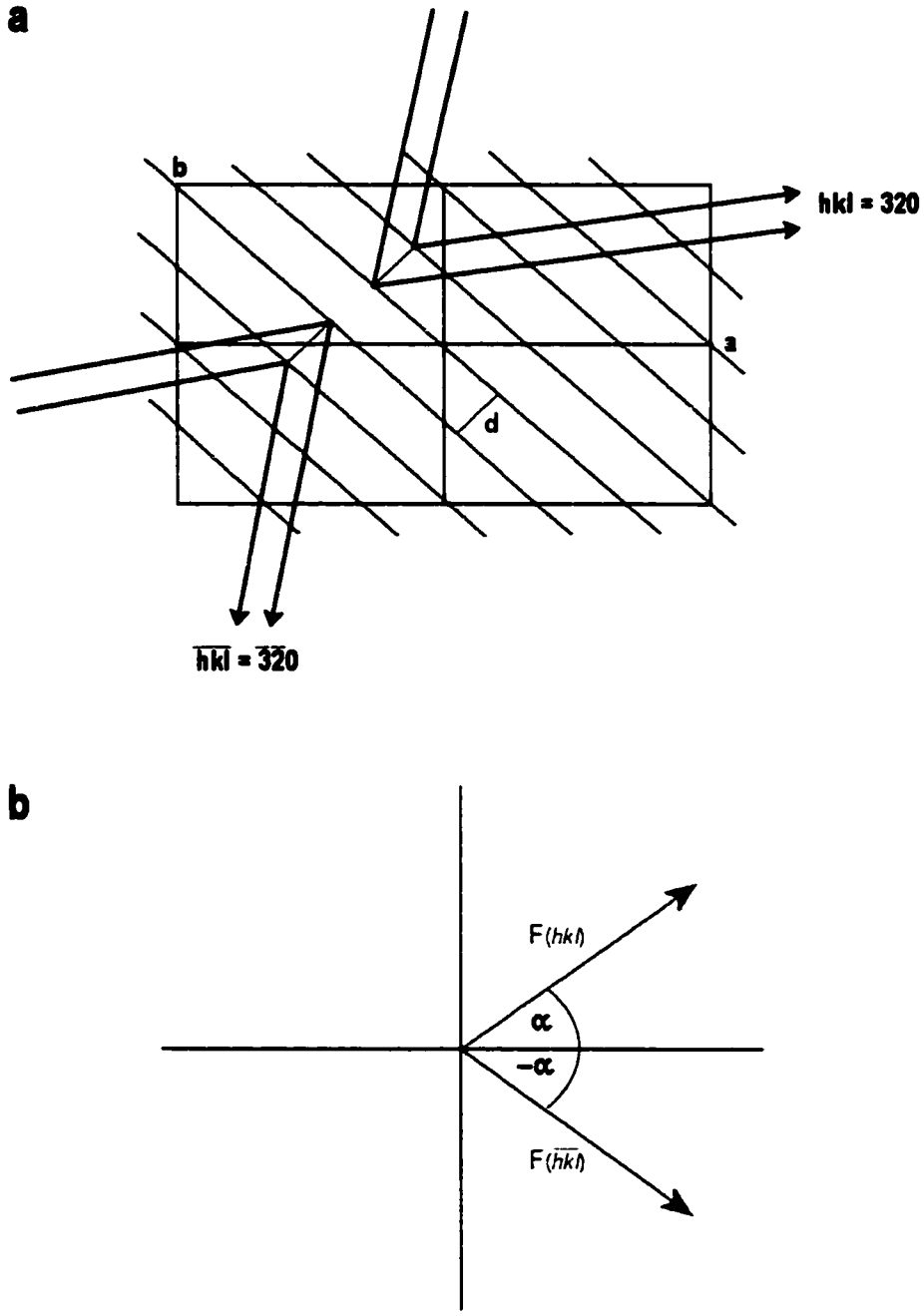

Figure 10. Friedel's law. The geometrical arrangement of Bragg planes (blue) given in a) allows the X-rays (red) to be reflected from opposite sides of the same set of planes upon a 180° rotation of the crystal. Friedel's law (b) states that the intensities of the resulting reflections will be equal ($F(hkl) = F(\overline{hkl})$), but their phases will be opposite ($\alpha(hkl) = -\alpha(\overline{hkl})$).

V- SCATTERING FACTOR

The X-ray scattering power of a given atom for a given reflection is largely dependent upon the number of electrons in that atom and is known as the scattering factor (f_o). It has been shown that the atomic scattering factor is influenced by Bragg angle and thermal motion. Thus, the corrected scattering factor (f) is given by equation 4.

$$f = f_o e^{-B(\sin^2\theta)/\lambda^2} \qquad (4)$$

where B is the correction for thermal motion and $(\sin^2\theta) / \lambda^2$ is the correction term for wavelength and scattering angle. Thermal motion is often called the B-factor and it is related to the mean square displacement (u^2) of a given atom by:

$$B = 8\pi^2 u^2 \qquad (5)$$

VI- PHASE PROBLEM

Once the intensity and phase of the reflections are available, an electron density map can be calculated. Unfortunately, while the intensity of each reflection of the diffraction pattern can be measured, all information on phases is usually lost. The determination of phases then becomes the central issue in crystallography and is known as the "phase problem". Many methods have been used to determine phases in crystallography, though only a few will be discussed here:

multiple isomorphous replacement (MIR) and multiwavelength anomalous dispersion (MAD) and molecular replacement (MR).

a- Patterson function

The Patterson function (Equation 6) is a very useful tool for solving the phase problem. Its calculation does not require phases, rather it utilizes the intensities (amplitudes squared) of each reflection as Fourier coefficients.

$$P(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|\mathbf{F}_{hkl}\right|^2 \cos 2\pi(hx + ky + lz) \qquad (6)$$

where x, y, and z are atomic coordinates. The these vectors can be plotted as a map, which exhibits a centre of symmetry and shows peaks that refer to interatomic distances. The distance between corresponding atoms in symmetry-related molecules of the unit cell have one or two constant coordinates, which leads to particular concentrations of vector points known as Harker sections. The information obtained from Harker sections of Patterson maps is useful for solving molecular structures by many methods, including MIR, MAD and MR.

b- Isomorphous replacement

In the early 1950s, Perutz and colleagues successfully established that isomorphous replacement could be used to solve the phases for protein structure analysis (Green *et al.*, 1954). In this method, a heavy atom is introduced into the protein crystal, most commonly by soaking. These electron-dense atoms will provide subtle but observable changes in the diffraction pattern.

By measuring the intensities of the native and derivative diffraction patterns, the position of the heavy atoms can be calculated and the phase problem resolved. The use of several such derivatives for phasing is called MIR. The general theory of the method follows.

The structure factor for the derivatized crystal (F_{PH}) becomes

$$F_{PH} = F_P + F_H \tag{7}$$

where F_P and F_H are the structure factors for the native protein and the heavy atoms, respectively (Figure 11). Thus if F_P and F_{PH} are measured, F_H can be calculated. Using the value F_H, a Patterson map can be computed from which the positions of the heavy atoms can be calculated and α_H can be determined. Figure 12a shows how the values of F_{PH}, F_H and α_H can be combined to determine the phase of F_P; the solution remains ambiguous, as two values are possible. To resolve this ambiguity, it is often required to utilize a second derivative (Figure 12b); this will lead to a single calculated value for α_P. The use of more than two derivatives is often necessary to reduce errors and obtain more accurate estimates of α_P.

For this method to be successful, however, the heavy atom chosen must neither alter the unit cell parameters, nor the arrangement of atoms within the cell. If these parameters change by more than 1%, the lack of isomorphism is generally considered too great and the data are probably unusable. Other problems with regards to MIR include: incomplete occupancy of heavy atom sites, non-specific binding of the heavy atoms to the protein surface, radiation damage accelerated by the presence of heavy atoms, and reduced diffraction by derivatized crystals. Another significant drawback of MIR is that numerous crystals are required for the experiment in order to accommodate the above limitations.

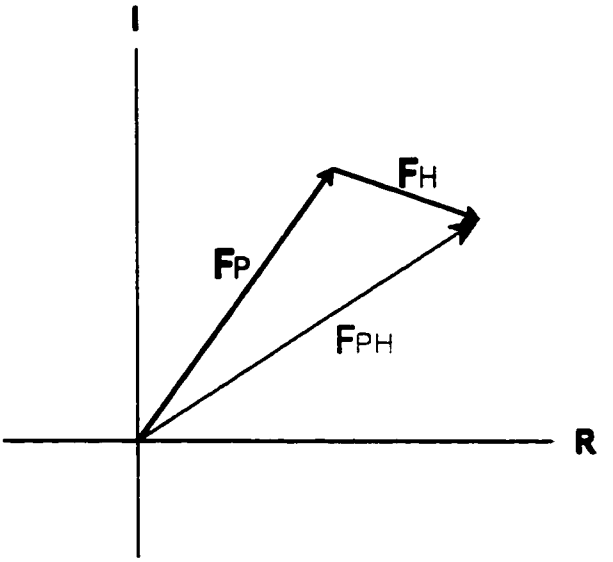

Figure 11. Contributions to the structure factor in isomorphous replacement. The contributions of the native protein (F_P, black) and the heavy atoms (F_H, red) to the observed structure factor (F_{PH}, blue) are shown. The knowledge of F_H leads to the calculation of the phase of F_P.

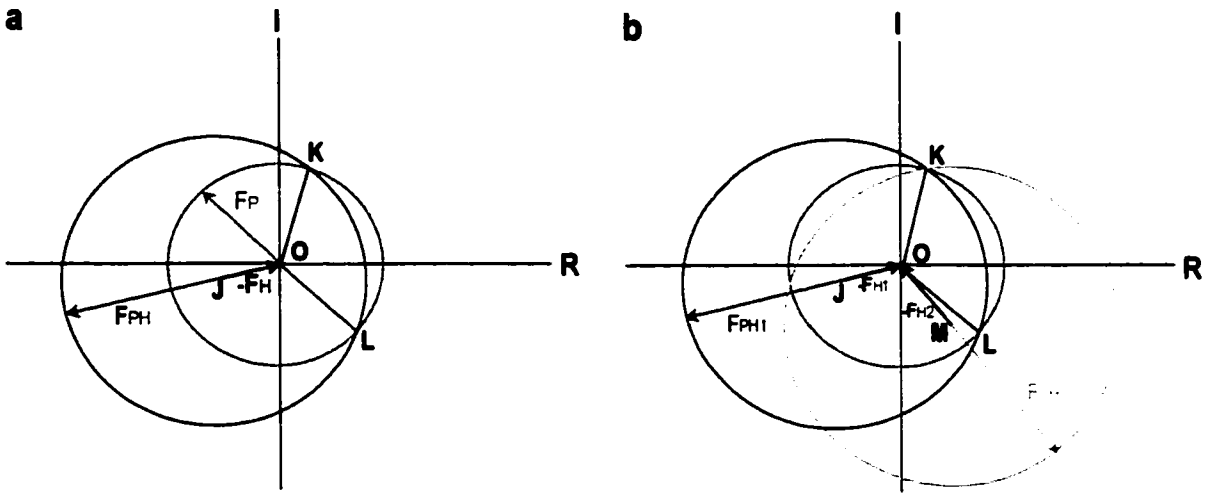

Figure 12. Harker constructions for the phase calculation using isomorphous replacement. The single isomorphous replacement situation shown in Figure 8 is represented in a). Subtraction of the heavy atom contribution (F_H, red) from the origin places the center of a circle of radius F_{PH} (green) at J. A circle of radius F_P (blue), centered about O, leads to two intersections and thus two possible values of F_P (OK and OL). This ambiguity can be resolved by MIR, shown in b). The contribution of a second heavy atom (F_{PH2}, yellow) is drawn with its center at M, leading to the unequivocal solution of F_P (OK). Unfortunately, this method is often complicated by lack of isomorphism and other errors that can occur in the experiment.

c- Multiwavelength Anomalous Dispersion

In the early 1990s, a second method for providing phase information became more common, that of multiwavelength anomalous dispersion (MAD). Here, the phase information is obtained from the anomalous scattering of X-rays by a heavy atom whose natural absorption frequency is close to the incident wavelength. This method became increasingly feasible as finely tuneable synchrotron radiation became more readily available to macromolecular crystallographers.

In order to understand how MAD works, it is first important to understand the basic phenomena that allow it. When a crystal is exposed to X-ray, the incident photon accelerates the atoms' electrons and scattering of the X-rays occurs. When the energy of the incident X-rays approaches the energy of an electronic transition, the energy of the absorbed X-ray promotes the transition of an electron from its ground state to an excited level; this transition amplifies the acceleration of the electron and results in anomalous scattering. The energies at which these transitions occur are characteristic to each type of atom and are known as absorption edges. The atoms found in macromolecules (such as C, N, O, S, P and H) have no electronic transitions in the wavelength range usually used in crystallography (generally around 1Å). Most heavy atoms, however, do possess edges corresponding to transitions at their K, L, and/or M edges. These are the atoms that are introduced into the crystal for MAD experiments.

Selenium. The most commonly used heavy atom for MAD is selenium (Se) due primarily to the pioneering efforts of Hendrickson and colleagues. This group developed a reliable method for the incorporation of selenium-labelled methionine (SeMet) into recombinant proteins (Yang *et al.*, 1990; Hendrickson, 1991). The amino acid analog is introduced by blocking methionine biosynthesis in the cells and incorporating SeMet in the growth medium. Methionine was chosen

as a target because its hydrophobic side chain is generally buried and well ordered; in addition, the sulphur atom can be substituted by selenium. The natural abundance of methionine in proteins (approximately 1 in 59 residues) yields a strong MAD signal.

In order to establish the characteristic absorption patterns of heavy atoms, theoretical absorption spectra can be calculated for isolated atoms from quantum mechanical equations (Figure 13a). Invariably, the details of the calculated spectra are incorrect near the absorption edges as the models break down and the calculations are incorrect. In addition, the effects of the chemical environment and the oxidation state of the heavy atom can cause shifts in energy of the edge. Though the calculated spectra can give an approximation for absorption edges, experimental spectra (Figure 13b) are essential for identification of fine detail. The experimental spectra are taken by measuring the fluorescence of the crystal as a function of X-ray energy. The values of fluorescence are directly related to the absorption of X-rays, as it is caused by the return of excited electrons to their ground states.

Anomalous scattering factor. Each heavy atom will exhibit characteristic wavelength-dependent scattering behaviour. The scattering of absorbed X-rays by the heavy atoms when the wavelength approaches an absorption edge is anomalous in the sense that correction terms must be applied to the scattering factor (f_0). The anomalous scattering factor (f_a) can be broken down into dispersion (f') and absorption ($\Delta f''$) components. $\Delta f'$ and $\Delta f''$ are the real and imaginary components of anomalous scattering, respectively (Figure 14a). Because $\Delta f'$ is real, it is collinear with f_0; for this reason, these variables can be combined into one term, f', the dispersive component (Equation 8).

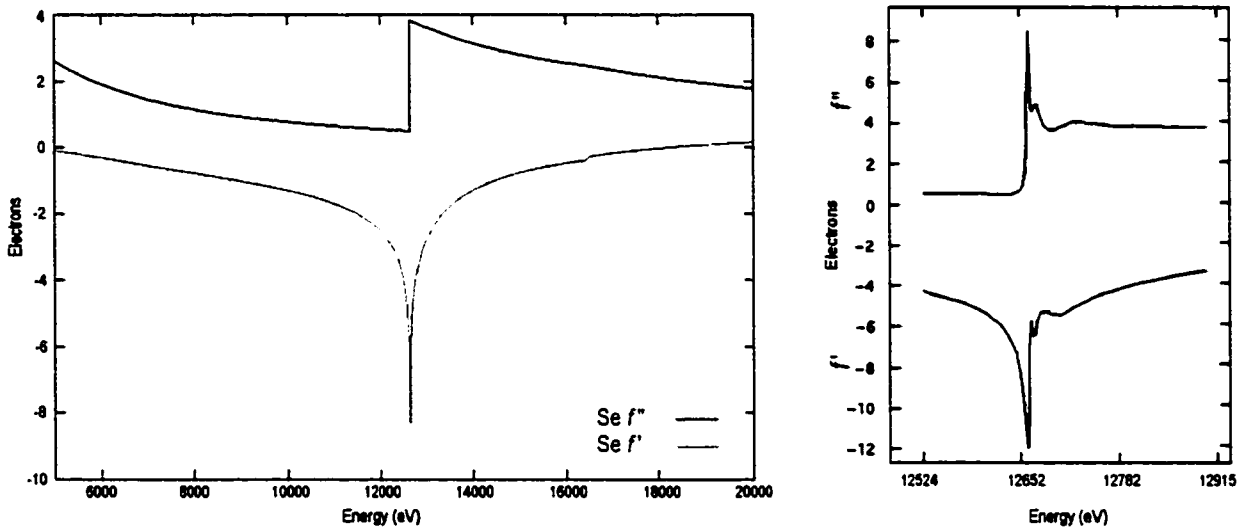

Figure 13. Anomalous absorption spectra of Selenium. The theoretical values of *f'* and *f"* for Se can be calculated using quantum mechanics equations and plotted, as shown in a). The experimental values must be determined by recording absorption spectra. This is accomplished by recording the fluorescence emitted by the crystal while it is exposed to a range of X-ray wavelengths. A high resolution scan for selenium is shown in b). The value of *f"* is related to fluorescence while the spectrum for *f'* is calculated from the *f"* values using the Kramers-Kronig equation. Units along the horizontal axis are the energy (E) in eV and can be related to the wavelength by the equation E (keV)=12.3985/λ.

$$f_a = f_o + \Delta f' + i\Delta f''$$

$$= f' + i\Delta f'' \qquad (8)$$

Figure 14 shows that the relationship between the dispersive and absorption components remains constant ($\pi/2$), even as the phase angle changes. For light atoms, the corrections for anomalous scattering are very small and usually ignored. In the case of heavy atoms, these contributions to the scattering factor can become significant.

Effect on structure factors. The effects of the anomalous scattering components on the structure factor for a given reflection can best be described by Argand diagrams (Figure 15). In this situation it is clear that $\mathbf{F}(hkl) \neq \mathbf{F}(\overline{hkl})$ and that $\alpha(hkl) \neq -\alpha(\overline{hkl})$; therefore, Friedel's law is broken in the case of anomalous scattering.

MAD approach. The differences in intensities between $\mathbf{F}(hkl)$ and $\mathbf{F}(\overline{hkl})$ at the same wavelength (Bijvoet differences) and those of the same $\mathbf{F}(hkl)$ at different wavelengths (dispersive differences) provide phase information for the anomalous scatterers. The basic concept in MAD can be demonstrated by an Argand phase diagram (Figure 16). If the positions of the anomalously scattering atoms in the unit cell can be determined, the corresponding phase angle α_A can be calculated. The most common way of solving the positions of the anomalously scattering atoms is by calculation of a Patterson map using Bijvoet differences, $\Delta\mathbf{F}$ (equation 9), which are space group dependent.

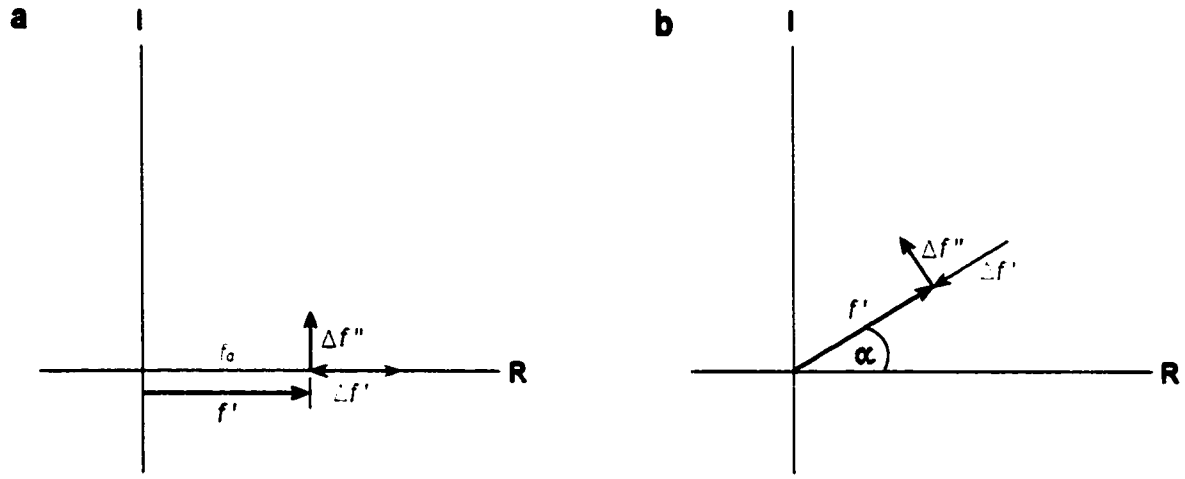

Figure 14. The effects of anomalous scattering on the scattering factor. a) The real component of the anomalous scattering factor ($\Delta f'$, green) is typically negative and collinear with the normal scattering factor (f^o, blue); these combined into the vector f' (black). The imaginary component of the anomalous scattering factor ($\Delta f''$, red) has a phase angle $\pi/2$ ahead of $\Delta f'$. b) The relationship between the various components of the scattering factor remains constant, regardless of phase angle (α).

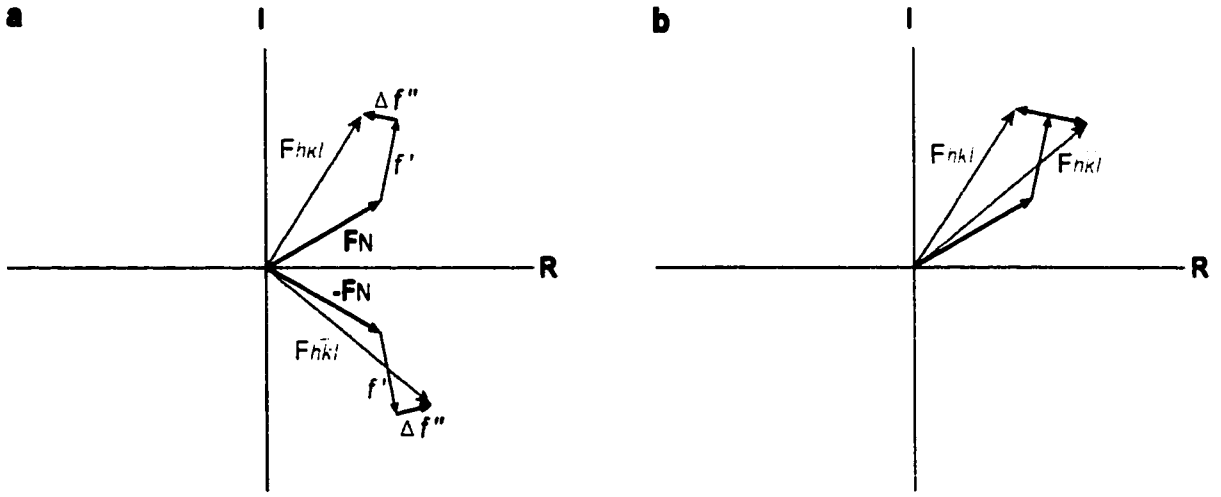

Figure 15. The effects of anomalous scattering on the structure factor. a) The contribution of f' (green) and $\Delta f''$ (red) are added to the resultant of the scattering from atoms without dispersion (F_N, black). The effects for $F(hkl)$ and $F(\bar{h}\bar{k}\bar{l})$ (blue) are both shown. b) $F(\bar{h}\bar{k}\bar{l})$ is reflected across the real axis to emphasize that Friedel's law does not hold for this case because $F(hkl) \neq F(\bar{h}\bar{k}\bar{l})$ and $\alpha(hkl) \neq -\alpha(\bar{h}\bar{k}\bar{l})$.

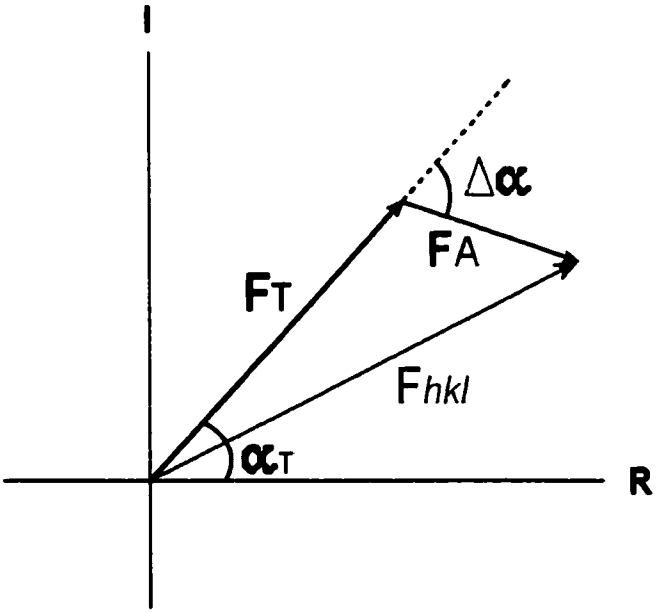

Figure 16. MAD phase diagram. The total observed scattering (**F**(*hkl*), blue) at a given wavelength is the result of the summation of the normal scattering contribution of all atoms (**F**T, black) and the anomalous scattering component of all atoms (**F**A, green). By calculating the phase angle (αA) for **F**A, an estimate for the phase of **F**T (αT)can be obtained.

$$\Delta F = |F^+| - |F^-| \qquad (9)$$

where

$$|F^+| \equiv |F hkl| = |F \bar{h} k \bar{l}| \qquad |F^-| \equiv |F \overline{hkl}| = |F h \bar{k} l|$$

Such a map should contain peaks corresponding to the interatomic vectors between pairs of anomalously scattering atoms.

Simultaneously, the phasing equations can be used to generate estimates for the normal scattering component of all atoms (F_T) and the difference in phase angle between normal and anomalously scattering atoms ($\Delta\alpha$). In order to obtain initial estimates for F_T and $\Delta\alpha$, the MAD phasing equations established by Karle in the early 1980s (Karle, 1980) can be used. Hendrickson (Hendrickson *et al.*, 1985) reworked the equation to the form seen in equation 10.

$$\left|{}^{\lambda}F(hkl)\right|^2 = F_T^2 + a(\lambda)F_A^2 + b(\lambda)F_T F_A \cos\alpha + c(\lambda)F_T F_A \sin\alpha \qquad (10)$$

where

$$a(\lambda) = \left(\Delta f''^2 + \Delta f'^2\right)/f^o$$

$$b(\lambda) = 2\left(\Delta f'/f^o\right)$$

$$c(\lambda) = 2\left(\Delta f''/f^o\right)$$

The above can be regarded as a system of simultaneous equations from which the quantities F_A, F_T and $\Delta\alpha$ can be obtained and where every measurement of F(*hkl*) at some wavelength would give one instance of the equation. To obtain values for the three unknowns, at least three observations are needed. To fulfill this requirement, it is necessary to collect data

using at least two wavelengths, as each wavelength yields two measurements, **F**(*hkl*) and **F**($\overline{hkl}$). It is preferable to collect data using at least three wavelengths to over-determine the values.

The estimate for $\Delta\alpha$ and the calculated value for α_A, can be combined to yield α_T, the phases of the normally scattering atoms, according to equation 11.

$$\alpha_T = \alpha_A + \Delta\alpha \qquad (11)$$

Fourier transforms using the amplitude of F_T and phase α_T will yield electron density maps for all atoms in the crystal. These maps provide a good approximation of the protein atom positions and can usually be used to trace the protein main chain and side chains, yielding an initial protein model.

Usually, three wavelengths are used in a MAD experiment: that at the peak of absorption, at the inflection point of the edge and at an energy remote from the peak. Data collected at the peak of absorption (maximal $\Delta f''$) will yield the maximal Bijvoet signal while data at the inflection point (maximal $\Delta f'$) will optimize the maximal dispersive signal. This latter signal is further optimized by collection of data at a remote wavelength.

There are several approaches to computing the values of all of these unknowns from the reduced data and a variety of programs are available for this purpose. When all the unknowns have been given initial values, one can calculate electron density and proceed with modelling and refinement of the protein structure.

Advantages of MAD. The advent of MAD as a tool for structure solution came at a time of great advances in crystallography, all of which contributed to its popularity. First and foremost, the increased availability of high intensity, finely tuneable X-rays from synchrotron sources made the MAD method possible to the crystallographer. Cryocrystallography introduced a means to protect

the crystals from radiation damage resulting from these intense beams, and improved the quality of the diffraction data. Further, improvements in detector technology and data collection software have also greatly increased the accuracy of data collection. All of these factors allow for the collection of better diffraction data in shorter time, making the application of MAD possible.

The use of a single crystal to collect all of the data prevents problems of non-isomorphism and minimizes systematic errors, resulting in more accurate phase angles. The accuracy of the phases is further increased by the strength of the anomalous signal at higher resolution. Both of these features lead to higher quality electron density maps, which make the model building easier and faster.

d- Molecular replacement

An alternative method for solving the phase problem is to estimate the phases using an atomic model. It is therefore important that the structure chosen as the search model be highly homologous to the molecule of interest. Sequence identity is generally thought to be correlated to the level of resemblance of two structures and values of at least 40% is usually required for successful molecular replacement experiments.

The MR approach requires orienting and positioning the search model in the unit cell, which often makes use of the Patterson function. The rotation function exploits the intramolecular vectors of the Patterson map, which depend on the orientation of the molecule. The translation function makes use of intermolecular vectors, which rely on the orientation and position of the molecule in the unit cell. The resulting rotation angles and three-dimensional translation place the search model in context; the next step is to transform to search model into the protein under study by inserting the correct amino acid sequence.

METHODS

In this chapter, the methods used to determine the various structures will be presented in two parts. The first part will describe experiments pertaining to the GTA and GTB proteins while the second will present methods for the anti-blood group A Fv structures. Unless otherwise stated, the work described in this section has been done by me.

PART I: GTA & GTB

1. PROTEIN EXPRESSION AND PURIFICATION

Gene constructs for GTA, GTB and GTB SeMet were prepared as previously described (Seto *et al.*, 1995; Seto *et al.*, 1997; Seto *et al.*, 1999) by Dr. Nina O.L. Seto at the Institute for Biological Sciences (IBS), National Research Council (NRC) of Canada. Table 2 shows the various constructs of GTA and GTB genes produced for crystallization. The amino acid sequence corresponding to the GTA or GTB sequence (Yamamoto *et al.*, 1990) was used to construct a synthetic gene using the *E. coli* preferred codon usage and inserted into the pUCE8 expression plasmid. The plasmids were then transformed into BL21 *E. coli* cells.

Table 2. Various GTA and GTB constructs prepared for crystallization trials

Protein	Amino acids	Notation
GTA	63-355	GTA-10
GTB	63-355	GTB-10
GTB	53-355 + ompA, cmyc (N-terminus); His_6 (C-terminus)	GTB
GTB	53-355	GTB-mut
GTB	58-355	GTB-15
GTB	58-355 with C80S, C196S, C209S	GTB-15 M3
GTB	68-355	GTB-5

Protein expression and purification were carried out as previously described by Dr. Monica Palcic, University of Alberta, (Seto *et al.*, 1995; Seto *et al.*, 1997; Seto *et al.*, 1999; Seto *et al.*, 2000). Briefly, GTA and GTB recombinant proteins were produced by growing the cells in Terrific broth medium (12 g tryptone, 24 g yeast extract, 4 ml glycerol/L culture) containing M9 salts supplemented with 0.4% casmino acids, 5 μg/ml vitamin B-1 and 100 μg/ml ampicillin. The SeMet derivatized GTB was produced by expressing either the GTB-10 or GTB-15 constructs in B834 *E.*

coli with SeMet (Sigma, Oakville, ON) added to the growth medium. Cells were grown at 30°C with vigorous shaking and induced by addition of 1 mM isopropylβ-D-thiogalacto-pyranoside at OD_{600} 0.6-0.8. The cells were harvested 21-24 hr after induction, were lysed by French Pressure cell, spun down (36,000 rpm, 1 hour) and the supernatant passed through a SP Sepharose Fast Flow column with 50 mM MOPS, pH 7.0, 0.5 M NaCl, 1 mM dithiothreitol (DTT). The active fractions, identified using the assay described in Seto *et al.* (2000), were combined and loaded onto a UDP-hexanolamine Sepharose column in the presence of manganese ion. GTA or GTB was eluted from the column with 50 mM MOPS, pH 7.0, 0.5 M NaCl, 1 mM DTT, 5 mM $MnCl_2$, 4~10 mM UDP. Fractions were then assayed and those containing enzyme were dialyzed and concentrated. Stock solutions for GTA-10 or GTB-10 contained 40-60 mg/mL protein, 20 mM N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) buffer pH 7.5, 50 mM NaCl, 2 mM $MnCl_2$. Stock solutions for GTB and GTB-5 were similar, with concentrations of ~25 mg/mL. The GTB-10 SeMet and GTB-15 SeMet stock solutions also had concentrations of ~25 mg/mL and contained 5-10 mM DTT.

2. CRYSTALLIZATION

Dr. Svetlana Borisova, University of Ottawa, crystallized all GTA and GTB proteins as well as their various complexes. Initial crystallization trials for GTB, GTB-5 and GTB-mut proteins proceeded using the hanging drop vapour diffusion method, where drops contained 10% polyethyleneglycol (PEG) 4000, 75 mM $(NH_4)_2SO_4$, 0.37 M sodium acetate (NaAc) pH 4.6 and 5-6 mg/mL protein and were suspended over a 1 mL reservoir (20% PEG 4000, 0.15M $(NH_4)_2SO_4$, 0.75 M NaAc pH 4.6).

Crystals of GTB-15 and GTB-15 M3 were obtained from drops containing 6-7% PEG 4000, 50 mM $(NH_4)_2SO_4$, 25 mM NaAc pH 4.6 and 5-6 mg/mL protein, suspended over a 1 mL reservoir (12-14% PEG 4000, 0.1 M $(NH_4)_2SO_4$, 0.05 M NaAc pH 4.6). Complexes with uridinediphospho-α-D-galactose-mercury (UDP-Gal-Hg) were obtained by co-crystallization or soaked with < 1 mM UDP-Gal-Hg. Some crystals were also soaked in 1-20 mM methylmercuryacetate (MMA).

Crystals of the GTA-10, GTB-10 and GTB-10 SeMet enzymes were grown as previously described (Patenaude *et al.*, 2002) at 18°C using the hanging drop vapour diffusion method. GTA-10 and GTB-10 were crystallized in 10 µL drops containing 6-8 mg/mL GTA or GTB, 70 mM N-(2-acetamido)-2-iminodiacetic acid (ADA) pH 7.5, 50 mM NaAc pH 4.6, 40 mM NaCl, 5-8 mM $MnCl_2$, 2.5% 2-methyl-2,4-pentanediol (MPD), 5% glycerol, 2% PEG 4000 and 0.3-0.5 mM 3-chloro-Hg-2-methoxy-propylurea that were suspended over 1 mL reservoir solution (50 mM ADA pH 7.5, 10 mM $MnCl_2$, 100 mM $(NH_4)_2SO_4$, 5% MPD, 10% glycerol and 8-10% PEG 4000). GTB-10 SeMet crystals were grown in the same conditions, except that glycerol was not added to the drops or the reservoir. Complexed forms of GTA-10 or GTB-10 were obtained by soaking the GTA or GTB crystals in a mixture of mother liquor (30 mM ADA pH 7.5, 30 mM NaAc pH 4.6, 5 mM $MnCl_2$, 10% PEG 4000) and substrate. Fucα(1→2)Gal-β-O-$(CH_2)_8$-CO_2CH_3 (H antigen, 7 mM) and UDP (10-

15 mM) were generally soaked for 5 days while other donor structures (UDP-Gal, UDP-GalNAc, UDP-Glc, UDP-GlcNAc) were soaked for 1-5 hours.

Unless otherwise stated, crystals were frozen in propane for storage and transportation. The cryoprotectant solution consisted of 30% glycerol added to the mother liquor.

3. DATA COLLECTION AND REDUCTION

I- DATA COLLECTION EQUIPMENT

Diffraction experiments were undertaken at various establishments with X-ray data collection capabilities, including beamlines X4A and X8C of the National Synchrotron Light Source (NSLS) at Brookhaven National Labs (BNL), and a home source at Queen's University. The equipment available at each of these establishments at the time of collection is outlined in Table 3.

Table 3. Data collection equipment used for GTA, GTB, GTB-SeMet and complexes

	Equipment	**Manufacturer**
	X8C (NSLS)	
Source	Synchrotron	-
Detector	ADSC Quantum 4R	Area Detector Systems Corp., Poway, CA
Base	MAR single-axis goniostat (June '99-March '01)	MAR Research, Evanston, IL
	Single-axis w/ 2θ arm (April '01-November '01)	Crystal Logic, Los Angeles, CA
Low temperature device	Oxford Cryosystems model 600	Oxford Cryosystems, Oxford, UK
Fluorometer	NaI single channel analyzer	Scintillation Counter, Newbury, OH
	X4A (NSLS)	
Source	Synchrotron	-
Detector	ADSC Quantum 4R	Area Detector Systems Corp., Poway, CA
Base	Huber 4-circle	Blake Industries, New Jersey, NJ
Low temperature device	Oxford Cryosystems model 600	Oxford Cryosystems, Oxford, UK
Fluorometer	Cyberstar X1000 scintillation detector	Oxford Instruments, Oxford, UK
	Queen's University	
Source	RU-300 rotating anode	Rigaku, Tokyo, Japan
Detector	MAR Research 300	MAR Research, Evanston, IL
Base	MAR Research single axis goniostat	MAR Research, Evanston, IL
Low temperature device	Oxford Cryosystems	Oxford Cryosystems, Oxford, UK

An overview of the numerous GTA and GTB crystals used in data collection is shown in Table 4 and summarized below. In order to understand the sequence of experiments, a brief explanation of the major data collection steps is given.

Table 4. Overview of crystals used for data collection

	Crystal type	No. crystals	No. datasets	Wavelengths (Å)	Average Resolution (Å)	Twinning
June 1999, X8C NSLS (BNL)						
1	GTB, Yb	3	2 x 3	1.3855, 1.3860, 1.3555		Yes
2	GTB	3	3	0.9057, 0.9789		Yes
3	GTB-mut, Hg^{2+}	1	1	0.9057	~2.00	Yes
4	GTB-mut	1	1	0.9057		Yes
5	GTB, Ho	1	1	0.9789		Yes
November 1999, X8C NSLS (BNL)						
6	GTB, Yb (c)	2	2	1.3869, 1.3855	~2.00	Yes
December 1999, Queen's University						
7	GTB, Yb	6	5	1.5418	~2.50	Yes
8	GTB-5	3	3	1.5418	~2.80	Yes
April 2000, Queen's University						
9	GTB-5, UDP-Gal, Mn^{2+}	3	1	1.5418	2.75	Yes
10	GTB-5, Mn^{2+}	4	0			Yes
11	GTB-5, UDP-Gal, Mn^{2+}	1	1	1.5418	2.80	No
12	GTB-5, UDP-Gal	1	1	1.5418	2.65	No
July 2000, Queen's University						
13	GTB-15, Mg^{2+} (c)	3	0			Yes
14	GTB-15, UDP-Gal-Hg (c)	3	0			Yes
15	GTB-15, UDP-Gal-Hg (c)	2	2	1.5418	2.65	No
16	GTB-15	2	0			Yes
17	GTB-15, UDP-Gal-Hg (c), MMA (s)	4	3	1.5418	2.70	No
18	GTB-5, UDP-Gal-Hg (s), MMA (s)	1	1	1.5418	2.75	No
19	GTB-5, UDP-Gal-Hg (s)	2	1	1.5418	2.80	No
October 2000, X8C NSLS (BNL)						
20	GTB-15, UDP-Gal-Hg	1	1	0.9796	2.00	No
21	GTB-15 SeMet, Co	5	0			Yes
22	GTB-15 SeMet, UDP-Gal-Hg	6	0			Yes
23	GTB-15 SeMet	2	0			Yes
24	GTB-15 M3	1	1	0.9796	1.50	No
25	GTB-15 M3, UDP-Gal-Hg	2	4	1.0834, 1.0092, 0.9930, 1.0102	1.50	No

Table 4, cont'd

			5	1.0095, 0.9933, 1.0080, 1.0099, 1.0087		
26	GTB-15 M3, UDP-Gal-Hg, MMA (s)	1	3	1.0095, 0.9600, 1.0088	1.50	No
December 2000, Queen's University						
27	GTB-10 SeMet	3	3	1.5418	2.65	No
28	GTB-10, Hg^{2+}	2	2	1.5418	2.70	No
January 2001, X8C NSLS (BNL)						
29	GTB-10, Hg^{2+}	2	6	0.9795	1.65	No
30	GTA-10, Hg^{2+}	2	3	1.200	1.80	No
31	GTB-10 SeMet, Hg^{2+}	3	3	0.9801, 0.9805, 0.9650	1.65	No
			5	0.9801, 0.9649, 0.9812, 1.0092, 0.9930	1.60	
			11	0.9800, 0.9647, 0.9801, 0.9802, 0.9649, 0.9803, 0.9801, 1.0091, 0.9930, 1.0101, 1.0096	1.75	
April 2001, X4A NSLS (BNL)						
32	GTB-10 SeMet, Hg^{2+}	1	5	0.9789, 0.9793, 0.9724, 0.9824	1.65	No
July 2001, X8C NSLS (BNL)						
33	GTB-10, UDP, Mn^{2+}, H antigen (B)	3	6	0.9795, 1.15	1.53	No
34	GTB-10, UDP-Gal, Mn^{2+} (A)	1	1	1.15	1.76	No
35	GTB-10, UDP-Gal, Mn^{2+} (B)	2	3	1.15	1.50	No
36	GTB-10, UDP-Gal-Hg, Mn^{2+} (A)	1	1	1.15	1.80	No
37	GTB-10, UDP-Gal-Hg, Mn^{2+} (B)	1	1	1.15	1.70	No
38	GTB-10, H antigen, Mn^{2+} (A)	1	1	1.15	1.85	No
39	GTB-10, UDP-Gal, Mn^{2+} (c)	1	1	1.15	1.60	No
July 2001, X12C NSLS (BNL)						
40	GTB-10, UDP-Gal, Mn^{2+} (c)	2	4	1.15	1.65	No
41	GTB-10, H antigen, Mn^{2+} (A)	1	2	1.15	1.80	No
November 2001, X8C NSLS (BNL)						
42	GTA-10, Mn^{2+}, UDP, H antigen	2	4	1.15	1.40	No
43	GTB-10, Mn^{2+}, UDP-Glc 5hr soak	1	2	1.15	1.29	No
44	GTA-10, Mn^{2+}, UDP-GlcNAc 5 hr soak	1	2	1.15	1.65	No
45	GTB-10, Mn^{2+}, UDP-Gal 1 hr soak	1	2	1.15	1.47	No
46	GTA-10, Mn^{2+}, UDP-GalNAc 1 hr soak	1	2	1.15	1.39	No
47	GTB-10, Mn^{2+}, UDP-GalNAc 1 hr soak	1	2	1.15	1.52	No
48	GTA-10, Mn^{2+}, UDP-Gal 1 hr soak	1	2	1.15	1.52	No
49	GTB-10, Mn^{2+}, UDP-Gal 3 hr soak	1	2	1.15	1.40	No
50	GTA-10, Mn^{2+}, UDP-GalNAc 3 hr soak	2	3	1.15	1.55	No
51	GTB-10, Mn^{2+}, H antigen	1	2	1.15	1.46	No
52	GTA-10, Mn^{2+}, H antigen	1	2	1.15	1.58	No
53	GTB-10, Mn^{2+}, UDP-Gal, H antigen	1	2	1.15	1.39	No
54	GTA-10, Mn^{2+}, UDP-GalNAc, H antigen	2	3	1.15	1.52	No
	Total	106	134			23

Early GTB and GTB-Ytterbium (Yb) crystals were mounted at beamline X8C at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratories (BNL) in June and November 1999. Data were collected on crystals cooled to 100 degrees Kelvin (K). Though a total of 14 datasets were collected, all were found to be a twinned C-centred lattice. Data reduction and subsequent structure solution were impossible. To test whether cryogenic conditions were responsible for twinning, crystals grown under the same conditions were mounted in glass capillaries and data were collected at room temperature in Dr. Zongchao Jia's laboratory at Queen's University in December 1999. The same pattern of twinning was observed.

In April 2000, data were collected (Queen's University) at room temperature on several crystals of GTB soaked with UDP-Gal and/or manganese. While most showed the twinning pattern previously observed, two C-centred orthorhombic datasets extending to 2.65 and 2.80 Å resolution were collected. In July of the same year, a number of crystals of GTB/GTB-5/GTB-15 obtained in gentler crystallization conditions were submitted to X-ray analysis at room temperature (Queen's University). In general, crystals of GTB-5 or GTB-15 containing UDP-Gal-Hg yielded orthorhombic data. Unliganded crystals of GTB-15, or those in the presence of magnesium alone yielded twinned data.

In September 2000, micro crystals of GTB-15 SeMet were mounted in a glass capillary and a fluorescence scan at λ=0.9795 Å was taken to evaluate the presence of selenium (X8C, NSLS). In October 2000, several crystals of GTB-15 SeMet, alone or soaked with UDP-Gal-Hg, were submitted to X-ray analysis (X8C, NSLS). While all crystals diffracted strongly, diffraction data were unsuitable for analysis. Crystals of the GTB M3 construct (C80S/C196S/C209S) in the unliganded form as well as co-crystallized with UDP-Gal-Hg were analyzed. 3 MAD experiments about the mercury edge were performed on GTB M3 in the presence of UDP-Gal-Hg as well as UDP-Gal-Hg and MMA.

In December 2000 (Queen's), GTB-10 SeMet crystals were submitted to X-rays, yielding improved data over the GTB-15 SeMet crystals. Crystals of unliganded GTB-10 (in the presence of Hg) were also analyzed. Two unliganded and three SeMet datasets were collected. In January 2001 (NSLS), datasets of the unliganded GTA-10 and GTB-10 (in the presence of Hg) were collected. In addition, 3 MAD experiments about the selenium edge and 2 MAD experiments about the mercury edge were collected using GTB-10 SeMet crystals.

III- GTB SE-MET HEAVY ATOM DETERMINATION

X-ray data for the SeMet derivatized GTB were collected at beamline X4A of the NSLS at BNL (April 2001). A strategy was devised based on the theoretical edges for mercury (Hg) and selenium (Figure 17b) whereby peak, inflection and remote datasets would be collected about the selenium edge and a last dataset collected at a wavelength with no selenium signal (i.e. only a mercury signal would be detected at this wavelength) to obtain a baseline signal for mercury alone. To determine the appropriate wavelengths for data collection, a fluorescence scan of the crystal was taken at wavelengths in the range of 0.9754 through 0.9832 Å (Figure 17b). Snapshots of the crystal were taken to determine the optimal crystal-to-detector distance and exposure time. The individual images were indexed using HKL2000 (Otwinowski & Minor, 1997) and the range of unique data to be collected was determined using this same program. Four datasets were then collected at wavelengths 0.9789 (pk1, pk2), 0.9793 (infl) and 0.9724 Å (rem) around the selenium edge, then at 0.9824 Å (hg) for a mercury signal. Each dataset was collected over 160° degrees in 0.75° degree intervals at a distance of 150 mm with 5s (pk2, infl, rem, hg) or 10s (pk1) exposures. Once collected, data were reduced and scaled using HKL2000.

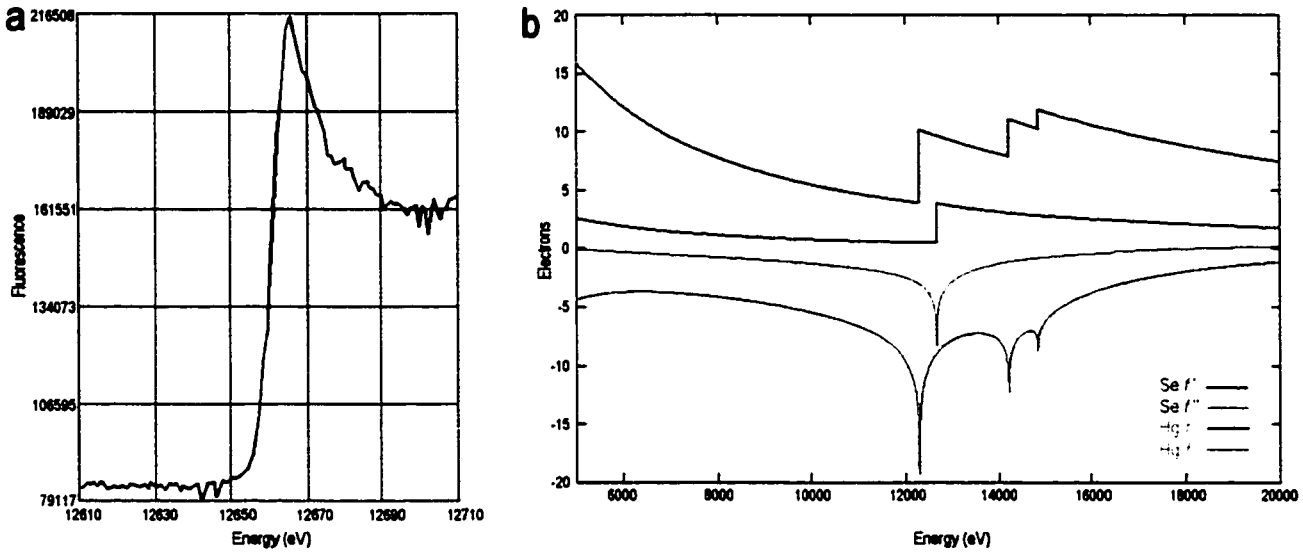

Figure 17. Strategy for data collection using GTB-10 SeMet Hg crystals. The experimental edge for selenium (a) was determined experimentally by taking a fluorescence scan at beamline X4A using X-rays at λ=0.9795 Å. Theoretical edges for selenium and mercury (b) were generated by the Edgeplots webtool (http://www.bmsc.washington.edu/scatter/). The strategic approach was to collect data at three wavelengths about the selenium edge to determine all heavy atom positions. Data at a wavelength exhibiting a low selenium signal would also be collected to allow identification of mercury positions. Peak, inflection and remote datasets were collected for selenium (λ=0.9789, 0.9793 and 0.9724 Å respectively) and a fourth wavelength at 0.9824 Å to provide a baseline signal for mercury. Units along the horizontal axis are the energy (E) in eV and can be related to the wavelength by the equation E (keV)=12.3985/λ.

IV- GTA, GTB AND COMPLEXES

A summary of data collection for the GTA, GTB, and complexed GTA or GTB is given in Table 5. Data were collected at beamline X8C at NSLS (BNL). Snapshots of the frozen crystals were taken to determine the optimal crystal-to-detector distance and exposure time. The individual images were indexed using HKL2000, and the range of unique data was determined using this same program. Once collected, data were reduced and scaled using HKL2000.

Table 5. Details of data collection for GTA-10, GTB-10 and various complexes

Crystal	Distance (mm)	Total width (°)	Interval (°)	Exposure time (s)	Wavelength (Å)	Maximal Resolution (Å)
GTA-10	135	100	1	180	1.20	1.80
GTB-10	150	65	1	45	0.980	1.65
	150	65	1	120	0.980	
GTA-10, UDP, H antigen, Mn^{2+}	200	100	1	20	1.15	1.35
	90	100	1	80	1.15	
GTB-10, UDP, H antigen, Mn^{2+}	100	120	1	10	1.15	1.40
GTB-10, UDP, H antigen, Mn^{2+}	100	120	1	10	1.15	1.32
	70	120	1	75	1.15	
GTA-10, H antigen, Mn^{2+}	120	110	1	20	1.15	1.58
	80	110	1	80	1.15	
GTA-10, H antigen, Mn^{2+}	200	100	1	20	1.15	1.46
	80	100	1	100	1.15	

4. Structure Solution

I- GTB-10 SeMet

Two programs were used in determining the heavy atom positions: Crystallography & NMR system (CNS) v.1.1 (Brünger *et al.*, 1998) and Shake-and-Bake (SnB) (Smith *et al.*, 1998). CNS generates Patterson coefficients and determines the heavy atom positions through alternating cycles of Patterson searches and Patterson-correlation refinement. SnB utilizes normalized structure factor magnitudes (E values) and generates trial phases based on randomly positioned atoms; reciprocal-space phase refinement is alternated with the application of phase constraints to find the positions of heavy atoms.

The structure solution of the GTB-10 SeMet is summarized in Figure 18. Using the data from GTB-10 SeMet pk1 (April 2001) to 2.0 Å, 2.5 Å and 3.0 Å, selenium atom positions were calculated. The hg data (April 2001) was also used in SnB to identify mercury positions. Positions of the heavy atoms were also calculated and refined using CNS in the following manner. The mercury atom positions identified using SnB were used for SAD phasing of hg data in CNS, producing Hendrickson-Lattman (HL) coefficients. The Se data (pk1, infl, rem) were then cross-phased using these HL coefficients; the Se sites were determined and subsequently refined. Heavy atom positions found using CNS and SnB were compared and those that correlated were deemed correct. Density modification was performed and electron density maps were calculated using all four wavelengths to resolve the hand ambiguity. The polypeptide chain was traced using the program ARP/wARP (Van Asselt *et al.*, 1998), inserting mainly Gly, Ser Val and Leu amino acid residues. The traced model was then mutated to correspond to the GTB amino acid sequence using SETOR (Evans, 1993).

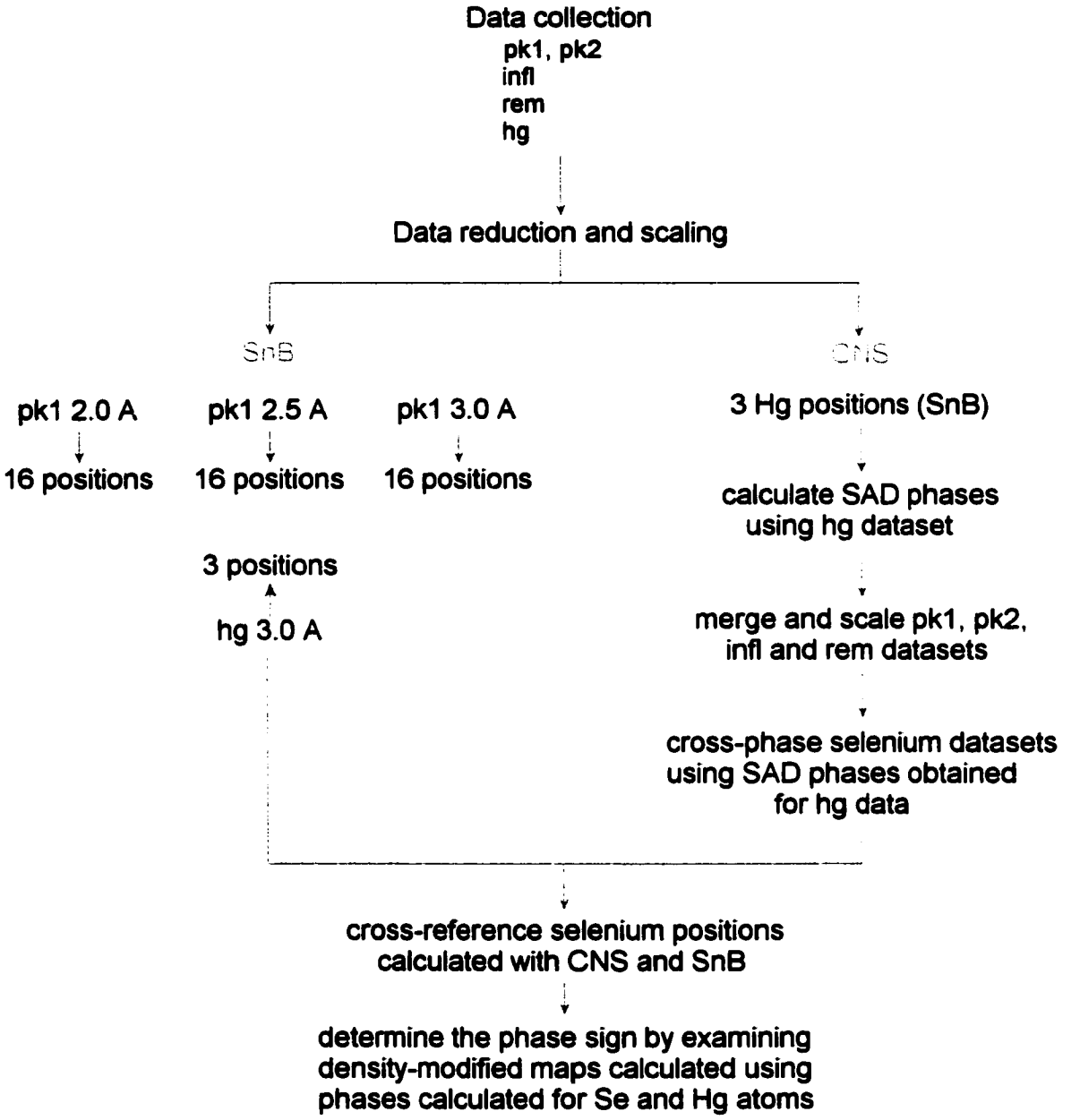

Figure 18. Solution of the heavy atom positions using MAD. Datasets were collected at wavelengths corresponding to the peak of selenium absorption (pk1, pk2), the inflection point (infl) and at a remote energy (rem). A fourth dataset (hg) was collected at a wavelength without selenium signal to obtain a baseline value for mercury absorption. Using SnB and the pk1 data, heavy atom sites were calculated using three resolution cutoffs; the position of 3 mercury atoms alone were also calculated. These mercury (Hg) positions were used in CNS to phase the hg data and calculate Hendrickson-Lattman (HL) coefficients. MAD data for the selenium signal (pk1, pk2, infl and rem) were then cross-phased using these HL coefficients, yielding 11 selenium sites. Selenium sites calculated using CNS and SnB were then compared, yielding 9 sites in addition to the 3 mercury sites. Density modification was applied to electron density maps calculated using phased data to determine the sign of the phases.

II- NATIVE GTB

The traced GTB SeMet model was used in solving the unliganded native GTB structure. Unless otherwise stated, determination of all subsequent GTA and GTB structures proceeded using CNS as follows. Molecular replacement using data in the 20.0-4.0 Å range and Fo >2σ was performed, followed by rigid body refinement in two steps: first with the molecule as a whole, then with the α-helices as rigid bodies. Positional refinement (20 cycles) and overall B-factor refinement (10 cycles) were alternated and performed at resolutions to 4.0, 3.5, 3.0 and 2.8 Å. At this last resolution, energy minimization (25 cycles) and individual B-factor refinement were performed to convergence (usually 2-3 cycles). Alternating cycles of positional refinement and individual B-factor refinement followed automated assignment of water molecules at 2.8 Å resolution. Using SETOR, the protein model and water molecules were fit to 3Fo-2Fc electron density maps. The adjusted model was submitted to energy minimization and individual B-factor refinement to 2.8 Å, and repeated at increased resolution (typically, an increase by 35-50% of data at the current resolution). The protein model was again submitted to manual intervention followed by positional and individual B-factor refinement. The resolution was then further increased, the model refined and automated assignment of water molecules performed. This method of manual intervention, refinement, increasing resolution with subsequent refinement and water divining was repeated until all the data in the 20.0-1.65 Å range were in use. The occupancy of mercury (and manganese, where applicable) ions was adjusted during refinement so that their B-factors reflected those of other atoms in the structure. At the maximal resolution, the geometry to X-ray weight was adjusted for increased weight on the X-ray terms, while keeping within acceptable boundaries of root mean square deviations (RMSD) and R_{free} (generally, RMSD bond length < 0.01 Å and RMSD bond angles < 2.00° and Rfree is within ~5-10% of the R-factor). This was followed by energy

minimization and individual B-factor refinement. The structure has been deposited in the Protein Data Bank (PDB) (accession code 1LZ7).

III- NATIVE GTA

The completed native GTB model was used to find a molecular replacement solution using the native GTA data. All other refinement proceeded as described in section ii. Final resolution limits for refinement for the native GTA model were 20.0-1.80 Å. The structure has been deposited in the PDB (accession code 1LZ0).

IV- GTB IN COMPLEX WITH H ANTIGEN AND UDP

The native GTB model was used to find a molecular replacement solution using the data for GTB in complex with H antigen and UDP. All other refinement proceeded as described in section ii. Two datasets were used and final resolution limits for the refinement for the models of GTB in complex with H antigen and UDP were 20.0-1.40 and 20.0-1.32 Å. The structure determined to 1.32 Å has been deposited in the PDB (accession code 1LZJ).

V- GTA IN COMPLEX WITH H ANTIGEN AND UDP

The native GTB model was used to find a molecular replacement solution using the data for GTA in complex with H antigen and UDP. All other refinement proceeded as described in section ii. Final resolution limits for the refinement of the model of GTA in complex with H antigen and UDP were 20.0-1.35 Å. The structure has been deposited in the PDB (accession code 1LZI).

VI- GTA IN COMPLEX WITH H ANTIGEN

The native GTB model was used to find a molecular replacement solution using the data for GTA in complex with H antigen. All other refinement proceeded as described in section ii. The final resolution limits for the refinement of the model of GTA in complex with H antigen were 20.0-1.58 Å.

VII- GTB IN COMPLEX WITH H ANTIGEN

The native GTB model was used to find a molecular replacement solution using the data for GTB in complex with H antigen. All other refinement proceeded as described in section ii. The final resolution range for the refinement of the model of GTB in complex with H antigen was 20.0-1.46 Å.

5. Molecular Modelling and Mutant Activity

Once all models were completed, the RMS differences between C^{α} of all proteins were calculated using CNS. Gal and GalNAc sugars were manually modelled into the GTA+H antigen+UDP and GTB+H antigen +UDP structures using a conservative approach that permitted only slight movements of glycosidic dihedral angles. The van der Waals radii of the atoms involved were considered in all cases. The nucleophile Glu 303 was modelled using a similar approach. Figures pertaining to GTA and GTB models were prepared using the SETOR package.

The E303A mutant was prepared and assayed by Dr. Monica Palcic (University of Alberta) using the same solutions of substrate, at the same concentrations as parallel experiments with the wild-type GTB enzyme (Seto *et al.*, 2000). Radiochemical assays utilized Sep-Pak C_{18} reverse phase cartridges (Waters, MA) using 500 μM Fuc-Gal-octyl acceptor, 270 μM UDP-Gal donor (UDP-[6-^{3}H]Gal, 60,000 dpm, American Radiolabeled Chemicals, MO) in 35 mM MOPS buffer pH 7.2 containing 20 mM $MnCl_2$ and 1 mg/mL BSA.

PART II: ANTI-BLOOD GROUP A Fv

1. PROTEIN EXPRESSION AND PURIFICATION

The gene for the anti-blood group A single-chain antigen-binding variable domain (scFv) was prepared, expressed and the protein purified as previously described (MacKenzie *et al.*, 1994; Patenaude *et al.*, 1998) by Rebbeca To at the IBS, NRC. Briefly, the AC1001 gene was synthesized with a linker $(GGGGS)_3$ between the variable light (V_L) and variable heavy (V_H) chains as well as *c-myc* (EQKLISEEDLN) immunodetection and His_5 (HHHHH) purification tags at the C-terminus. The protein was expressed in *E. coli* cells and purified by metal-ion affinity chromatography using an imidazole gradient. Eluted fractions containing AC1001 scFv were identified by optical density readings using UV light at a wavelength of 280 nm and subsequently pooled. Treatment of the scFv by subtilisin was necessary to remove aggregates, which also partially cleaved the C-terminal extension (*c-myc* and His_5). The digest was purified using size exclusion chromatography. The fractions corresponding to AC1001 Fv were pooled and concentrated using Centricon®10 (Millipore, Bedford, MA) to a final concentration of 7 mg/mL (calculated using the formula $c = OD_{280}/1.50l$, where l is the pathlength).

2. CRYSTALLIZATION

The crystallization of the AC1001 Fv was accomplished by Rebecca To, IBS NRC, and has been previously described (Patenaude *et al.*, 1998). Crystals were grown at room temperature using the hanging drop vapour diffusion method. Briefly, crystals were seeded into a mixture of 4.5 µL of AC1001 Fv, 1 µL 900 mM GalNAc and 4.5 µL of reservoir solution (30% PEG 4000, 0.1 M HEPES pH 7.0, 10 mM $CaCl_2$) and suspended over a 1 mL reservoir. The initial seed crystals were obtained under similar conditions, except at higher concentration of PEG 4000.

3. DATA COLLECTION AND REDUCTION

I- DATA COLLECTION EQUIPMENT

Diffraction experiments were undertaken at the A1 station of the Cornell High Energy Synchrotron Source (CHESS), and on a home source in the laboratory of Dr. Louis Delbaere at the University of Saskatchewan. The equipment available at each establishment at the time of collection is outlined in Table 6.

Table 6. Data collection equipment used for AC1001 Fv crystals

	Equipment	Manufacturer
	University of Saskatchewan	
Source	FR751 rotating anode	ENRAF-Nonius, Delft, Netherlands
Detector	FAST Area Detector	ENRAF-Nonius, Delft, Netherlands
Base	Kappa axis 4-circle goniostat	
	A1 (CHESS)	
Source	Synchrotron	
Detector	ADSC Quantum 1	Area Detector Systems Corp., Poway, CA
Base	Huber single axis goniostat	Blake Industries, New Jersey, NJ
Low temperature device	Oxford Cryosystems	Oxford Cryosystems, Oxford, UK

II- LOWER RESOLUTION DATA (UNIVERSITY OF SASKATCHEWAN)

Room temperature data were collected at a wavelength of 1.5418 Å on a single 0.5 mm x 0.3 mm x 0.15 mm anti-A Fv crystal mounted in a glass capillary onto the goniostat. The dataset was collected in intervals of 1° over a total of 180° with 240s exposures and at a distance of 200

mm. Data were reduced and scaled using the CCP4 suite (Collaborative Computational Project, 1994). Data were collected and processed by Dr. Leslie Tari.

III- Higher Resolution Data (CHESS)

High resolution X-ray data were collected at the A1 station of the Cornell High Energy Synchrotron Source (CHESS) at Cornell University. One crystal 0.4 mm x 0.2 mm x 0.1 mm was briefly dipped into cryoprotectant solution, consisting of mother liquor with 20% of the water replaced by glycerol, then mounted in a rayon loop on an extended-arc goniometer head (Charles Supper Co., Natick, MA) and cooled to 140 K in a liquid nitrogen stream. All four datasets were collected with 4s exposures at a distance of 60 mm and using intervals of 1°. The narrow coverage of the detector required that several scans were carried out using different chi angles using the extended-arc goniometer head. Goniometer head angles and total width of data collected, respectively, are: -20° and 142°, 20° and 50°, 60° and 120°, 60° and 140°. Data were reduced and scaled using the DENZO suite (Otwinowski & Minor, 1997).

4. STRUCTURE SOLUTION

I- LOWER RESOLUTION DATA (A)

The VL and VH dimer of the antibody YsT9.1 (Evans *et al.*, 1994) was used as a probe in molecular replacement experiments. Reflections in the resolution range 15-4.5 Å were used in the rotation and translation function searches performed using X-PLOR (Brünger, 1992). The YsT9.1 Fv probe was adjusted using rigid body refinement and the domain sequences were changed to those of AC1001 (Chen *et al.*, 1987). Electron density maps were calculated then displayed using the program FRODO (Jones, 1978). Refinement consisted of several rounds of manual intervention alternating with positional refinement, overall B-factor refinement and increasing resolution (typically, an increase by 35-50% of data at the current resolution). A first round of automatic water placement was performed using X-PLOR with resolution limits of 6.0-2.7 Å; another round was performed to 2.2 Å in later refinement rounds. Final structure refinements were done using the X-PLOR program with data in the 6.0-2.2 Å range. The structure has been deposited in the PDB (accession code 1JV5).

II- HIGHER RESOLUTION DATA (B)

The lower resolution room temperature model of the AC1001 Fv (A) (Thomas *et al.*, 2002) was used to perform molecular replacement using the high resolution low temperature AC1001 Fv data (B). Molecular replacement using data in the 20.0-4.0 Å range was performed, followed by rigid body refinement in two steps: first with the molecule as a whole, then with the VL and VH domains as rigid bodies. Positional refinement (20 cycles) and overall B-factor refinement (10 cycles) were alternated and performed at resolutions to 4.0, 3.5, 3.0 and 2.8 Å. At this resolution,

energy minimization (25 cycles) and individual B-factor refinement were performed to convergence (usually 2-3 cycles). Alternating cycles of positional refinement and individual B-factor refinement followed automated assignment of water molecules to 2.8 Å. Using SETOR, the protein model and water molecules were fit to electron density maps. The adjusted model was submitted to energy minimizations and individual B-factor refinement to 2.8 Å, and repeated at increased resolution (typically, an increase by 35-50% of data at the current resolution). The protein model was again submitted to manual intervention followed by positional and individual B-factor refinement. The resolution was further increased, the model refined and automated water placement performed. This method of manual intervention, refinement, resolution increase with subsequent refinement and water determination was repeated using data up to 1.3 Å. At 1.5 Å resolution, the geometry to X-ray weight was adjusted for increased weight on the X-ray terms, while keeping within acceptable boundaries of RMSD and R_{free}.

At 1.3 Å resolution, CNS was replaced by SHELX-97 (Sheldrick & Schneider, 1997) for refinement. Ten cycles of refinement with isotropic B-factors was performed, followed by 20 cycles of refinement with restrained individual anisotropic B-factor refinement. Manual intervention to adjust protein atoms was followed by automatic water placement and 20 cycles of refinement. Riding hydrogen atoms were added for two rounds of refinement (10 cycles each) alternated with minor adjustments to the model. The weighting scheme was adjusted to values suggested by SHELX, followed by a final round of refinement (10 cycles).

5. Modelling of the Trisaccharide antigen and mutant production

The surface of the anti-blood group A Fv was generated using MS (M. Connolly, 1983; M.L. Connolly, 1983) and the trisaccharide A antigen was manually modelled into the antigen-binding pocket of the 2.2 Å model using SETOR. A conservative approach was employed in that only small movements of glycosidic and side chain dihedral angles were permitted. The antigen conformation was based on the structure of the trisaccharide observed in the *Dolichos biflorus* lectin (Casset *et al.*, 1996) and the orientation was selected to correspond to the shape of the observed binding pocket.

As this complex is a model and not an observed structure, it was used as a predictive tool only, to identify residues that were likely to strongly affect trisaccharide recognition. The designed mutations were introduced by site-directed mutagenesis into an scFv species with a short linker between the V_L and V_H domains to create dimeric species whose higher avidity would aid panning and surface plasmon resonance analysis. The preparation and analysis of the site-directed mutants have been described in detail (Thomas, 1999; Thomas *et al.*, 2002).

Once both models (A and B) were completed, the RMSD between C^{α} of all proteins were calculated using CNS. Overlap of the protein main chain and figures were prepared using SETOR.

RESULTS AND DISCUSSION

In this chapter, the various structures solved will be presented and discussed in two parts. The first part will incorporate several GTA and GTB structures while the second will include the anti-blood group A Fv structures and the mutational analysis of the binding site.

PART I: GTA & GTB

1. EARLY RESULTS

In X-ray diffraction experiments, initial crystals obtained using GTB, GTB-5 and GTB-15 proteins all showed the same pattern of twinning (Figure 19a-c), indexing to a twinned C-centred monoclinic lattice. Though several twinned datasets were collected, the two lattices could not be deconvoluted. Surprisingly, a few single C-centred orthorhombic data could be reproducibly obtained by crystallizing GTB-15 in the presence of the UDP-Gal or UDP-Gal-Hg donor substrate. Reproducible single crystals of both the liganded and unliganded structures, however, could only be obtained when the GTB-10 protein was used.

Initial MAD experiments were conducted about the mercury edge using co-crystals of GTB-10 and UDP-Gal-Hg subsequently soaked in MMA. The presence of mercury was confirmed by fluorescence scans such as that shown in Figure 20. Using difference Patterson maps calculated with CNS, it was possible to identify two heavy atom positions corresponding to mercury atoms. Ambiguous electron density maps were obtained using the phases of these atoms and further analysis of the data was difficult.

The presence of selenium in GTB-10 SeMet and GTB-15 SeMet micro crystals was verified by taking fluorescence scans using λ=0.9795 at beamline X8C of the NSLS. The characteristic profile of the scan and the peak at X-ray energy of 12.6565 keV (Figure 21) confirmed that these crystals indeed contained selenium. Once suitable crystals of GTB-10 SeMet had been obtained, several MAD experiments about the selenium and mercury edges were conducted. Ambiguous heavy atoms positions were obtained that yielded uninterpretable electron density maps. The reason for this failure was not clear, but the assumption was made that in most cases, this was due to having missed the absorption edge. One experiment

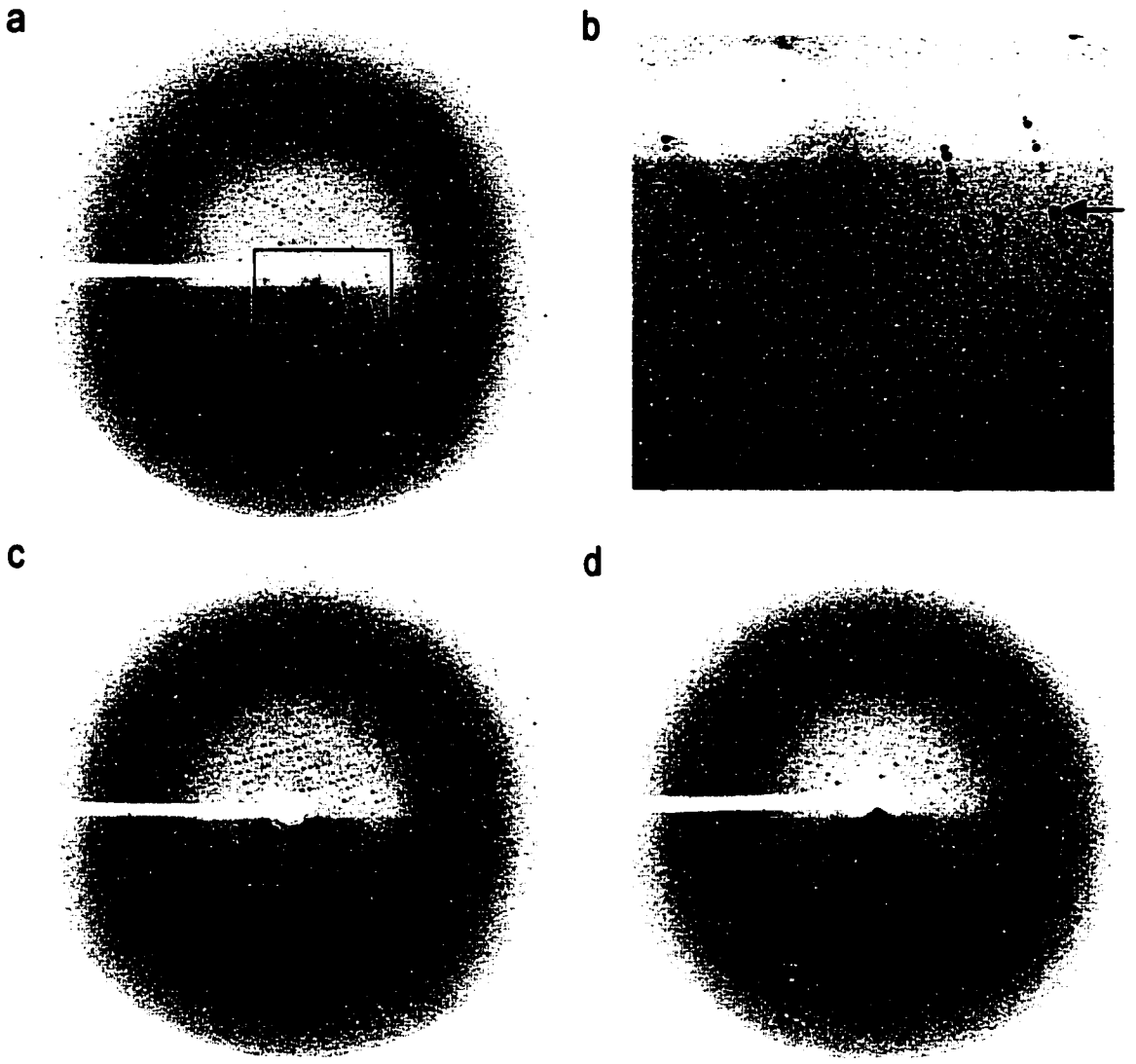

Figure 19. Typical diffraction data for early GTB crystals. a) Early crystals of GTB, GTB-5, GTB-mut and GTB-15 generally showed a twinned monoclinic lattice; this is emphasized in the enlarged view (b), where two lattices are visible. c) Diffraction pattern collected at 90° to a). Data corresponding to a C-centered orthorhombic space group was obtained for GTB-15 when crystals were grown in the presence of UDP-Gal-Hg (d).

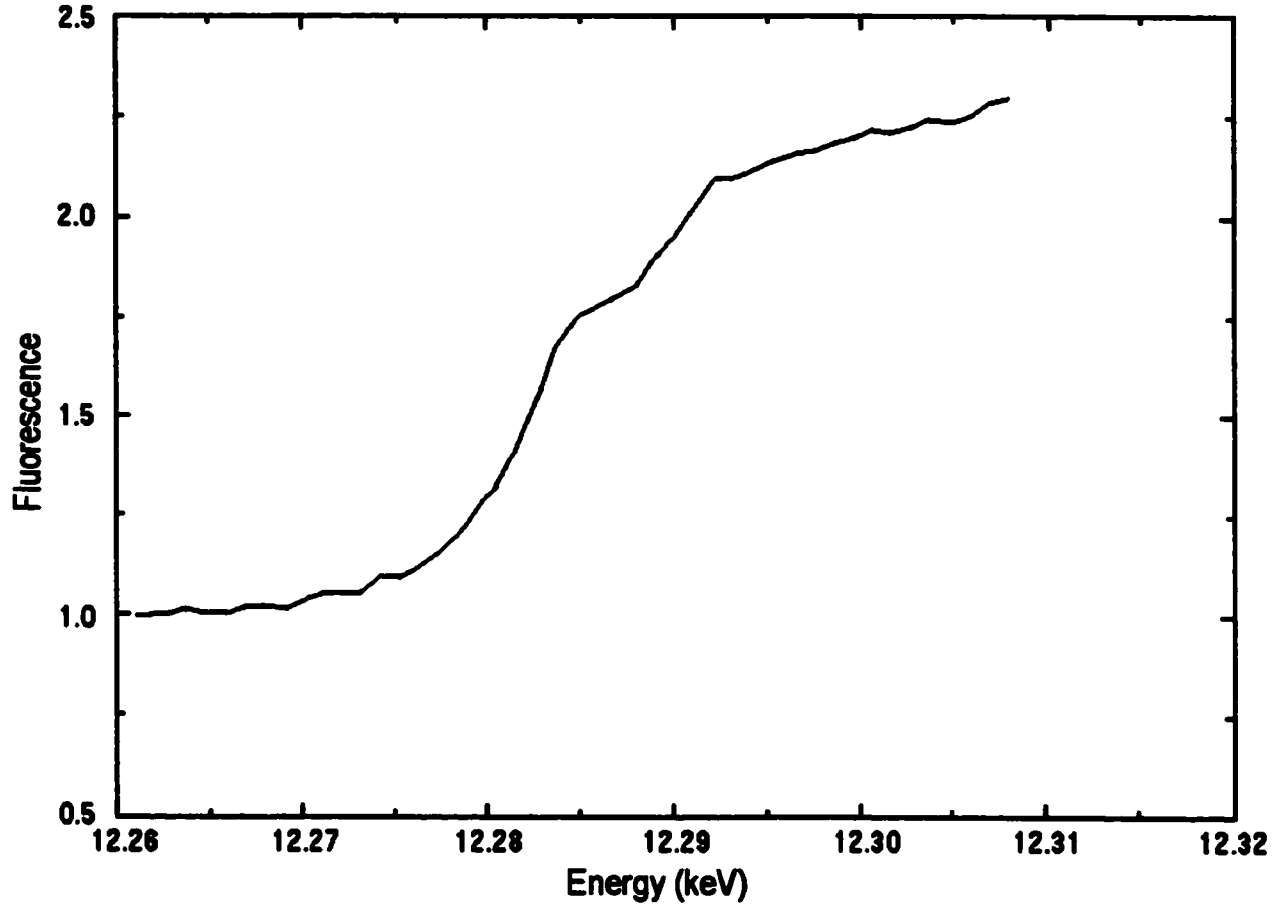

Figure 20. Fluorescence scan of GTB-10 at the mercury edge. Fluorescence scans of GTB-10 co-crystallized with UDP-Gal-Hg and subsequently soaked in methylmercuryacetate were taken to determine wavelengths useful for MAD data collection on initial crystals. Units along the horizontal axis are the energy (E) in keV and can be related to the wavelength by the equation E=12.3985/λ.

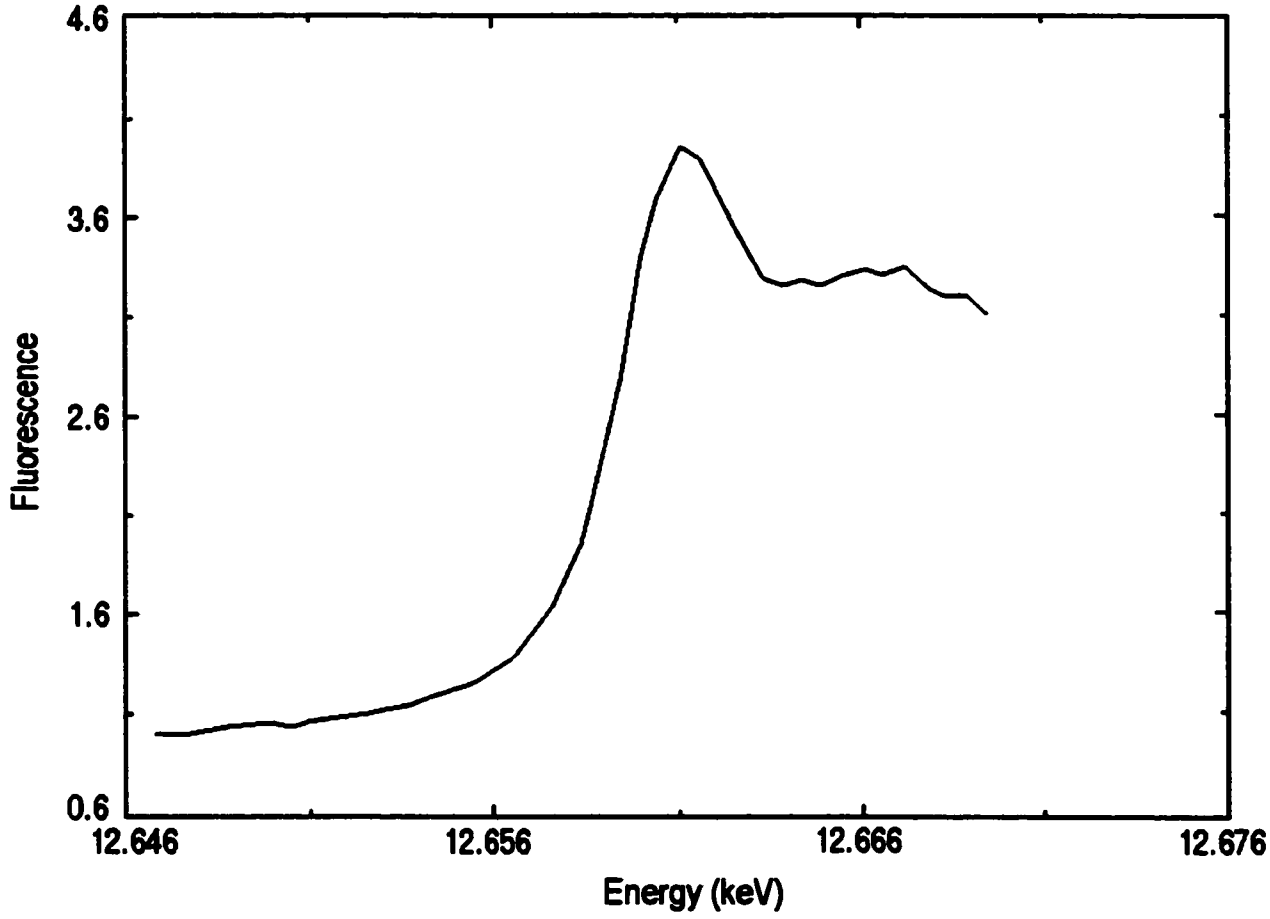

Figure 21. Fluorescence scan of GTB-10 SeMet microcrystals. Microcrystals of GTB-10 SeMet were mounted in a glass capillary and a fluorescent scan was taken at beamline X8C of the NSLS, BNL. The characteristic profile of the scan confirms the presence of selenium in the crystals. Units along the horizontal axis are the energy (E) in keV and can be related to the wavelength by the equation $E=12.3985/\lambda$.

(section 2), however, led to a clear solution for the mercury and selenium atom positions.

The GTA-10, GTB-10 and GTB-10 SeMet constructs were chosen over the GTB–15 constructs because of their ease of manipulation through the late stages of purification and crystallization. Further, substrate compounds could more easily be introduced to crystals of native GTA–10 and GTB-10 proteins. Typical crystals of various GTA and GTB constructs, liganded and unliganded, are shown in Figure 22.

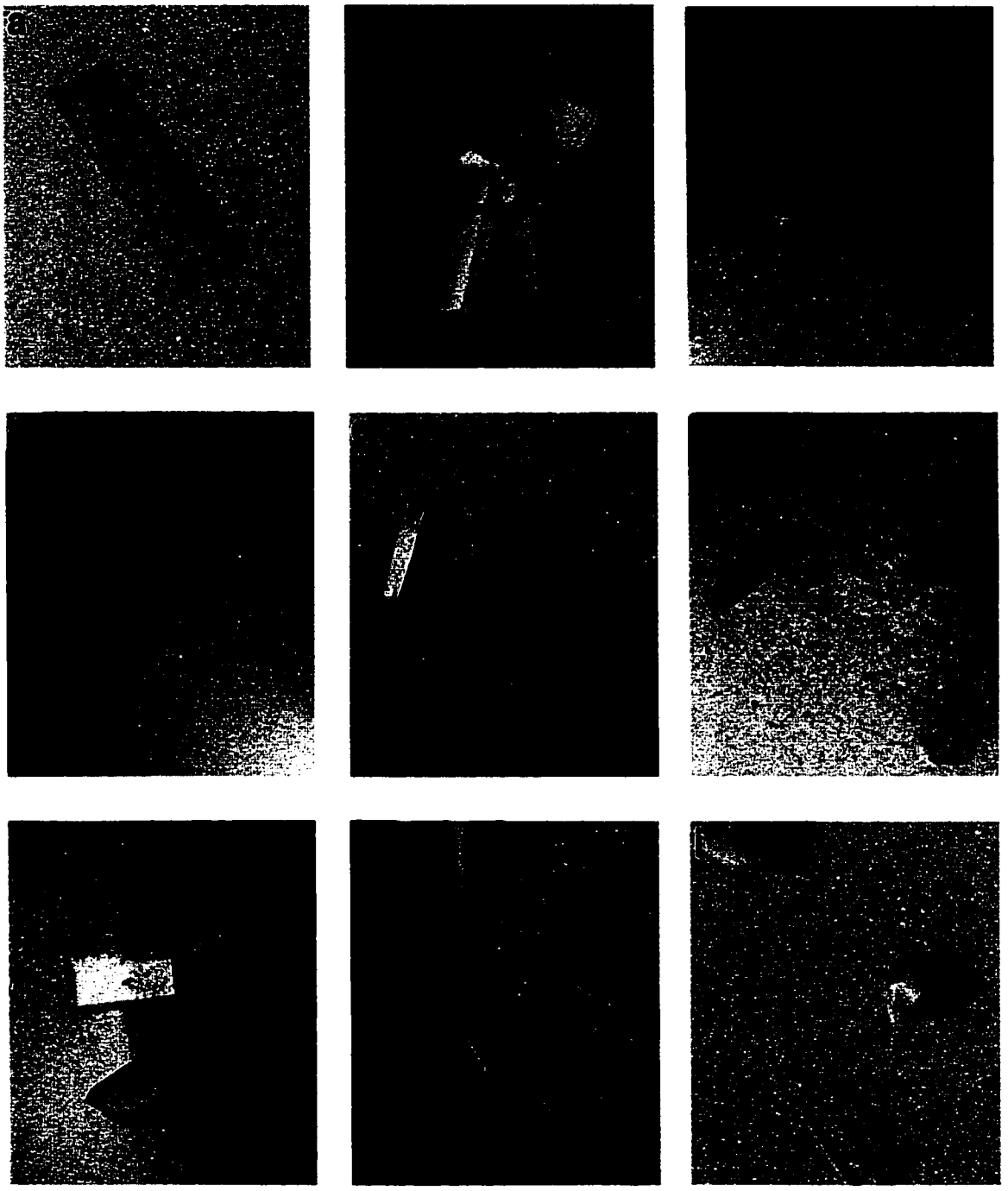

Figure 22. Typical crystals of the GTB and GTA proteins. Crystals of various GTA/GTB constructs were grown by Dr. Svetlana Borisova. Crystals of unliganded GTB-5 (a) and GTB-15 (b) yielded twinned data. The remaining crystals, GTB-10 (c-d), GTB-10 soaked with UDP and H antigen (5 days) (e), GTB-10 soaked with UDP-Gal-Hg (5 days) (f), and GTA-10 (g-l) yielded orthorhombic data.

2. STRUCTURE SOLUTION USING GTB-10 SEMET

Using crystals of GTB-10 SeMet, data were collected about the selenium edge at four wavelengths: 0.9789 (pk1, pk2), 0.9793 (infl), 0.9724 (rem) and 0.9824 (hg). Data collection statistics for these datasets can be found in Table 7. Though peak, inflection point and remote data for selenium carry a strong mercury signal, the dataset collected at λ=0.9824 (in the absence of a selenium signal) provides an accurate baseline for the mercury signal alone, allowing for the identification of the mercury atoms among calculated heavy atom positions.

Table 7. Data collection statistics for GTB-10 SeMet MAD experiment

	pk1	pk2	infl	rem	hg
Wavelength (Å)	0.978854	0.978854	0.979297	0.972432	0.982432
Energy (keV)	12.666	12.666	12.661	12.750	12.620
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell dimensions					
a (Å)	52.8	52.9	52.9	52.9	52.9
b (Å)	150	150	150	150	150
c (Å)	79.8	79.9	79.9	79.9	79.9
Resolution (Å)	20.0-1.65	20.0-1.75	20.0-1.80	20.0-1.84	20.0-1.80
Unique reflections	38,147	32,397	29,872	28,013	29,996
Completeness (%)	99.1 (99.8)	99.7 (100)	99.8 (100)	99.8 (100)	99.7 (100)
Rmerge	0.083 (0.32)	0.078 (0.41)	0.082 (0.39)	0.074 (0.39)	0.081 (0.41)
I/σ	31.1 (4.9)	33.9 (5.0)	27.4 (5.6)	32.8 (5.6)	26.6 (5.4)

Where applicable, statistics for the last resolution shell are shown in parentheses.

SnB and CNS were both used to determine the position of selenium atoms in the unit cell. Data for pk1 were submitted to SnB in three batches, using different resolution limits for each; this yielded 16 heavy atom positions for each batch. The "hg" dataset was also submitted to SnB, which revealed 3 sites to be mercury. CNS was used to determine selenium atom sites by cross-

phasing the pk1, infl and rem datasets with the mercury atom phases found using SnB; 11 selenium atom positions were found. Comparison of the calculated heavy atom sites from CNS and SnB revealed that 9 of the selenium locations could be correlated. Despite the different approaches used by the two programs, similar positions were found and therefore confidence could be placed in the selenium and mercury positions calculated. The ambiguous nature of the heavy atom solutions was resolved by calculating electron density maps using both signs of the heavy atom phases. Visual inspection of density-modified maps revealed that the positions of the 3 mercury and 9 selenium atoms shown in Table 8 were correct.

Table 8. Heavy atom positions determined using CNS and SnB

Atom	x	y	z
Hg 1	-19.6	-68.3	-24.2
Hg 2	-15.6	-46.3	-3.1
Hg 3	-3.4	-25.8	-38.7
Se 1	-10.5	-52.4	-15.5
Se 2	-12.5	-56.1	-31.9
Se 3	-9.7	-59.9	-0.01
Se 4	-9.8	-13.8	-39.0
Se 5	-4.2	-28.6	-34.9
Se 6	-3.2	-27.0	-32.6
Se 7	-16.1	-9.0	-24.3
Se 8	-24.2	-52.4	-24.2
Se 9	-13.5	-48.1	-0.65

High resolution electron density maps to 1.65 Å resolution were unambiguous and allowed tracing of the protein main chain using ARP/wARP; the correct amino acid sequence for GTB (Figure 23) was inserted using SETOR. This yielded an initial model corresponding to amino acid

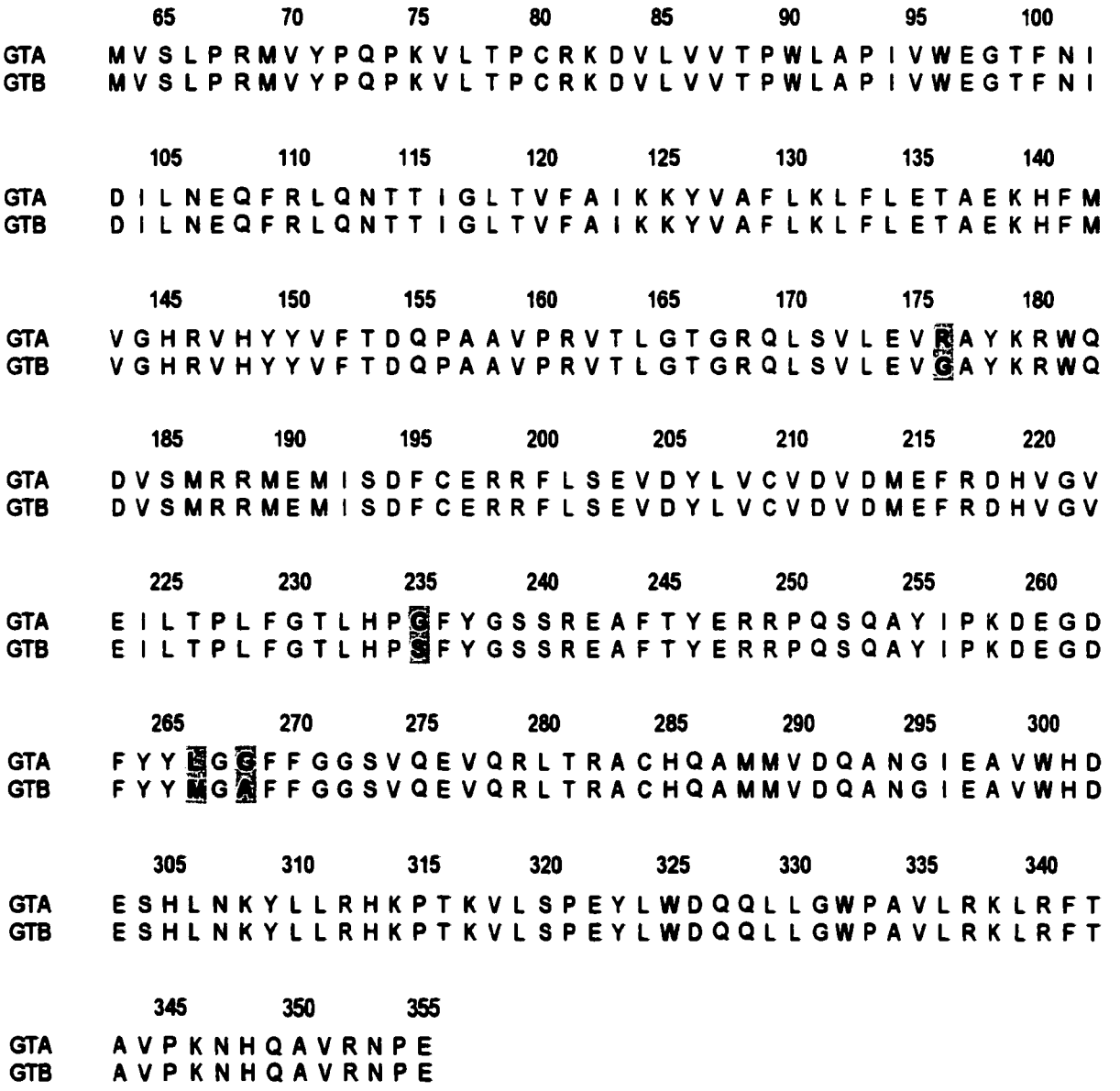

Figure 23. GTA and GTB sequences. The sequences of the GTA-10 and GTB-10 proteins are shown with the four critical amino acid residues highlighted in yellow. Constructs for both proteins begin at amino acid 63, as the transmembrane and stem regions were deleted from the gene construct to increase solubility.

residues 66-175 and 199-344. The loop corresponding to residues 176-198 and the C-terminal 11 amino acids could not be traced due to a lack of electron density for these regions. Interestingly, the 176-198 region contains three methionine residues, derivatized to SeMet, which should have given a strong MAD signal. The fact that the positions of these three Se atoms remain unknown strongly suggests that this region is disordered, which resulted in a weak and diffuse selenium signal. The final model was used in searching for a molecular replacement solution using native GTB data. The native GTB model was in turn used for molecular replacement solutions for other GTB and GTA structures (using the appropriate sequence, Figure 23).

3. General protein fold of GTA and GTB

The topology of GTA and GTB proteins resembles that of the related bovine α(1-3)Gal transferase structures reported earlier (Gastinel *et al.*, 1999; Boix *et al.*, 2001) (Figure 24). This was not surprising as the proteins show ~40% sequence identity. Overlap of the alpha carbon (C^{α}) atoms of GTB and the α3GalT structures (Figure 24) reveals that the three-dimensional arrangement of the secondary structure elements is also very similar, except at the N-terminus where significant differences can be observed. However, due to high thermal displacement and unsupported observations, the first structure (Gastinel *et al.*, 2001) of α3GalT is much less reliable than the later structure solved by Boix *et al.* (2001). The GTA and GTB enzymes can be considered to form part of the same superfamily as the α3GalT, SpsA, β4GalT1, GlcAT-I, LgtC and GnTI (Charnock & Davies, 1999; Gastinel *et al.*, 1999; Pedersen *et al.*, 2000; Unligil *et al.*, 2000; Boix *et al.*, 2001; Gastinel *et al.*, 2001; Persson *et al.*, 2001).

The overall structure for all of the GTA and GTB enzymes determined in this study remains constant, with dimensions of approximately 43 Å x 54 Å x 55 Å. The polypeptide chain is organized in two domains separated by a cleft containing the putative active site. The N-terminal domain comprises residues 63-227 and 338-345, including a four-stranded parallel β-sheet (β3-β6, defined in Table 9), flanked by 3 α-helices (α1-α3) and arranged in a distorted Rossmann fold. The C-terminal domain has α/β topology consisting of residues 228-337. A large central mixed β-sheet is composed of 9 strands originating from both the C- and N-terminal domains. The active site cleft is approximately 13 Å wide in GTA and GTB and contains all four critical amino acid residues. The substrate binding site is located in the cleft, flanked by the top of the central β-sheet on one side and the 2-stranded N-terminal antiparallel β-sheet (β7, β11) on the other. One of the most remarkable aspects of the GTA and GTB, which is common to other glycosyltransferases, is the open nature of the active site clefts. This results in few enzyme residues being positioned to

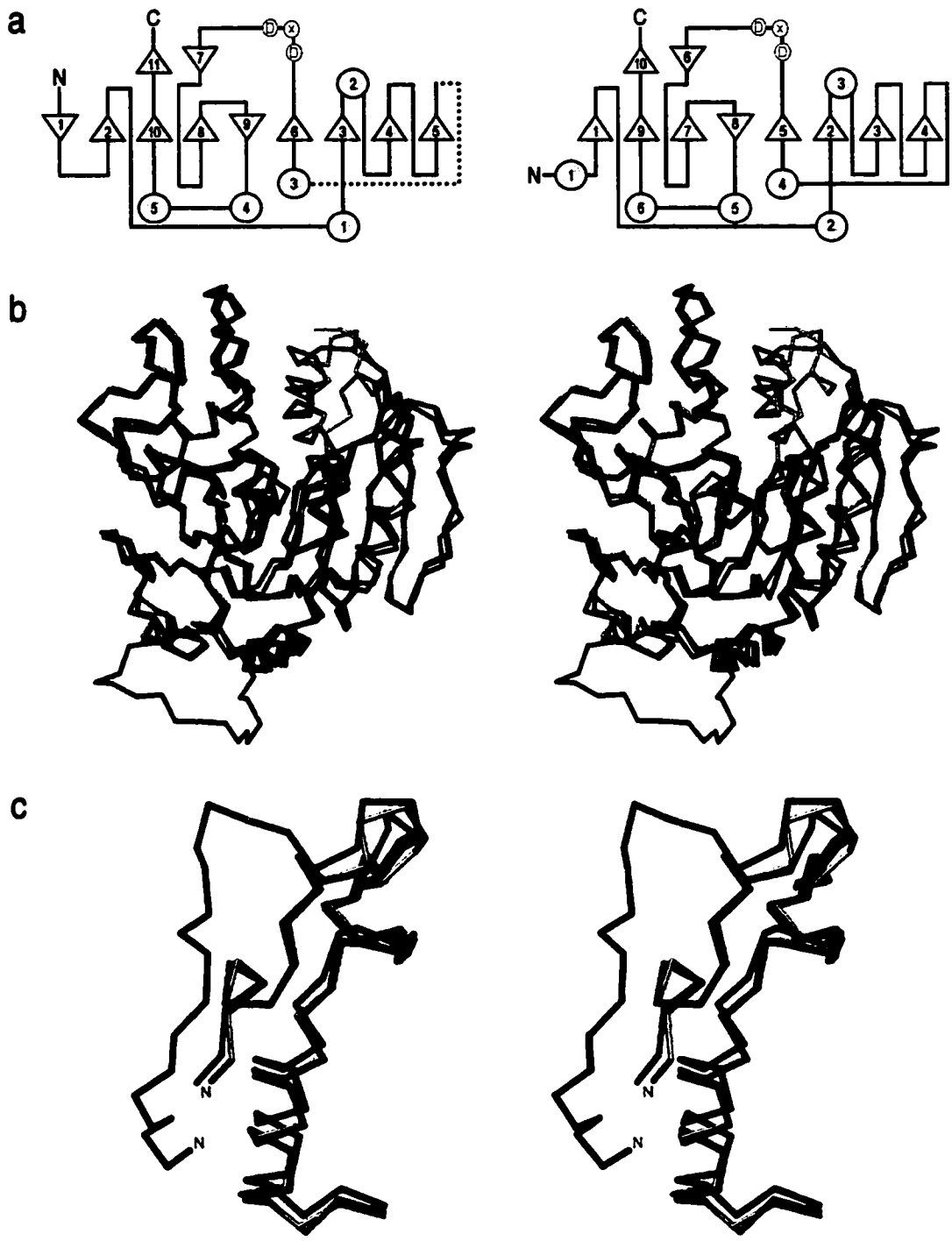

Figure 24. Comparison of GTA and GTB with α3GalT. a) Topology diagrams of GTA/GTB (left) and α3GalT (Gastinel *et al.*, 2001) (right) show much similarity in the secondary structure elements. α-helices (circles) and β-strands (triangles) are colored according to the domain to which they belong: yellow for N-terminal and blue for C-terminal domains. The disordered portion of GTA/GTB is shown by a dashed line. A nine-stranded mixed β-sheet is formed by strands from both the N-terminal and C-terminal domains and a small 2-stranded antiparallel β-sheet is present in the N-terminal domain. b) The C^{α} overlap of GTB (red) and α3GalT (Gastinel *et al.* (2001) green, Boix *et al.* (2001) white) reveals that the three-dimensional arrangement of secondary structure elements for both enzymes is very similar over most of the main chains. Significant differences arise, however, in the N-terminal region (c).

contact and recognize donor substrates. The DXD motif is defined by residues Asp211 Val212 Asp213 and is located in the cleft. In addition, all four of the critical amino acid residues are located in the cleft, with residues 266 and 268 opposite the DXD. The RMS differences resulting from the superposition of 264 C^{α} from the various liganded and unliganded structures are shown in Table 10. These values show that there exists very little variation in the protein backbone atoms between the various models.

Table 9. β-strands and α-helices in the GTA and GTB structures

β-strands	Amino acid residues	Domain	α-helix	Amino acid residues	Domain
β1	86-89	N-terminal	α1	101-112	N-terminal
β2	91-96	N-terminal	α2	125-140	N-terminal
β3	114-122	N-terminal	α3	~196-203	N-terminal
β4	146-154	N-terminal	α4	275-293	C-terminal
β5	167-175	N-terminal	α5	301-313	C-terminal
β6	205-211	N-terminal			
β7	213-218	N-terminal			
β8	228-233	C-terminal			
β9	268-273	C-terminal			
β10	317-320	C-terminal			
β11	339-344	N-terminal			

Secondary structure elements were identified using hydrogen-bonding patterns.

Table 10. RMS differences between C^{α} of various GTA and GTB protein models

	GTA	GTB	GTA+H	GTB+H	GTA+ H+UDP	GTB+H+ UDP (A)	GTB+H+ UDP (B)
GTA	-	0.21	0.20	0.21	0.23	0.30	0.26
GTB	0.21	-	0.19	0.16	0.19	0.24	0.15
GTA+H	0.20	0.19	-	0.15	0.13	0.25	0.22
GTB+H	0.21	0.16	0.15	-	0.19	0.22	0.21
GTA+H+UDP	0.23	0.19	0.13	0.19	-	0.22	0.21
GTB+H+UDP (A)	0.30	0.24	0.25	0.22	0.22	-	0.24
GTB+H+UDP (B)	0.26	0.16	0.22	0.21	0.21	0.24	-

C^{α} for residues 63-176 and 196-345 were used in calculating RMS differences (given in Å).

All enzyme structures presented here exhibit excellent electron density over most of the polypeptide chain, except for a disordered loop consisting of residues ~177-195 (may vary slightly between models) adjacent to the active site, and disorder in the final 10 residues of the C-terminus. Similar disordered polypeptide loops have been observed in other glycosyltransferases (Vrielink *et al.*, 1994; Charnock & Davies, 1999; Gastinel *et al.*, 1999; Unligil *et al.*, 2000; Gastinel *et al.*, 2001; Mulichak *et al.*, 2001). It has been proposed that flexible loops in proximity to the active site limit access to the active site, preventing the premature hydrolysis of the donor (Charnock & Davies, 1999). Alternatively, Unligil *et al.* (2000) suggests that such structures could play an important role in the release mechanism of the enzyme products. Critical amino acid 176 is located adjacent to this flexible loop and its mutation from Arg (in GTA) to Gly (in GTB) results in a 11-fold increase in turnover number, indicating that this region may indeed play a role in product release.

The intramolecular interactions observed in all GTA and GTB models include many hydrogen bonds in the β-sheet and α-helix structures, though some hydrogen bonding in the loop regions is present. There are also several intermolecular interactions between symmetry-related molecules and these are outlined in Table 11. The most interesting feature of crystal packing is at the N-terminus, which is intertwined with the N-terminus of a symmetry-related partner (Figure 25). Interestingly, two main-chain hydrogen bonds are observed in this region (between Leu85 and Val 76, Asn 101 and Pro 89). The position of amino acid Cys 80 at the centre of this interaction suggests that slight rearrangement of the main chain could result in disulfide bond formation between symmetry-related molecules, providing an explanation for the aggregation of the enzymes in the absence of mercury ions. In most structures, a partially occupied mercury ion is observed to be coordinated to Cys 80.

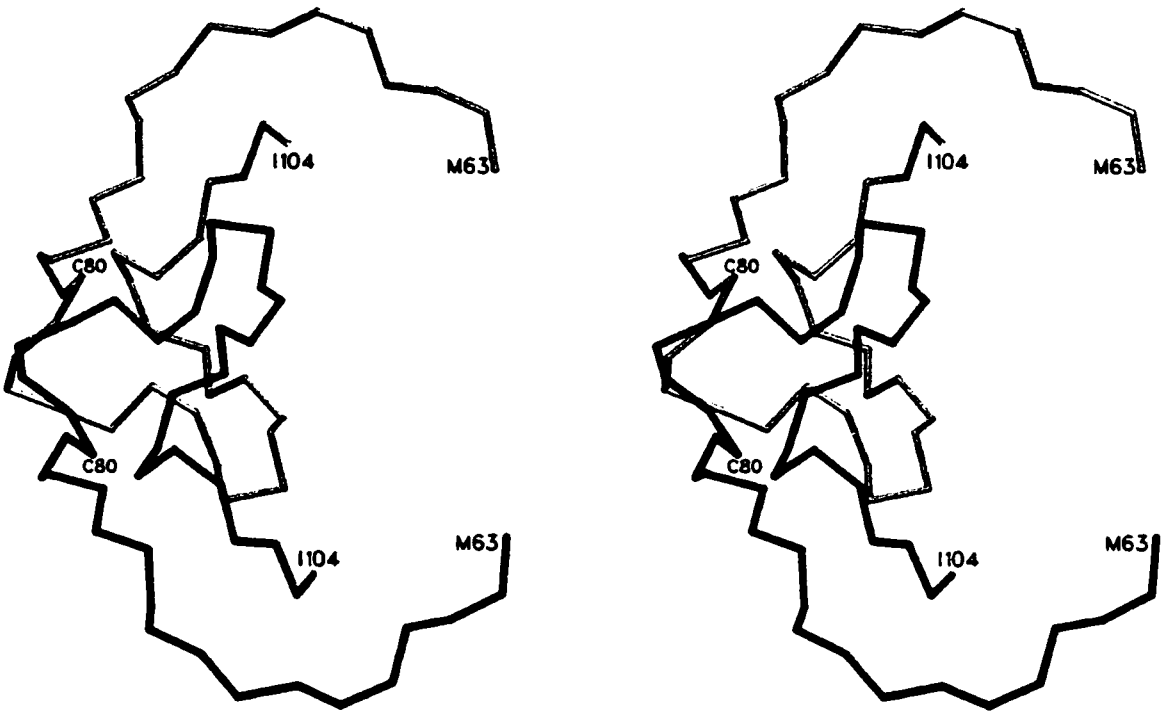

Figure 25. Packing of the GTA and GTB N-terminus. One of the most interesting features of crystal packing and intermolecular interactions is the intertwining of the N-terminus. The cysteine residue present at amino acid 80 is positioned to form a disulfide bond with its symmetry-related partner, providing an explanation for the formation of aggregates in the absence of mercury ion.

Table 11. Intermolecular interactions in the GTA and GTB structures

GTA		Distance (Å)	GTB		Distance (Å)
OG Ser 65	NE2 His 313	3.09	OG Ser 65	NE2 His 313	3.21
OG Ser 65	CE1 His 313	2.99	OG Ser 65	CE1 His 313	3.02
N Leu 66	O Arg 312	2.93	N Leu 66	O Arg 312	2.94
NH1 Arg 68	OD2 Asp 259	2.86	NH1 Arg 68	OD2 Asp 259	2.93
NH2 Arg 68	OE1 Glu 260	2.59	NH2 Arg 68	OE1 Glu 260	2.84
NH2 Arg 68	OD1 Asp 259	3.02	NH2 Arg 68	OD1 Asp 259	2.88
OH Tyr 71	OD1 Asp 262	2.44	OH Tyr 71	OD1 Asp 262	2.56
NE2 Gln 73	O Ser 239	3.04	NE2 Gln 73	O Ser 239	3.09
N Leu 85	O Val 76	2.85	N Leu 85	O Val 76	2.84
N Asn 101	O Pro 89	2.89	N Asn 101	O Pro 89	2.86
NE Arg 110	O Leu 164	2.88	NE Arg 110	O Leu 164	2.97
NH1 Arg 110	O Leu 164	3.09	NH1 Arg 110	O Leu 164	3.12
			NZ Lys 131	O Arg 68	3.16
OG1 Thr 163	OE1 Glu 107	2.57	OG1 Thr 163	OE1 Glu 107	2.50
N Thr 166	O Gly 144	3.04	N Thr 166	O Gly 144	3.11
			OH Tyr 178	SD Met 63	3.02
N Arg 241	OE1 Gln 73	2.69	N Arg 241	OE1 Gln 73	2.63
NH1 Arg 241	O Tyr 71	2.94	NH1 Arg 241	O Tyr 71	2.96
NH2 Arg 241	O Try 71	3.03	NH2 Arg 241	O Tyr 71	3.02
NE Arg 279	OE2 Glu 174	2.85	NE Arg 279	OE1 Glu 174	3.02
NH1 Arg 279	OE1 Glu 174	2.73	NH1 Arg 279	OE2 Glu 174	2.86
OE1 Gln 286	O Val 172	2.93	OE1 Gln 286	O Val 172	3.25
NH1 Arg 312	O Leu 66	2.87	NH1 Arg 312	O Leu 66	2.21
NZ Lys 314	OH Tyr 71	2.80	NZ Lys 314	OH Tyr 71	2.90
N Trp 332	O Gln 328	2.97	N Trp 332	O Gln 328	2.97
NH1 Arg 68	OD2 Asp 259	2.86	NH1 Arg 68	OD2 Asp 259	2.93
NH2 Arg 68	OD1 Asp 259	3.02	NH2 Arg 68	OD1 Asp 259	2.88
NH2 Arg 68	OE1 Glu 260	2.59	NH2 Arg 68	OE1 Glu 260	2.84
NH1 Arg 279	OE1 Glu 174	2.73	NH1 Arg 279	OE2 Glu 174	2.86

Residues 337-345 of the C-terminus are located in the N-terminal domain, forming strand β11. This final β-strand of the enzymes lines the active site cleft near the UDP binding site. Though the final 10 residues of the C-terminus are disordered, the present structures suggest that they may be positioned to interact with bound substrates. The formation of a C-terminal "lid", which interacts strongly with the UDP portion of the donor, has been previously observed in other structures (Unligil *et al.*, 2000; Boix *et al.*, 2001). The high structural homology of GTA and GTB

with one of these structures (α3GalT) further supports this type of configuration for the C-terminal residues.

4. NATIVE GTA AND GTB STRUCTURES

The structures of native GTA and GTB were refined to 1.80 and 1.65 Å, respectively. Data collection and refinement statistics for both structures are shown in Table 12.

Table 12. Data collection and refinement statistics for native GTA and GTB structures

Data collection statistics	Native GTA	Native GTB
Wavelength (Å)	1.2	0.9795
Space group	$C222_1$	$C222_1$
Cell dimensions		
a (Å)	52.5	52.8
b (Å)	149	150
c (Å)	79.4	80.0
Resolution (Å)	20-1.8	20-1.65
Unique reflections	26,921	37,031
Completeness (%)	91.5 (74.3)	95.8 (79.5)
Rmerge	0.049 (0.30)	0.054 (0.38)
I/σ	33.7 (4.27)	35.8 (4.76)
Refinement statistics		
Resolution (Å)	20-1.8	20-1.65
Rwork (%)	17.5	18.1
Rfree (%)	21.2	20.5
Number of atoms		
Protein	2,174	2,176
Water	250	292
Ion	4 (Hg^{2+})	3 (Hg^{2+})
Average B-factor (Å^2)		
Protein	25.6	21.7
Solvent	39.0	38.0
Ion	37.9	38.5
RMS deviations		
Bonds (Å)	0.00852	0.00864
Angles (°)	1.46	1.54

Where applicable, statistics for the last resolution shell are shown in parentheses.

The structure of the native GTA and GTB proteins is shown in Figure 26 and are as described in section 3 (General protein fold of GTA and GTB). Though several disordered side chains were observed in both models, alternate conformations could only be modelled for residue Cys 284.

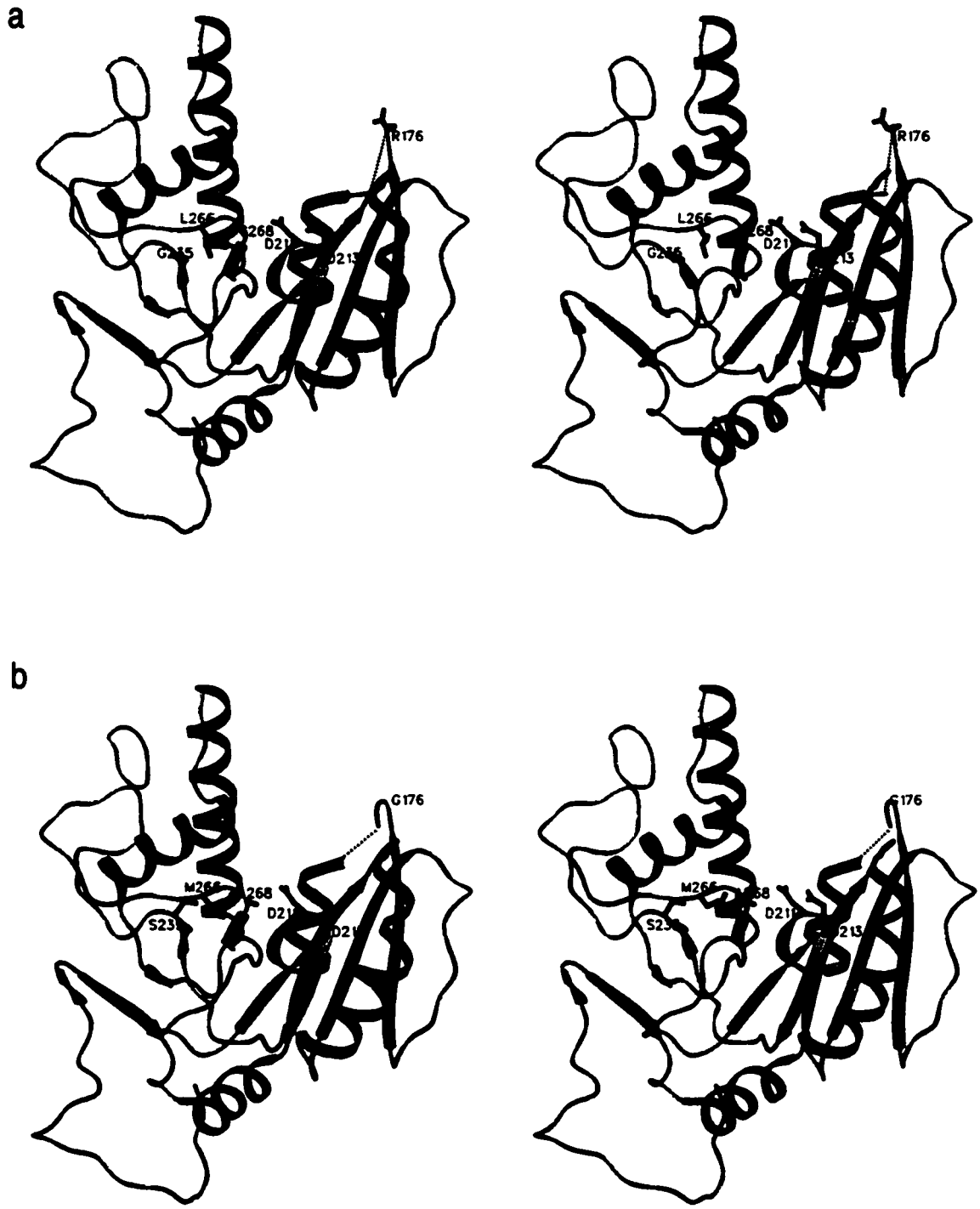

Figure 26. Native GTA and GTB structures. The GTA (a) and GTB (b) enzymes are divided into 2 domains, each composed of β-strands (blue) flanked by α-helices (green). The active site is formed by the mixed central β-sheet and the small antiparallel β-sheet and contains all four critical amino acid residues as well as the DxD motif (red). The black dashed line corresponds to the disordered region (amino acids ~177-196).

5. GTA and GTB in the presence of H antigen and UDP

The structure of GTB in complex with the H antigen and UDP was refined to 1.40 (A) and 1.32 Å (B), while the structure of GTA in the presence of substrates was refined to 1.35 Å. Data collection and refinement statistics can be found in Table 13. As the structures of liganded GTB are very similar, our discussion will be limited to the 1.32 Å (B) model.

Table 13. Data collection and structure refinement statistics for GTA and GTB in complex with H antigen and UDP

Data collection statistics	**GTA+H UDP**	**GTB+H UDP (A)**	**GTB+H+UDP (B)**
Wavelength (Å)	1.15	1.15	1.15
Space group	$C222_1$	$C222_1$	$C222_1$
Cell dimensions			
a (Å)	52.8	52.7	52.7
b (Å)	150	150	150.5
c (Å)	80.0	79.5	79. 5
Resolution (Å)	50-1.35	50-1.40	50-1.32
Unique reflections	66,876	59,846	72, 904
Completeness (%)	95.5 (69.5)	96.2 (74.7)	98.4 (99.3)
Rmerge	0.050 (0.22)	0.045 (0.39)	0.047 (0.40)
I/σ	42.7 (4.99)	37.5 (2.01)	46.2 (4.17)
Refinement statistics			
Resolution (Å)	20-1.35	20-1.4	20-1.32
Rwork (%)	17.9	17.4	18. 5
Rfree (%)	20.0	19.6	20.0
Number of atoms			
Protein	2198	2179	2177
Water	394	373	388
Ion	5 (Hg^{2+}), 1 (Mn^{2+})	4 (Hg^{2+}), 1 (Mn^{2+})	4 (Hg^{2+}), 1 (Mn^{2+})
Substrate	28 (H antigen), 25 (UDP)	28 (H antigen), 25 (UDP)	28 (H antigen), 25 (UDP)
Average B-factor ($Å^2$)			
Protein	15.7	17.1	17.7
Solvent	32.1	33.5	35.6
Ion	26.4 (Hg^{2+}) , 32.1 (Mn^{2+})	29.9 (Hg^{2+}), 23.1 (Mn^{2+})	34.9 (Hg^{2+}), 22.2 (Mn^{2+})
Substrate	16.6 (H antigen), 47.2 (UDP)	17.5 (H antigen), 24.2 (UDP)	19.6 (H antigen), 22.5 (UDP)
Rms deviations			
Bonds (Å)	0.0093	0.0073	0.0096
Angles (°)	1.5	1.4	1.5

Where applicable, statistics for the last resolution shell are shown in parentheses.

The resolution to which the data extend for these structures allowed the modelling of several alternate conformations for the side chains of the following amino acid residues: Leu 280, Cys 284, His 285 and Gln 328 in GTA; Val 87, Cys 284, Asp 302 and Gln 328 in GTB (Figure 27).

In both GTA+H+UDP and GTB+H+UDP, the 177-198 loop and the C-terminal 10 amino acids remain disordered. In some structures (Vrielink *et al.*, 1994; Moréra *et al.*, 1999; Unligil *et al.*, 2000), binding of the donor substrate results in these loops becoming ordered. Persson *et al.* (2001) showed that binding of the donor substrate results in conformational rearrangement with concomitant formation of the acceptor binding site. However, GTA and GTB in complex with UDP and the acceptor disaccharide display the same disordered loops as the native structures, which indicates that the acceptor binding site in GTA and GTB is independent of the ordering of these loops. It has been postulated that the flexible loop may be important in product release (Unligil & Rini, 2000; Unligil *et al.*, 2000), and this is consistent with kinetic data of the GTA enzyme where the R176G mutation causes a 11-fold increase in enzyme turnover with no significant change in donor specificity (Seto *et al.*, 1997). The current structures also show that product release in GTA and GTB is probably aided by steric factors, as the nascent trisaccharide B antigen would conflict with UDP.

Unambiguous electron density is observed for the H antigen and UDP in both GTA and GTB complexes (Figure 28). Most of the aliphatic group substituted at position C1 on the acceptor Gal residue, used as an aid for enzyme assays, is also visible in the electron density maps. Interestingly, the aliphatic tail adopts different conformations in the GTA *versus* the GTB complexes (Figure 29). The lack of side chain bulk at amino acid residue 235 in GTA (Gly 235) may allow the tail to position itself along the protein backbone while in GTB, the presence of Ser 235 forces it into a different conformation.

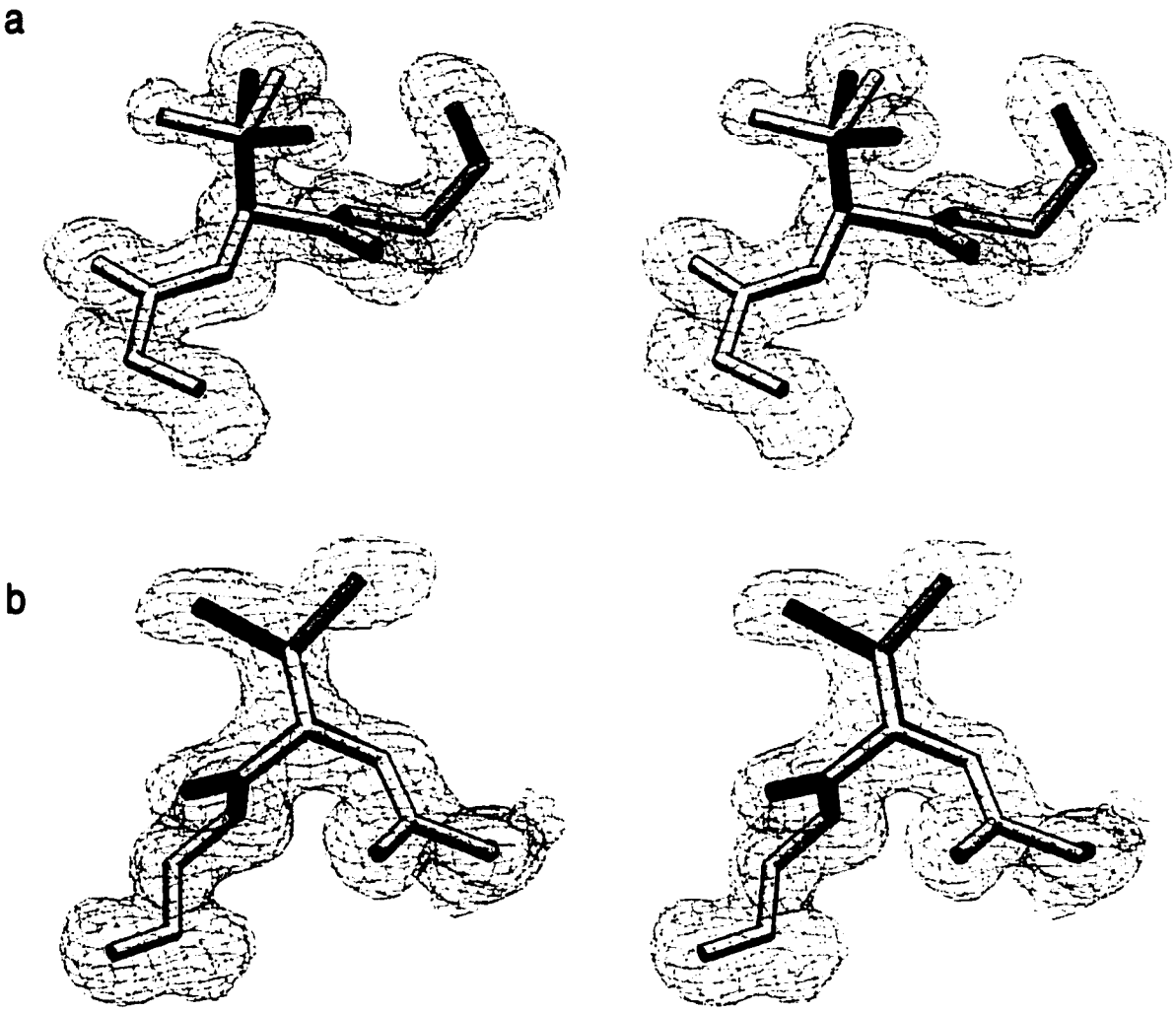

Figure 27. Modelled alternate conformations for side chains in the GTB+H+UDP structure. Electron density is shown in blue, with alternate conformations for the residues shown in white and red. Unambiguous electron density could be observed for several residues in both GTA and GTB. a) and b) show the modeled positions of the GTB Val 87 and Cys 284, respectively.

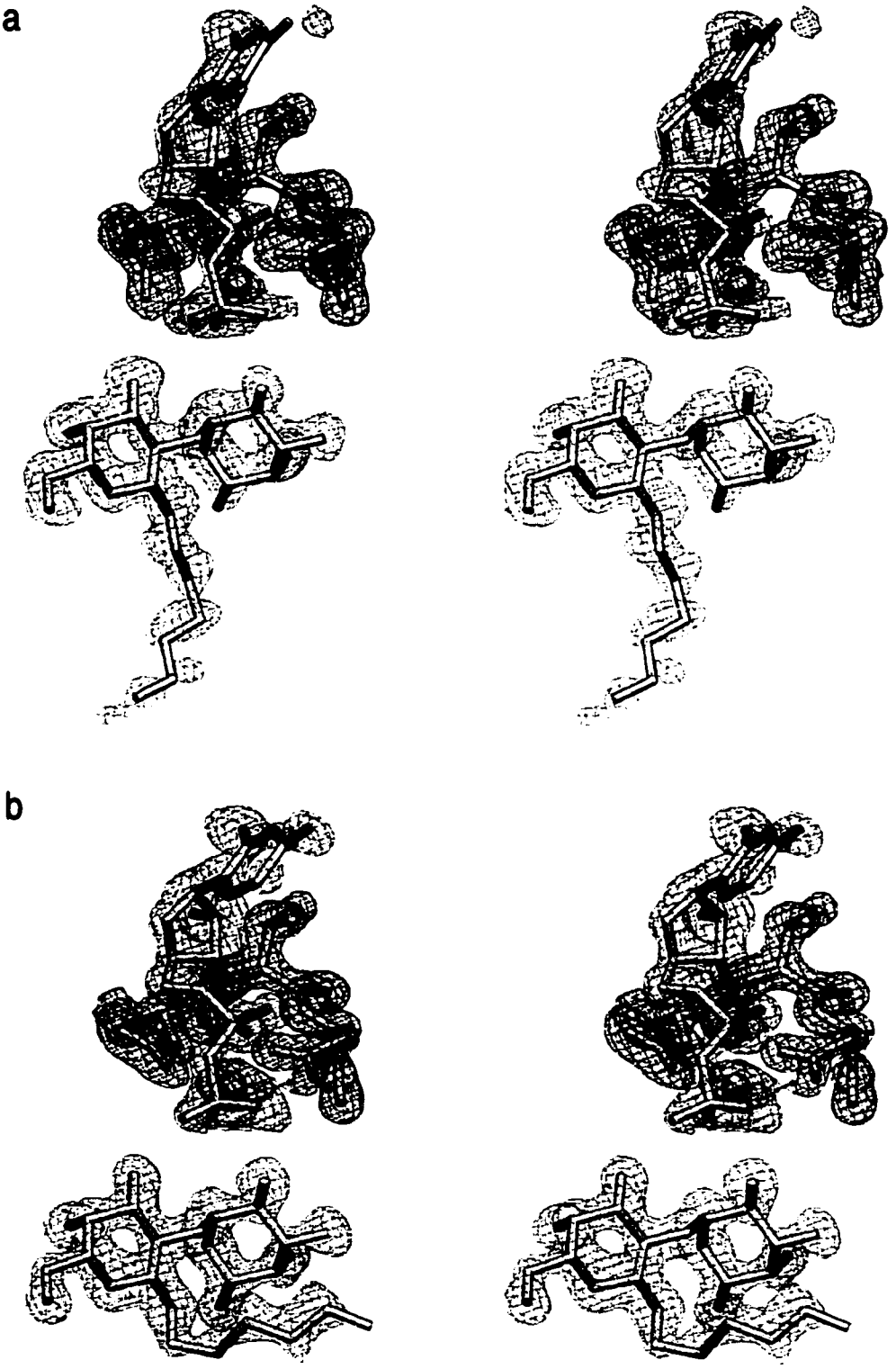

Figure 28. Electron density for the enzymes in complex with H antigen and UDP. Electron density for the DXD motif (Asp 211 Val 212 Asp 213) and manganese is shown in red, UDP in green and H antigen in blue. The aliphatic group substituted at position C1 on the acceptor Gal has assumed a mostly ordered conformation in the crystal and can be seen to adopt different conformations in GTA (a) and GTB (b) Excellent electron density is observed over the entire course of the polypeptide chain with the exception of a disordered loop between residues 177-194, and the final 10 residues of the C-terminus.

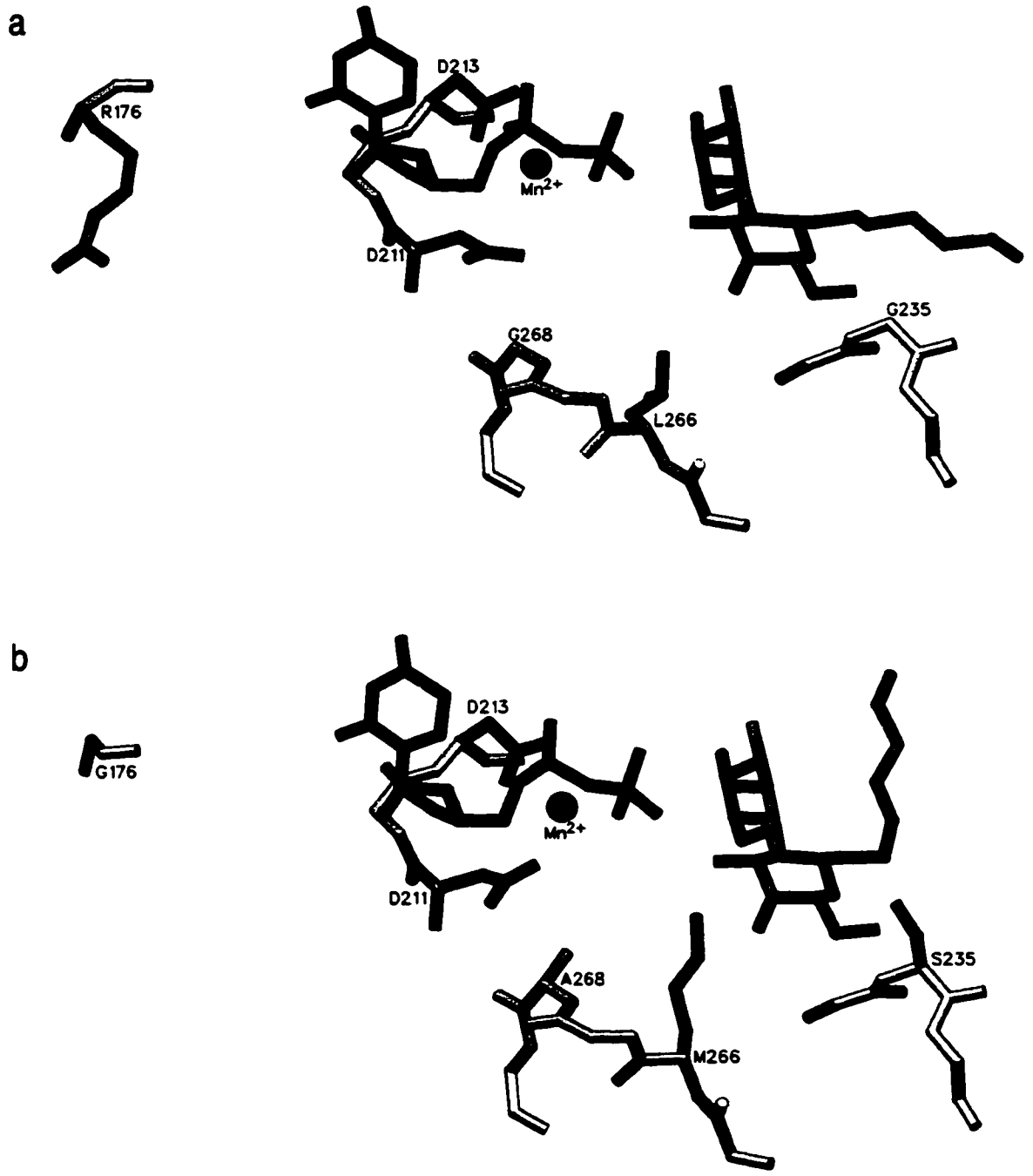

Figure 29. GTA and GTB in complex with UDP and H antigen. The substrates (green) are bound in the active site cleft in proximity of three of the four critical amino acid residues (blue). UDP is bound primarily through coordination with the manganese ion and the DXD motif (red). Arg/Gly 176 in GTA (a) and GTB (b) is removed from the active site and is therefore unlikely to contact the donor sugar. Again, the difference in the conformations of the H antigen aliphatic tail is prominent and is likely due to the critical residue Gly/Ser at position 235. This raises the possibility that Gly/Ser 235 may select between different glycoconjugate structures displaying the H antigen acceptor and may account for the observed three-fold reduction of acceptor Km for GTA compared to GTB.

I- PROTEIN-SUBSTRATE INTERACTIONS

The most significant protein-substrate contacts observed in the antigen binding site of GTB are depicted in Figure 30 (the interactions observed in GTA are identical except there is no contact from Leu 266) and listed in Table 14. The acceptor interacts with the enzyme primarily through the galactose 4-hydroxyl group, making hydrogen bond interactions with the side chains of His 233 and Glu 303, while the Gal 6-hydroxyl group hydrogen bonds to Thr 245 and the Fuc 4'-hydroxyl to Asp 326. These observations agree with kinetic studies using acceptor analogues that the 4-hydroxyl group of Gal is a key polar group in binding of the acceptor (Lowary & Hindsgaul, 1993; Lowary & Hindsgaul, 1994). Similar studies with the acceptor analogues with modified Fuc residues indicated that the 2'-hydroxyl group is also crucial for recognition by both enzymes (Mukherjee *et al.*, 2000). In addition to the interactions listed, the Gal residue of the H antigen was observed to be in stacking interactions with the aromatic side chain of Trp 300. Interactions of the acceptor with UDP further support kinetic findings. Both sugar residues of the acceptor (3-hydroxyl group of Gal and 2'-hydroxyl group of Fuc) make strong hydrogen bonds with the β-phosphate group of the UDP, demonstrating how these groups are crucial to recognition. Further, this suggests that the binding of the donor is a key preliminary step in the formation of the acceptor binding site, agreeing with studies on these enzymes that showed that donor binding must precede acceptor binding (Kamath *et al.*, 1999). In GTB the acceptor makes a van der Waals contact with the third critical residue (Met 266), which is involved in the selection of donor. A large portion of the aliphatic group substituted at position C1 on the H antigen Gal residue, is ordered in both GTA and GTB complexes. It is not surprising that this group makes no contact with the protein, as C1 is the substitution point for the H-antigen to larger glycoproteins and glycolipids that are the natural substrates for GTA and GTB. The UDP substrate not only interacts with main chain protein atoms, but also with the side chain of Asp 213, which forms part of the DXD motif.

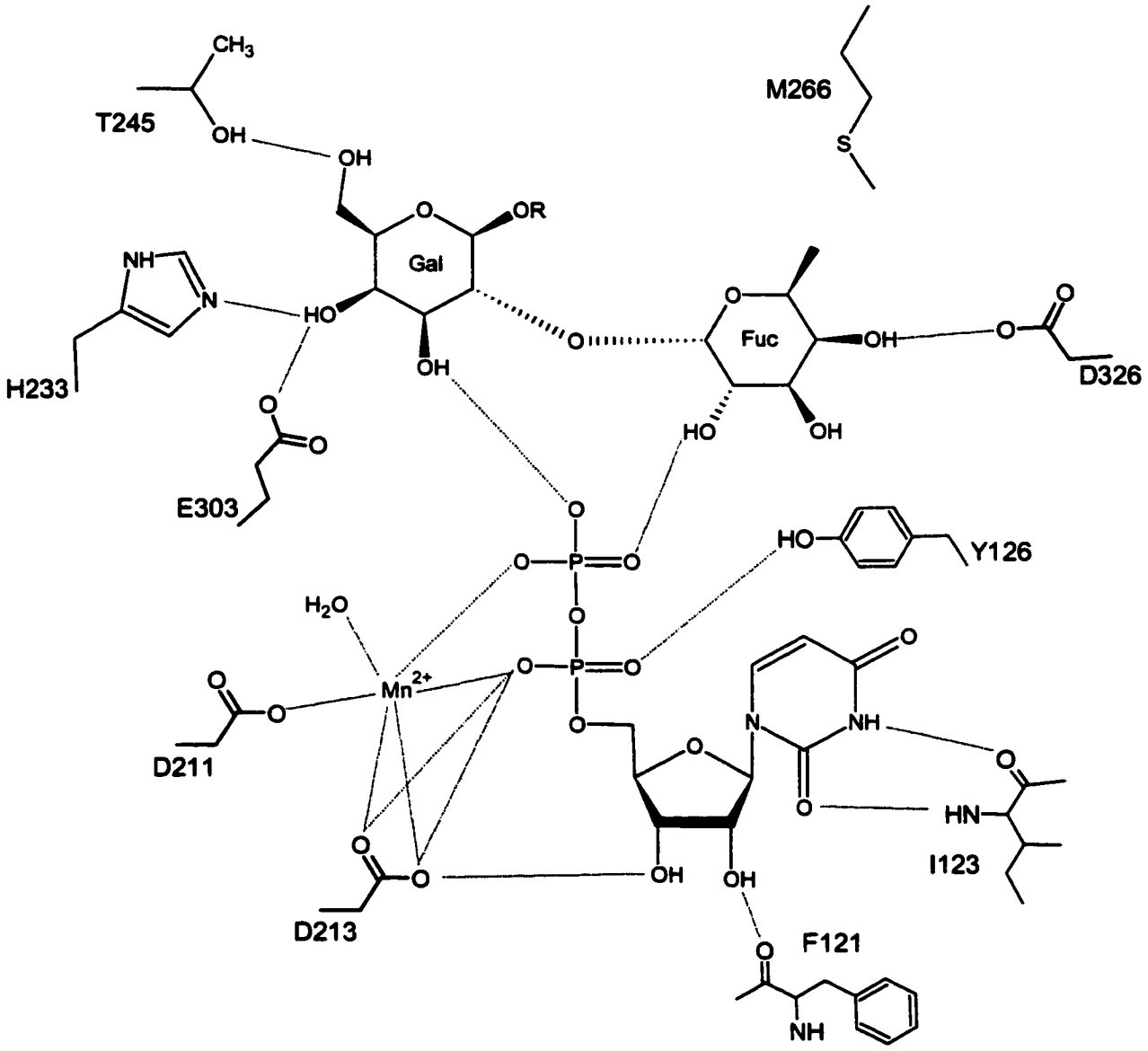

Figure 30. Protein-substrate contacts for GTB. Significant contacts observed between GTB, the H antigen disaccharide and UDP in the 1.32 Å resolution structure. Contacts between UDP, H antigen and GTA are similar, except that residue Leu 266 makes no contact with the substrate. In addition to the interactions shown here, aromatic stacking is observed between the uracyl moiety of UDP and Tyr 126 as well as the Gal of the H antigen and Trp 300.

Table 14. Interactions observed in the structures of GTA and GTB in complex with UDP and H antigen

GTA +H antigen+UDP		Distance (Å)	GTB +H antigen+UDP (B)		Distance (Å)
Mn^{2+}	OD2 Asp 211	3.28	Mn^{2+}	OD2 Asp 211	2.33
Mn^{2+}	OD1 Asp 213	2.81	Mn^{2+}	OD1 Asp 213	2.62
Mn^{2+}	OD2 Asp 213	2.44	Mn^{2+}	OD2 Asp 213	2.38
Mn^{2+}	O1α UDP	2.24	Mn^{2+}	O1α UDP	2.19
Mn^{2+}	O1β UDP	2.33	Mn^{2+}	O3β UDP	2.23
Mn^{2+}	OH2	3.07	Mn^{2+}	OH2	2.17
Gal O3	O2β UDP	2.69	Gal O3	O1β UDP	2.56
Fuc O2	O3β UDP	2.70	Fuc O2	O2β UDP	2.79
Gal O4	NE2 His 233	2.82	Gal O4	NE2 His 233	2.85
Gal O4	OE1 Glu 303	3.21	Gal O4	OE1 Glu 303	3.21
Gal O4	OE2 Glu 303	2.66	Gal O4	OE2 Glu 303	2.68
Gal O5	ND2 His 233	3.18	Gal O5	ND2 His 233	3.09
Gal O6	OG1 Thr 245	2.79	Gal O6	OG1 Thr 245	2.78
Gal O6	NE1 Trp 300	3.39	Gal O6	NE1 Trp 300	3.36
Fuc O4	OD2 Asp 326	2.68	Fuc O4	OD2 Asp 326	2.66
			Fuc O5	CE Met 266	3.13
N3 UDP	O Ile 123	2.97	N3 UDP	O Ile 123	2.91
O2 UDP	O Phe 121	3.33	O2 UDP	O Phe 121	3.47
O2 UDP	O Ile 123	2.97			
O2* UDP	O Phe 121	2.66			
O3* UDP	OD1 Asp 211	3.38			
O3* UDP	N Val 212	3.06	O3* UDP	N Val 212	3.15
O3* UDP	OD2 Asp 213	3.49	O3* UDP	OD2 Asp 213	2.87
O1α UDP	OD1 Asp 213	2.86	O1α UDP	OD1 Asp 213	2.96
O1α UDP	OD2 Asp 213	2.99	O1α UDP	OD2 Asp 213	2.94
O2α UDP	OH Tyr 126	2.53	O2α UDP	OH Tyr 126	2.58

The DXD motif is located in the middle of the active site cleft where it serves to coordinate Mn^{2+}. In addition to coordinating with the β-phosphate group of the UDP-donor, the manganese ion has been suggested to have a role in catalysis (Busch *et al.*, 1998). In GTA and GTB complexes, the manganese is in an octahedral coordination state, interacting with both phosphate groups of the donor and a water molecule. In addition, both aspartate residues (Asp 211, Asp 213) interact with the manganese ion, including a bidentate coordinate bond to Asp 213 (Figure 31).

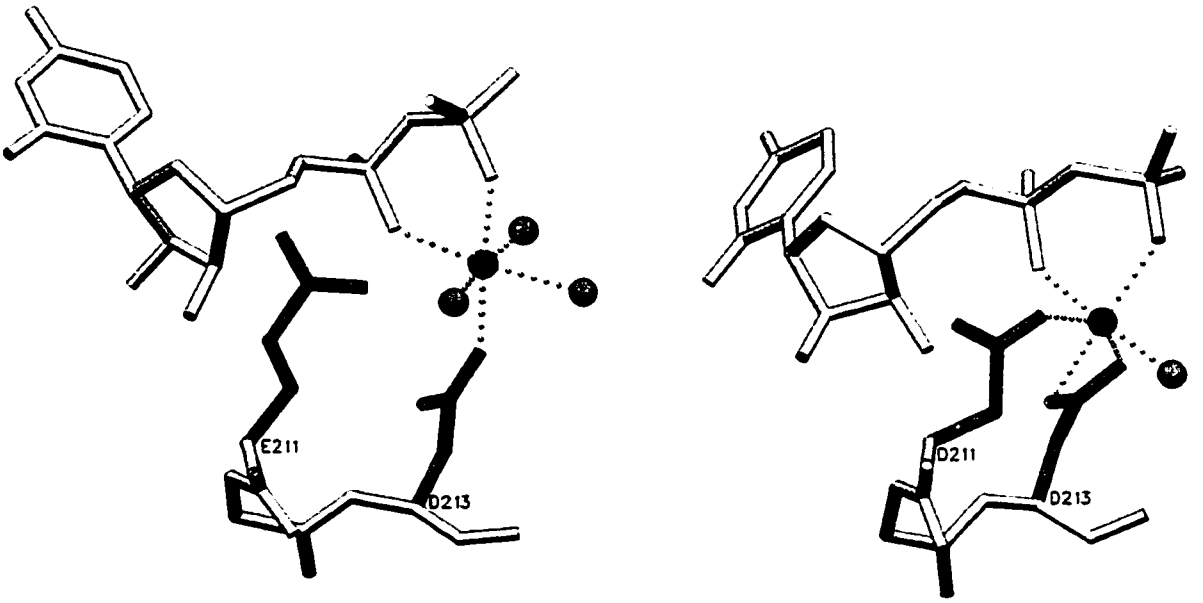

Figure 31. Coordination of the manganese ion by the DXD motif in inverting and retaining glycosyltransferases. The GTB (right) DXD coordination of manganese (red) is typical of retaining enzymes, where both aspartates side chains (green) contact the ion, including a bidentate coordinate bond between the second aspartate and manganese. In inverting enzymes, such as GnTI (left), the first residue is not highly conserved, usually contacting the donor sugar or the ribose of UDP, and only the second contacts the manganese ion. The protein main chain and UDP are shown in white, DXD side chains in green and water molecules in blue.

This precise coordination of the Mn^{2+} ion with the aspartate residues of the DXD has also been observed in the structures of the two other retaining glycosyltransferases that have been solved (Gastinel *et al.*, 2001; Persson *et al.*, 2001). However, an examination of the inverting glycosyltransferase structures crystallized in the presence of manganese (Charnock & Davies, 1999; Unligil *et al.*, 2000; Pederson *et al.*, 2000) shows that only one of the residues of the DXD motif interact with the manganese ion while the other contacts either the ribose or sugar moieties (Figure 31). A water molecule or low-occupancy Mn^{2+} ion in β4Gal-T1 (Gastinel *et al.*, 1999) shows similar coordination. The observation that the first residue of the DXD motif in inverting galactosyltransferases is not as highly conserved as the second residue (Unligil & Rini, 2000) suggests that these different coordination patterns may be characteristic of inverting *versus* retaining glycosyltransferases. The different coordination of the Mn^{2+} ion will impact on the catalytic action of the enzyme. For example, the single aspartic acid residue complexed to Mn^{2+} in inverting glycosyltransferases might make the metal ion a stronger Lewis acid which, when coordinated to the phosphate group, would make UDP a better leaving group than in the retaining enzymes where both aspartic acid residues of the DXD are coordinated to Mn^{2+}.

II- GTA & GTB Enzyme Specificity

As a preliminary step in determining enzyme specificity, our collaborators (Monica Palcic, University of Alberta; Nina Seto, National Research Council of Canada) have combined mutagenesis studies with sensitive kinetic analysis to determine and quantify the source for donor specificity. Ole Hindsgaul's group at the University of Alberta have determined key groups required for recognition of the H antigen by GTA and GTB through the synthesis of H acceptor analogues

with sugar hydroxyl groups replaced with other moieties. Their findings are summarized in the paragraphs below.

a- Enzyme kinetics and donor specificity

In an effort to understand the influence of each of the critical amino acid residues in GTA and GTB on enzyme kinetics and donor specificity, Seto and colleagues expressed synthetic genes encoding the catalytic domain of GTA, GTB or "chimeric" mutants in *E. coli* (Seto *et al.*, 1997; Seto *et al.*, 1999). The high yields of enzyme achievable by this method allowed for more precise kinetic analysis of the blood group glycosyltransferases. The results of these investigations of the activity of GTA, GTB and various mutants are summarized in Table 15.

Table 15. Mutational analysis of the GTA and GTB donor specificity and kinetic activity

Critical amino acid identity				UDP-GalNAc				UDP-Gal			
				K_A (μM)	K_B (μM)	k_{cat} (s^{-1})	K_{ia} (μM)	K_A (μM)	K_B (μM)	k_{cat} (s^{-1})	K_{ia} (μM)
A^{176}_{Arg}	A^{235}_{Gly}	A^{266}_{Leu}	A^{268}_{Gly}	15	13	4.9	4.9	23	6.3	0.020	23
B	A	A	A	43	126	55	12	68	43	0.037	76
B	B	A	A	206	51	24	121	708	36	0.030	644
A	B	A	A	44	22	5.8	39	529	18	0.010	568
A	A	B	A	35	75	7.4	8	30	5.6	0.95	24
A	B	B	A	251	184	12	42	292	87	4.4	67
B	A	B	A	67	508	5.5	12	100	55	3.4	27
A	B	A	B[1]	367	43	2	-	1058	128	3.0	-
B	B	A	B[1]	636	129	15	-	1130	183	3.1	-
B	B	B	A	348	253	10	139	420	56	2.1	105
B^{176}_{Gly}	B^{235}_{Ser}	B^{266}_{Met}	B^{268}_{Ala}	281	285	0.3	122	54	34	6.5	27

[1] Monica Palcic, personal communication

Wild-type GTA, mutants BAAA, BBAA and ABAA have the third and fourth amino acids derived from the A enzyme and show predominantly A activity, with B activity less than 1% of A (Seto *et al.*, 1999). Surprisingly, mutant BAAA showed a 11-fold increase in the catalytic constant (k_{cat}) compared to the wild-type A enzyme (Seto *et al.*, 1997), while its donor specificity was not significantly altering. It was proposed that the removal of the Arg side chain at amino acid 176 more effectively stabilized the transition state (Seto *et al.*, 1997), indicating a role in enzyme turnover. Hybrid enzymes AABA, ABBA, BBBA and BABA have third and fourth amino acids derived from GTB and GTA, respectively. These enzymes all possess dual A/B specificity and can catalyze both A and B reactions (Seto *et al.*, 1999). Interestingly, AABA possesses only one B amino acid, yet shows significant B activity. Conversely, mutant BBAB shows significant A activity. Together, these findings serve to underline the importance of residue 266, and possibly 268, in determining enzyme specificity (Seto *et al.*, 1999).

Based on these results, the authors hypothesized that the high degree of donor specificity shown by the A and B glycosyltransferases must arise from a high degree of complementarity between the protein surface and the substrate structure. The difference in size between the Gal and GalNAc is correlated by the difference in size between the Leu and Met at position 266 of GTA and GTB, respectively (Seto *et al.*, 1999). Further, the Leu 266 side chain of GTA could form a pocket of the appropriate size and shape to accommodate the acetamido group of the GalNAc sugar (Figure 32).

b- Acceptor specificity

Hindsgaul and colleagues performed kinetic evaluation on synthesized acceptor analogs to identify the functional groups of the H acceptor that are important in its recognition by the A and B

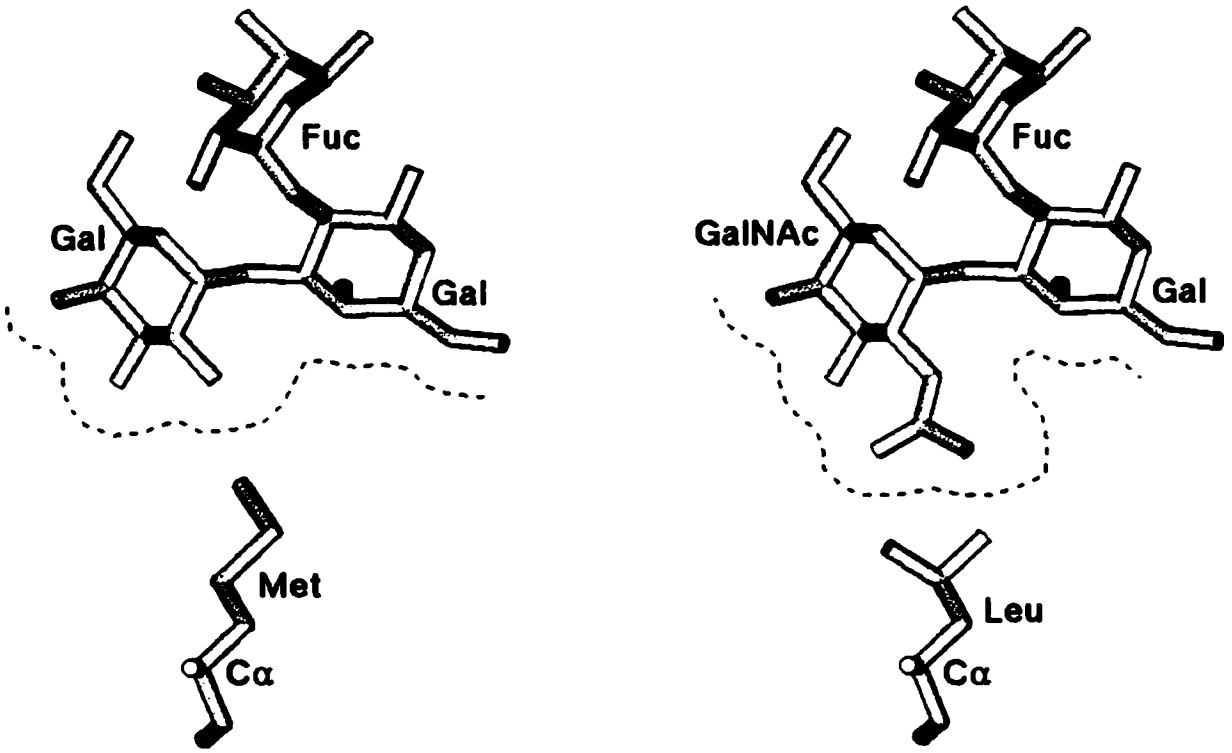

Figure 32. Model of the contact between residue Leu/Met 266 in GTA/GTB and their respective antigens. The van der Waals surface of the antigens are shown by a dashed line. GTA (right) contains Leu 266, which is smaller than Met 266 in GTB (left). The resulting pocket that is formed is of the appropriate size and shape to accommodate the acetamido group on the terminal GalNAc of the A antigen (Seto *et al.*, 1999).

glycosyltransferases. For the Gal residue of the H antigen, only disaccharides with an intact 3,4-diol in a *galacto* configuration were recognized as substrates, though modifications at the Gal 3- and 6-hydroxyl positions were tolerated (Lowary & Hindsgaul, 1993; Lowary & Hindsgaul, 1994). Modifications at the 4-hydroxyl position of Gal, however, abolished the recognition of the disaccharide, suggesting that it is a key polar group crucial in binding to the enzyme (Lowary & Hindsgaul, 1993; Lowary & Hindsgaul, 1994). The 2'-hydroxyl group of Fuc was also revealed to be crucial in binding to the GTA and GTB enzymes (Mukherjee *et al.*, 2000). In the same study, it was shown that while the Fuc 4'-hydroxyl and C-6' groups were crucial for recognition by GTB, GTA tolerated modifications to these same groups. In general, it was found that GTB is less tolerant than GTA to modifications of the acceptor structure (Lowary & Hindsgaul, 1993; Lowary & Hindsgaul, 1994; Mukherjee *et al.*, 2000).

The authors postulate that the status of the acceptor functional groups is related to their role in protein-substrate interactions. The 3'-hydroxyl and 4'-hydroxyl groups of Fuc were proposed to be hydrogen bond acceptors in GTA and GTB based on the kinetic studies, while the 2'-hydroxyl group is likely to be both a hydrogen bond donor and acceptor in the active site (Mukherjee *et al.*, 2000). A negatively charged residue is postulated to be in the vicinity of Gal 3-hydroxyl in the acceptor binding site (Lowary & Hindsgaul, 1994).

As discussed in section I (PROTEIN-SUBSTRATE INTERACTIONS), the structural data agrees with the conclusions of Hindsgaul and colleagues. The key polar group Gal 4-hydroxyl makes hydrogen bond interactions with the side chains of His 233 and Glu 303, while the Gal 6-hydroxyl group hydrogen bonds to Thr 245 and the Fuc 4'-hydroxyl to Asp 326. The 3-hydroxyl group of Gal is positioned within hydrogen bonding distance to the negatively charged diphosphate group of UDP, which also interacts with the key polar 2'-hydroxyl group of Fuc.

c- Structure-function relationship of the critical amino acids

One of the main objectives of this study was to determine the location of each critical amino acid and how it affects donor and acceptor recognition. Because both enzymes show constitutive UDP-donor hydrolase activity even in the absence of acceptor (Monica Palcic, personal communication), it was not possible to co-crystallize the wild-type enzymes with their respective donors. However, the structures of GTA and GTB in complex with the H antigen acceptor and with UDP (Figure 29) allow the position of the donor sugars to be modelled unambiguously (Figure 33) and reveals that only two of the four critical amino acid residues can make contact with the donor. As a consequence of the size of the Met 266 in GTB, the conformation of the Gal donor sugar is rotated slightly compared to the conformation of the GalNAc donor in GTA.

The first of the critical residues (Arg/Gly 176) is the furthest from the substrate binding site, and is located near the 177-198 disordered loop. Though this residue is unlikely to contact the donor sugar, its mutation from Arg to Gly causes a 11-fold increase in enzyme turnover without significantly changing donor specificity, suggesting a role in product release. The second critical amino acid residue (Gly/Ser 235) is also not positioned to directly contact the donor, although this residue does force the aliphatic tail on the acceptor to assume different conformations in GTA and GTB (Figure 33). The different conformations might account for the observed three-fold reduction of acceptor Km for GTA compared to GTB (Seto *et al.*, 1997; Seto *et al.*, 1999), and raises the possibility that Gly/Ser 235 may select between different glycoconjugate structures displaying the H antigen acceptor.

Just two (Leu/Met 266 and Gly/Ala 268) of the four critical amino acid residues are observed to occupy positions in the active sites of GTA and GTB that can contact the nucleotide-sugar residues. This agrees with kinetic studies using GTA/GTB chimeric enzymes that demonstrated that Leu/Met 266 dominates A/B donor substrate specificity

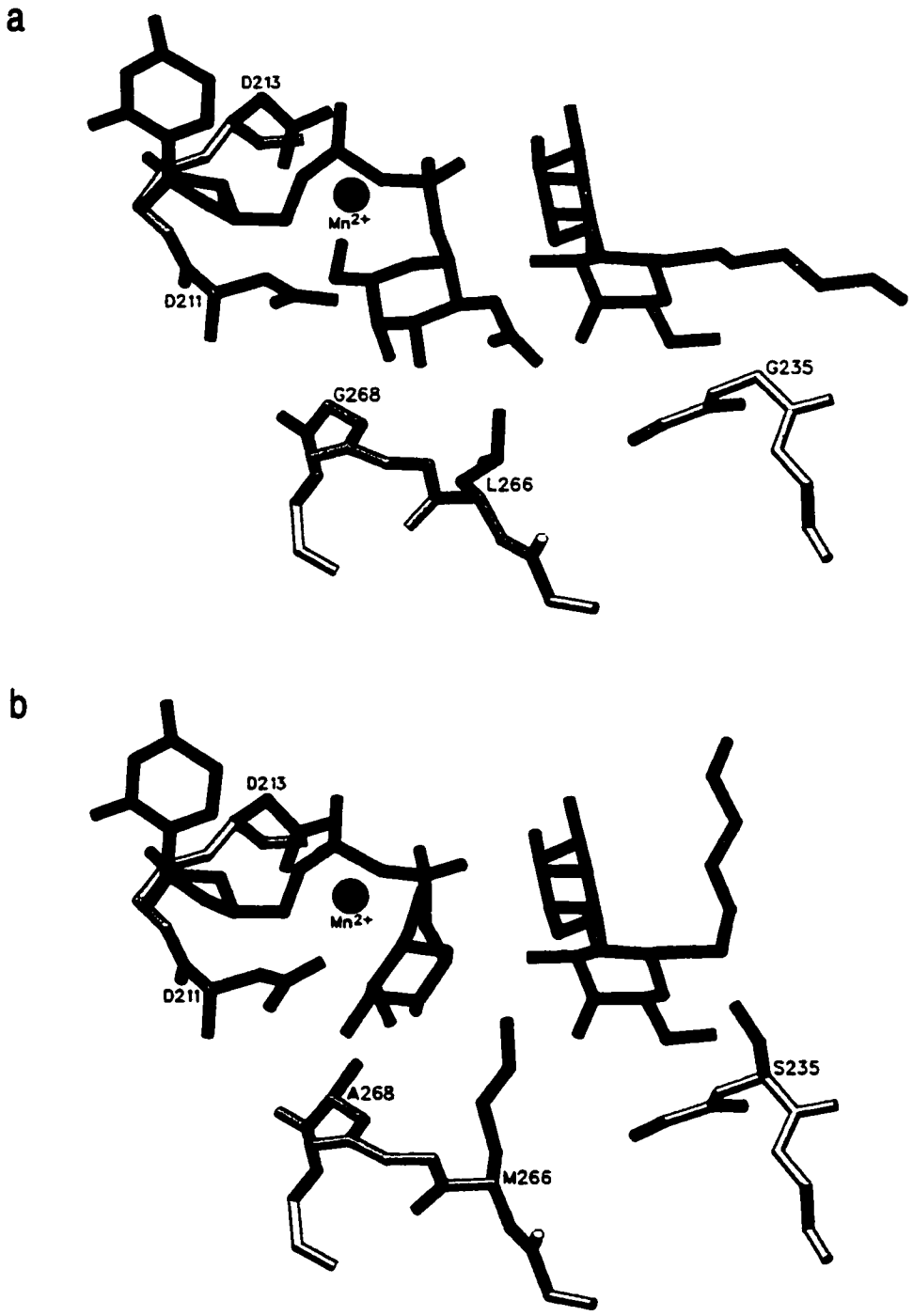

Figure 33. Model of GTA and GTB in complex with donor and acceptor substrates. Based on the structures of GTA/GTB in complex with H antigen and UDP, the respective donor sugars (GalNAc, Gal) were modeled into the active site. Location of the substrates (green) in the GTA (a) and GTB (b) binding pockets with respect to 3 of the 4 "critical" amino acids (blue) that differ between the two enzymes and which serve to differentiate between the two sugar-nucleotide donors. Clearly, only residues Leu/Met 266 and Gly/Ala 268 are positioned to interact with the donor sugar.

(Seto *et al.*, 1997; Seto *et al.*, 1999), with Gly/Ala 268 having lesser effect. The present structures show that only Leu/Met 266 is positioned to contact the characteristic acetamido/hydroxyl groups and so distinguish between UDP-GalNAc and UDP-Gal. This is a clearly complementary interaction where the smaller Leu 266 in GTA accommodates the larger acetamido group in UDP-GalNAc while the corresponding smaller hydroxyl group on UDP-Gal is accommodated by the larger Met 266 in GTB. Gly/Ala 268 can only contact moieties of the donor sugar residues that are identical in UDP-GalNAc and UDP-Gal (specifically hydroxyl groups O3 and O4), which does not explain how this residue could contribute to substrate specificity. Indeed, while it is easy to see how Met 266 of GTB excludes the bulky acetamido group of UDP-GalNAc, closer examination was required to understand how the larger active site cleft in GTA selects against smaller UDP-Gal.

Figure 34 depicts the two critical amino acid residues Leu/Met 266 and Gly/Ala 268 positioned relative to the modelled closest approach of their respective sugar donor residues as determined by the observed position of UDP in GTA and GTB. Residues Met 266 and Ala 268 in GTB are both bulkier than the corresponding residues in GTA, with the result that the floor of the active site is raised in GTB with respect to GTA. This provides GTB with a significantly smaller and therefore more conformationally restricted active site cleft, such that these residues serve to exclude UDP-GalNAc and position UDP-Gal. In GTA, the floor of the active site is lower, providing a larger active site cleft that allows a high degree of motion to a potential UDP-Gal substrate; i.e. the active site cleft in GTA is essentially a large void to UDP-Gal where it has little possibility of forming the complementary interactions that are required for substrate recognition and enzyme action. In addition, the lowered floor of the active site cleft in GTA reveals His 233 that can form a hydrogen bond with the acetamido group of UDP-GalNAc and thus allows the sugar residue to be selectively positioned.

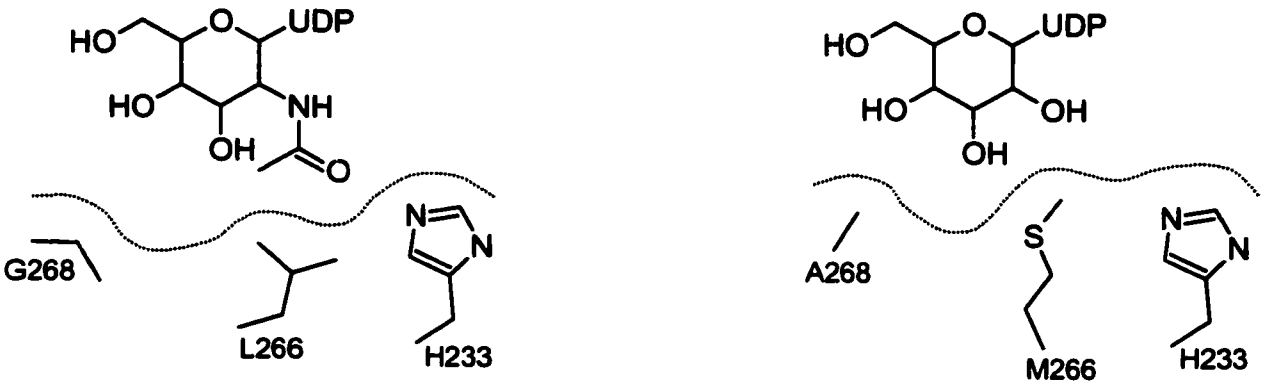

Figure 34. Specificity mechanism. Schematic view of the active sites of GTA (left) and GTB (right) showing how the enzymes achieve substrate specificity. Each enzyme shows a high degree of complementarity with the modeled positions of its respective donor sugar residue. GTB can easily exclude bulkier A donor (UDP-GalNAc) as the floor of the active site cleft contains the larger critical residues Met 266 and Ala 268 that restrict the space available to UDP-donor sugar moieties. In contrast, the smaller critical residues Leu 266 and Gly 268 in GTA form a larger active site cleft which would lack complementarity with the smaller B donor (UDP-Gal).

III- NUCLEOPHILE AND MECHANISM

Classified as "retaining" glycosyltransferases, GTA and GTB preserve the stereochemistry of the donor sugar-UDP bond upon transfer to acceptor. However, as these reactions are thought to proceed through nucleophilic attack that requires inversion of stereochemical configuration, it has been postulated that retaining enzymes utilize a two-step mechanism whereby the donor sugar is transferred first to an enzyme nucleophile and subsequently to the acceptor (Sinnott, 1990; Davies *et al.*, 1998; Takayama *et al.*, 1999). Several candidate residues have been put forward for various GTs (Vrielink *et al.*, 1994; Charnock & Davies, 1999; Gastinel *et al.*, 1999; Moréra *et al.*, 1999; Pedersen *et al.*, 2000; Unligil *et al.*, 2000; Gastinel *et al.*, 2001; Mulichak *et al.*, 2001; Persson *et al.*, 2001). The closest nucleophilic residue to the transfer site in GTA and GTB is Asp 211, which is one of the residues involved in the Mn^{2+} metal-binding DXD motif; however, the resulting enzyme-sugar complex would position the sugar residue at least 6 Å from the observed position of the bound acceptor. Further, to undertake transfer to Asp 211 the donor sugar would be required not to have significant contact with Leu/Met 266 and Gly/Asp 268, which are known to determine enzyme donor specificity. Figure 35 shows the relative positions observed for the acceptor and UDP together with the modelled positions of donor sugar and Glu 303 side chain. It is evident that the donor is positioned for inverting transfer to Glu 303, and the resulting complex (Figure 35) would subsequently be positioned for a second inverting transfer to the acceptor. However, the most persuasive argument for proposing Glu 303 as the enzyme nucleophile is that it would require the donor sugar residue to be positioned relative to the critical residues Leu/Met 266 and Gly/Ala 268 such that selectivity can be applied not only during the first transfer, but also during the second transfer. This assignment is supported by radiochemical enzyme assays that show that the mutation in GTB of Glu 303 to Ala yields a protein that displays the same behaviour during purification and crystallization as the native enzymes, yet shows 30,000-fold decrease in

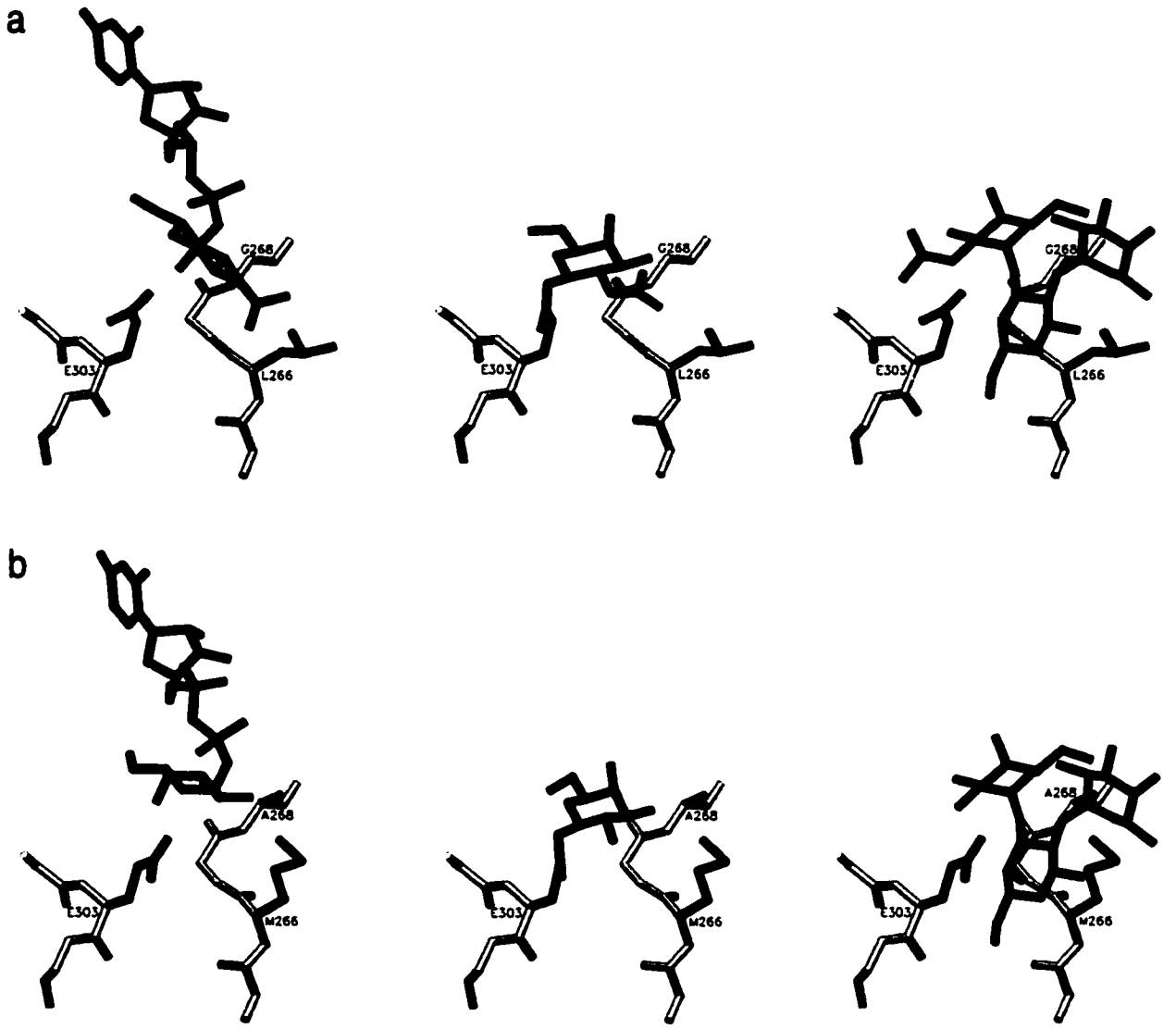

Figure 35. Location of the proposed nucleophile. The left panel shows the observed position of the UDP (green) in GTA (a) GTB (b) together with the modeled positions of the GalNAc and Gal moieties and candidate nucleophile Glu 303 (red). GTA and GTB are "retaining" glycosyltransferase enzymes and transfer of donor to acceptor probably proceeds through a two-step mechanism where the donor is first transferred with inversion to a nucleophile (middle panel), and then transferred again with inversion to the acceptor to yield a product with conserved stereochemistry (right panel). The position of the nucleophile with respect to residue 266 and 268 (blue) allows for specific recognition of the donor sugar not only in the first transfer, but also in the second. The different modelled positions for Glu 303 in GTA and GTB is a measure of the space surrounding this residue.

specific activity. Though the mutant activity was carefully measured, trace activities may be present due to contamination by wild-type enzymes evolved from used affinity columns and/or errors in translation during protein over-expression that generate small amounts of wild-type or active mutant enzymes (Ly & Withers, 1999). For these reasons, steps were taken to avoid cross-contamination of wild-type and mutant samples.

6. GTA AND GTB IN THE PRESENCE OF H ANTIGEN

The structures of GTA and GTB in complex with the H antigen alone were solved to 1.58 and 1.46 Å, respectively. Data collection and refinement statistics can be found in Table 16.

Table 16. Data collection and structure refinement statistics for GTA and GTB in complex with H antigen

Data collection statistics	GTA+H antigen	GTB+H antigen
Wavelength (Å)	1.15	1.15
Space group	$C222_1$	$C222_1$
Cell dimensions		
a (Å)	52.7	52.7
b (Å)	150	151
c (Å)	79.8	79. 5
Resolution (Å)	50-1.58	50-1.46
Unique reflections	43,304	54,816
Completeness (%)	98.9 (100)	98.9 (99.8)
Rmerge	0.062 (0.21)	0.058 (0.41)
I/σ	47.2 (14.3)	37.1 (3.49)
Refinement statistics		
Resolution (Å)	20-1.58	20-1.46
Rwork (%)	16.5	18.1
Rfree (%)	18.4	19.7
Number of atoms		
Protein	2208	2197
Water	359	360
Ion	4 (Hg^{2+})	4 (Hg^{2+}), 1 (Mn^{2+})
Substrate	28 (H antigen)	28 (H antigen)
Average B-factor (Å^2)		
Protein	16.6	18.7
Solvent	32.4	36.1
Ion	28.2 (Hg)	40.7 (Hg), 47.1 (Mn)
Substrate	19.1 (H antigen)	20.8 (H antigen)
Rms deviations		
Bonds (Å)	0.0094	0.0078
Angles (°)	1.5	1.4

Where applicable, statistics for the last resolution shell are shown in parentheses.

Unambiguous electron density was present for the protein and the acceptor structure (Figure 36). The position and conformation of the bound H antigen are the same as those observed in the enzyme-H-UDP complexes. Again, the region corresponding to residues ~177-196 was disordered in both structures.

In the GTA+H complex, the side chains of amino acid residues Asn 113, Gln 275, Cys 284, Asp 302 and Gln 328 were observed to have alternate conformation while in GTB+H, alternate conformations were modelled for Met 214, Cys 284, His 285 and Gln 328.

Though previous kinetic analysis revealed that binding of the donor substrate must precede acceptor binding for catalysis (Kamath *et al.*, 1999), these structures surprisingly show that the enzymes are capable of binding acceptor in the absence of UDP donor. It has been reported for several GTs that the acceptor binding site is formed only after the UDP donor is bound. Similarly, it had earlier been thought that UDP formed an essential part of the acceptor binding site for GTA and GTB. However, it now appears that while interactions between UDP and H antigen are not essential for binding of the acceptor, they may serve to strengthen the affinity with which the enzyme binds this substrate. Interactions between the GTA/GTB and the H antigen are shown in Table 17. Manganese is present in the GTB complex while it is absent in the GTA structure, indicating that binding of the H antigen is also independent of the coordination of manganese. In addition, the manganese coordinate bonds are longer than in the GTB+H antigen+UDP complex, indicating that tight coordination of Mn^{2+} may require the presence of UDP.

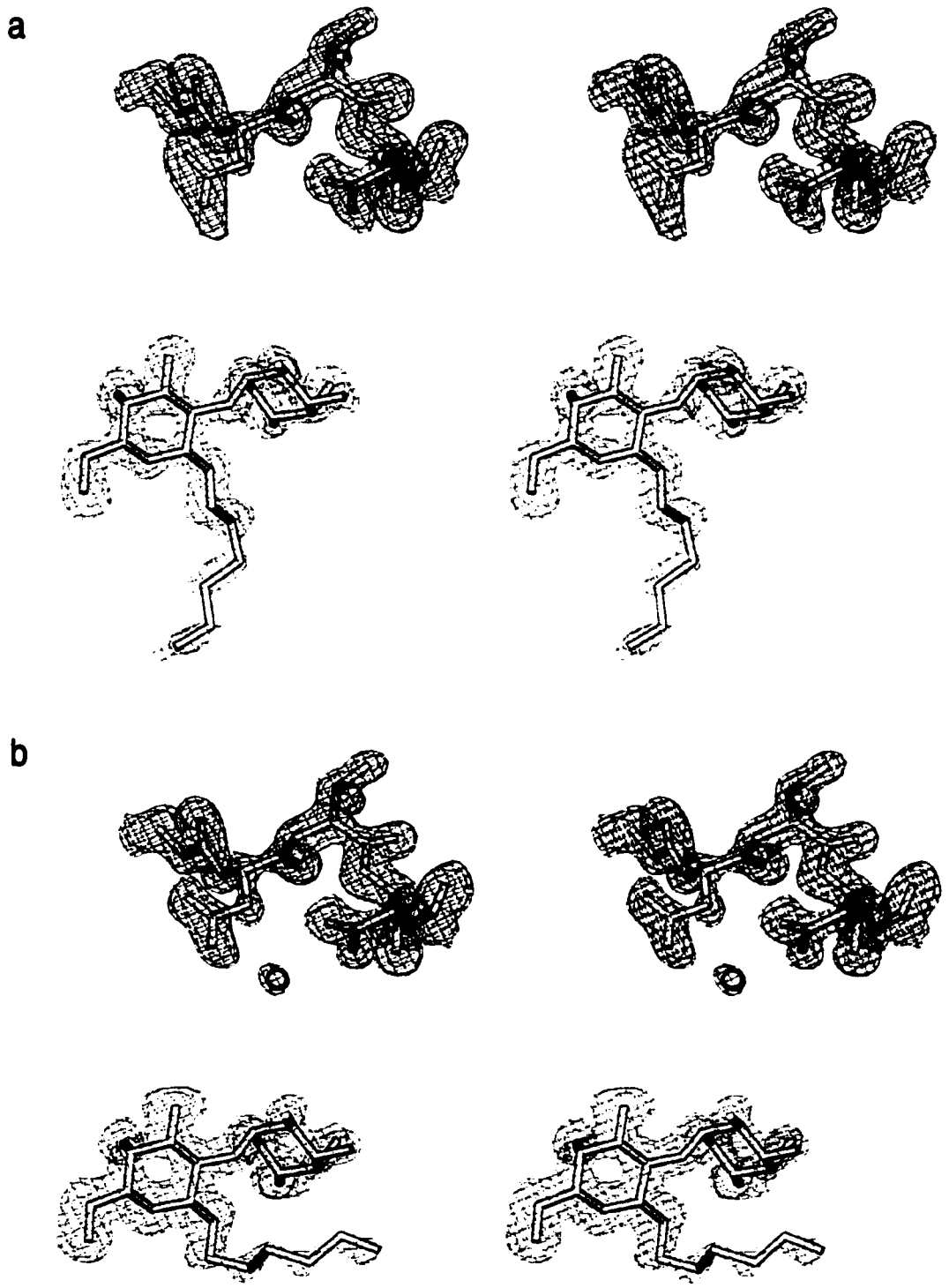

Figure 36. Electron density for the H antigen complexes. Electron density for the DXD motif (and manganese) is shown in red and the H antigen in blue. The aliphatic tail of the acceptor, which binds in the same manner and position as in the enzyme+UDP+H, adopts different conformation in GTA (a) and GTB (b). Despite important interactions of the H antigen with UDP in the previous GTA/GTB complexes and kinetic data showing that donor binding must precede acceptor binding in the complexes, it is evident that acceptor can bind independently of UDP-donor. In addition, manganese coordination is not necessary for the formation of the enzyme-H antigen complex, as shown by its absence in GTA (a).

Table 17. Significant contacts observed in the structures of GTA and GTB in complex with H antigen

GTA +H antigen		Distance (Å)	GTB +H antigen		Distance (Å)
			Mn^{2+}	OD2 Asp 211	2.71
			Mn^{2+}	OD2 Asp 213	2.90
			Mn^{2+}	CE Met 214	2.76
Gal O4	ND2 His 233	2.80	Gal O4	NE2 His 233	2.86
Gal O4	OE1 Glu 303	3.20	Gal O4	OE1 Glu 303	3.28
Gal O4	OE2 Glu 303	2.67	Gal O4	OE2 Glu 303	2.77
Gal O5	ND2 His 233	3.18	Gal O5	ND2 His 233	3.19
Gal O6	OG1 Thr 245	2.78	Gal O6	OG1 Thr 245	2.77
Gal O6	NE1 Trp 300	3.38	Gal O6	NE1 Trp 300	3.34
Fuc O4	OD2 Asp 326	2.72	Fuc O4	OD2 Asp 326	2.69
			Fuc O5	CE Met 266	3.18

Part II: ANTI-BLOOD GROUP A Fv

1. OVERALL STRUCTURE

The structure of the anti-blood group A Fv was determined to 2.2 (A) and 1.1 Å (B) using data collected at the University of Saskatchewan and CHESS, respectively. A typical crystal for the AC1001Fv is shown in Figure 37. Table 18 shows data collection and refinement statistics for both structures. The lower resolution structure was determined first and subsequently used to solve the structure from data collected at CHESS.

Table 18. Data collection and structure refinement statistics for anti-blood group A Fv

Data collection statistics	University of Saskatchewan (A)	CHESS (B)
Wavelength (Å)	1.5418	0.908
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a (Å)	51.7	50.9
b (Å)	58.0	57.4
c (Å)	83.6	83.3
Resolution (Å)	20-2.2	40-1.0
Unique reflections	12,348	94,005
Completeness (%)	93.0 (71.8)	94.3 (86.1)
Rmerge	0.10	0.065 (0.31)
I/σ	-	29.7 (7.49)
Refinement statistics		
Resolution (Å)	6.0-2.2	20.0-1.1
Rwork (%)	18.3	11.3
Rfree (%)	22.3	14.1
Number of atoms		
Protein	1737	1931
Water	62	436
Average B-factor (Å^2)		
Protein	27.4	11.6
Solvent	37.1	32.0
Rms deviations		
Bonds (Å)	0.0062	0.015
Angles	1.4°	0.033 Å

Where applicable, statistics for the last resolution shell are shown in parentheses.

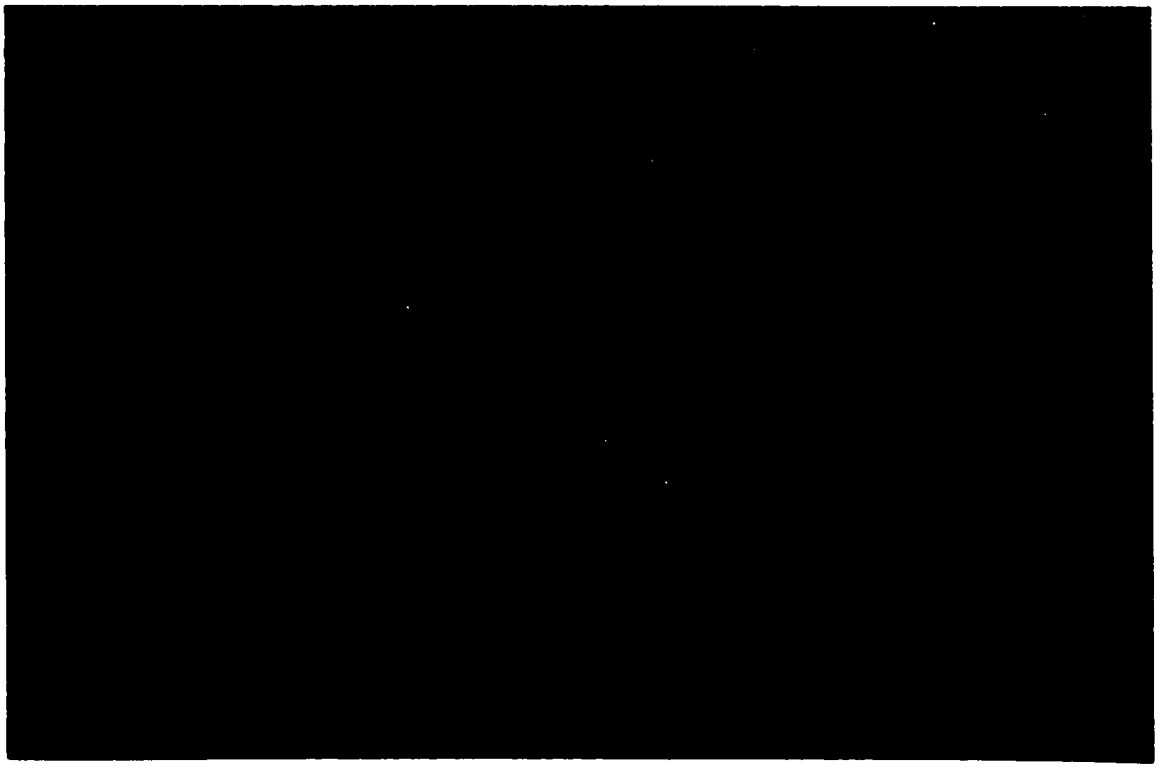

Figure 37. AC1001 anti-blood group A Fv crystal. Rebecca To (NRC) produced AC1001 Fv crystals by seeding smaller crystals into newly mixed protein drops. A small seed crystal can be seen to the left of the main crystal.

The overall structure of the lower (A) and high (B) resolution structures of the AC1001 Fv are very similar, with only small differences. The Fv structure shows a typical antibody fold, consisting primarily of β-strands, arranged in two domains (Figure 38). Both the V_L and the V_H domains are composed of 11 β-strands arranged in anti-parallel sheets. Intramolecular hydrogen bonds are primarily found in the β-sheet structures of individual domains. However, inter-domain hydrogen bonds were observed between the side chains of Gln L38 and Tyr H91 as well as the side chain of Gln H105 and the main chain oxygen atom of Asp L41 (where L and H denote the light or heavy chain). Both of these hydrogen bonds are observed in the lower (A) and higher (B) resolution models of the Fv. Packing of the Fv in the unit cell is in a head-to-tail fashion typical of antibody fragments and allows for a number of intermolecular interactions (Table 19). While some of these interactions are conserved between models (A) and (B), several new intermolecular hydrogen bonds appear in the model determined to higher resolution and can be attributed to ordering of side chains and domain movement due to cryogenic conditions of data collection.

The CDRs of the light and heavy chains (defined in Figure 39) adopt canonical conformations (Chothia & Lesk, 1987; Al-Lazikani *et al.*, 1997), except CDR H3 for which canonical structures are undetermined. Superposition of the two Fv structures over 224 C^{α} atoms yields a RMS difference of 0.42, indicating that the models are very similar.

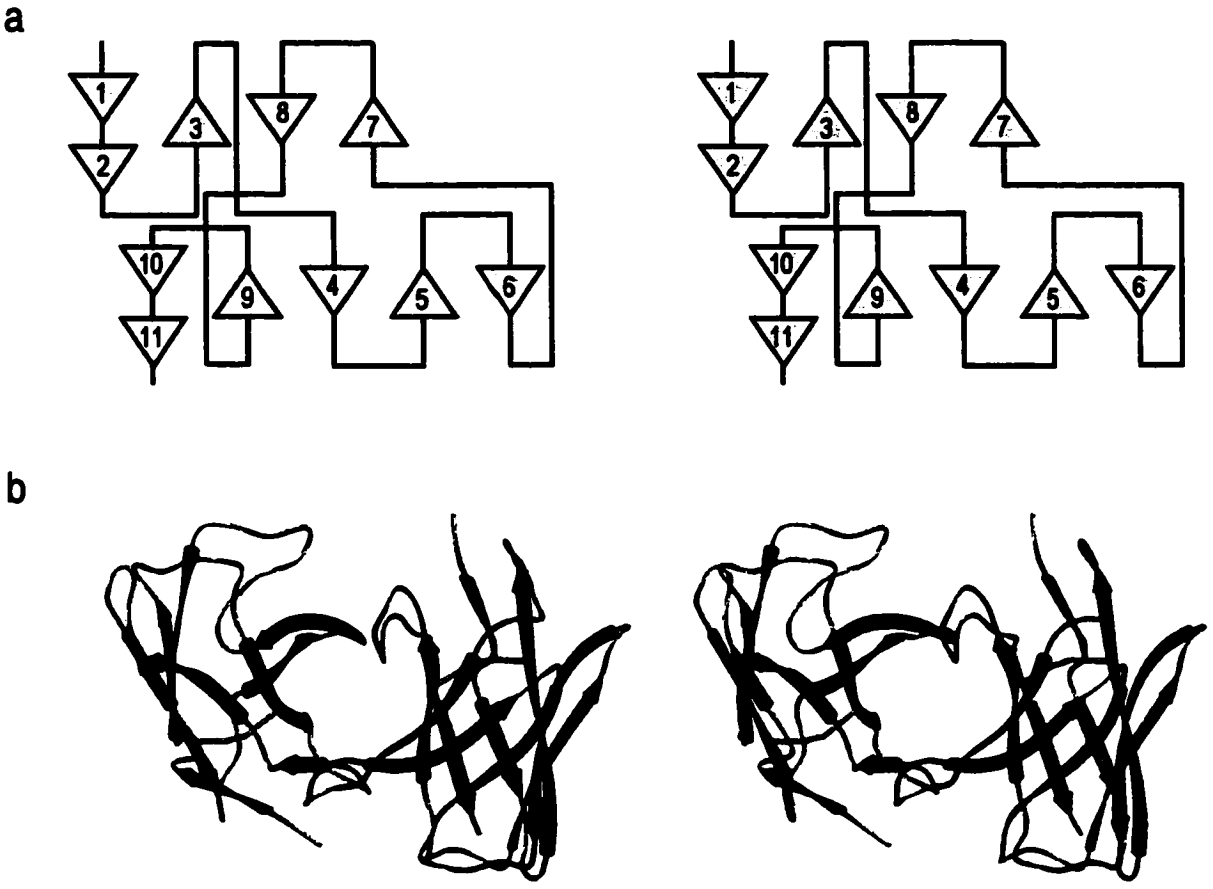

Figure 38. Overall structure of the anti-blood group A Fv. a) Topology diagrams of the VH (blue) and VL (green) domains with β-strands shown as triangles. Both chains show the typical variable domain antibody fold. b) A ribbon diagram of the Fv viewed down its pseudo two-fold axis. The antigen binding site is at the top of the barrel and is formed by CDR L1, L2, L3 and H3. Both Fv structures show the same topology and structure.

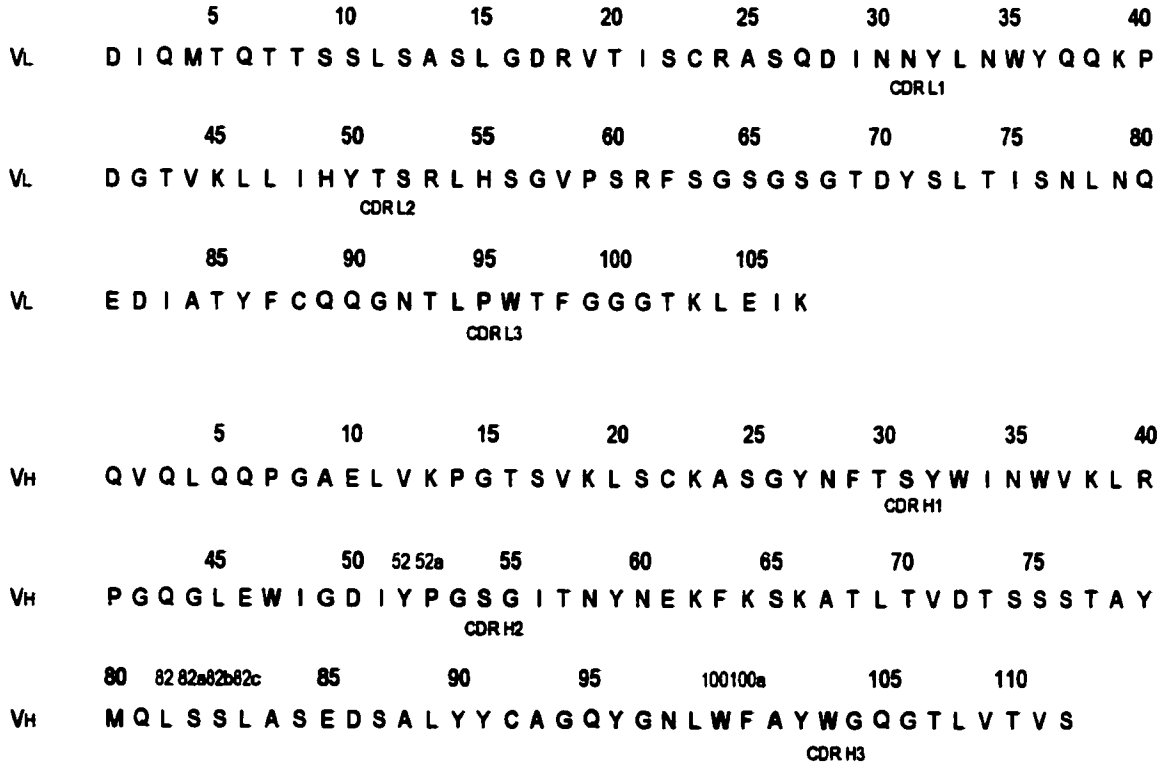

Figure 39. AC1001 anti-blood group A Fv sequence. The sequence of the VL and VH chains of AC1001 anti-blood group A Fv are shown separately. Heavy and light chain CDR are highlighted and labeled. The amino acid sequence is numbered using Kabat numbering as described by Chen *et al.* (1987).

Table 19. Intermolecular interactions in the lower (A) and high (B) resolution models of the anti-blood group A Fv

Anti-blood group A Fv (A)		Distance (Å)	Anti-blood group A Fv (B)		Distance (Å)
			OG1 Thr L5	OG Ser H65	2.87
			N Ser L9	OE2 Glu H61	3.04
			OG Ser L9	OE2 Glu H61	3.15
			NH1 Arg L24	OG Ser H65	2.88
			NH2 Arg L24	O Lys H62	2.72
OH Tyr L32	O Gln H6	2.79	OH Tyr L32	O Gln H6	2.84
NE2 His L49	O Ala H9	3.07	NE2 His L49	O Ala H9	3.16
NH1 Arg L53	OE1 Glu H10	2.63	NE Arg L53	OE1 Glu H10	2.77
N Thr L69	OE2 Glu H85	3.27	N Thr L69	OE2 Glu H85	2.87
OG1 Thr L69	OE1 Giu H85	2.62	OG1 Thr L69	OE1 Glu H85	2.83
			NE2 Gln L80	O Ser H54	2.97
			OG Ser H17	OD2 Asp L17	2.86
			OG Ser H65	OG1 Thr L5	2.87
			NE2 Gln H81	O Gly L16	2.93
			NE2 Gln H81	O Ser L76	2.99
			NE2 Gln H95	O Pro L40	3.11
OD1 Asn H98	O Gly H8	2.44	OD1 Asn H98	O Gly H8	3.05

2. LOWER RESOLUTION STRUCTURE (A)

Of the CDRs, L1, L2, L3, and H3 interact to form a pronounced pocket on the surface of the antibody at the antigen binding site (Figure 40). The pocket is V-shaped: 2.8 Å wide and 11 Å deep at its deepest point, increasing to a width of 4.7 Å at the surface. Given these dimensions, the antigen binding site can easily accommodate the terminal GalNAc of the A trisaccharide antigen.

I- MODELLING OF THE ANTIBODY-ANTIGEN COMPLEX

By careful consideration of van der Waals radii and stereochemical restrictions of the antigen, the A trisaccharide was modelled into the antigen-binding pocket of the AC1001 Fv (A) (Figure 41). A number of potential complementary interactions were observed, such as between the side chain of Asn-L34 and the 3- and 4-hydroxyl groups of the GalNAc residue of the antigen; potential stacking interactions could be modelled between Tyr L36, the Gal residue of the A trisaccharide and Trp H100. The acetamido group of the GalNAc was within hydrogen-bonding distance of the side chain of Gln-L89 in the model. These interactions suggest a reasonable basis for differentiation between the A and B antigens in which the hydroxyl group of the latter would be unable to bridge the distance necessary for hydrogen-bonding these side chains.

II- SITE-DIRECTED MUTANTS

Based on the model of the blood group A-trisaccharide with the AC1001 Fv (Figure 41), amino acid mutations were proposed that would alter polarity, surface complementarity, and side chain aliphatic character to yield antibody fragments with potentially improved affinity for the

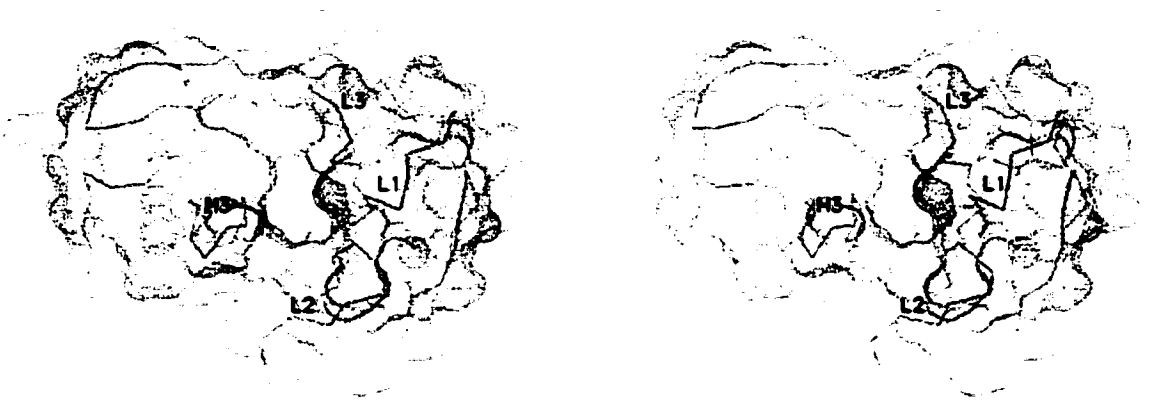

Figure 40. Molecular surface of the anti-blood group A Fv. The antigen-binding site of the anti-blood group A Fv is formed by CDRs L1, L2, L3 and H3, and is manifested by a pronounced pocket at the antibody surface. This pocket is of an appropriate size to accommodate the terminal GalNAc residue of the A trisaccharide antigen.

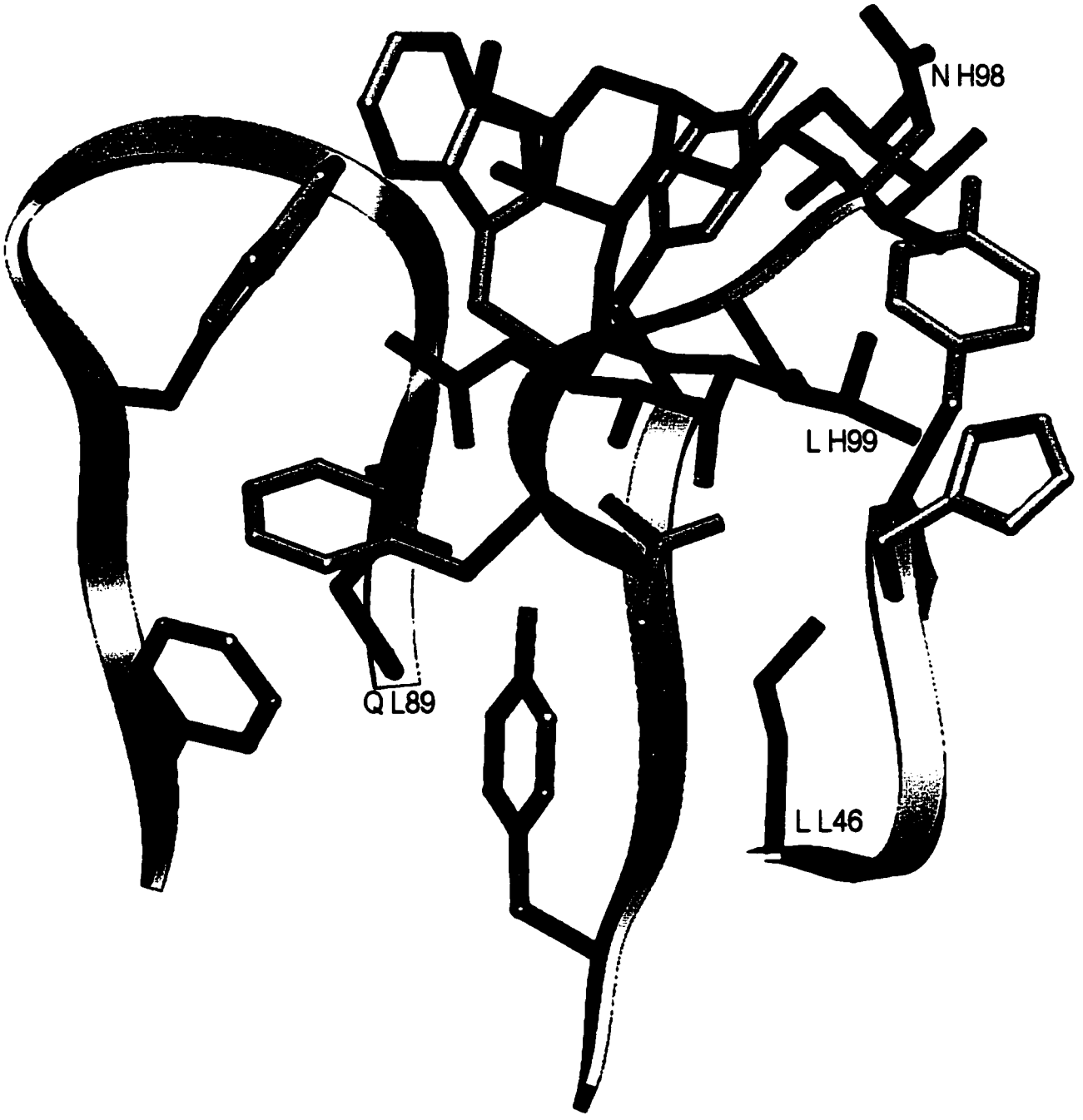

Figure 41. Modelled anti-blood group A Fv-A trisaccharide complex. The A trisaccharide antigen (red) was modeled into the binding pocket of the 2.2 Å structure of the Fv. Residues lining the binding site (white) are predominantly aromatic in nature. Residue Gln 89 (yellow) of the light chain is positioned for specific recognition of the acetamido group of the terminal GalNAc. Other amino acids lining the binding site (blue), including Asn 98 and Leu 99 of the heavy chain as well as Leu 46 of the light chain, were proposed for single point-mutation to improve the binding affinity of the antibody.

trisaccharide epitope. Mutational studies and binding analysis were performed by Ms. Roula Thomas of Dr. Evans' group. The mutants were prepared and submitted to kinetic analysis using a BIACORE™ apparatus. The binding by the scFv to antigen was entirely bivalent at the surface density of the bovine serum albumin (BSA) -trisaccharide employed. Only the higher affinity mutants gave data that allowed for the determination of dissociation constant (K_D) values from rate constants while the K_D values for remaining mutants were derived from Scatchard analyses of equilibrium binding (Table 20). Most mutants, including L(99H)V, were tested for binding to BSA-B-trisaccharide, though no affinity was observed.

Table 20. Binding affinities of wild-type and anti-blood group A Fv mutants

ScFv mutant	scFv[a]		Pentasaccharide[a]	Trisaccharide[a]
	Kinetic analysis	Equilibrium analysis		
	K_D (10^{-6} M)	K_D (10^{-6} M)	K_D (10^{-5} M)	K_D (10^{-5} M)
WT	-	26	6.9	20
L46(L)N[b]	-	16	-	-
L46(L)M	-	18	-	-
N98(H)Q	-	22	-	-
L99(H)I	1.9	1.5	1.9	11
L99(H)V	0.8	0.86	1.1	6.5
L46(L)N/L99(H)I	2.0	2.8	-	-
L46(L)M/L99(H)I	2.1	1.4	-	-
N102(H)Q/L99(H)I	1.2	1.2	-	-
L46(L)M/N98(H)Q/L99(H)I	1.8	1.4	-	-
L46(L)N/N98(H)Q/L99(H)I	1.3	1.5	-	-

Immobiliizations on the sensor chips were done using amine coupling.
[a]Molecule (analyte) injected over the sensor chip.
[b]The one-letter code for amino acids is used with the L and H in brackets indicating light and heavy chain, respectively.

The monovalent scFv-trisaccharide interaction is near the lower limit of the affinity range of BIACORE measurements using the antibody fragment as the analyte or injected molecule. The

K_D for the bivalent binding of wild-type scFv dimer to trisaccharide antigen was determined to be 26 μM as compared with a value of 290 μM for the binding of Fv monomer to the same antigen (Patenaude *et al.*, 1998).

For mutants with measurable dissociation rates, binding was homogeneous and all bivalent. The increase in binding affinity for the N(98H)Q, L(46L)M, and L(46L)N mutants is small (1.2-, 1.4-, and 1.6-fold, respectively, Table 20), showing only marginal improvements in affinity. However, the effects of mutations at residue Leu H99 were quite dramatic. The L(99H)I and L(99H)V mutations increased the K_D for bivalent binding by 15- and 31-fold, respectively, relative to the wild-type. Although Leu-H99 lines the binding site, its side chain would probably not contact the trisaccharide antigen, and a conservative mutation at this position would not be expected to greatly alter interactions with the antigen. A closer look at the affinities shows that the Leu-Ile-Val progression at position H99 is coupled with a decrease in flexibility of the side chain, indicating a potential entropic component in binding. Typically, higher entropy in the binding site residues would be expected to be unfavourable, as a thermodynamic penalty would be incurred if antigen binding required residue immobilization. The reduced side chain flexibility of the L(99H)I and L(99H)V mutants thus provides a potential explanation for the increase in binding. Significantly, these scFv mutants did not display detectable binding to the B antigen, demonstrating that the increase in the affinity of the mutant for the blood group A antigen is accompanied by retention of the fine specificity of the native antibody, consistent with the model that suggests that those residues do not contact the acetamido/hydroxyl groups.

An attempt was made to produce mutants with even higher affinity by combining the single point mutations into double and triple point mutations. The L(99H)I mutation in combination with mutations at positions L46 and H102 resulted in marginal improvements in some instances. Surprisingly, the L(46L)N/L(99H)I, L(46L)M/L(99H)I, and L(46L)M/N(98H)Q/L(99H)I mutants did

not show significant increases in binding affinity over the L(103H)I single mutant, indicating that their respective contributions to binding were not cooperative. In contrast, mutants N(98H)Q/L(99H)I and L(46L)N/N(98H)Q/L(99H)I both display a slight increase in binding affinity over the L(99H)I mutant.

The affinities of wild-type and mutant scFvs for trisaccharide and pentasaccharide antigen were determined by binding free monovalent antigen to immobilized scFv (Table 20). This method gave a K_D of 200 μM for the wild-type scFv trisaccharide interaction, which is in good agreement with the previously reported value of 290 μM for the binding of Fv to immobilized BSA-A-trisaccharide (Patenaude *et al.*, 1998). The affinity of wild-type scFv for the pentasaccharide antigen was 3-fold higher. Relative to the wild type, the L(99H)I mutation improved the affinity for trisaccharide and pentasaccharide by 2- and 3-fold, respectively. As with scFv dimer binding to immobilized BSA-A-trisaccharide, the L(99H)V mutation led to further 2-fold improvements. A comparison of the K_D values for bivalent *versus* monovalent scFv binding to BSA-A-trisaccharide indicated that the increase in affinity due to valency was 75-fold for the L(99H)I and L(99H)V mutants.

The site-directed mutagenesis results from AC1001 Fv are in marked contrast to those obtained previously with another anti-carbohydrate antibody (Brummell *et al.*, 1993). The anti-*Salmonella* carbohydrate antibody, Se155–4, was extensively mutated in its CDR H3 by replacing each of the four native residues with all 19 other amino acids, though none of the mutants showed any improvement in oligosaccharide binding (Brummell *et al.*, 1993). Because the native Se155–4 antibody is an immunoglobulin G (IgG) and had a higher initial binding affinity, it is reasonable to propose that the difference in behaviour arises from the AC1001 antibody being an IgM of lower affinity, which has undergone little if any somatic mutation. The process of mutating a limited number of amino acids in order to explore questions of affinity and specificity mimics *in vitro* the

process of affinity maturation that antibodies can undergo when challenged with certain classes of antigen. Reducing the conformational flexibility of the antigen binding site by the somatic mutation of a limited number of residues has been shown to be a key factor in affinity maturation (Manivel *et al.*, 2000) and antibody affinity (Sharp, 1998). This concept extends beyond the paratope as the free and peptide-bound structures of antibody 48G7 and its germline predecessor (Wedemayer *et al.*, 1997) showed that none of the nine mutated amino acid residues in the mature antibody directly contact the antigen but still cooperate in reducing entropic effects.

3. HIGHER RESOLUTION STRUCTURE (B)

While the electron density was good for the lower resolution model (A), it is unambiguous for most of this model to 1.1 Å (Figure 42). Electron density for individual atoms is clearly emerging and positioning of protein atoms is exceptional. In addition, several side chains with alternate conformations have been identified in the high resolution model (Table 21, with residues numbered as outlined in Figure 39).

Table 21. Side chains showing alternate conformations.

V_L residues		V_H residues	
Gln	3	Leu	11
Thr	5	Ser	17
Thr	8	Lys	23
Ser	9	Asn	35
Leu	11	Val	37
Arg	18	Asp	50
Ser	65	Asn	58
Ser	72	Ser	65
Leu	73	Lys	64
Ile	75	Ser	82a
Lys	103	Ser	82b
Lys	107	Ser	84
		Tyr	102

The overall structure and topology of this model is the same as that for the lower resolution structure (A). The RMS difference calculated for the overlap of 224 C^{α} from the two models is 0.42. Though this value may seem elevated at first glance, it can be easily explained by a number of contributing factors. First, two regions of poor electron density from the first model (A) have been placed in good electron density using higher resolution data; these two regions (Lys 39-Gly 42 of the light chain, Arg 40-Gly 44 of the heavy chain) are largely different between the two

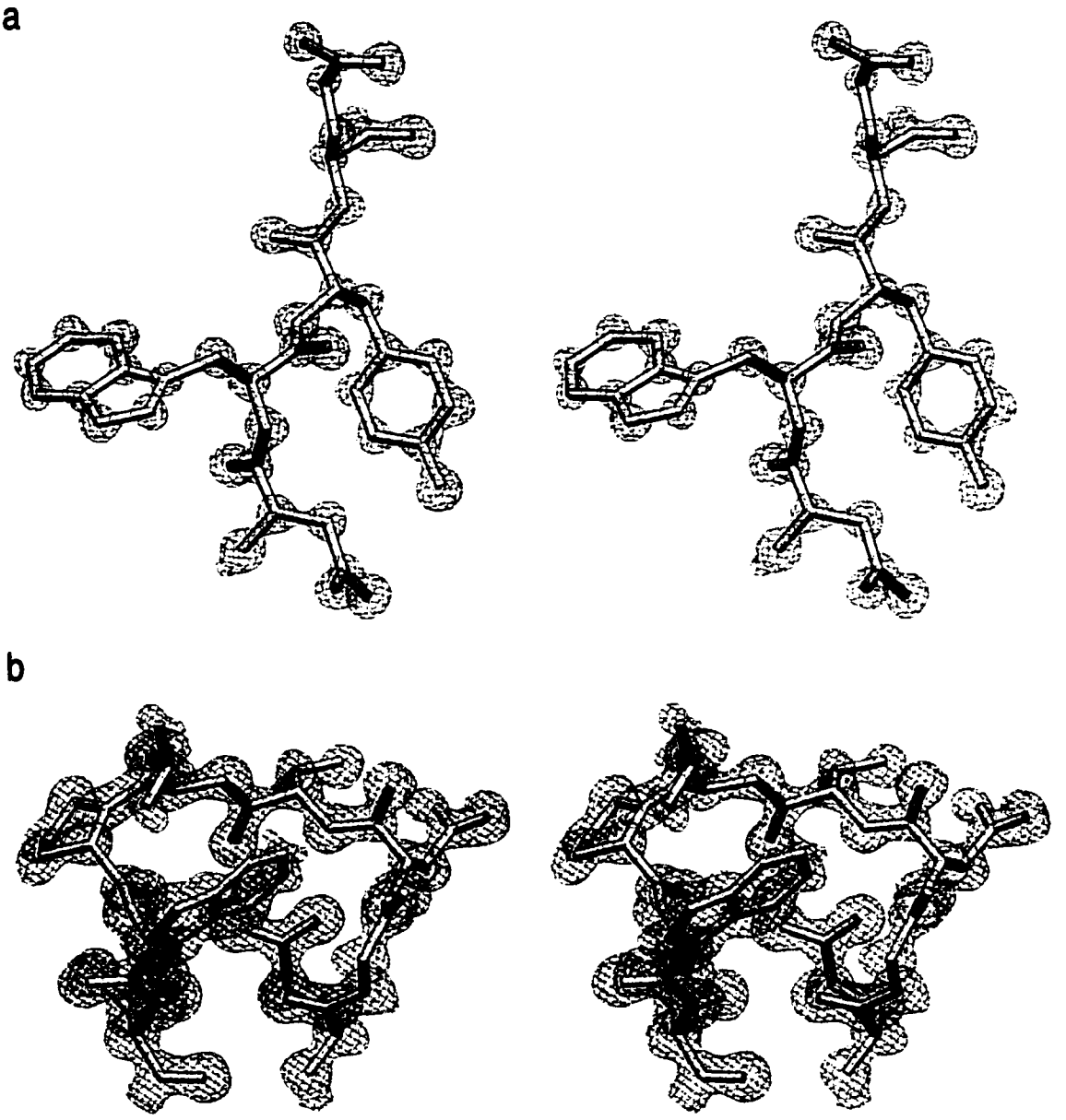

Figure 42. Electron density for the high resolution model (B) of the AC1001 Fv. a) Stereoview of the electron density contoured to 3.5 σ for residues Asn L34 Gln L37. b) Electron density contoured to 2.0 σ for residues found in CDR L2.

models and if they are excluded from calculations, the RMS difference over 215 C^{α} becomes 0.35. In addition, overlap of the individual domains is good, with RMS differences of 0.29 and 0.24 (excluding Lys L39-Gly L42) for the V_L and 0.38 and 0.28 (excluding Arg H40-Gly H44) for the V_H. When the V_L domains of both models are overlapped (Figure 43a), the V_H domain of the higher resolution model is shifted, forming a more compact Fv structure. Cryogenic temperatures may bring about this more compact tertiary structure, which in turn reduces the size of the unit cell *versus* the room temperature unit cell (A) (Table 18).

Usually, the mean square displacement of an atom (u^2), related to the temperature (B) factor by the equation $B = 8\pi^2 u^2$, is approximated to be spherical and is therefore "isotropic". Due to the number of reflections available in the high resolution dataset, individual anisotropic temperature factors can be refined. This extends the description of the temperature factor from one parameter (isotropic) to six (anisotropic), which allows for a more detailed description of the ellipsoid defining the probability distribution of each atom. Refinement of the anisotropic B-factors is possible only at very high resolution, generally beyond 1.3 Å provided that the number of reflections in the dataset is greater than the number of parameters to be solved. When restrained anisotropic temperature factor refinement was introduced to the anti-blood group A Fv structure at 1.2 Å resolution, the working and free R-factors dropped by 3.76% and 2.07%, respectively. While working and free R-factors benefit from the introduction of anisotropy, computation time is significantly increased.

Another factor to consider in structure determination to high resolution is the number of water molecules to be added. Data collection at low temperature (100K) leads to a larger ordered hydration sphere around the protein and thus to an increased number of water molecules to be placed. This can be increased further when data extend to very high resolution. In this case, 436 water molecules were included in the model, most of which show very good electron density.

Figure 43. Overlap of the high (B) and lower (A) resolution models of the AC1001 Fv. The high resolution model (B) is shown in red while the lower resolution model (A) is shown in white. a) The V_L domains of both models (A, B) were overlapped using SETOR. Good correlation could be obtained for the V_L domains, while significant differences could be observed for the V_H domains. b) Overlap of the residues found in the binding site reveals very little difference in side chain positions for the V_L domains (left) while the V_H domains (right) exhibit significant changes in residues Trp H33 and Trp H100 as well as alternate conformations for residues Asn H35, Val H37 and Asp H50.

The residues lining the binding site in this model are the same as those of the previous structure. The side chain atom positions are very similar in both cases, as shown in Figure 43b. The most significant differences in the conformation of side chains occur in the V_H domain, particularly at residues Trp H33, Trp H100. Despite these changes, the size and shape of the antigen binding site is unlikely to be significantly affected. Trp H33 is adjacent to CDR H1 and has limited influence in the binding site. Trp H100, however, is found in CDR H3 and located in the centre of the binding site. Further, the observed alternate conformations for the side chains of residues Asn H35, Val H37 and Asp H50 (which do not contact the antigen) support the notion that binding of antigen in this system involves antibody residues that do not display high motion, and is therefore "lock and key".

CONCLUSION

The biosynthesis and immune recognition of the human ABO(H) blood group antigens represents an advantageous system for the analysis of structure-function relationships. The carbohydrate antigens A and B differ only in an acetamido *versus* a hydroxyl group, respectively, on the terminal sugar yet these very similar antigens are synthesized by glycosyltransferases (GTA and GTB) that differ by only four amino acids, and are differentiated by highly specific non cross-reactive antibodies (anti-A and anti-B).

In order to explore the basis of substrate specificity, the structures of the human ABO(H) blood group glycosyltransferases have been determined to high resolution, both unliganded and in the presence of substrates. GTA and GTB share the same structure and show homology to a number of previously determined structures and can be classified in the same GT superfamily as SpsA, GnTI and LgtC. The large, open, solvent-exposed active site allows just a few enzyme residues to contact and recognize the donor substrates.

Of the four known "critical" amino acid residues in GTA and GTB, only two (Leu/Met 266 and Gly/Ala 268) were found to occupy positions in the active sites that can contact the donor sugar and one (Arg/Gly 176) was so distant from the active site that its participation in specific recognition of UDP-Gal or UDP-GalNAc is unlikely. Interestingly, the second critical amino acid residue (Gly/Ser 235) is also not positioned to directly contact the donor, although this residue does force the aliphatic tail on the acceptor to assume different conformations in GTA and GTB and may select between different glycoconjugate structures displaying the H antigen acceptor. The relatively low number of protein-carbohydrate interactions formed in these complexes confers specificity to the blood group glycosyltransferases in different ways. Firstly, several residues participate in specific hydrogen bonds that recognize and bind the H antigen acceptor disaccharide

as well as the UDP moiety of the donor. Secondly, critical residues Leu/Met 266 and Gly/Ala 268 are involved in selecting between very similar donor-sugar substrates, where the GTB active site excludes the A donor based on size while the larger GTA active site does not permit specific interactions with UDP-Gal.

These structures also provide novel insight into the Mn^{2+} coordination. In the GTA and GTB complexes with UDP and acceptor substrates, both phosphate groups of the donor and both aspartate residues of the DXD motif, including a bidentate coordinate bond, are liganded to the manganese ion. The interaction of the Mn^{2+} with the DXD motif in this manner has also been observed in the structures of the two other retaining glycosyltransferases (Gastinel *et al.*, 2001; Persson *et al.*, 2001) while the inverting glycosyltransferase structures (Charnock & Davies, 1999; Gastinel *et al.*, 1999; Unligil *et al.*, 2000; Pederson *et al.*, 2000) shows that only the second aspartate residue of the DXD motif interacts with the manganese ion while the first aspartate (not highly conserved in inverting GTs) contacts either the ribose or sugar moieties of the donor. These different coordination patterns may be characteristic of inverting *versus* retaining glycosyltransferases and probably influence the catalytic action of the enzyme.

The structures of GTA and GTB in complex with the H antigen substrate alone were also determined. Interestingly, these structures are the first of GTs in complex with acceptor substrate alone and show that, contrary to expectations, the H antigen is capable of binding the enzymes even in the absence of UDP-sugar donor.

In contrast to the open GTA and GTB active sites, the high resolution structure of the AC1001 Fv shows that the CDRs define a deep pocket that can accommodate the terminal GalNAc of the A trisaccharide antigen. Several aromatic residues line the binding site and provide a largely solvent-excluded environment for the antigen. The Fv-trisaccharide model shows that the terminal GalNAc residue of the trisaccharide antigen is probably engulfed by the pocket on the antibody,

allowing interaction with all parts of the sugar residue. While the affinity of these interactions is relatively weak compared to antibody recognition of proteins, the antibody shows no measurable affinity for the B antigen.

The determination of the structures of both the enzymes responsible for the human blood group antigen biosynthesis and of an antibody specific for the human blood group A antigen allows a comparison of the two different classes of molecules with functions centred about the same substrate. The anti-blood group A antibody and the A and B glycosyltransferases have fundamentally different roles related to the ABO(H) histo-blood groups: the GTs synthesize and release the antigens while the antibodies recognize and bind them. The biological necessity of these proteins to faithfully perform the synthesis and recognition of these antigens begs the question: do the antibodies and GTs show any common mechanisms for the recognition of their respective substrates?

The anti-blood group A antibody and the A and B glycosyltransferases clearly exhibit different types of binding sites, which contribute to their different philosophies in substrate recognition. While the glycosyltransferases show relatively few interactions between the carbohydrate substrates and the enzymes, the Fv-A trisaccharide model suggests that several interactions form between the antibody and its antigen. Both classes of molecules rely on specific recognition of the acetamido group of the GalNAc residue in large part by a single amino acid residue. However, where the anti-A Fv shows no cross-reactivity with the B antigen, a certain degree of cross-reactivity of the donor sugars has been observed in GTA and GTB, as well as detectable activity with corresponding glucose-based donors.

This work represents a first step in the establishment of the structure-function relationship of various components of the ABO(H) histoblood groups. Experiments underway will provide a more complete picture of the molecular interactions and their role in binding specificity. These

include kinetic characterization and crystallization of Glu 303 mutants and of a naturally occurring mutant (P234S) for GTB, which alters its specificity to that of the A enzyme. Amino and deoxy analogs of the H-antigen acceptor, as well as a UDP-2F-Gal analog of the UDP-Gal donor have been synthesized by our collaborators and soaking experiments with GTA-10, GTB-10 and the mutants described above are planned. By preventing the constitutive hydrolysis of the donor by the GTA and GTB enzymes it may be possible to obtain structure in which the donor sugar is present and ordered.

REFERENCES

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

Amado, M., Bennett, E. P., Carneiro, F. & Clausen, H. (2000). Characterization of the histo-blood group O(2) gene and its protein product. *Vox Sang* **79**, 219-26.

Aminoff, D. & Morgan, W. T. J. (1948). Hexosamine components of the human blood group substances. *Nature* **162**, 579-80.

Aminoff, D., Morgan, W. T. J. & Watkins, W. M. (1950). Studies in Immunochemistry VIII. The isolation and properties of the human blood group A substance. *Biochemical Journal* **46**, 426-39.

Annison, E. F. & Morgan, W. T. J. (1952). Studies in Immunochemistry XI. The isolation and properties of human blood group H substance. *Biochemical Journal* **52**, 247-60.

Barondes, S. H., Cooper, D. N., Gitt, M. A. & Leffler, H. (1994). Galectins. Structure and function of a large family of animal lectins. *J Biol Chem* **269**, 20807-10.

Barrett, C. S. (1952). *Structure of metals, crystallographic methods, principles, and data*. 2nd edit, McGraw-Hill, New York, 661 p.

Bernstein, F. (1924). Ergebnisse einer biostatischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen. *Klin Wschr* **3**, 1495-7.

Blundell, T. L. & Johnson, L. N. (1976). *Protein crystallography*, Academic Press, New York, 565 p.

Boix, E., Swaminathan, G. J., Zhang, Y., Natesh, R., Brew, K. & Acharya, K. R. (2001). Structure of UDP Complex of UDP-galactose:β-Galactoside-α-1,3-galactosyltransferase at 1.53-Å Resolution Reveals a Conformational Change in the Catalytically Important C Terminus. *J Biol Chem* **276**, 48608-14.

Boyd, W. C. & Reguera, R. M. (1949). Haemagglutinating substances for human cells in various plants. *J. Immunol.* **62**, 333-9.

Bragg, W. L. (1913). The diffraction of short electromagnetic waves by a crystal. *Proc. Camb. Phil. Soc.* **17**, 43-57.

Bray, H. G. & Stacey, M. (1946). *Biochem. J.* **40**, 124.

Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A. & Imberty, A. (1998a). Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *J Biochem (Tokyo)* **123**, 1000-9.

Breton, C. & Imberty, A. (1999). Structure/function studies of glycosyltransferases. *Curr Opin Struct Biol* **9**, 563-71.

Breton, C., Oriol, R. & Imberty, A. (1998b). Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology* **8**, 87-94.

Brummell, D. A., Sharma, V. P., Anand, N. N., Bilous, D., Dubuc, G., Michniewicz, J., MacKenzie, C. R., Sadowska, J., Sigurskjold, B. W., Sinnott, B., Young, N. M., Bundle, D.R. & Narang, S.A. (1993). Probing the combining site of an anti-carbohydrate antibody by saturation-mutagenesis: role of the heavy-chain CDR3 residues. *Biochemistry* **32**, 1180-7.

Brünger, A. T. (1992). *X-PLOR, Version 3.1: a system for X-ray crystallography and NMR*, Yale University Press, New Haven, 382 p.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr* **D54**, 905-21.

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

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D. & Aktories, K. (1998). A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. *J Biol Chem* **273**, 19566-72.

Campbell, J. A., Davies, G. J., Bulone, V. & Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J* **326**, 929-39.

Campbell, J. A., Davies, G. J., Bulone, V. V. & Henrissat, B. (1998). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **329**, 719.

Capon, S. M. & Goldfinger, D. (1995). Acute hemolytic transfusion reaction, a paradigm of the systemic inflammatory response: new insights into pathophysiology and treatment. *Transfusion* **35**, 513-20.

Caputto, R., Leloir, L. F., Cardini, I. & Paladini, A. C. (1950). Isolation of the coenzyme of the galactose phosphate-glucose phosphate transformation. *J Biol Chem* **184**, 333-50.

Casset, F., Peters, T., Etzler, M., Korchagina, E., Nifant'ev, N., Perez, S. & Imberty, A. (1996). Conformational analysis of blood group A trisaccharide in solution and in the binding site of *Dolichos biflorus* lectin using transient and transferred nuclear Overhauser enhancement (NOE) and rotating-frame NOE experiments. *Eur J Biochem* **239**, 710-9.

Charnock, S. J. & Davies, G. J. (1999). Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380-5.

Chen, H. T., Kabat, E. A., Lundblad, A. & Ratcliffe, R. M. (1987). Nucleotide and translated amino acid sequences of cDNA coding for the variable regions of the light and heavy chains of mouse hybridoma antibodies to blood group A and B substances. *J Biol Chem* **262**, 13579-83.

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

Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L. & Hakomori, S. (1990). Isolation to homogeneity and partial characterization of a histo-blood group A defined Fucα1→2Galα1→3-N-acetylgalactosaminyltransferase from human lung tissue. *J Biol Chem* **265**, 1139-45.

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr* **D50**, 760-63.

Connolly, M. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-13.

Connolly, M. L. (1983). Analytical molecular surface calculation. *J Appl Crystallogr* **16**, 548-58.

Côté, R. H. & Morgan, W. T. J. (1956). Some nitrogen containing disaccharides isolated from human blood group A substance. *Nature* **191**, 149-50.

Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-Active Enzymes server, http://afmb.cnrs-mrs.fr/~cazy/CAZY/.

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

Datta, A. K. & Paulson, J. C. (1997). Sialylmotifs of sialyltransferases. *Indian J Biochem Biophys* **34**, 157-65.

David, L., Leitao, D., Sobrinho-Simoes, M., Bennett, E. P., White, T., Mandel, U., Dabelsteen, E. & Clausen, H. (1993). Biosynthetic basis of incompatible histo-blood group A antigen expression: anti-A transferase antibodies reactive with gastric cancer tissue of type O individuals. *Cancer Res* **53**, 5494-500.

Davidsohn, I. & Stejskal, R. (1972). Tissue antigens A, B and H in health and disease. *Haematologia* **6**, 177-84.

Davies, G. & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853-9.

Davies, G., Sinnott, M. L. & Withers, S. G. (1998). Glycosyl Transfer. In *Comprehensive Biological Catalysis: A Mechanistic Reference* (Sinnott, M. L., ed.), Vol. 1, pp. 119-209. Academic Press, San Diego.

Davies, G. J. (2001). Sweet secrets of synthesis. *Nat Struct Biol* **8**, 98-100.

Drickamer, K. (1995). Multiplicity of lectin-carbohydrate interactions. *Nat Struct Biol* **2**, 437-9.

Drickamer, K. (1997). Making a fitting choice: common aspects of sugar-binding sites in plant and animal lectins. *Structure* **5**, 465-8.

Eisler, M. (1930). Über ein gemeinsames, Antigen in den Zellen des Menschen und in Shigabazillen. *Zeitscrift für Immunforsch* **67**, 38.

England, P., Nageotte, R., Renard, M., Page, A. L. & Bedouelle, H. (1999). Functional characterization of the somatic hypermutation process leading to antibody D1.3, a high affinity antibody directed against lysozyme. *J Immunol* **162**, 2129-36.

Epstein, A. & Ottenberg, R. (1908). Simple method of performing serum reactions. *Proceedings of the NY Pathology Society* **8**, 117-23.

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

Evans, S. V., Rose, D. R., To, R., Young, N. M. & Bundle, D. R. (1994). Exploring the mimicry of polysaccharide antigens by anti-idiotypic antibodies. The crystallization, molecular replacement, and refinement to 2.8 Å resolution of an idiotope-anti-idiotope Fab complex and of the unliganded anti-idiotope Fab. *J Mol Biol* **241**, 691-705.

Feizi, T. (1991). Carbohydrate differentiation antigens: probable ligands for cell adhesion molecules. *Trends Biochem Sci* **16**, 84-6.

Finne, J., Krusius, T., Rauvala, H. & Jarnefelt, J. (1980). Molecular nature of the blood-group ABH antigens of the human erythrocyte membrane. *Rev Fr Transfus Immunohematol* **23**, 545-52.

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

Fukumori, Y., Ohnoki, S., Yoshimura, K., Nagao, N., Shibata, H., Tomita, T., Okubo, Y. & Yamaguchi, H. (1996). Rapid detection of the *cis*AB allele consisting of a chimera of normal A and B alleles by PCR-RFLPs. *Transfus Med* **6**, 337-44.

Garratty, G. (1995). Blood group antigens as tumour markers, parasitic/bacterial/viral receptors, and their association with immunologically important proteins. *Immunological Investigations* **24**, 213-232.

Gastinel, L. N., Bignon, C., Misra, A. K., Hindsgaul, O., Shaper, J. H. & Joziasse, D. H. (2001). Bovine α1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *Embo J* **20**, 638-49.

Gastinel, L. N., Cambillau, C. & Bourne, Y. (1999). Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *Embo J* **18**, 3546-57.

Geremia, R. A., Harduin-Lepers, A. & Delannoy, P. (1997). Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action. *Glycobiology* **7**, v-vii.

Gibbons, R. A. & Morgan, W. T. J. (1954). Studies in Immunochemistry 14. The isolation and properties of substances of human origin possessing blood group B activity. *Biochemistry Journal* **57**, 283-95.

Glusker, J. P. & Trueblood, K. N. (1972). *Crystal structure analysis: a primer*, Oxford University Press, Oxford, 192 p.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). The structure of haemoglobin. IV. Sign determination by the isomorhpous replacement method. *Proc. Roy. Soc.* **A225**, 287-307.

Greenwell, P., Yates, A. D. & Watkins, W. M. (1986). UDP-N-acetyl-D-galactosamine as a donor substrate for the glycosyltransferase encoded by the B gene at the human blood group ABO locus. *Carbohydr Res* **149**, 149-70.

Ha, S., Walker, D., Shi, Y. & Walker, S. (2000). The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci* **9**, 1045-52.

Hagen, F. K., Hazes, B., Raffo, R., deSa, D. & Tabak, L. A. (1999). Structure-function analysis of the UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. *J Biol Chem* **274**, 6797-803.

Hakomori, S. (1999). Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer. *Biochim Biophys Acta* **1473**, 247-66.

Hakomori, S. & Strycharz, G. D. (1968). Investigations on cellular blood-group substances. I. Isolation and chemical composition of blood-group ABH and Le^b isoantigens of sphingoglycolipid nature. *Biochemistry* **7**, 1279-86.

Haldane, J. B. S. (1937). The biochemistry of the individual. In *Perspectives in Biochemistry* (Needham, J. & Green, D. E., eds.), pp. 1-10. Cambridge University Press, Cambridge.

Harrap, G. J. & Watkins, W. M. (1964). Characterisation of the enzyme form *Trichomonas* foetus that destroys the serological specificity of blood group A substances. *Biochemical Journal* **93**, 9P.

Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J Mol Biol* **234**, 958-64.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-8.

Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). Direct phase determination based on anomalous scattering. *Methods Enzymol* **115**, 41-55.

Henrissat, B. & Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**, 637-44.

Ignatovich, O., Tomlinson, I. M., Jones, P. T. & Winter, G. (1997). The creation of diversity in the human immunoglobulin V_λ repertoire. *J Mol Biol* **268**, 69-77.

International Tables for X-ray Crystallography. (1974). (Ibers, J. A. & Hamilton, W. C., Eds.), IV, The Kynoch Press, Birmingham, England.

International Union of Biochemistry and Molecular Biology (1992). *Enzyme nomenclature: recommendations of the nomenclature committee of the International Union of Biochemistry and Molecular Biology*, Academic Press, San Diego.

Iseki, S., Furukawa, K. & Yamamoto, S. (1959). B substance decomposing enzyme produced by anarobic bacterium. II. Chemical action of B decomposing enzyme. *Proceedings of the Imperial Academy of Japan* **35**, 460-5.

Janeway, C. A., Jr. & Travers, P. (1997). *Immunobiology: the immune system in health and disease.* 3rd edit., Current Biology Limited, London.

Janin, J. & Chothia, C. (1990). The structure of protein-protein recognition sites. *J Biol Chem* **265**, 16027-30.

Jeffrey, P. D., Bajorath, J., Chang, C. Y., Yelton, D., Hellstrom, I., Hellstrom, K. E. & Sheriff, S. (1995). The x-ray structure of an anti-tumour antibody in complex with antigen. *Nat Struct Biol* **2**, 466-71.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.

Kabat, E. A. & Leskowitz. (1955). Immunochemical studies on blood groups XVII. Structural units involved in blood group A specificity. *J Am Chem Soc* **77**, 5154-69.

Kamath, V. P., Seto, N. O., Compston, C. A., Hindsgaul, O. & Palcic, M. M. (1999). Synthesis of the acceptor analog αFuc(1→;2)αGal-O(CH2)7 CH3: a probe for the kinetic mechanism of recombinant human blood group B glycosyltransferase. *Glycoconj J* **16**, 599-606.

Karle, J. (1980). Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *Int J Quant Chem: Quantum Biology Symposium* **7**, 357-67.

Kim, S. C., Singh, A. N. & Raushel, F. M. (1988). Analysis of the galactosyltransferase reaction by positional isotope exchange and secondary deuterium isotope effects. *Arch Biochem Biophys* **267**, 54-9.

Koscielak, J. (1963). Blood group specific glycolipids from human erythrocytes. *Biochim Biophys Acta* **78**, 313-28.

Landsteiner, K. (1900). Zur Kenntniss der antifermentativen, lytischen und agglutinierenden Wikungen des Blutserums und der Lymphe. *Zentbl. Bakt. Parasitkde* **27**, 357-366.

Landsteiner, K. (1901). Uber Agglunationserscheinungen normalen menschlichen Blutes. *Wein Klin Wschr* **14**, 1132-4.

Landsteiner, K. & Harte, R. A. (1940). *Journal of Experimental Medicine* **71**, 551.

Leloir, L. F. (1974). The Fifth Feodor Lynen lecture 'I hate to bore people with my recollections'. In *Biology and Chemistry of Eucaryotic Cell Surfaces. Miami Winter Symposium* (Lee, E. Y. C. & Smith, E. E., eds.), Vol. 7, pp. 1-20. Academic Press, New York.

Lloyd, K. O., Kabat, E. A., Layug, E. J. & Gruezo, F. (1966). Immunochemical studies on blood groups. XXXIV. Structures of some oligosaccharides produced by alkaline degradation of blood group A, B, and H substances. *Biochemistry* **5**, 1489-501.

Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-98.

Lowary, T. L. & Hindsgaul, O. (1993). Recognition of synthetic deoxy and deoxyfluoro analogs of the acceptor α-L-Fuc p-(1→2)-β-D-Gal p-OR by the blood-group A and B gene-specified glycosyltransferases. *Carbohydr Res* **249**, 163-95.

Lowary, T. L. & Hindsgaul, O. (1994). Recognition of synthetic O-methyl, epimeric, and amino analogues of the acceptor α-L-Fuc p-(1→2)-β-D-Gal p-OR by the blood-group A and B gene-specified glycosyltransferases. *Carbohydr Res* **251**, 33-67.

Ly, H. D. & Withers, S. G. (1999). Mutagenesis of glycosidases. *Annu Rev Biochem.* **68**, 487-522.

MacKenzie, C. R., Sharma, V., Brummell, D., Bilous, D., Dubuc, G., Sadowska, J., Young, N. M., Bundle, D. R. & Narang, S. A. (1994). Effect of C_λ-C_κ domain switching on Fab activity and yield in *Escherichia coli*: synthesis and expression of genes encoding two anti-carbohydrate Fabs. *Biotechnology (N Y)* **12**, 390-5.

Maenaka, K., Matsushima, M., Song, H., Sunada, F., Watanabe, K. & Kumagai, I. (1995). Dissection of protein-carbohydrate interactions in mutant hen egg-white lysozyme complexes and their hydrolytic activity. *J Mol Biol* **247**, 281-93.

Manivel, V., Sahoo, N. C., Salunke, D. M. & Rao, K. V. (2000). Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* **13**, 611-20.

McPherson, A. (1999). *Crystallization of biological macromolecules*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 586 p.

Meldgaard, P., Johnson, P. H., Langkilde, N. C., Wolf, H. & Orntoft, T. F. (1995). Loss of ABH antigen expression in bladder cancer is not caused by loss of heterozygosity of the ABO locus. *Int J Cancer* **63**, 341-4.

Meyer, J. E. & Schulz, G. E. (1997). Energy profile of maltooligosaccharide permeation through maltoporin as derived from the structure and from a statistical analysis of saccharide-protein interactions. *Protein Sci* **6**, 1084-91.

Mifsud, N. A., Watt, J. M., Condon, J. A., Haddad, A. P. & Sparrow, R. L. (2000). A novel *cis*-AB variant allele arising from a nucleotide substitution A796C in the B transferase gene. *Transfusion* **40**, 1276-7.

Miyake, M., Taki, T., Hitomi, S. & Hakomori, S. (1992). Correlation of expression of H/Le(y)/Le(b) antigens with survival in patients with carcinoma of the lung. *N Engl J Med* **327**, 14-8.

Moréra, S., Imberty, A., Aschke-Sonnenborn, U., Rüger, W. & Freemont, P. S. (1999). T4 phage β-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J Mol Biol* **292**, 717-30.

Moréra, S., Lariviere, L., Kurzeck, J., Aschke-Sonnenborn, U., Freemont, P. S., Janin, J. & Ruger, W. (2001). High resolution crystal structures of T4 phage β-glucosyltransferase: induced fit and effect of substrate and metal binding. *J Mol Biol* **311**, 569-77.

Morgan, W. T. & Watkins, W. M. (2000). Unravelling the biochemical basis of blood group ABO and Lewis antigenic specificity. *Glycoconj J* **17**, 501-30.

Morgan, W. T. J. (1960). The Croonian Lecture: a contribution to human biochemical genetics; The chemical bases of blood group specificity. *Proceedings of the Royal Society Series B* **151**, 308-47.

Morgan, W. T. J. & Watkins, W. M. (1948). The detection of a product of of the blood group O gene and the relationship of the so-called O-substance to the agglutinogens A and B. *British Journal of Experimental Pathology* **29**, 159-73.

Morgan, W. T. J. & Watkins, W. M. (1953). The inhibition of the haemagglutinins in plant seeds by human blood group substances and simple sugars. *British Journal of Experimental Pathology* **34**, 94-103.

Mourant, A. E., Kopec, A. C. and Domainiewska-Sobezak, K. (1971) Blood-groups and blood-clotting. *Lancet* **1**, 223-7.

Mukherjee, A., Palcic, M. M. & Hindsgaul, O. (2000). Synthesis and enzymatic evaluation of modified acceptors of recombinant blood group A and B glycosyltransferases. *Carbohydrate Research* **326**, 1-21.

Mulichak, A. M., Losey, H. C., Walsh, C. T. & Garavito, R. M. (2001). Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure (Camb)* **9**, 547-57.

Murray, B. W., Wittmann, V., Burkart, M. D., Hung, S. C. & Wong, C. H. (1997). Mechanism of human α-1,3-fucosyltransferase V: glycosidic cleavage occurs prior to nucleophilic attack. *Biochemistry* **36**, 823-31.

Ogasawara, K., Yabe, R., Uchikawa, M., Saitou, N., Bannai, M., Nakata, K., Takenaka, M., Fujisawa, K., Ishikawa, Y., Juji, T. & Tokunaga, K. (1996). Molecular genetic analysis of variant phenotypes of the ABO blood group system. *Blood* **88**, 2732-7.

Olsson, M. L. & Chester, M. A. (2001). Polymorphism and recombination events at the ABO locus: a major challenge for genomic ABO blood grouping strategies. *Transfus Med* **11**, 295-313.

Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L. & Breton, C. (1999). Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* **9**, 323-34.

Oriol, R., Samuelsson, B. E. & Messeter, L. (1990). ABO antibodies-serological behaviour and immuno-chemical characterization. *J Immunogenet* **17**, 279-99.

Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscilation mode. *Methods Enzymol* **276**, 307-26.

Padlan, E. A. (1990). On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins* **7**, 112-24.

Painter, T. J., Watkins, W. M. & Morgan, W. T. (1965). Serologically active fucose-containing oligosaccharides isolated from human blood-group A and B substances. *Nature* **206**, 594-7.

Painter, T. J., Watkins, W. M. & Morgan, W. T. J. (1962). Isolation of a B specific disaccharide from human blood group B substance. *Nature* **193**, 1042-1044.

Painter, T. J., Watkins, W. M. & Morgan, W. T. J. (1963). Isolation of two serologically active trisaccharides from human B substance. *Nature* **199**, 282-3.

Palcic, M. M., Seto, N. O. & Hindsgaul, O. (2001). Natural and recombinant A and B gene encoded glycosyltransferases. *Transfus Med* **11**, 315-23.

Patenaude, S. I., MacKenzie, C. R., Bilous, D., To, R. J., Ryan, S. E., Young, N. M. & Evans, S. V. (1998). Production, crystallization and diffraction to atomic resolution of an antibody Fv specific for the blood-group A oligosaccharide antigen. *Acta Crystallogr D Biol Crystallogr* **54**, 1456-9.

Patenaude, S. I., Seto, N. O. L., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M. and Evans, S. V. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat Struct Biol* **9**, 685-90.

Paulson, J. C. & Colley, K. J. (1989). Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J Biol Chem* **264**, 17615-8.

Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A. & Negishi, M. (2000). Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J Biol Chem* **275**, 34580-5.

Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G. & Strynadka, N. C. (2001). Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nat Struct Biol* **8**, 166-75.

Quiocho, F. A. (1986). Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. *Annu Rev Biochem* **55**, 287-315.

Quiocho, F. A. (1988). Molecular features and basic understanding of protein-carbohydrate interactions: the arabinose-binding protein-sugar complex. *Curr Top Microbiol Immunol* **139**, 135-48.

Quiocho, F. A. (1989). Protein-carbohydrate interactions: basic molecular features. *Pure Appl Chem* **61**, 1293-1306.

Quiocho, F. A., Vyas, N. K. & Spurlino, J. C. (1991). Atomic interactions between proteins and carbohydrates. *Trans Am Crystallogr Assoc* **25**, 23-35.

Rademacher, T. W., Parekh, R. B. & Dwek, R. A. (1988). Glycobiology. *Annu Rev Biochem* **57**, 785-838.

Renkonen, K. O. (1948). Studies on haemagglutinins present in seeds of some representatives of the family *Leguminosae*. *Annals of Medical and Experimental Biology Fenn.* **26**, 66-72.

Rydberg, L. (2001). ABO-incompatibility in solid organ transplantation. *Transfus Med* **11**, 325-42.

Sands, D. E. (1969). *Introduction to Crystallography*, Dover Publications, Inc., Mineola, N.Y, 165 p.

Schachter, H., Michaels, M. A., Tilley, C. A., Crookston, M. C. & Crookston, J. H. (1973). Qualitative differences in the N-acetyl-D-galactosaminyltransferases produced by human A1 and A2 genes. *Proc Natl Acad Sci U S A* **70**, 220-4.

Schiff, F. (1927). Über den serologischen Nachweis der Blutgruppeneigenschaft O. *Klin Wschr* **6**, 303.

Schiff, F. (1934). *Zeitscrift für Immunitatsforsch* **82**, 302.

Seto, N. O., Compston, C. A., Evans, S. V., Bundle, D. R., Narang, S. A. & Palcic, M. M. (1999). Donor substrate specificity of recombinant human blood group A, B and hybrid A/B glycosyltransferases expressed in *Escherichia coli*. *Eur J Biochem* **259**, 770-5.

Seto, N. O., Compston, C. A., Szpacenko, A. & Palcic, M. M. (2000). Enzymatic synthesis of blood group A and B trisaccharide analogues. *Carbohydr Res* **324**, 161-9.

Seto, N. O., Palcic, M. M., Compston, C. A., Li, H., Bundle, D. R. & Narang, S. A. (1997). Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes. *J Biol Chem* **272**, 14133-8.

Seto, N. O., Palcic, M. M., Hindsgaul, O., Bundle, D. R. & Narang, S. A. (1995). Expression of a recombinant human glycosyltransferase from a synthetic gene and its utilization for synthesis of the human blood group B trisaccharide. *Eur J Biochem* **234**, 323-8.

Seymour, G. J., Savage, N. W. & Walsh, L. J. (1995). *Immunology: an Introduction for the Health Sciences*, McGraw-Hill Book Company, Roseville, Australia, 152 p.

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

Sharon, N. & Lis, H. (1990). Carbohydrate-protein interactions. *Chem Brit* **26**, 679-682.

Sharp, K. A. (1998). Calculation of HyHel10-lysozyme binding free energy changes: effect of ten point mutations. *Proteins* **33**, 39-48.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: High-resolution refinement. *Methods Enzymol* **277**, 319-343.

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

Sinnott, M. L. (1990). Catalytic mchanisms of enzymic glycosyl transfer. *Chemical Reviews* **90**, 1171-1202.

Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). The use of SnB to determine an anomalous scattering substructure. *Acta Crystallogr D Biol Crystallogr* **54**, 799-804.

Spurlino, J. C., Rodseth, L. E. & Quiocho, F. A. (1992). Atomic interactions in protein-carbohydrate complexes. Tryptophan residues in the periplasmic maltodextrin receptor for active transport and chemotaxis. *J Mol Biol* **226**, 15-22.

Stout, G. H. & Jensen, L. H. (1989). *X-ray structure determination : a practical guide*. 2nd edit, Wiley, New York, 453 p.

Takayama, S., Chung, S. J., Igarashi, Y., Ichikawa, Y., Sepp, A., Lechler, R. I., Wu, J., Hayashi, T., Siuzdak, G. & Wong, C. H. (1999). Selective inhibition of β-1,4- and α-1,3-galactosyltransferases: donor sugar-nucleotide based approach. *Bioorg Med Chem* **7**, 401-9.

Tarbouriech, N., Charnock, S. J. & Davies, G. J. (2001). Three-dimensional Structures of the Mn and Mg dTDP Complexes of the Family GT-2 Glycosyltransferase SpsA: A Comparison with Related NDP-sugar Glycosyltransferases. *J Mol Biol* **314**, 655-61.

Taylor, G. (1996). Sialidases: structures, biological significance and therapeutic potential. *Curr Opin Struct Biol* **6**, 830-7.

Thomas, R. (1999). The effect of point mutations on the binding affinity of anti-blood group A antibody AC1001, University of Ottawa.

Thomas, R., Patenaude, S. I., MacKenzie, C. R., To, R., Hirama, T., Young, N. M. & Evans, S. V. (2002). Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. *J Biol Chem* **277**, 2059-64.

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

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* **302**, 575-81.

Toone, E. J. (1994). *Curr Opin Struct Biol* **4**, 719.

Topping, M. D. & Watkins, W. M. (1975). Isolectric points of the human blood group A-1, A-2 and B gene-associated glycosyltransferases in ovarian cyst fluids and serum. *Biochem Biophys Res Commun* **64**, 89-96.

Unligil, U. M. & Rini, J. M. (2000). Glycosyltransferase structure and mechanism. *Curr Opin Struct Biol* **10**, 510-7.

Unligil, U. M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H. & Rini, J. M. (2000). X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *Embo J* **19**, 5269-80.

Van Asselt, E. J., Perrakis, A., Kalk, K. H., Lamzin, V. S. & Dijkstra, B. W. (1998). Accelerated X-ray structure elucidation of a 36 kDa muramidase/transglycosylase using wARP. *Acta Crystallogr D Biol Crystallogr* **54**, 58-73.

Van Die, I., Bakker, H. & Van den Eijnden, D. H. (1997). Identification of conserved amino acid motifs in members of the β1→4-galactosyltransferase gene family. *Glycobiology* **7**, v-viii.

Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130.

Villeneuve, S., Souchon, H., Riottot, M. M., Mazie, J. C., Lei, P., Glaudemans, C. P., Kovac, P., Fournier, J. M. & Alzari, P. M. (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**, 8433-8.

Vogel, F., Pettenkofer, H. J., and Helmbold, W. (1960) Uber die Populationsgenetik der ABO-Blugruppen. 2. Mitteilung. Genhaüfigkeit und epidemische Erkrankunge. *Acta Genet Statis Med* **10**, 267.

von Decastello, A. & Sturli, A. (1902). Über die isoagglutinine im serum gesunder und kranker Menschen. *Munchen Medicine Wchnschr* **49**, 1090-5.

von Dungern, E. & Hirszfeld, L. (1910). Über Vererbung gruppenspezifischer Strukturen des Blutes. *Zeitscrift für Immunforsch* **6**, 284-92.

Vrielink, A., Rüger, W., Driessen, H. P. C. & Freemont, P. S. (1994). Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *Embo J* **13**, 3413-22.

Vyas, N. K. (1991). Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol* **1**.

Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1991). Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and Functional Similarity. *J Biol Chem* **266**, 5226-37.

Watkins, W. M. (1956). The appearance of H specificity following the enzymic inactivation of blood group B substance. *Biochemical Journal* **64**, 21P.

Watkins, W. M. (1962). Changes in the specificity of blood group mucopolysaccharides induced by enzymes from *Trichomonas foetus*. *Immunology* **5**, 245-66.

Watkins, W. M. (1965). Relationship between structure, specificity and genes within the ABO and Lewis blood-group systems. *Bibl Haematol* **23**, 443-52.

Watkins, W. M. (1966). Blood group substances. *Science* **152**, 172-81.

Watkins, W. M. (2001). The ABO blood group system: historical background. *Transfus Med* **11**, 243-65.

Watkins, W. M. & Morgan, W. T. J. (1952). Neutralisations of the anti-H agglutinin in eel serum by simple sugars. *Nature* **169**, 825-6.

Watkins, W. M. & Morgan, W. T. J. (1955). Inhibition by simple sugars of enzymes which decompose blood-group substances. *Nature* **175**, 676-7.

Watson, J. D. & Crick, F. H. C. (1953). Molecular structure of the nucleic acids: a structure of deoxynucleic acid. *Nature* **171**, 964-7.

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

Weis, W. I. (1997). Cell-surface carbohydrate recognition by animal and viral lectins. *Curr Opin Struct Biol* **7**, 624-30.

Wiggins, C. A. & Munro, S. (1998). Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci USA* **95**, 7945-50.

Witebsky, E. & Klendsshoj, N. C. (1941). The isolation of an O specific substance from gastric juice of secretors and carbohydrate-like substances from gastric juice of non-secretors. *Journal of Experimental Medicine* **73**, 655-67.

Woolfson, M. M. (1997). *An introduction to X-ray crystallography.* 2nd edit, Cambridge University Press, New York, NY, USA, 402 p.

Yamakawa, T. & Iida, T. (1953). Immunochemical study on the red blood cells. I Globoside as the agglutinogen of the ABO system on erythrocytes. *Japanese Journal of Experimental Medicine* **23**, 327.

Yamamoto, F. (2001). Cloning and regulation of the ABO genes. *Transfus Med* **11**, 281-94.

Yamamoto, F., Clausen, H., White, T., Marken, J. & Hakomori, S. (1990). Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229-33.

Yamamoto, F. & Hakomori, S. (1990). Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. *J Biol Chem* **265**, 19257-62.

Yamamoto, F. & McNeill, P. D. (1996). Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferases. In vitro mutagenesis study. *J Biol Chem* **271**, 10515-20.

Yamamoto, F., McNeill, P. D. & Hakomori, S. (1992). Human histo-blood group A2 transferase coded by A2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal. *Biochem Biophys Res Commun* **187**, 366-74.

Yamamoto, F., McNeill, P. D., Kominato, Y., Yamamoto, M., Hakomori, S., Ishimoto, S., Nishida, S., Shima, M. & Fujimura, Y. (1993a). Molecular genetic analysis of the ABO blood group system: 2. *cis*-AB alleles. *Vox Sang* **64**, 120-3.

Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S., Bromilow, I. M. & Duguid, J. K. (1993b). Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele. *Vox Sang* **64**, 175-8.

Yamamoto, F., McNeill, P. D., Yamamoto, M., Hakomori, S. & Harris, T. (1993c). Molecular genetic analysis of the ABO blood group system: 3. A(X) and B(A) alleles. *Vox Sang* **64**, 171-4.

Yang, P. L. & Schultz, P. G. (1999). Mutational analysis of the affinity maturation of antibody 48G7. *J Mol Biol* **294**, 1191-201.

Yang, W., Hendrickson, W. A., Kalman, E. T. & Crouch, R. J. (1990). Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from *Escherichia coli*. *J Biol Chem* **265**, 13553-9.

Yates, A. D., Feeney, J., Donald, A. S. & Watkins, W. M. (1984). Characterisation of a blood-group A-active tetrasaccharide synthesised by a blood-group B gene-specified glycosyltransferase. *Carbohydr Res* **130**, 251-60.

Yates, A. D. & Watkins, W. M. (1982). The biosynthesis of blood group B determinants by the blood group A gene-specified α-3-N-acetyl-D-galactosaminyltransferase. *Biochem Biophys Res Commun* **109**, 958-65.

Zdanov, A., Li, Y., Bundle, D. R., Deng, S. J., MacKenzie, C. R., Narang, S. A., Young, N. M. & Cygler, M. (1994). Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-Å resolution. *Proc Natl Acad Sci U S A* **91**, 6423-7.

Zechel, D. L. & Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* **33**, 11-8.

Sonia I. Patenaude
1 Thornbury Cres.
Ottawa, Ontario
K2G 6C6
(613) 274-0618

EDUCATION

May 1996 - present

Doctoral Degree
Department of Biochemistry, Microbiology & Immunology
University of Ottawa
Title: Structural Studies of the Biosynthesis and Recognition of the Human ABO(H) Blood Group Antigens

My graduate research focused on the determination of the three-dimensional structure of carbohydrate-binding proteins in order to determine their mechanism and the nature of their specificity. This project involved several protein expression and purification techniques, including: tissue culture (insect cells), bacterial culture (*E. coli*), protease digest, HPLC and FPLC (incl. affinity, size exclusion and ion exchange chromatography), polyacrylamide gel electrophoresis of proteins and Western blotting. I have extensive experience in protein crystallization, X-ray data collection (home and synchrotron sources), structure determination methods (including molecular replacement, isomorphous replacement, and multiwavelength anomalous dispersion) and molecular modelling. I have also gained some experience in molecular biology.

1992-1996

B. Sc. (Honours) Biochemistry, *Magna cum laude*
Department of Biochemistry
University of Ottawa
Title: The Determination of the Crystal Structure of an Antigen-Binding Fragment Specific to the Blood Group A Oligosaccharide

As an Honours' student, I solved the structure of an antibody fragment using molecular replacement, modelling and computational methods. In addition, mutants were designed to probe the specificity and affinity of the antibody.

EXPERIENCE

I. Teaching experience

September - December 1997 September - December 1996	Department of Biochemistry University of Ottawa Laboratory teaching assistant in the Physical Biochemistry section of the 3rd year laboratory course.
January - April 1996	Department of Biochemistry University of Ottawa Laboratory teaching assistant in the 2nd year laboratory course.

II. Scientific

May 1995 - August 1995	University of Ottawa Department of Biochemistry As a summer student, I set up and maintained a new laboratory. I was also introduced to three-dimensional structure solution using various programs on Silicon Graphics hardware.

III. Other

June 1992 - October 1997	Canadian Armed Forces Reserves 28 Service Battalion Supply Technician

CONTRIBUTIONS TO RESEARCH

I. Articles in refereed journals

Patenaude, S. I., Seto, N. O. L., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M. and Evans, S. V. (2002) The Structural Basis for Specificity in Human ABO(H) Blood Group Biosynthesis. *Nat Struct Biol* **9**, 685-90.

Thomas, R., **Patenaude, S.I.**, MacKenzie, C.R., To, R., Hirama, T., Young, N.M. & Evans, S.V. (2002) Structure of an Anti-blood Group A Fv and Improvement of Its Binding Affinity without Loss of Specificity. *J Biol Chem* **277**, 2059-64

Müller-Loennies, S., MacKenzie, C.R., **Patenaude, S.I.**, Evans, S.V., Kosma, P., Brade, H., Brade, L. And Narang, S. (2000) Characterization of High Affinity Monoclonal Antibodies Specific for Chlamydial Lipopolysaccharide. *Glycobiology* **10**, 121-130.

Patenaude, S.I., Vijay, S.M., Yang, Q.-L., Jennings, H.J., and Evans, S.V. (1998) Crystallization and Preliminary X-ray Diffraction Analysis of Antigen-binding Fragments Which are Specific for Antigenic Conformations of Sialic Acid Homopolymers. *Acta Crystallogr* **D54**, 1005-1007.

Patenaude, S.I., MacKenzie, C.R., Bilous, D., To, R.J., Ryan, S.E., Young, N.M., and Evans, S.V. (1998) Production, Crystallization and Diffraction to Atomic Resolution of an Antibody Fv Specific for the Blood-group A Oligosaccharide Antigen. *Acta Crystallogr* **D54**, 1456-1459.

II. Non-refereed contributions (Presentations)

Structure-Function Relationship of the Human ABO Blood Group Glycosyltransferases. **Sonia I. Patenaude,** Nina O.L. Seto, Svetlana Borisova, Adam Szpacenko, Monica M. Palcic and Stephen V. Evans. American Crystallographic Association Annual Meeting, San Antonio, TX, 2002.

Generation of Higher Affinity Mutants of a Blood Group A Specific Antibody and Their Characterization by Surface Plasmon Resonance. R. MacKenzie, R. To, R. Wehbi, T. Hirama, **S. Patenaude**, M. Young, J.R. Brisson, and S.V. Evans. XXth International Carbohydrate Symposium, Hamburg, Germany, 2000.

A single Point-Mutation Yields a 30-Fold Increase in Binding Affinity of a Carbohydrate-Specific scFv. **Sonia I. Patenaude**, Roula Thomas, C. Roger MacKenzie, N. Martin Young and Stephen V. Evans. BioContact Québec, Québec City, QC, 1999.

Single Point-Mutation Yields 30-Fold Increase in Binding Affinity of a Carbohydrate-Specific scFv. **Sonia I. Patenaude**, Roula Thomas, C. Roger MacKenzie, N. Martin Young and Stephen V. Evans. XVIIIth Congress of the International Union of Crystallography, Glasgow, Scotland, 1999.

Antibody Recognition of the Blood Group A Oligosaccharide Antigen. **Sonia I. Patenaude**, Rebecca J. To, Doris Bilous, N.Martin Young and Stephen V. Evans. American Crystallographic Association Annual Meeting, St. Louis, MO, 1997.

The Binding Site of an Antibody Fv Antigen-Binding Fragment Specific to the Blood Group A Polysaccharide Antigen Displays Two Distinct Conformations. **Sonia I. Patenaude**, Rebecca J. To, Les W. Tari, Doris Bilous, N. Martin Young and Stephen V. Evans. XVIIth Congress of the International Union of Crystallography, Seattle, WA. 1996.

AWARDS

2002	American Crystallographic Association Pauling Prize
2001-present	Ontario Graduate Scholarship, Ministry of Education and Training
1998-present	University of Ottawa Excellence Scholarship
1998-2001	Doctoral Research Award, Medical Research Council of Canada (now CIHR)
1998	Ontario Graduate Scholarship, Ministry of Education and Training (Declined)
1998	Postgraduate Scholarship, Natural Sciences and Engineering Research Council of Canada (Declined)
1992	Governor-General's Award - Bronze Medal
1992	Scholarship for Academic Excellence, University of Ottawa
1992	Ontario Scholar
1992	Canada Scholar

OTHER

Bilingual
Qualified Level "C" CPR and First Aid
Working knowledge of Unix, Windows and various Microsoft tools

REFERENCES

References are available upon request.