

University of St Andrews



Full metadata for this thesis is available in
St Andrews Research Repository
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

Structural Studies of Thyroid Peroxidase and Thyroid Peroxidase Antibodies

Elaine Hendry

A thesis submitted for the degree of Doctor of Philosophy in the University of St.
Andrews, September 2000.



π 738

To Jane and Martin

Declaration

- i) I, Elaine Hendry, hereby certify that this thesis, which is approximately 50 000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Signature

Date

19th Sept 2000

- ii) I was admitted as a research student in May 1999 and as a candidate for the degree of PhD in May 1999 at the University of St. Andrews; the higher study for which this is a record was carried out in the University of Bath between October 1997 and April 1999, and in the University of St. Andrews between May 1999 and September 2000.

Signature

Date

19th Sept 2000

- iii) I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature of Supervisor

Date

19th Sept. 2000

Copyright

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Signature

Date

19th Sept 2000

Acknowledgements

This work is the result of an MRC Industrial Collaborative Studentship with RSR Ltd, Cardiff, and I am grateful to the MRC and RSR for funding. I would like to thank RSR for allowing me to work at the company for five months, and I am particularly grateful to Drs Wisia Furmaniak and Fiona Grennan Jones for their supervision during this period.

Various people at RSR have assisted with this study, and I would like to thank them for their contributions. They are Drs Kasia Ziemnicka and Fiona Grennan Jones who grew the original TPO crystals, Drs Mike Powell and Nabuki Hayakawa who sequenced the 4F5 Fab DNA, and Sara Nute and Joanne Godfrey who carried out the 4F5/2G4 binding studies.

I am grateful to Drs Anne Dell and Stuart Haslam of Imperial College, London for carrying out the mass spectroscopy of TPO glycans, and to Dr Catherine Botting and Paul Talbot of St. Andrews University for providing the mass spectroscopy and N-terminal sequencing services.

My sincere thanks go to Professor Garry Taylor for his supervision and friendship. He has helped to make the past three years an enjoyable experience and I wish him the best for his new role at St. Andrews. I am also grateful to all the members of Garry's group who have provided assistance along the way. Finally, I would like to thank Mum, Dad, and Rob for their love and support.

Abstract

Thyroid peroxidase (TPO) catalyses the production of thyroid hormones and is a major autoantigen in autoimmune thyroid disease (AITD). The majority of TPO autoantibodies bind to an immunodominant region consisting of two overlapping domains. Precise location of these domains would help the understanding of AITD pathogenesis. The aim of this study was to use X-ray crystallography and molecular modelling to investigate the interaction of TPO with autoantibodies.

Myeloperoxidase (MPO) was used as a phasing model to solve the structure of TPO to 7 Å. A potential solution indicates that TPO uses the same interchain disulphide bond as MPO, but that the dimers are at a different orientation. The resolution limits of the TPO crystals could not be extended and heterogeneity at the N-terminus was shown to be a possible cause. A model of TPO was built, and provides information on regions of the structure that may be accessible to autoantibodies.

The antibodies 2G4 and 4F5, react with two major autoantigenic regions on TPO. 2G4 is an autoantibody isolated from Hashimoto's disease patients, and 4F5 is a mouse monoclonal antibody. The Fab fragments of 2G4 and 4F5 were purified and their binding to TPO characterised. The 4F5 Fab fragment was crystallised and the structure solved to 1.9 Å using molecular replacement. The refined structure has an R factor of 19.5% and a free R factor of 23.9%. The combining site of 4F5 is flat and is rich in tyrosine residues. Crystal contacts may be representative of 4F5-TPO interactions. Comparison of the structure of 4F5 with that of a TPO autoantibody Fab, TR1.9, shows that the two antibodies are not likely to recognise the same region on TPO.

In summary, a model of TPO has been built and structure of an antibody that binds to the immunodomain of TPO has been solved. These have been analysed to provide structural insights into the interaction of TPO with autoantibodies.

Contents

Chapter 1	Introduction	
1.1	Immunology	
1.1.1	Overview of the Immune System	1
1.1.2	Antibody Structure	5
1.1.3	Generation of Antibody Diversity	12
1.1.4	Antibody-Antigen Structure	14
1.1.5	Epitopes	18
1.2	Autoimmunity	
1.2.1	Autoimmune Disease	20
1.2.2	Mechanisms of Tolerance	21
1.2.3	Breakdown of Tolerance	22
1.2.4	Autoantibodies	25
1.3	Autoimmune Thyroid Disease	
1.3.1	Clinical Aspects	26
1.3.2	Predisposition	27
1.3.3	Pathogenesis	29
1.3.4	Autoantibodies	32
1.4	Thyroid Peroxidase	
1.4.1	Peroxidase Family	33
1.4.2	Molecular Biology	35
1.4.3	Structure	38
1.4.4	Catalytic Role	42
1.4.5	Epitope Mapping	45
1.4.6	Thyroid Peroxidase Autoantibodies	48
1.5	Experimental Aims	52

Chapter 2	Thyroid Peroxidase Characterisation	
2.1	Introduction	55
2.2	Production and Purification	56
2.3	Thyroid Peroxidase Assays	57
2.4	SDS-PAGE Analysis	59
2.5	IEF Analysis	62
2.6	N-terminal Sequencing	64
2.7	Mass Spectroscopy	68
2.8	Deglycosylation Studies	70
2.9	Summary	73
Chapter 3	Thyroid Peroxidase Crystallography	
3.1	Introduction	77
3.2	TPO-Fab Complex Purification	78
3.3	Crystallisation	80
3.4	Data Collection	84
3.5	Molecular Replacement	90
3.6	Refinement	98
3.7	Summary	98
Chapter 4	Thyroid Peroxidase Modelling	
4.1	Introduction	100
4.2	Protein Sequence Alignments	101
4.3	Peroxidase Domain	106
4.4	CCP and EGF Domains	120
4.5	N-terminal Domain	122
4.6	Quaternary Structure	127
4.7	Summary	127
Chapter 5	4F5 and 2G4 Characterisation and Fab Production	
5.1	Introduction	130
5.2	4F5 Production and Sequencing	131

	5.3	4F5 Sequence Analysis	133
	5.4	4F5 and 2G4 IgG Characterisation	139
	5.5	4F5 and 2G4 Fab Fragment Production	142
	5.6	4F5 and 2G4 Fab Fragment Characterisation	145
	5.7	Summary	152
Chapter 6		4F5 Crystallography	
	6.1	Introduction	153
	6.2	Crystallisation	154
	6.3	Data Collection	158
	6.4	Molecular Replacement	162
	6.5	Refinement	165
	6.6	Validation	171
	6.7	Summary	176
Chapter 7		4F5 Structure Analysis	
	7.1	Introduction	177
	7.2	Framework Regions	178
	7.3	Complementarity Determining Regions	185
	7.4	Comparison with TR1.9	201
	7.5	Summary	204
Chapter 8		Discussion	
	8.1	Thyroid Peroxidase Heterogeneity	208
	8.2	Thyroid Peroxidase Structure	211
	8.3	4F5 Structure	217
	8.4	Summary	220
Appendix A		Protein Crystallography Theory	221
Appendix B		Amino Acid Sequences of TPO Autoantibodies	235
References			245

Abbreviations

ADCC	Antibody Dependent Cell Mediated Cytotoxicity
AITD	Autoimmune Thyroid Disease
AMPS	Ammonium Persulphate
APC	Antigen Presenting Cell
ASA	Accessible Surface Area
Bi	Biotin
Bp	Base pair
BSA	Bovine Serum Albumin
Calc	Calculated
CCD	Charge Coupled Device
CCO	Cytochrome c oxidase
CCP	Complement Control Protein
CD	Cluster of Differentiation
cDNA	Complementary DNA
CDR	Complementarity Determining Region
D	Diversity
DIT	Diiodotyrosine
DNA	Deoxyribonucleic Acid
EAT	Experimental Autoimmune Thyroiditis
EBV	Epstein-Barr Virus
EDTA	Ethylenediamine tetra-acetate
EGF	Epidermal Growth Factor
ELISA	Enzyme-linked immunosorbant assay
Endo	Endoglycosidase
EPO	Eosinophil peroxidase
F	Structure Factor
Fab	Fragment antibody binding
Fc	Fragment crystallisable
Fuc	Fucose

FWR	Framework Region
Gu	Guaicol unit
HC	Heavy Chain
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
Hex	Hexose
Hexnac	N-acetyl Hexose
HLA	Human Leukocyte Antigen
IDDM	Insulin-Dependent Diabetes Mellitus
IEF	Isoelectric Focusing
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
J	Joining
LATS	Long-Acting Thyroid Stimulator
LC	Light Chain
LPO	Lactoperoxidase
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization- Time of Flight
MHC	Major Histocompatibility Complex
MIT	Monoiodotyrosine
MPO	Myeloperoxidase
MS	Mass Spectroscopy
MS	Multiple Sclerosis
M _w	Molecular weight
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Center for Biotechnology Information
NIBSC	National Institute for Biological Standards and Control
NK	Natural Killer
NMR	Nuclear Magnetic Resonance
Obs	Observed
OD	Optical Density
PBS	Phosphate Buffered Saline
PDB	Protein Data Bank
PEG	Polyethylene Glycol

PGHS	Prostaglandin H2 synthase
PNGase	Peptide-N-Glycosidase
PSF	Protein Structure File
R:S	Replacement: Silent
RA	Rheumatoid Arthritis
RIA	Radioimmunoassay
RF	Rheumatoid Factor
Rmsd	Root mean squared deviation
RNA	Ribonucleic Acid
SA-POD	Streptavidin-peroxidase
ScFv	Single chain Fragment variable
SDR	Specificity Determining Region
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
T ₃	Thyroxine
T ₄	Triiodothyronine
Tc	Cytotoxic T cell
TCR	T-cell receptor
TEMED	Tetramethylethylenediamine
Tg	Thyroglobulin
Th	T helper cell
TM	Transmembrane
TMB	Tetramethyl benzidine
TPO	Thyroid peroxidase
TRIS	Tris-(hydroxymethyl)-methylamine
TSHR	Thyroid Stimulating Hormone Receptor
V	Variable

Units

International System of Units, except for:

Å	Ångström 1Å = 0.1 nm
Da	Daltons, 1Da = 1 gram/mole

Amino Acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartate	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamate	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Chapter 1

Introduction

1.1 Immunology

1.1.1 Overview of the Immune System

Immunology as an experimental science dates to Edward Jenner's successful vaccination against smallpox in 1796 [Jenner 1798]. One hundred years later, Ehrlich and Morgenroth showed that animals did not produce antibodies against their own erythrocytes [Ehrlich & Morgenroth 1901], and the possibility of the immune system attacking the host seemed inconceivable. However, in 1906, Donath and Landsteiner isolated an autoantibody associated with a human disease [Donath & Landsteiner 1904]. Today autoimmune disease is recognised as a widespread phenomenon, and the study of immunology has led to a growing understanding of the mechanisms responsible for autoimmunity [Ollier & Symmons 1992].

The immune system has been found to be a rapid and effective weapon against disease [Roitt 1994]. The receptors of the immune system are capable of binding to an almost infinite number of previously unencountered molecules. They have the ability to distinguish the subtle differences between benign self molecules and potentially damaging foreign, or non-self, molecules. In a healthy individual, an encounter with a non-self molecule leads to the triggering of powerful effector mechanisms that eliminate the associated pathogen. However, on occasion, the attack becomes directed towards self molecules and results in autoimmune disease.

The immune system can be classified as innate or acquired, and the immune responses as cell-mediated or humoral. The innate immune system, first discovered in marine starfish in the late 1890s [Metchnikoff 1893], is composed of an array of cells and molecules of a broad specificity with important roles during the early stages of infection. In contrast, acquired immunity is mediated by small mononuclear blood cells, called lymphocytes [Gowans 1962], with high specificity and the ability to form immunological memory. Molecules that are recognised by lymphocyte receptors are known as antigens and knowledge of self antigens is learned during development by the process of tolerance [Billingham 1953]. Each lymphocyte expresses a receptor of single specificity and reaction with self antigens in the thymus and bone marrow leads to deletion or silencing. Surviving lymphocytes are stimulated by contact with non-self antigen and are expanded by clonal selection [Burnet 1959].

There are two classes of lymphocytes with specific receptors for antigens, T- and B-cells, which mediate cell-mediated and humoral immunity respectively. T-cells possess antigen binding molecules called T-cell receptors (TCRs) [Hendrick 1984]. These recognise and bind peptides derived from antigens that are presented by the major histocompatibility complex (MHC) [Zinkernagel & Doherty 1974] on the professional antigen presenting cells (APCs). Dendritic cells, macrophages and B-cells are the most common APCs. Recently, the interaction of TCRs and APCs has been shown to occur at immunological synapses composed of a central cluster of TCRs surrounded by an outer ring of adhesion molecules [Monks 1998]. Binding leads to the transmission of signals from the TCR to the cytoplasm via a cell surface molecule called CD3 [Weissman 1988]. Molecules with a CD prefix are cell surface molecules named according to clusters of differentiation identified by monoclonal antibody binding.

There are two types of T-cells – CD4⁺ helper cells (Th) and CD8⁺ cytotoxic T-cells (Tc) [Kehry & Hodgkin 1993; Gotch 1996]. Each interact with different MHC molecules, CD4⁺ with MHC class II molecules and CD8⁺ with MHC class I

molecules [Moss 1992]. Activated cytotoxic T-cells degranulate to release perforin and other cytotoxic mediators that induce apoptosis of the target cell. Helper T-cells release molecules called cytokines that control the function of immune cells. Th cells fall into two classes that recognise different pathogens and are defined by the specific cytokine cocktail they release [Mosman 1986]. CD4⁺ Th1-cells are activated by phagocytosis and intracellular microbes and produce interferon- γ (IFN- γ) that promotes phagocytosis. In contrast, extracellular infections induce the differentiation of CD4⁺ T-cells into the Th2 subset that produce interleukin (IL) 4 and 5 and promote immunoglobulin (Ig) E and eosinophil-mediated killing. This results in distinct T-cell profiles in inflammatory autoimmune disease (mediated by Th1) and allergies (mediated by Th2) [Charlton & Lafferty 1995].

MHC molecules, with over 100 alleles, are highly polymorphic and allow T-cells to sample antigens from different cell compartments [Germain 1994]. Class I molecules are found on all nucleated cells, and present a selection of endogenous peptides for surveillance by the T-cells. The peptides are cleaved in the cytoplasm by the proteasome to eight to nine amino acids long [Townsend & Bodmer 1989]. After association with the MHC molecule and β -2 microglobulin they are transported to the cell surface as a stable complex. CD8⁺ T-cells are therefore alerted to the presence of intracellular viral and tumor antigens which leads to their elimination. Class II molecules are present on APCs and may be induced on other cells. APCs internalize peptides derived from extracellular antigens following binding via surface Igs and Ig receptors. The peptides are degraded in acid endosomes to 13-17 residues and bind to an invariant chain which dissociates on binding MHC [Hunt 1992].

Antibodies were first discovered in 1890 by Behring [Behring & Kitasato 1890]. It has since been shown that each B-cell expresses about 10⁶ antibodies of the same specificity on the cell surface [Abler 1992]. Antigen binding to a surface expressed antibody causes the B-cell to internalise the antigen and present it in the context of MHC II molecules. Most B-cells are dependent on T-cell help for their

proliferation, although some T-independent B-cells can be activated directly by antigen binding [Parker 1993]. Activated B-cells migrate to the germinal centres where affinity maturation and class switching take place [Benner 1981]. B-cells that bind antigen more tightly are selected for, and the cells mature into plasma cells which secrete antibodies into the serum. B-cell memory lasts for more than ten years in humans and depends on stores of antibody in the secondary lymphoid tissue [Gray 1992]. On a second encounter with the antigen the immune system can respond in a more rapid and extensive way.

Antibodies contribute to the immune response by promoting agglutination, stimulating phagocytosis and antibody dependent cell mediated cytotoxicity (ADCC), and activating a group of serum proteins known as complement [Kulberg 1972]. ADCC is triggered by the binding of natural killer (NK) cells to antibody coated infected cells via Fc receptors, and leads to death of the infected cell. The classical complement pathway can be activated by antigen-bound IgG or IgM, or directly by charged groups on bacteria and viruses. Activation leads to phagocytosis, release of anaphylotoxins and leukocyte regulatory molecules and lysis of the pathogenic cell.

The structure of many of the molecules of the surface of lymphocytes have been solved by X-ray crystallography or nuclear magnetic resonance (NMR) [Bork 1994; Maenaka & Jones 1999; Rudd 1999] (Figure 1.1). They tend to have a modular structure, are often glycosylated, and a common feature is the presence of the Ig fold, first identified in antibodies [Poljak 1973]. The TCR is a disulphide-linked transmembrane heterodimer composed of either α and β chains or γ and δ chains [Garcia 1996], each made up of two Ig domains. The MHC molecules achieve a binding site by using the groove between two α -helices resting on a platform of eight antiparallel β -strands [Bjorkman 1987]. The structure of MHC:TCR:peptide complexes [Garboczi 1996; Garcia 1996; Reinhertz 1999] indicates that only a few residues of the MHC associated peptide are recognised by the TCR.

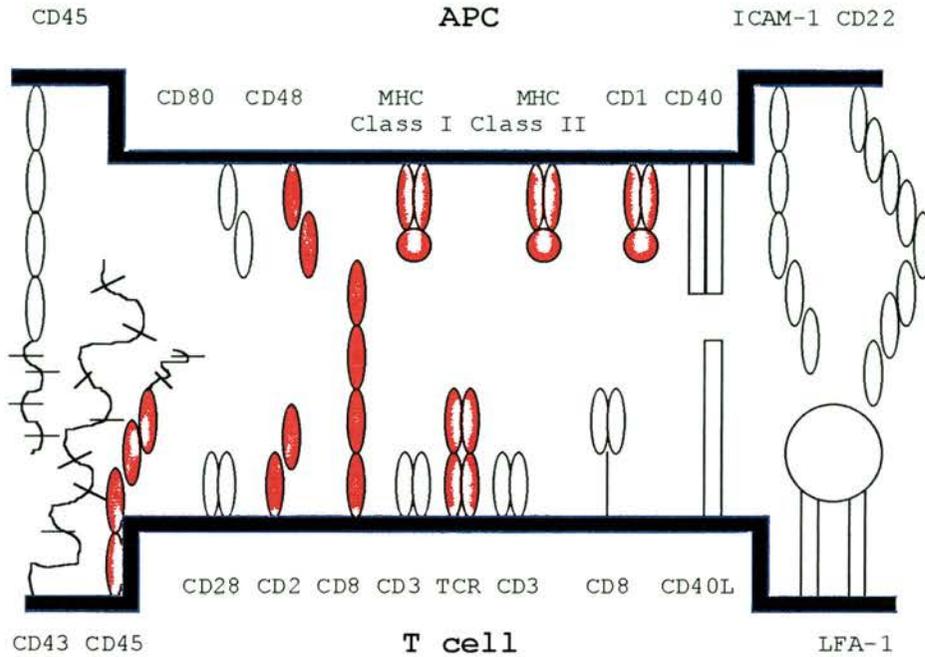


Figure 1.1

The structure of the immunological synapse, known crystal structures are shaded red.

1.1.2 Antibody Structure

There is more structural information about antibodies than any other protein. As of 1 September 2000, the total number of antibody structures is 330, representing 2.5% of all structures in the protein structure database [ABG 2000]. 296 of the structures have been solved by crystallography, four by NMR and 30 are derived from molecular modelling. 47% of the structures are uncomplexed, and 53% are bound to an antigen. The majority of structures are of fragments of antibodies that

contain the antigen interaction site. Antibody structure has been reviewed in detail [Padlan 1994].

Antibodies are composed of a heavy chain (HC) and a light chain (LC). The LC of human antibodies can be either λ or κ . The HC determines the class or isotype, and can be M, G (four subclasses), A (two subclasses), E or D (Table 1.1). The IgG molecule consists of three domains – two identical Fabs (fragment antibody binding) and one Fc (fragment crystallisable), linked by a flexible hinge region. The Fab fragments are made up of the V_L , C_L , V_H and C_{H1} domains, the Fc is a dimer of the C_{H2} and C_{H3} domains. A single chain of variable domains (ScFv) is an artificial construct made by attaching the V_L and V_H domains by a short linker. The domain structure is illustrated schematically in Figure 1.2 and the structure of an intact antibody molecule is shown in Figure 1.3.

IgG	Most abundant Ig in internal body fluids, fixes complement, crosses the placenta and binds to macrophages and polymorphs
IgA	Most abundant Ig in sero-mucous secretions, fixes complement and binds to macrophages and polymorphs
IgM	Produced early in the immune response, effective agglutinator, fixes complement
IgD	Present on lymphocyte surface
IgE	Protection of external body surfaces, fixes to mast cells and basophils, recruits antimicrobial agents, raised in parasitic infections, responsible for allergy

Table 1.1

The properties of the different antibody isotypes.

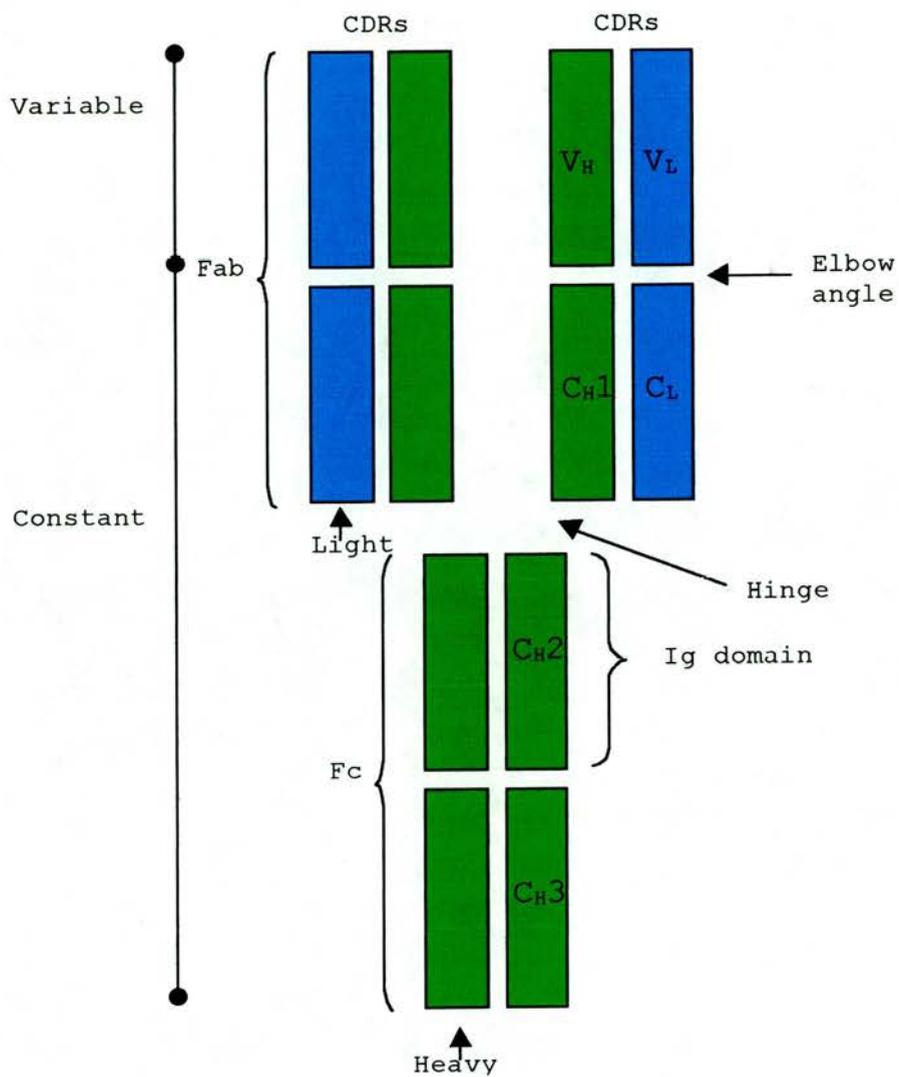


Figure 1.2

A schematic diagram of the structure of an IgG molecule.

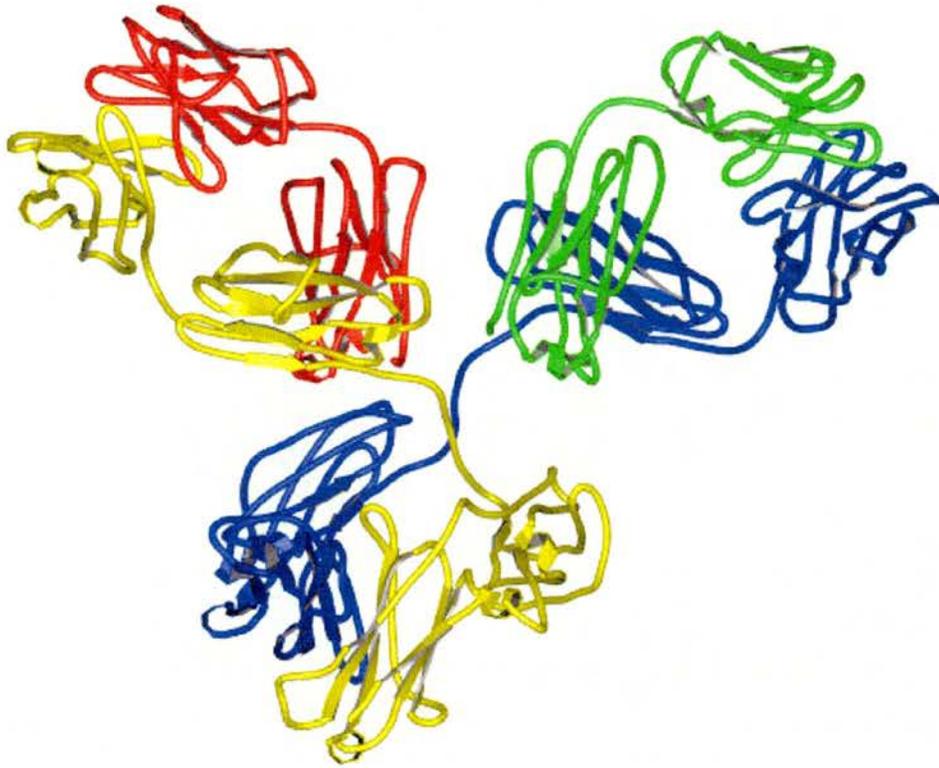


Figure 1.3

The crystal structure of an intact antibody IgG1 molecule, 1IGY [Harris 1998]. LCs are in red and green and HCs in blue and yellow.

The individual domains form a compact bilayer of two antiparallel β -sheets arranged in an Ig fold. The β -sheets form a framework structure that is essentially invariable in structure between antibodies from different classes and origins. The N-terminal domains of the HC and LC are known as variable domains, and each contain a sandwich of a four-stranded and a five-stranded β -sheet, shown in Figure 1.4A. The remaining C-terminal LC domain (C_L) and the three HC domains (C_{H1-3}) are called constant domains, and are composed of a sandwich of two β -sheets of three and four strands (Figure 1.4B). Antigen specificity is derived from six loops

with hypervariable amino acid sequences [Wu & Kabat 1970], called complementarity determining regions (CDRs), that link the β -strands of the N-terminal domain. The loops are shown schematically in Figure 1.5. The Fc domain uses the β -sheet as a binding surface. It is involved in the effector functions of the antibody such as binding to complement factors and receptors.

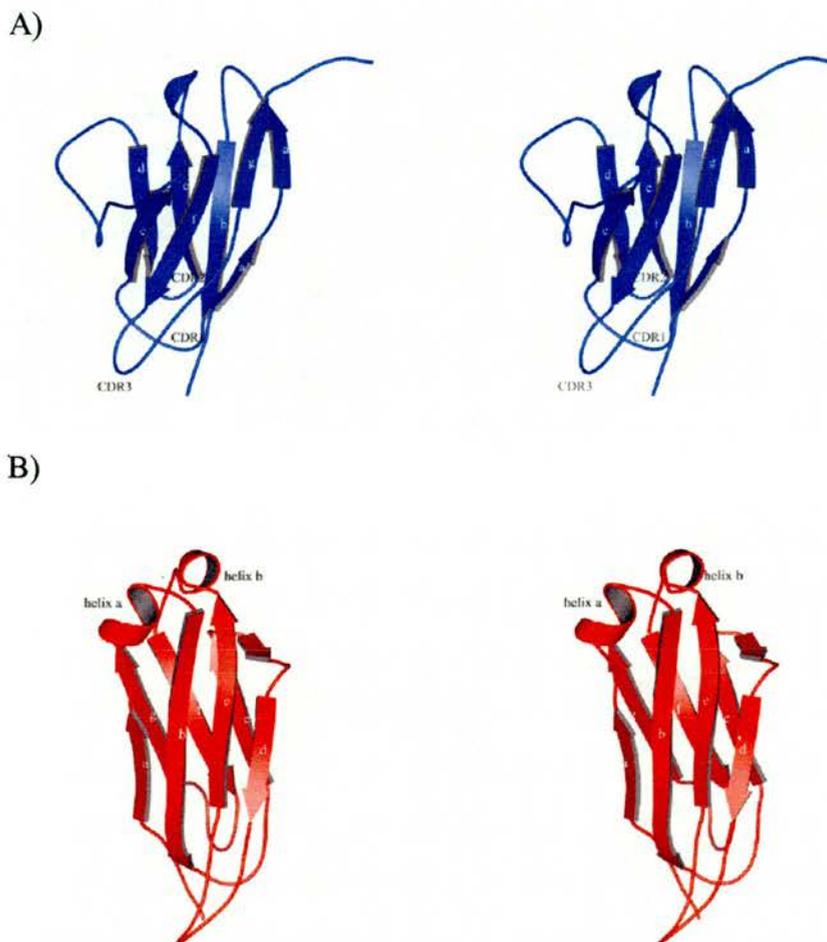


Figure 1.4

Stereo diagrams of A) a variable domain and B) a constant domain of the LC of an IgG molecule.

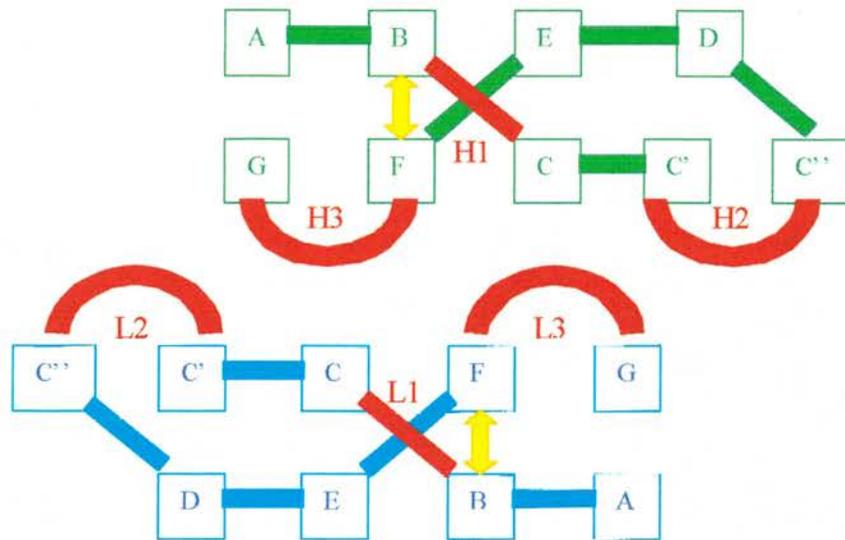


Figure 1.5

A schematic diagram of the variable domains of the HC (green) and LC (blue), looking down onto the CDRs. Yellow arrows indicate disulphide bridges and red lines are the CDRs.

The first crystal structure of an antibody-antigen complex showed the CDRs from both the LC (L1-3) and HC (H1-3) are involved in contacting the antigen [Amit 1986]. The nature of the antibody binding site is determined by the relative position, sequence and length of the CDRs. The relative position of the CDRs is determined by the position of the V_L and V_H domains with respect to one another. Five CDRs, in combination with the important framework amino acids, form a restricted set of loop structures known as canonical forms [Chothia & Lesk 1987; Chothia 1989; Al-Lazikani 1997]. In contrast, H3 is highly variable in structure, varying in length from 2 to 26 amino acids in man. However, rules have been described for the relations between the sequence of H3 and the conformation of the stem region [Morea 1998b]. The definition of residues making up the CDRs can be based on structure [Chothia & Lesk 1987], however, sequence variability based, or Kabat, definitions are still the most widespread.

The pair of V domains and the pair of C domains are tightly associated with each other via pseudo-diads. The link region between the V and C domain is called the switch, the relative disposition of the two domains is defined by the angle between the V_L - V_H and C_L - C_H pseudo-dyads. This angle is called the elbow bend and has been observed to vary from 127° (1bbd) [Tormo 1992] to 176° (2f19) [Lascombe 1989].

There are structures of two human IgG1 Fcs [Deisenhofer 1981; Corper 1997], a rabbit Fc [Sutton & Phillips 1983], and a rat Fc [Burmeister 1994], and the Fc region has been visualised in the structure of intact IgG antibodies [Harris 1992; Harris 1997; Harris 1998]. The C_{H3} domains are closely associated to each other, burying about 2000 \AA^2 of surface area, while the C_{H2} domains are further apart with N-linked carbohydrates at the interface. The crystal structures of complexes of Fc with various Fc receptors have been solved [Burmeister 1994; Garman 2000; Sondermann 2000].

The solution of the structure of full-length antibodies allowed the visualisation of the hinge regions. The hinge is made up of three parts: a flexible upper part that permits movement of the Fab arms and determines their separation, a firm central spacer, and a flexible base allowing Fc to rotate. The relative sizes of the three regions vary between antibodies of different class, with human IgD possessing only one central residue, while IgG3 having 40. The IgG3 hinge, not easily accessible to crystallography due to its inherent flexibility, has been modelled as a stiff rod, 140 \AA long, with the structure of a polyproline double helix [Marquart 1980]. The variability in the amino acid sequences of the hinge regions of different antibody classes allows binding to a wide repertoire of differently spaced epitopes.

1.1.3 The Generation of Antigen Binding Diversity

It is estimated that there are 10^{10} antibody molecules with different antigen specificities in a single mouse [Golub 1987]. The diversity is generated by the mechanisms of V(D)J recombination [Kurosawa & Tonegawa 1982], and somatic hypermutation [Tonegawa 1977]. There are three classes of genes involved in creating specificity – variable (V), diversity (D) (HC only) and joining (J). V(D)J recombination is illustrated in Figure 1.6 and the generation of diversity is quantified in Table 1.2.

The V genes make up CDRs 1 and 2, whereas CDR3 is composed of the V, D and J genes. The murine germline κ chain locus comprises at least 320 V_{κ} genes, classified into 18 families, and four J genes [Thiebe 1999]. The murine H chain locus has been less well studied, but so far 15 V gene families, four J genes and 13 D genes have been identified [Igbblast 2000]. The genes are combined by recombination during maturation to produce a single HC or LC. The process of allelic exclusion ensures that B-cells produce only one set of functional receptor gene rearrangements, as successful rearrangement at one allele causes DNA recombination to cease elsewhere [Davie 1971].

V(D)J recombination is the only site specific recombination reaction known in vertebrates, and leads to the assembly of the individual exons. There are two sets of conserved, symmetrical, AT rich, recognition signal sequences that flank the joining sites and interact with RAG recombinase [Lewis 1994]. Orientation of the recombination sequences with respect to each other determines whether the rearrangement is deletional or inversional. During joining the loss and addition of one to 10 nucleotides, called N nucleotides, at the coding junctions adds further diversity. The effector functions of antibodies can be modulated by class switching, leading to the different antibody classes (see Table 1.1). Switch regions are highly repetitive sequences found directly upstream of C_H genes that act as signals for the rearranged V domains to associate with C_H genes.

Somatic hypermutation in the germinal centres is a further mechanism for generating diversity [Klein & Zachau 1995]. The mutations are concentrated at hotspots around the CDRs, in particular CDR H1 [Chang & Casali 1994]. Almost exclusively point mutations occur, only a small proportion lead to an increase in affinity for antigen and undergo positive selection.

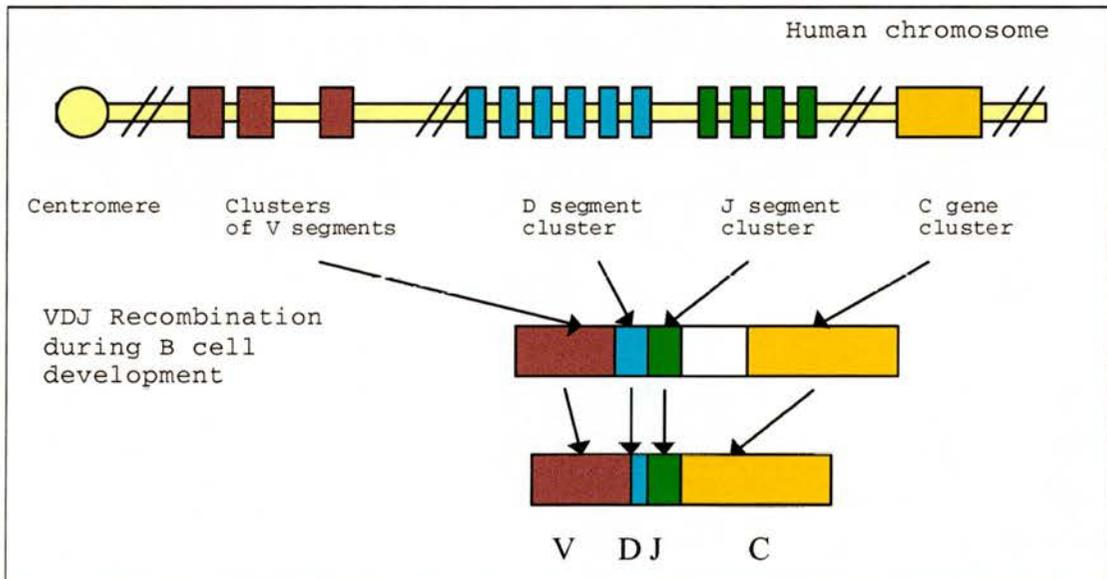


Figure 1.6

Generation of antibody HC diversity

	H	L	
		κ	λ
V gene segments	~100	~70	25
D gene segments	~4	-	-
J gene segments	<u>6</u>	<u>5</u>	<u>8</u>
Random combinatorial joining	<u>2400</u>	<u>350</u>	<u>200</u>
Combinatorial heterodimers		8.4×10^5	4.8×10^5
Junctional diversity, somatic mutation		$\sim 9 \times 10^9$	$\sim 5 \times 10^9$

Table 1.2

Calculation of human V gene diversity.

1.1.4 Antibody-Antigen Structure

Antibodies bind antigen with high affinity (10^{-8} to 10^{-10} M) by using multiple non-covalent interactions between sterically and electrostatically complementary surfaces [Davies & Padlan 1990]. There are many crystal structures of antibody-antigen complexes, the combined antigens are commonly peptides or haptens [Wilson & Stanfield 1994]. Fewer structures of antibodies in complex with protein antigens have been solved, of which lysozyme complexes are the most numerous [Davies & Cohen 1996]. The structure of a Fab fragment bound to lysozyme is shown in Figure 1.7. The overall surface features of the antibody are dependent on the size of the bound antigen. The database of ScFv and Fab structures has been analysed in order to classify antigen-contacting residues and combining site shapes [MacCallum 1996]. Surfaces can be clustered into three topographical classes: concave (mostly hapten binders), ridged (mostly peptide binders), and planar (mostly protein binders).

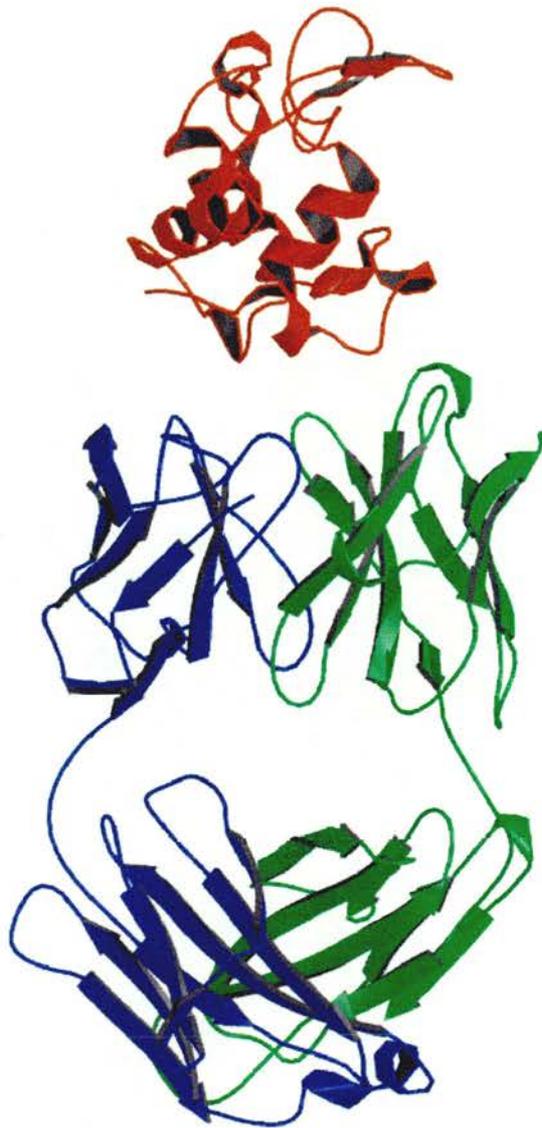


Figure 1.7.

The structure of the Fab fragment D1.3 in complex with lysozyme, lysozyme is at the top of the picture in red.

The interaction of three hypervariable loops dominate the overall binding of antibodies to antigens, L3 (21%), H2 (23%), H3 (29%) compared with L1 (9%), L2 (4%), H1 (10%) and the framework (4%) [Wilson & Stanfield 1994]. Study of antigen-contacting residues in antibody-antigen complex structures has led to the proposition of new definitions of CDR residues [MacCallum 1996]. The N-terminus of L1 and the C-terminus of H2 are shorter than those identified by sequence comparisons. CDRs defined by contact analysis are compared with those derived from sequence analysis (Kabat) and from the structures of uncomplexed antibodies (Chothia) in Table 1.3. Residues showing high sequence variability have been termed specificity determining residues (SDRs) and have been shown to be involved in interaction between antibodies and protein antigens [Padlan 1995]. These variability values depend on the source on the antibody (mouse/human) and the length of the CDR.

CDR	Kabat	Chothia	Contact
L1	L24 - L34	L24 - L34	L30 - L36
L2	L50 - L56	L50 - L56	L46 - L55
L3	L89 - L97	L89 - L97	L89 - L96
H1	H31 - H36	H26 - H32/34	H30 - H36
H2	H50 - H65	H52 - H56	H47 - H58
H3	H95 - H102	H95 - H102	H93 - H101

Table 1.3

A comparison of the different definitions of CDRs, using Kabat numbering.

Study of the amino acids found in CDRs show that aromatic residues and asparagines are common [Padlan 1990]. The side chains of tryptophan, tyrosine and phenylalanine tend to be exposed to the solvent, compared to their more frequent location in protein hydrophobic cores. The preference for large aromatics can be explained by their extensive surface area for hydrophobic interactions, their large polarizability for van der Waals interactions and their ability to hydrogen bond through side chain polar atoms. In addition, their relative rigidity means there is little loss of conformational entropy on complexation. Asparagines have been shown to play an important role in maintaining the structure of the interface through hydrogen bonding to framework residues.

Crystal structures of complexed and uncomplexed antibodies show there can be a conformational change on complexation [Wilson & Stanfield 1994]. This is similar to the induced fit binding of enzymes to substrates, suggesting that the combining site is not rigid. Induced fit has been observed to arise from small movements in side chains [Davies & Padlan 1990], movements of loops [Rini 1992] and changes in the relative position of the variable domains [Bhat 1990]. More dynamic structural information can be provided by high resolution NMR studies. For example, the structure of a trisaccharide-antibody complex has been solved by NMR and crystallography [Bundle 1994]. The NMR structures show two trisaccharide conformations, one of which was observed in the crystal structure.

The complex of the antibody D1.3 with lysozyme has been solved to 1.8 Å [Braden 1996] and studied extensively [Braden 1998]. It buries 1234 Å² of interaction area and contains 17 hydrogen bonds, a large number of van der Waals interactions and about 50 water molecules bound in and around the interface. On forming the complex there is a small displacement of the V_H and V_L domains relative to their position in the free F_V, and a marked decrease in the mobility of H3, but no major rearrangement of side chains. Calorimetry studies have shown the reaction to be enthalpy driven. The negative entropy is due to conformational stabilization in the antibody and the presence of solvent networks bridging the antibody-antigen

interface. Alanine scanning mutagenesis has shown that the energetics of binding are dominated by only a few, juxtaposed, residues on each side of the interface. Mutations leading to the removal of specific hydrogen bonds and van der Waals interactions in the interface show that the complex is structurally and thermodynamically tolerant to the loss of contacts, mainly due to compensatory changes in the solvent structure. It was proposed that this tolerance of mutations may be a general mechanism by which antibodies can maintain recognition during mutation of foreign antigens.

Some antibodies have been identified that are polyreactive and have been called natural autoantibodies (NAAs) [Logtenberg 1990]. They have the ability to bind, albeit with low affinity (10^{-4} to 10^{-7} M), to more than one antigen. Such polyreactive antibodies are usually immature, of IgM class and still in their germline sequence. Their properties illustrate the inherently sticky nature of the CDRs. NAA may have a role against common environmental pathogens in early life [Huang 1992] and there is evidence that anti-Ig autoantibodies may be involved in immune complex clearance [Terness & Opelz 1998].

The crystal structure of a germline Fab fragment and its complex with a hapten has been solved, and the structures compared to the structures of the affinity matured antibody. [Wedemayer 1997]. Maturation leads to a 30 000 times higher affinity for hapten, and is a result of nine somatic mutations that occur up to 15 Å from the bound hapten. The germline antibody binds to the hapten by an induced fit mechanism, whereas the mature antibody binds by lock-and-key, suggesting that polyreactivity is achieved by adopting more than one combining site configuration.

1.1.5 Epitopes

The antibody combining site is known as the paratope and the part of the antigen which is in contact with the paratope is called the epitope [Jerne 1960]. Crystal

structures of antigen complexes have showed that epitopes recognised by antibodies are discontinuous or conformational [Wilson & Stanfield 1994] whereas TCRs recognises continuous regions of structure [Garcia 1996]. B-cell epitopes are defined by the footprint of a single antibody on a fully folded molecule. Thus an antigenic protein contains as many epitopes as antibodies that can be raised against it. Methods of B-cell epitope mapping that disrupt the native structure of the antigenic molecule should therefore be treated with caution [Laver 1990]. Crystal structure complexes have shown that between 15 and 22 amino acids residues from distinct regions of the polypeptide chain are within close contact with the antibody. These residues are brought together in the folded protein and are called a structural epitope [Davies & Cohen 1996]. Thermodynamic studies using antigen mutants have showed that only three to five of these residues contribute significantly to the binding energy and form a so-called functional epitope [Benjamin & Perdue 1996].

Conformational epitopes have the potential to cover the whole surface of a protein in an overlapping continuum [Benjamin 1984]. Multiple structures of lysozyme-antibody complexes have shown that the entire lysozyme surface is capable of binding antibodies [Bentley 1996]. There is also evidence that, like antibodies, antigenic proteins are not static partners in the interaction. The structure of two complexes antibodies with neuraminidase have been solved, in which the antibodies are closely overlapping [Malby 1994]. The study demonstrated that amino acid residues on the antigen may be in quite different chemical environments, and may also adopt different conformations in complexes with overlapping antibodies.

Immunodominant epitopes are defined as those to which the immune response is directed most frequently [Atassi 1984]. Of the many thousands of peptides encoded by a antigen, only a small fraction are presented as T-cell epitopes, predominantly due to the restrictions of antigen processing [Yewdell & Bennick 1999]. Clusters of B-cell epitopes that define immunodominant domains, have been shown to exist by epitope mapping [Mackay & Gershwin 1989], although their existence has been disputed [Benjamin 1995]. However, due to the time-consuming nature and

technical difficulties of obtaining multiple crystal structures of antibodies in complex with each of the epitopes in such a cluster, the domains have not yet been defined in detail. Immunodominant domains on B-cell antigens could be a consequence of a structural feature of the antigen, or may result from the mechanisms of immune regulation [Benjamin 1984]. For example, due to the need of most B-cells for T-cell activation, restricted antigen presentation may influence the B-cell repertoire [Parker 1993].

1.2 Autoimmunity

1.2.1 Autoimmune Disease

Autoimmunity affects 5 to 7 % of the population and leads to a diverse range of diseases such as insulin-dependent diabetes mellitus (IDDM), multiple sclerosis (MS), rheumatoid arthritis (RA) and autoimmune thyroid disease (AITD) [Ollier & Symmons 1992]. Autoimmune diseases can be broadly classified as organ-specific or multisystem. The organ-specific diseases are characterised by the presence of autoantibodies directed to molecules from a single organ, whereas systemic diseases involve autoantibodies to a whole range of nuclear and cytoplasmic targets. Classification can also occur based on the effector mechanism, either T-cell mediated, for example IDDM and RA, or B-cell driven, for example Graves' disease.

Autoimmunity arises when there is a breakdown in the mechanisms of tolerance. The causes of this failure are as varied as their symptoms, involving the combination of genetic predispositions and environmental factors. Most autoimmune diseases are liable to spontaneously exacerbate and remiss, indicating a delicate balance between opposing control mechanisms.

Autoimmune diseases are usually treated by manipulation of the immune system [Roitt 1994]. For example, anti-inflammatory drugs are employed to treat patients with RA and systemic lupus erythematosus (SLE). Immunosuppressive drugs have been successfully used with many autoimmune diseases, for example Crohn's disease and myasthenia gravis. For many organ specific diseases, metabolic control can be used. For example, insulin is used to treat IDDM, and vitamin B₁₂ to treat pernicious anemia.

1.2.2 Mechanisms of Tolerance

Autoimmunity is normally prevented by many degenerate mechanisms of tolerance [Sinha 1990]. Burnet's clonal selection theory [Burnet 1959] proposed that T-cell clones with unique receptors exist before antigen exposure and that only cells with specific receptors are selected by antigen for activation. Harmful T-cells recognising self-MHC bound to self-peptides are removed early on by negative selection [Nossal 1994], while T-cells with some affinity for self MHC are cloned by positive selection [von Boehmer 1994]. Such mechanisms, known as central tolerance, are important in the thymus for removing harmful and useless rearrangements of TCR genes and selecting for the useful ones. Similar mechanisms occur in the germinal centers of the bone marrow, the site of Ig production of the primary and secondary response.

Central tolerance only involves the subset of self-molecules present in the thymus or bone marrow during early development. Therefore, additional mechanisms of peripheral tolerance are needed. There are numerous mechanisms for generating tolerance in peripheral T-cells. Self reactive T-cells may become non-responsive, a state called anergy, in the absence of co-stimulation, they may be apoptosed by persistent self-antigen stimulation, and they may be suppressed by suppressor cells or a combination of cytokines [Zuniga-Pflucker 1995]. Lymphocyte receptors are capable of delivering negative signals, for example, activated T-cells express a

receptor called CTLA-4 that recognises B7 molecules on the APC and inhibits signalling [Lee 1998]. Peripheral B-cell tolerance also involves clonal deletion, anergy and immunosuppression. In addition, the absence of T-cell help may block Ig secretion [Hodgkin 1991], and regulation by idiotype-antidiotypic networks may occur [Coutinho 1989]. These mechanisms are summarised in Figure 1.8.

1.2.3 Breakdown of Tolerance

Autoimmunity occurs only in the case of simultaneous disruption of the multiple mechanisms of self-tolerance. The redundancy explains why autoimmune diseases may follow a slowly progressive, transient or relapsing course [Ollier & Symmons 1992]. This implies that several non-mutually exclusive etiological mechanisms are likely to be responsible for autoimmune disease. These can be broadly categorised as either defects in tolerance induction or conventional immune responses against self-molecules to which, under normal circumstances, tolerance need not be established [Sinha 1990]. The presence of multiple genetic loci predisposing to autoimmunity suggests that each locus may perform a beneficial role and have been conserved through evolution [Schwartz & Cohen 2000]. Pathogenic effects may result from fortuitous grouping of such alleles.

1. Weaknesses in central or peripheral tolerance

Some individuals have inherited weaknesses in tolerance mechanisms. For example, susceptibility to murine and human IDDM correlates strongly with MHC class II I-A or HLA DQ alleles that lack an aspartate at position $\beta 7$ [Todd 1987]. The crystal structure of a mouse MHC I-A lacking this aspartate was solved in complex with a high affinity peptide from the autoantigen glutamic acid decarboxylase [Corper 2000]. The peptide binding groove has features that allow high promiscuity, it is wide and has an oxyanion hole due to the loss of the aspartate.

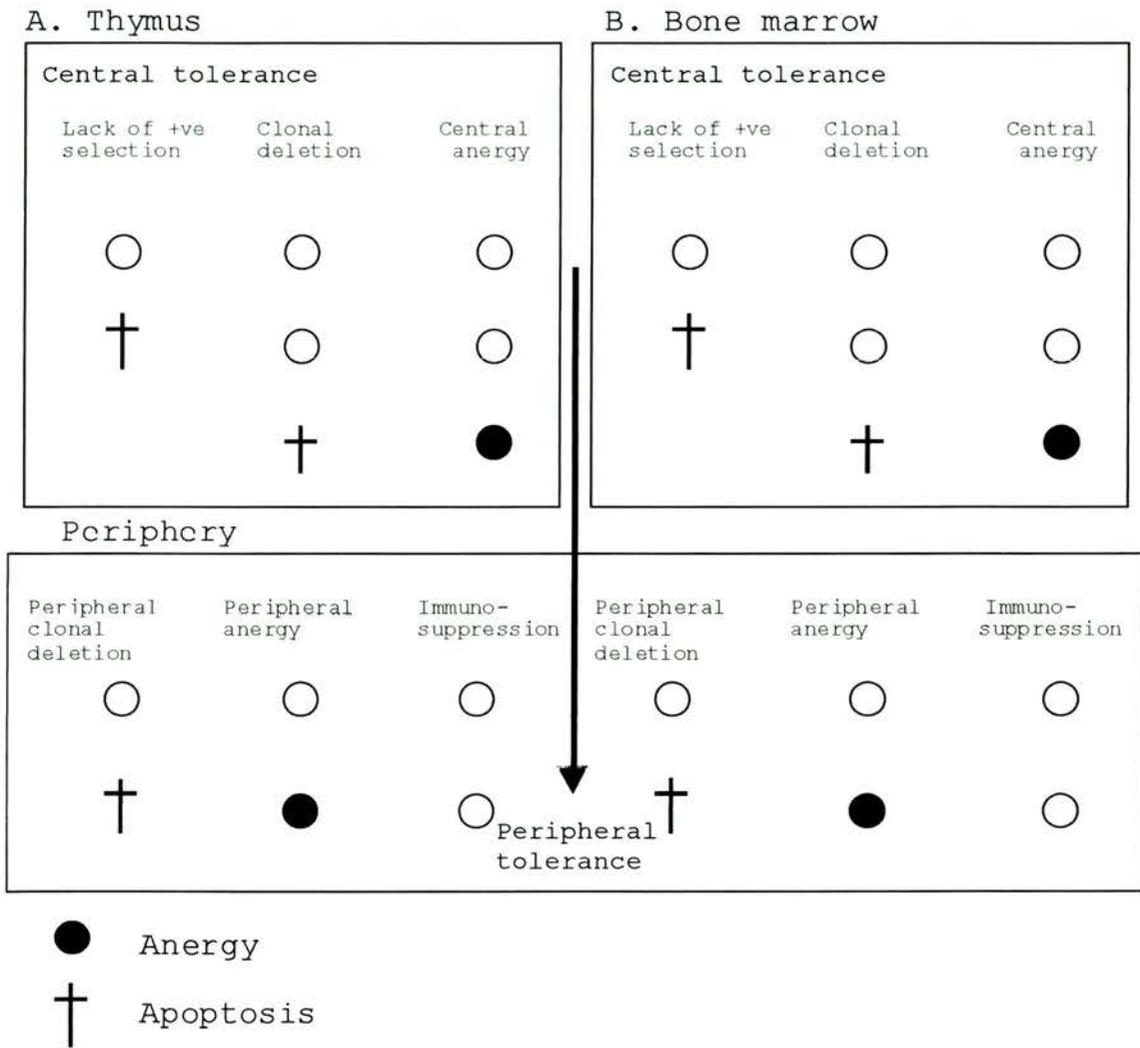


Figure 1.8

Summary of tolerance mechanisms of A) T-cells and B) B-cells.

2. Release of anatomically sequestered antigens

Antigens associated with peripheral tissue may not come into contact with the developing T- or B-cell repertoire and therefore, tolerance may be unnecessary. Induction of autoimmune disease following contact with antigen of such so-called 'immunologically privileged' sites is exemplified by the development of sympathetic ophthalmia following eye injury [Chan & Mochizuki 1999].

3. Infectious agents

Infection may act as a trigger of autoimmunity in predisposed individuals. Infectious agents have been shown to participate in the induction of autoimmunity by tissue damage and immunological activation [Tomer & Davies 1993]. For example, some individuals have developed type I diabetes following congenital rubella infection, and this was thought to be due to the induction of immunological abnormalities by the virus [Rabinowe 1986]. There is evidence for molecular mimicry as an explanation for multiple sclerosis [Wucherpfennig & Strominger 1995], with T-cell clones to myelin basic protein activated by viral and bacterial peptides. It has also been shown that infectious agents can have a protective effect on Th1 cell mediated diseases [Bach 1994].

4. Cryptic self and modified self hypotheses

Some self-peptides fail to get processed and presented by the MHC, although TCRs specific to such peptides are readily found in the normal repertoire [Theofilopoulos 1995a]. These T-cells exist in a state of ignorance and the peptides are called cryptic. The presentation failure may be due to abnormality in the APC, the antigen, or the processing events themselves. Alterations in self-antigen presentation and priming with certain self peptides or foreign molecular mimics leads to the display of previously cryptic self-determinants which may trigger autoimmune disease. For example, murine APCs cannot present the full length cytochrome c molecule, whereas immunization with the C-terminal peptide elicits a strong T-cell response [Mamula 1993].

5. Immunoregulatory disturbances

There is evidence to suggest that autoimmune diseases are associated with immune dysregulation as various immunological manipulations lead to spontaneous disease [Wen 1994]. Overexpression of cytokines following an infection can lead to loss of self-tolerance. For example, the expression of IFN- γ in the islet cells of transgenic mice leads to the development of type I diabetes [Sarvetnick 1990] possibly by activating quiescent autoreactive T-cells. The reciprocal counter-regulatory function of Th1 and Th2 cells has led to the hypothesis that their imbalance leads to disease [Charlton & Lafferty 1995]. Th1 cell mediated disease is inhibited by Th1 inhibitors and accelerated by Th1 stimulators or Th2 inhibitors.

1.2.4 Autoantibodies

In 1956, Witebsky and Rose induced AITD in rabbits by injecting thyroglobulin in the presence of Freud's adjuvant [Rose & Witebsky 1956]. This experiment was the first to show that it was possible to produce autoantibodies against a non-sequestered autoantigen. Autoimmune diseases are characterised by the presence of autoantibodies [Salvi 1988] which play an important diagnostic role, with the level of autoantibody often corresponding to the disease severity [Druet 1992]. Some autoantibodies, for example those specific for thyroid stimulating hormone receptor (TSHR), initiate disease [Rees Smith 1988]. In other cases autoantibodies may be an epiphenomenon in which they are the result, rather than the cause, of the initial condition [Karjalainen 1986].

Autoreactive B-cells producing non-pathogenic NAAs constitute a substantial part of the B-cell repertoire [Dighiero 1986]. It has been proposed that uncontrolled stimulation of NAAs by cytokines leads to the production of pathogenic autoantibodies [Matechak 1996]. For example, an autoantibody found in patients with SLE has a germline V_H sequence [Sanz 1989]. However, the isolation of many somatically mutated autoantibodies, for example, in pathogenic anti-DNA

autoantibodies [Marion 1992] suggests antigen driven selection is normally involved.

The crystal structure of an antibody-autoantigen complex of a rheumatoid factor (RF) Fab bound to its low affinity autoantigen IgG Fc [Corper 1997; Sutton 1998], has been solved. The antibody binding residues are located at the edge of the antibody combining site and involve residues that have undergone somatic mutation. This indicates that the antibody may be crossreactive and bind a different antigen with the conventional combining site. The presence of somatic mutations suggests that antigen-driven selection has taken place.

1.3 Autoimmune Thyroid Disease

1.3.1 Clinical Aspects

Clinical descriptions of AITD date to the 1780s [Volpe 1988] when the disorder known as Graves' disease was first reported. In areas of the world where dietary iodine intake is adequate, autoimmune mechanisms are the commonest cause of thyroid disorders, causing disease in 70-80% of patients [McGregor, 1992]. AITD affects approximately 1% of the population [Weetman 1992] and symptoms vary considerably. The disease manifests itself clinically as either thyroid overactivity (hyperthyroidism) with or without changes in the eyes (ophthalmopathy) or thyroid underactivity/failure (hypothyroidism) [Braverman & Utiger 1996]. In both situations the thyroid may be enlarged to form a goitre. Hyperthyroidism with diffuse enlargement of the thyroid and ophthalmopathy is called Graves' disease, whereas thyroid failure with lymphocytic infiltration is called goitrous (Hashimoto's) thyroiditis or atrophic thyroiditis (primary myxedema) if the thyroid is not palpable. Between 5-9% of women develop postpartum thyroid dysfunction,

transient hyperthyroidism or hypothyroidism, in the year after delivery [Gernstein 1990].

AITD has long been a model system for the understanding of autoimmune diseases. This is due to the prevalence of the disease, the easy access to the thyroid gland, well-established animal models and historical precedence. AITD is clinically benign and there are a range of safe and effective therapeutics available, directed towards reducing or replacing thyroid hormone. In general hypothyroidism is treated with thyroid hormones and hyperthyroidism with inhibitors of thyroid hormone synthesis, radioactive iodine or surgery. However, there is a debate about whether thyroid drugs have direct immunosuppressive actions or if their actions are a consequence of the normalisation of thyroid function [Paschke 1995]. It hoped that study of AITD will be applicable to more serious and less easy to treat autoimmune diseases.

1.3.2 Predisposition

The likelihood of developing AITD is determined by a combination of genetic, constitutional and environment factors with the relative influence of each factor varying between diseases and individuals [Weetman 1992].

1) Genetic Factors

The most clearly linked gene to AITD is the MHC, or HLA in man, particularly class II molecules [Weetman 1992]. This association is shared with many other autoimmune diseases, although presence of the loci does not inevitably lead to disease [Theofilopoulos 1995b]. Population based studies have linked primary myxoedema and Hashimoto's thyroiditis with loci on HLA-DR molecules, particularly 3, 4 and 5 but the linkages are weak and inconsistent [Farid 1981; Bogner 1992; Boehm 1993].

The concordance rate for developing Graves' disease in monozygotic twins is 50% whereas in HLA-identical siblings it is 7-16%, suggesting a role for non-HLA genes in susceptibility [Stenszky 1985]. However, search for such genes has so far produced inconclusive results. Experiments show that anti-thyroid peroxidase (TPO) autoantibodies have restricted heavy and light chain combinations [Chazenbalk 1993b], suggesting that Ig genes are important. Segregation analysis has shown that the inheritance of thyroid autoantibodies is a Mendelian dominant trait in women with reduced penetrance in men [Phillips 1990; Phillips 1991]. The high prevalence of autoimmune hypothyroidism in Down's syndrome and in kindreds with familial Alzheimer's disease suggests the presence of an unidentified susceptibility gene on chromosome 21 [Ewins 1991]. Study of animals with spontaneously occurring experimental autoimmune thyroiditis (EAT) suggests that genes involved in T-cell regulation and target organ responsiveness are important [Weetman 1991], but it is not clear how directly this can be related to humans.

2) Non-genetic Factors

Susceptibility increases with age, due to longer environmental exposure and changes in immunoregulation. As with all autoimmune diseases, AITD is more frequent in females and it has been shown that oestrogen exacerbates EAT whereas testosterone ameliorates it [Weetman 1991]. 18% of healthy female blood donors have anti-TPO antibodies which rises to 30% for the age group of 55-64 years, compared to 10% of males [Prentice 1990]. Remission of AITD symptoms is seen during the last trimester of pregnancy, however hormone fluctuations following birth can cause post-partum thyroiditis.

Diet can also play a role, in EAT it has been shown that a high iodine diet exacerbates thyroiditis whereas a low iodine diet causes remission [Cohen & Weetman 1988]. There appears to be a link between fetal nutritional or hormonal factors, reflected in birth weight, and the presence of thyroglobulin (Tg) and TPO antibodies [Phillips 1993a]. Stressful life events may precipitate Graves' disease [Radosavljevic 1996], probably due to repression of the immune system.

1.3.3 Pathogenesis

The identification and characterization of autoantigens are important in the understanding of the pathogenesis of autoimmune diseases [Furmaniak 1999]. Not only does the identification of autoantigens confirm the autoimmune nature of the disease, but it also allows the study of lymphocyte responses, autoantibody production and epidemiology. Autoantibodies to Tg were first reported in 1956 [Roitt 1956], and to a distinct microsomal antigen later shown to be TPO in 1959 [Belyavin & Trotter 1959]. Long-acting thyroid stimulator (LATS) were found in 1956 [Roitt 1956] and shown to bind the thyroid stimulating hormone receptor [Rees Smith & Hall 1974]. TPO, Tg and TSHR molecules have all been cloned and large amounts of pure protein are essential to further study. There is evidence that six other molecules may be autoantigens, a second colloid autoantigen [Balfour 1961], ATRA-1 [Hirayu 1987], a 70 kDa protein [Chan 1989], a 64 kDa protein [Dong 1991] and the Na⁺/I⁻ symporter [Spitzweg 1997].

Multiple mechanisms can lead to the same overall clinical manifestations of AITD, making it unrealistic to search for a single etiological explanation [Weetman & McGregor 1994]. AITD is likely to be both cell-mediated and antibody mediated, with the relative contribution from each arm of the immune system varying between diseases. Pathogenic mechanisms include ADCC, complement attack, stimulating or blocking autoantibodies, cytotoxic cell killing and immune complex formation. Possible initiation and amplification mechanisms are depicted in Figure 1.9. Mechanisms mediated by autoantibodies are discussed in section 1.3.4.

Thyroid autoreactive T-cells have shown to be present in the peripheral blood and thyroid infiltrate of Hashimoto's and Graves' patients [Weetman & McGregor 1984]. Evidence exists for the presence of cytotoxic T-cells [Wu 1994], however, it appears that the majority of T-cells are CD4⁺ expressing activation markers such as MHC class II molecules [Aichinger 1995]. There is evidence that the T-cell repertoire in AITD is biased, thought to be a reflection of a restricted number of

antigenic epitopes [Martin 1999]. An impairment in T-cell mediated immune suppression has been proposed to play a major role in pathogenesis [Volpe 1993]. In addition, cytokines produced by T-cells may have a pathogenic role. For example, externally administered IFN- γ has been shown to induce AITD by exacerbating preexisting thyroiditis [Frohman 1991]. More recently, evidence has been found for a crucial role of IL-1 β in the induction and perpetuation of tissue damage via suicide Fas/Fas ligand-mediated interaction on the thyrocytes [Paolieri 1999].

Epstein-Barr virus (EBV) infection has been implicated in the development of AITD [Coyle 1989], and a correlation has been shown between *Helicobacter pylori* infection and incidence of AITD [de Luis 1998]. However consistent pathogen identification is lacking due to the time lag between infection and autoimmune symptoms, and the likelihood that disease may be a cumulative effect of many separate infections. Potentially, an infectious agent could cause disease by causing non-specific induction of the immune system, or by acting as a molecular mimic to self-molecules.

Thyrocytes express adhesion molecules such as intercellular cell adhesion molecule which enhance their susceptibility to cell-mediated injury [Weetman 1989b]. It has been proposed that thyrocytes may act as APCs by expressing MHC class II molecules on their surface [Hanafusa 1983]. It is possible that MHC expression could perpetuate the immune response, although it is not yet clear if thyrocytes have the ability to deliver the second signal. Alternatively, class II expression in the absence of costimulation may have the opposite result of inducing peripheral tolerance [Weetman 1994].

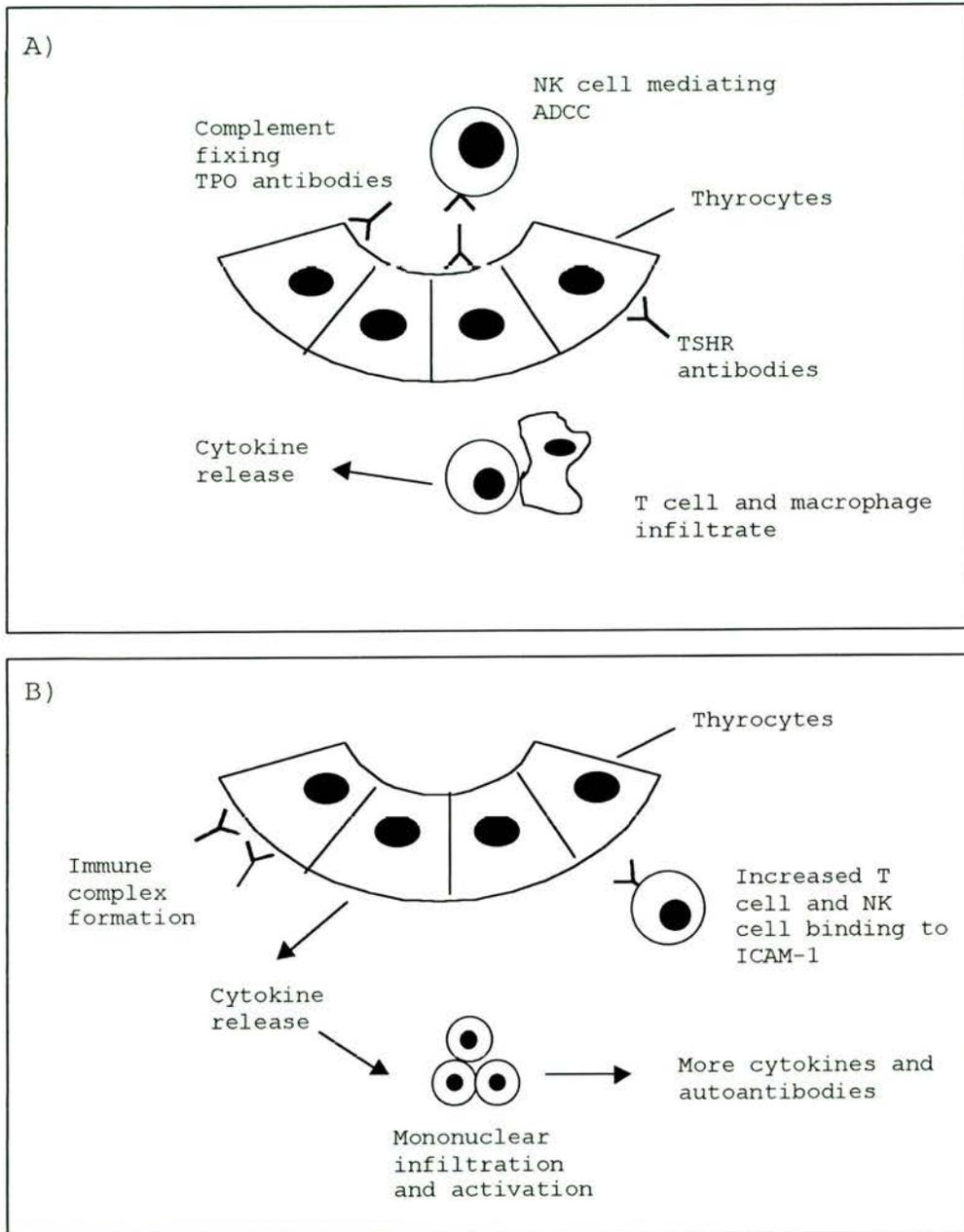


Figure 1.9

Effector mechanism in autoimmune hypothyroidism.

A) Initiation events, B) amplification events.

1.3.4 Autoantibodies

The sera of patients with AITD frequently contain anti-TPO, -Tg, and -TSHR autoantibodies, which may be detected before clinical symptoms appear [Furmaniak 1987]. The autoantibody response to thyroid antigens, and to TPO in particular, is amongst the best characterised of any autoimmune disease. Antibodies to TPO have been historically measured by passive tanned erythrocyte hemagglutination [Mariotti 1990]. However, alternative methods have been developed, including enzyme-linked immunosorbent assays [Schardt 1982] and highly sensitive assays for autoantibodies to TPO and Tg, based on the interaction between the antibodies and ¹²⁵I-labelled antigen and precipitation with solid-phase protein A [Beever 1989].

Autoantibodies to TPO, Tg and TSHR have been studied extensively. The results of the studies of autoantibodies to TPO are covered in section 1.4.6. Two methods have been used, the first, which has many practical limitations, is the production of monoclonal antibodies by heterohybridoma formation of immortalised B-cells [Horimoto 1992]. The second, and more widely used, is the methodology of phage display combinatorial libraries [Hexham 1994; Rapoport 1995]. Random combinations of heavy and light chains are cloned and expressed in bacteriophage and selected by screening with antigen.

The thyroid stimulating hormone receptor (TSHR) is a 120 kDa transmembrane protein and a member of the G-protein coupled receptor family. Almost all patients with Graves' disease and about 10-20% of patients with autoimmune hypothyroidism have antibodies to TSHR [Weetman & McGregor 1994]. Thyroid stimulating autoantibodies to the TSHR elicit hyperthyroidism in Graves' disease by mimicking the action of TSH following binding to the receptor. Conversely, blocking antibodies to the TSHR may result in atrophic thyroiditis, which is often accompanied by hypothyroidism. The sites of interaction of TSH and TSHR

autoantibodies have not been identified, but there is considerable evidence that they bind to the same region of the TSHR [Rapoport 1998].

Thyroglobulin (Tg) is a water-soluble dimer of 2 x 330 kDa that is iodinated as an intermediate in production of thyroid hormone. Anti-Tg autoantibodies have been found in up to 70% of patients with Graves' disease and 95-100% of patients with Hashimoto's thyroiditis, and with a background level of 18% in healthy females [Furmaniak 1999]. The epitopes are restricted in number, to two or three, and are species specific [Chan 1987; Bresler 1990]. Studies with human monoclonal antibodies have indicated that the epitopes recognised by patients are different from those in disease-free individuals [Prentice 1995]. Antibodies to Tg do not fix complement, probably due to the wide spaced epitopes [Adler 1984], however the potential to mediate ADCC has been demonstrated *in vivo* [Weetman 1989a].

ADCC is mediated by NK cells via their interaction with antibody bound Fc receptors. The mechanism is thought to be important in Hashimoto's disease [Bogner 1984], probably through antibody binding to TPO, although a novel antigen may be involved [Bogner 1990]. More recently, cytotoxic antibody activity was shown to be 2.5 times greater in patients with AITD compared with controls [Bogner 1995]. However, it has been reported that soluble complement receptor 1 had no effect on EAT, implying that complement does not have a role in the initial phase of tissue injury [Metcalf 1996].

1.4 Thyroid Peroxidase

1.4.1 Peroxidase Family

Peroxidases catalyze the oxidation of a variety of molecules using hydrogen peroxide [Poulos 1993]. The classic peroxidase reaction cycle involves three steps. First, hydrogen peroxide reacts with the enzyme in a two electron reaction to give

compound I, then the substrate reduces compound I to compound II. In the final step, compound II is reduced back to the resting state and the product is released. Peroxidases can be classified into the superfamilies plant, fungal, bacterial and animal, based on the structure of representatives from each superfamily [Taurog 1999]. The most well studied peroxidases are small monomers with non-covalently bound heme groups. The paradigm is cytochrome c peroxidase and a very high resolution crystal structure of the yeast enzyme has been available for some time [Finzel 1984].

TPO is a member of the animal family of peroxidases that includes the mammalian peroxidases and several invertebrate peroxidases [Taurog & Wall 1998]. All animal peroxidases show a high degree of sequence identity to each other, and have key catalytic residues conserved. The mammalian peroxidases so far identified are myeloperoxidase (MPO), lactoperoxidase (LPO), eosinophil peroxidase (EPO) and salivary peroxidase (SPO). All are large proteins (> 700 amino acids long) with a covalently bound heme. The invertebrate peroxidases include peroxinectin from crayfish blood [Johansson 1995], peroxidases in the squid light organ [Tomarev 1993], peroxidasin, an enzyme-matrix protein of *Drosophila* development [Nelson 1994], and a potential peroxidase on chromosome III of the fully sequenced *C. elegans* genome [Wilson 1994]. More recently, TPO homologues have been cloned from the endostyle organ of sedentary marine invertebrates such as *Ciona intestinalis* and *Halocynthia roretzi* [Ogasawara 1999]. The endostyle secretes mucus proteins for internal filter feeding and is thought to be homologous to the thyroid gland in higher vertebrates. Endostyle homologues show an overall identity of 41% to human TPO, which extends over all the domains present in the human enzyme.

There is evidence that the precursor of the animal peroxidase family may have been an adhesion molecule [Taurog 1999]. Peroxinectin combines adhesion and peroxidase functionalities [Johansson 1995] and it has been shown that MPO mediates cell adhesion via the alpha M beta 2 integrin [Johansson 1997].

MPO catalyses the oxidation of Cl^- to HOCl in neutrophils, which then acts as an antibacterial agent. The structure of the canine and human enzymes have been solved [Zeng & Fenna 1992] [Davey & Fenna 1996], and more recently the human structure extended to 1.8 Å [Fiedler 2000]. MPO is a dimer of a large (64 kDa) and a small (14 kDa) chain and contains 19 α -helices, and the structure is shown in Figure 1.10. Each monomer contains a covalently attached heme (protoporphyrin IX), and a structural calcium ion. The N-terminal 166 amino acids are thought to play a role in processing and sorting of MPO and are cleaved by posttranslational processing [Andersson 1998]. Although there is no sequence or fold similarity between MPO and cytochrome c peroxidase, the active sites of both molecules can be superimposed and shown to be very similar, thought to be an indication of convergent evolution to a common enzyme mechanism [Taurog 1999]. Further, the structure of MPO in complex with a bisubstrate analogue inhibitor which binds to H95 and R339 has been solved [Davey & Fenna 1996], and suggests that the mechanisms of compound I formation are similar in MPO and cytochrome c peroxidase.

1.4.2 Molecular Biology

cDNAs encoding TPO have been isolated from man, [Kimura 1987; Libert 1987a; Magnusson 1987; Seto 1987] pig [Magnusson 1986], rat [Derwahl 1989] and mouse [Kotani 1993] and all show a high degree of sequence similarity. Human TPO cDNA consists of 2799 nucleotides and encodes 933 amino acids. Somatic cell hybrid mapping and fluorescence *in situ* hybridization show that the TPO gene is located on the short arm of chromosome 2 [Endo 1995]. The gene contains 17 exons and 16 introns and covers 150 Kbp, a TATA box is found 25 bp upstream of the start site. Comparison with the human MPO gene [Morishita 1987], shows that the position of the 3rd to 11th exon-intron junctions in MPO coincide exactly with the 2nd to 11th junctions in TPO, except that the 7th junction is missing in MPO

[Kimura 1989]. The first 735 amino acids of the TPO sequence are 42% identical to human MPO, indicating that they share a common ancestor [Kimura & Ikeda-Saito 1988]. However, the MPO gene is 10 kb long, that is less than one tenth of the length of the TPO gene. In TPO, exon 13 encodes a complement control protein (CCP) homologue, exon 14 an epidermal growth factor (EGF) homologue and the 15th exon a transmembrane (TM) domain. There is also homology to the heme-binding region of cytochrome c oxidase subunit I in the middle of TPO which indicates that the gene may include some mitochondrial genetic material [Libert 1987b].

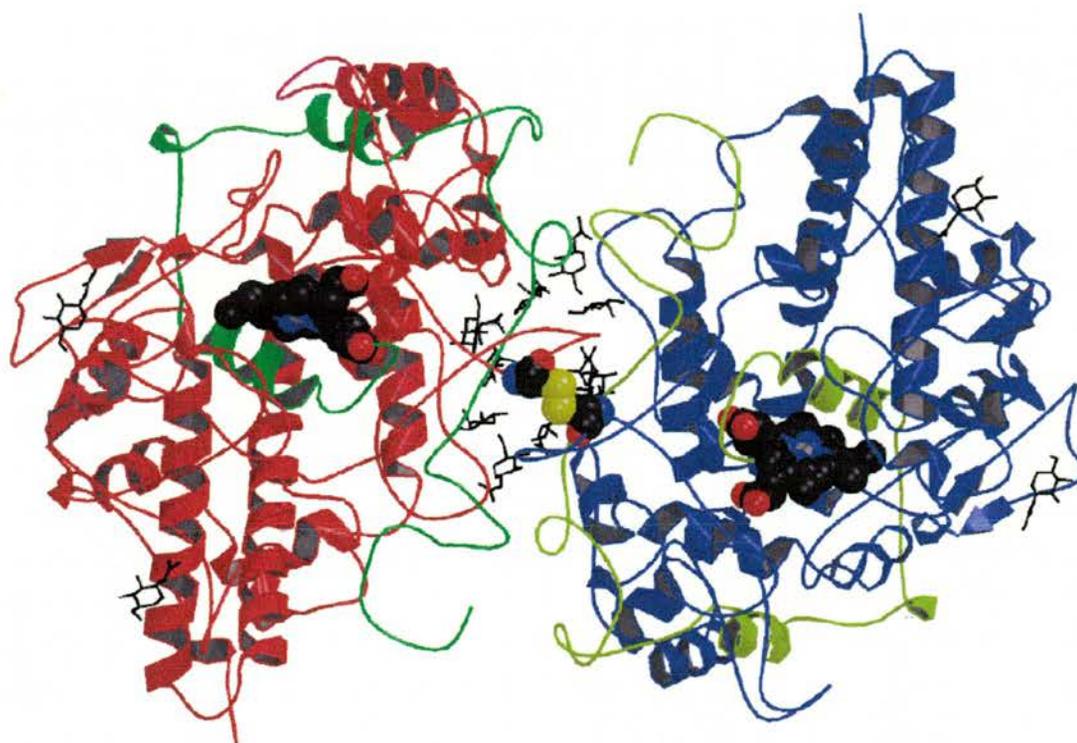


Figure 1.10

The structure of the MPO dimer. Light chains (green, yellow), heavy chains (red, blue). Carbohydrate residues are shown clustering at the dimer interface, the heme moiety is in space filling representation at the centre of each monomer, and the interchain disulphide bridge cysteines are shown in space filling representation at the centre of the dimer.

A second cDNA, TPO-2, has been identified which lacks 171 base pairs in the middle of the sequence corresponding to the 10th exon [Kimura 1987]. Donor and acceptor splice sites surround the deleted regions suggesting that alternative splicing is taking place. The resulting 876 amino acid protein is inactive due to lack of heme incorporation [Niccoli 1997] and the missing peptide includes aspartic acid 579, thought to be a key active site residue. The presence of both TPO-1 and TPO-2 has been shown in normal and Graves' disease tissue [Elisei 1991], but due to improper folding TPO2 is trapped in the endoplasmic reticulum and rapidly degraded [Niccoli 1997]. The physiological or pathophysiological significance of the alternative splicing are unknown.

Study of the TPO promoter by DNAase I footprinting and mobility shift assay [Kikkawa 1990; Abramowicz 1992b] have shown the presence of thyroid transcription factor-1, 2 and Pax-8 binding sites in the sequence -170 to +1. TSH stimulation leads to the activation of the transcription factors, causing TPO synthesis. Mutations in the TPO gene result in both quantitative and qualitative abnormalities, and are thought to be the most prevalent causes of inborn errors of thyroid metabolism. For example the insertion of 4 bp in the 8th exon lead to a frameshift that produced a truncated protein with reduced enzyme activity [Abramowicz 1992a].

TPO was originally purified from thyroid microsomes by trypsin digestion and detergent treatment [Hosoya & Morrison 1965; Taurog 1970]. N-terminal sequencing showed this led to trypsin cleavage after residue 109 of porcine [Yokoyama & Taurog 1988] and human [Taurog 1990] TPO. Recombinant TPO has been expressed in Chinese hamster ovary (CHO) cells [Hata 1989; Foti 1990; Kaufman 1990; Kaufman 1991]. The TPO is enzymatically active and recognises TPO autoantibodies, but only small amounts of protein have been produced [Guo 1998b]. Glycosylated TPO has been expressed in yeast, but the levels of protein produced are also very low [Wedlock 1993].

Larger amounts of TPO with autoantibody reactivity have been produced in baculovirus-insect cell system [Haubruck 1993; Kendler 1993; Seto 1993], but the protein was enzymatically inactive. Addition of a heme precursor led to the production of full-length TPO with low activity [Fan 1996] and of the extracellular fragment with activity comparable to the native enzyme [Grennan Jones 1996]. However, two studies have reported that TPO produced in insect cells is insoluble at the high concentrations required for crystallisation, thought to be due to differences in the properties of native and recombinant proteins [Guo 1998b; Gardas 1999]. The distinct properties of many recombinant preparations may be explained by incorrect heme incorporation. Recently it has been shown that the heme plays an important role in intracellular trafficking of TPO and that H₂O₂ generated at the apical membrane is important in autocatalytic covalent heme binding [Fayadat 1999].

1.4.3 Structure

TPO is a membrane bound, glycosylated, heme containing dimer of 2 x 110 kDa. A schematic diagram of the dimer is shown in Figure 1.11. The monomer consists of a single peptide chain, with a short C-terminal cytoplasmic tail, and a large, extracellular N-terminal region that contains the enzymatic and antibody binding activity. The native protein is a disulphide linked dimer [Baker 1994b] that can be cleaved from the membrane using trypsin. The TPO sequence contains an animal peroxidase domain, an EGF domain, and a CCP domain [Kimura 1989]. The peroxidase domain contains the enzyme active site, the CCP and EGF domains are likely to act as physical spacers, and the N-terminal domain, by analogy to MPO, may have a role in trafficking [Andersson 1998]. The EGF, CCP and transmembrane domains are only present in TPO and the ascidian homologues, all the other proteins in the animal peroxidase family are soluble. The presence of a

CCP and an EGF domain separating the functional domain from the membrane is also found in the selectins [Barlow & Campbell 1994]. A secondary structure model of TPO was built before the crystal structure of MPO became available [Banga 1990] and models of the CCP and EGF domains [Estienne 1999a] and the CCP, EGF and peroxidase domains have recently been built [Hobby 2000].

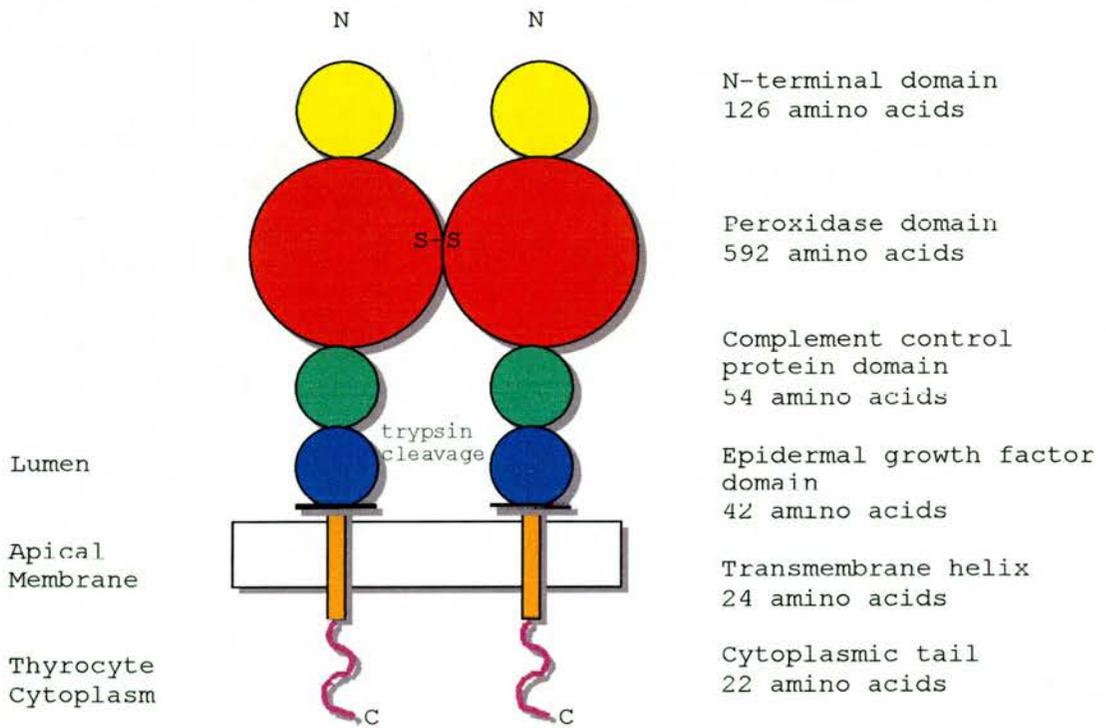


Figure 1.11

A domain model of TPO.

The peroxidase domain contains 592 residues and includes the heme group and the active site. All the key structural residues in MPO are conserved in the sequence of TPO, suggesting the fold is likely to be similar. Out of the 24 residues in MPO that are within 4.5 Å of the heme atom, all but three are in a similar position in the alignment with TPO [Zeng & Fenna 1992]. This indicates that the heme is likely to be bound in the same environment in both structures leading to similar catalytic mechanisms, with the three changes being responsible for the different substrate specificity.

The EGF domain contains about forty amino acids. It is part of a sequence family found originally in epidermal growth factor, and subsequently shown to be present in a large number of membrane bound, extracellular, eukaryotic proteins [Campbell & Bork 1993]. EGF domains can act as spacer units on cell surface proteins and can mediate protein-protein interactions. Many EGF domain containing proteins require calcium for their biological function and a calcium-binding site has been found to be located at the N-terminus. The structure of an EGF domain pair has shown that the domains are fixed in an extended conformation, stabilised by calcium ligation [Downing 1996].

The CCP domain is also known as the Sushi, SCR (short consensus repeat) or c4b-beta 2 glycoprotein domain [Kato & Enjyoji 1991]. It is found in a range of proteins involved in protein-protein interactions at the cell surface, for example key complement components C3b, C4b and Factor H [Klickstein 1988]. CCP domains are often found in multiple copies separated by three to five amino acids. They are thought to ensure an appropriate spatial arrangement of functional modules and to impart a degree of flexibility to the stems of receptors and cell adhesion proteins.

There are four potential N-glycosylation sites in the human TPO sequence (asparagines 129, 307, 342, 569) and native TPO has been estimated to contain 12.6 kDa of carbohydrate [Grennan Jones 1996]. The location and nature of carbohydrates have been determined for porcine TPO, which contains about 10%

carbohydrate [Rawitch 1992]. Four out of the five potential glycosylation sites are occupied and the sugars are of the high mannose type. A study of the nature of the bound carbohydrates on human TPO has been carried out by treating TPO with a cocktail of glycosidases followed by analysis of sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) mobility and autoantibody binding [Kiso 1992]. It was found that deglycosylation has no clear effect on autoantibody binding, and that high mannose type residues were associated with the lower band of the TPO doublet and complex type residues with the upper band. However, there is evidence for glycosylation being necessary for activity, as treatment of human TPO with endo-H leads to inhibition of the enzyme [Giraud 1992].

The folding of TPO in CHO cells has been monitored by antibody binding [Fayadat 1998] and showed that inhibition of N-glycosylation by tunicamycin led to a 95% decrease in enzymatic activity. This implies that proper folding and transportation to the cell surface requires correct glycosylation. The same study reported that the inhibition of O-glycans and complex structures with phenyl- α -GalNAc and deoxymannojirimycin respectively, did not influence trafficking or activity, indicating that only N-linked sugars are involved.

TPO consistently appears as a doublet after immunoprecipitation and western blot analysis [Czarnocka 1985; Katjita 1985; Kotani 1986; Ohtaki 1986a], and it has been suggested that the lower band represents the alternative spliced form. However the same doublet is observed with recombinant TPO generated from a single cloned cDNA that lacks introns [Kaufmann 1989], and monoclonal antibodies that recognise the spliced residues react with both bands of the doublet [Cetani 1995]. It was suggested that trypsin cleavage during purification gave rise to the doublet. However, purification of native porcine [Nakagwa 1985] and human TPO [Ohtaki 1986b] using detergent solubilisation, in the absence of proteases, showed there was still a doublet present. It was therefore suggested that the doublet may be due to endogenous proteolysis in the thyroid cells. There is also

evidence that different glycoforms may be responsible for the doublet, as native TPO showed marked changes in the double band pattern after treatment with different glycosidases [Kiso 1992]. However, treatment of recombinant TPO with enzymatic deglycosylation increased in parallel the mobility of both domains of the doublet [Foti & Rapoport 1990] indicating that glycosylation was not responsible in this case.

1.4.4 Catalytic Role

Thyroid hormone is an iodine containing steroid hormone produced by the thyroid gland. The major forms of the hormone, thyroxine (T_4) and triiodothyronine (T_3), have a pleiotropic stimulatory role on metabolism and growth [Braverman & Utiger 1996]. The structure of the hormones is shown in Figure 1.12, and the site of synthesis in Figure 1.13. The hormones binds to thyroid hormone receptor in the nucleus which then acts as a transcriptional activator [Baniahmad 1993]. The structure of the ligand binding domain of thyroid hormone receptor has been solved by crystallography [Ralff 1998].

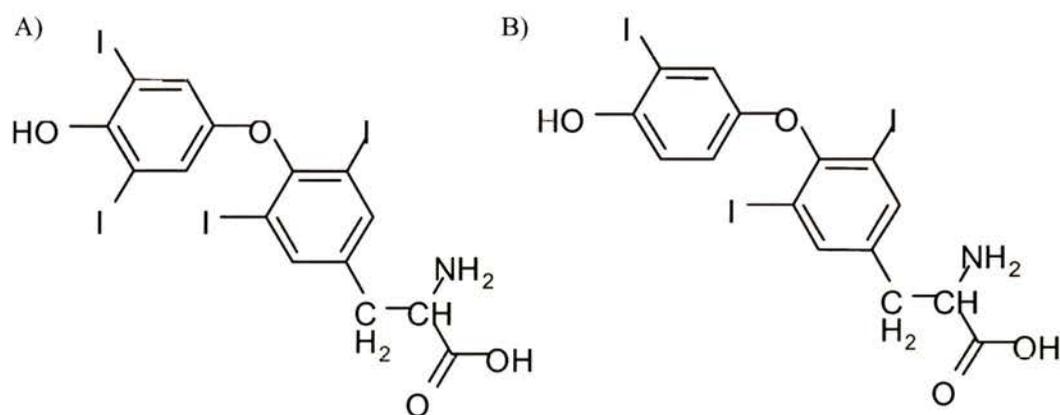


Figure 1.12

The structure of thyroid hormones, A) T_4 , B) T_3 .

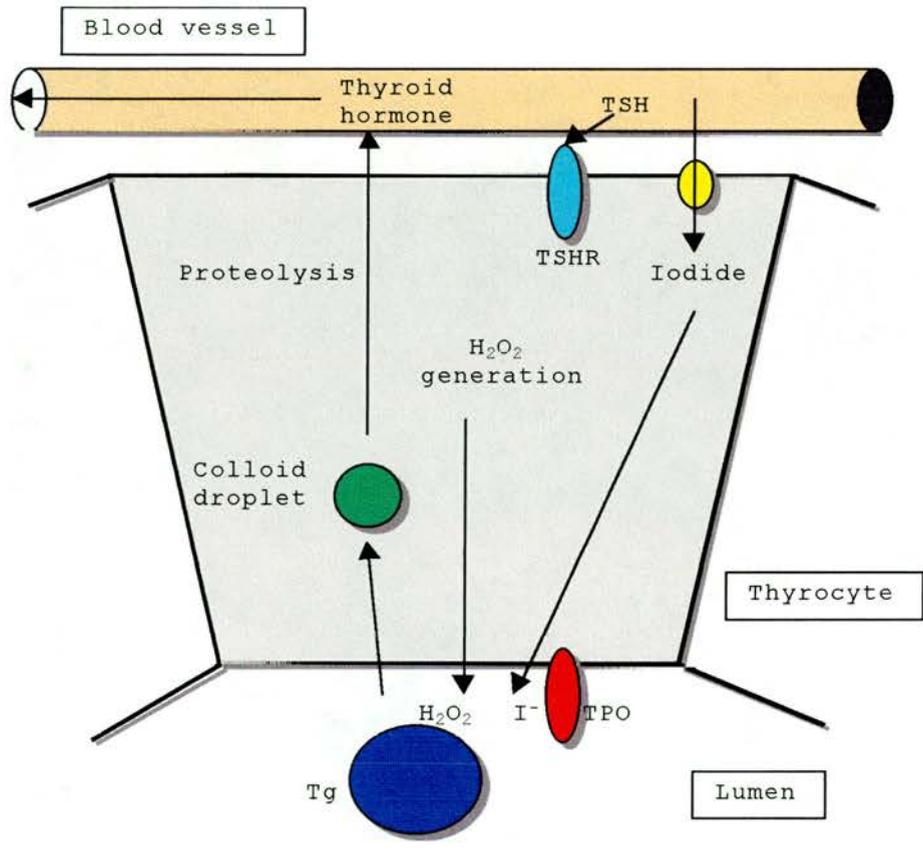


Figure 1.13

Summary of the major steps in thyroid hormone biosynthesis.

The iodination and coupling of thyroid hormone occur on Tg, a water-soluble, glycosylated dimer of 2 x 330 kDa synthesized by the thyrocytes and secreted into the thyroid. Electron microscopy shows that the iodination of Tg occurs at the cell-colloid interface, close to the apical membrane [Ekholm & Bjorkman 1984]. Human Tg contains 134 tyrosyl residues per dimer [Malthiery & Lissitzky 1987], but only three different tyrosyl residues have been found to be iodinated (5, 2553 and 2520) [Xiao 1996]. Proteins other than Tg can act as substrates for TPO, albeit at lower efficiency, and it appears that selectivity is determined by tyrosine accessibility. Indeed, the same tyrosines are iodinated in chemical and enzymatic iodination [Xiao 1996]. However, this study indicated that TPO was necessary for the preferential formation of DIT rather than MIT.

About two to four molecules of T₄ are made per 660 kDa Tg, and it is unlikely that after proteolysis the Tg can be reused. This seemingly inefficient process has been well conserved through evolution and may confer some, as yet unknown, advantage. It has been postulated that insoluble Tg multimers, formed by disulphide and dityrosine bonds may have a role in iodine storage [Baudry 1997]. The colloid, containing T₄ and T₃ bound to Tg, is resorpted into the thyrocyte by macropinocytosis and micropinocytosis. Proteolysis liberates the thyroid hormone which is secreted across the basement membrane and into the bloodstream.

The reaction mechanism of TPO has been proposed to be similar to that of other peroxidases [Taurog 1996]. First is thought that ferric TPO forms an intermediate with H₂O₂, known as Compound I, which contains two oxidising equivalents. Compound I then oxidises inorganic iodide to form an iodinating species, most likely an enzyme bound hypoiodite, which iodinates tyrosine residues via the formation of Compound II. There is evidence for the iodination reaction being mediated by radicals [Pommier 1976], and also for a two-electron oxidation involving I⁺ [Ohtaki 1981]. More recently, it has been suggested that both the iodination and coupling reactions are mediated by the porphyrin π -cation radical form of compound I (Figure 1.14) [Taurog 1996]. A study of the kinetics of TPO

catalysis with different substrates showed that the pH optima of the reaction depends on both the electron donor and H₂O₂ concentrations [Kootstra 1993].

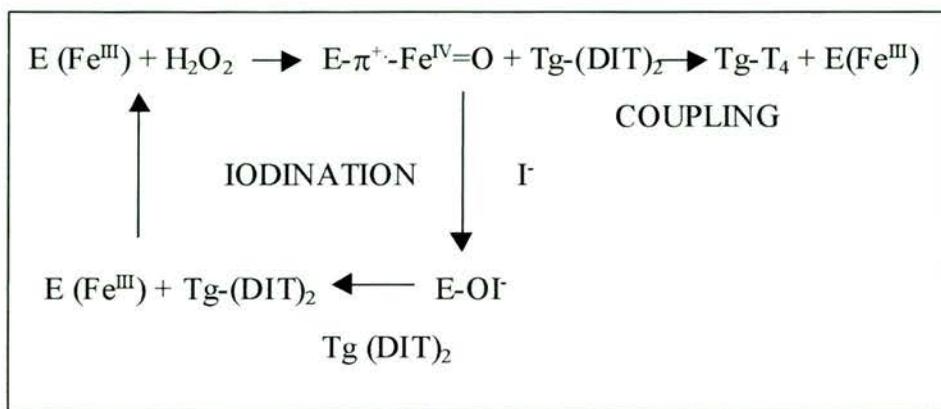


Figure 1.14.

A proposed mechanism of action of TPO [Taurog 1996].

1.4.5 Epitope Mapping

Epitopes to TPO have been extensively studied using recombinant peptides [Arcott 1996], mutated TPO [Nishikawa 1996], recombinant TPO autoantibodies [McLachlan & Rapoport 1995] and chimeric molecules of TPO and MPO [Nishikawa 1993]. Reduction of TPO decreases the ability to bind autoantibodies [Finke 1990; Gardas 1990] implying that, in common with epitopes recognised by the majority of antibodies, TPO epitopes are conformational. Indeed, human monoclonal autoantibodies that preferentially recognise denatured TPO could not be isolated from an immunoglobulin gene library despite selection with denatured protein [Guo 1999]. However, there is evidence that there are some autoantibodies in patient sera that recognise linear TPO epitopes [Hamada 1987]. These autoantibodies are thought to be in the minority and may be elicited against

degraded proteins released from damaged thyroid tissue [Nagayama 1995]. A linear epitope, C21 (amino acids 710-722), that binds to a murine monoclonal antibody called m47 has been identified [Ruf 1989; Finke 1991; Libert 1991], but its significance has been disputed. The C21 epitope was not found to overlap with epitopes recognised by more than 80% of the human autoantibody repertoire [Chazenbalk 1993a]. In addition, replacement of the C21 epitope with corresponding regions of MPO did not alter binding to autoantibodies [Nishikawa 1996]. A further study showed that TPO-reactive Fab did not inhibit the binding of m47, but that m47 did inhibit the binding of Fab [Czarnocka 1997].

The membrane form of TPO is a dimer and this is the only form immunoprecipitated from thyroid cell lines [Baker 1994b]. It has been shown that a TPO Fab inhibits equally the binding of autoantibodies to a TPO monomer and dimer [Nishikawa 1994b], showing that the contact regions between dimers is not immunodominant. The *in vivo* structure of TPO may be important as it has been shown that cellular TPO and not purified TPO and adjuvant, induces antibodies in mice that resemble autoimmune antibodies [Jaume 1999b]. Replacement of the N-terminal 121 residues of TPO with the N-terminal 146 residues of MPO did not alter the binding of TPO autoantibodies [Nishikawa 1993; Nishikawa 1994b], indicating that autoantibodies do not interact with the N-terminal domain of TPO.

A number of separate studies have identified two overlapping, conformational domains that are involved in autoantibody binding. The first study [Ruf 1989] used a panel of murine monoclonal antibodies and found four interacting domains, A-D, with the binding of patient serum autoantibodies inhibiting binding to A and B. Domain A was shown to map to the linear C21 epitopes. A different study [Chazenbalk 1993b] used TPO-reactive Fabs from phage display combinatorial libraries of patients with Graves' disease which defined two neighbouring domains, A and B each subdivided into two subdomains based on cross-inhibition experiments. A study was carried out to compare the murine monoclonal panel and Fab fragments from phage display combinatorial libraries derived from AITD

patients [Guo 1998a]. This study showed that the immunodominant domains are recognized by autoantibodies in all 195 patient's sera studied, and by approximately 80% of TPO autoantibodies within each serum. In addition, they provided evidence that the domain called A by Chazenbalk *et al.* corresponds to B in Ruf *et al.*. Here, the domain nomenclature according to Ruf *et al.* will be used throughout. A separate study of mouse monoclonals in competition with patient autoantibodies has shown that an immunodominant region, defined by a single monoclonal antibody, binds to 80-90% of serum autoantibodies in patients with moderate autoantibody levels and at least 65% in patients with high autoantibody levels [Czarnocka 1996].

TPO autoantibody epitopic fingerprints can be defined, and there is evidence that they may be inherited in patients with Hashimoto's thyroiditis [Jaume 1994; Nishikawa 1994a]. Using epitopic fingerprints, it has been shown that the autoimmune response to TPO is not related to thyroid status and is stable, without epitope spreading, for at least 15 years [Jaume 1995a; Jaume 1995b; Jaume 1996a; Czarnocka 1998]. However, analysis of recombinant TPO fragments expressed in bacteria indicates that there are differences in the autoantibody response in Hashimoto's and Graves' diseases [Bermann 1993]. In addition, the study of a series of truncated and deleted TPO from an *in vitro* transcription/translation system suggests that there may be some differences in epitope recognition between diseased and healthy individuals [Grennan Jones 1999].

Proteolytic peptides of TPO generated by enzymatic hydrolysis in nondenaturing and nonreducing conditions [Estienne 1998], showed that the C-terminal CCP and EGF domains (residues 742-848) contain at least one conformational B-cell epitope. Peptide binding was influenced by iodination, suggesting that a tyrosine residue is involved in the epitope. A putative tyrosine at the domain interface was identified in a model of the CCP and EGF domains. Monoclonal antibody reactivity showed that there was a calcium induced conformational change in the

epitope, and this was explained by a structural calcium that is predicted to stabilise the domains [Estienne 1999a]. The epitope is more frequently recognised by Hashimoto's than Graves' sera, indicating that the epitope is disease specific.

A model of TPO has been built and used to identify prominent surface structures [Hobby 2000]. Antisera were raised to these structure and identified an epitope (residues 599-617) which was shown by competition with mouse monoclonal antibodies to be part of the B domain. Antibodies to this epitope inhibit the binding to TPO of human autoantibodies in the majority of serum samples from 65 patients with AITD.

It has been suggested that restricted B-cell epitopic recognition may result from the ability of B-cells to capture and influence peptide presentation to T-cells [Guo 1996]. Experiments have shown that B-cell epitopes targetted by membrane bound antibodies can enhance or suppress the peptides that are presented to T-cells [Simitsek 1995]. Recently, evidence has been provided that TPO-specific B-cells can present antigen to sensitized T-cells in mice injected with fibroblasts co-expressing TPO and MHC class II [Guo 2000]. B-cells may play this role in other autoimmune diseases, for example, B-cells have been shown to be important APCs in the non-obese diabetic mouse model of IDDM type I [Falcone 1998]. However, the role played by B-cell presentation of antigen to T-cells in the establishment of dominant B- and T-cell epitope specificity, and in maintaining the reported stability of the specificity [Jaume 1996a] is unclear.

1.4.6 Thyroid Peroxidase Autoantibodies

Antibodies present in Hashimoto's disease patients capable of fixing complement were identified in 1957 [Trotter 1957] and were named thyroid microsomal antibodies after their characterisation in cell fractionation experiments. Strong

autoantibody binding along the apical, microvillar border was demonstrated in 1964 [Roitt 1964] indicating that microsomal antigen was also located on the cell surface. In 1985 it was discovered that TPO and thyroid microsomal antigen were in fact the same protein [Czarnocka 1985; Portmann 1985] and that microsomal antibodies were recognising TPO.

Antibodies to TPO are found in over 90% of patients with Graves' disease and in almost all of those with Hashimoto's thyroiditis [Prentice 1990]. The concentration of anti-TPO autoantibodies in the serum of AITD patients can vary between 50-1400 $\mu\text{g/ml}$ and may represent up to 10% of the total serum immunoglobulin [Beever 1989]. The antibodies are polyclonal, found in all four IgG subclasses and distributed between both light chains classes [Doble 1988]. More recently, low titre of IgE class TPO autoantibodies have been detected [Guo 1997]. There is evidence Tg autoantibodies can cross react with TPO [Kohno 1988], and these so-called TGPO autoantibodies appear to be associated with AITD [Ruf 1994; Estienne 1999b].

The presence of TPO autoantibodies is closely associated with thyroid destruction and hypothyroidism, suggesting they have a pathogenic role. TPO autoantibodies have been found to be complement fixing and cytotoxic [Khoury 1981] and it has been shown by some investigators [Okamoto 1989], but not others [Kohno 1991] that antibodies to TPO inhibit enzyme activity. However, 18% of UK females have antibodies to TPO but show no signs of disease [Prentice 1990], and these autoantibodies have all the characteristics of TPO autoantibodies found in AITD patients [Jaume 1995a]. There is evidence that TPO may not initiate disease as immunisation of mice with porcine TPO led to production of autoantibodies but there was no evidence of thyroiditis [McLachlan 1990]. However, it has recently been shown that immunisation with cellular TPO induces antibodies that resemble autoantibodies, suggesting that conventional immunisation may not be representative of human disease [Jaume 1999b].

Segregation analysis has indicated that the inheritance of TPO autoantibodies is a dominant Mendelian trait [Prentice 1993] and that there is genetic transmission of TPO autoantibody epitopic fingerprints [Jaume 1999a]. However, it has not been possible to distinguish between a single locus and a multifactorial model of inheritance [Phillips 1993b].

TPO autoantibodies may be found in patients with non-thyroid autoimmune diseases, consistent with the observation that autoimmune attack is often directed to more than one organ [Ollier & Symmons 1992]. For example, IDDM is frequently associated with AITD and there are indications of an association between the disease and the presence of anti-TPO antibodies [Chang 1998]. In addition, measurement of Tg and TPO antibodies in pregnant women seems useful as a marker for risk of miscarriage [Roberts 1996]. Alteration of TPO, followed by immunodetection of antibodies, have been shown to be an early marker of thyroid follicular tumours, closely related to acceleration of tumor growth in the first stages of malignant transformation [Garcia 1998]. Finally, the presence of TPO autoantibodies was found to be associated with a better prognosis in breast carcinoma [Smyth 1998], suggesting a possible biological link between the diseases.

Many TPO autoantibodies have been cloned by phage display techniques [Rapoport 1995], and one human monoclonal antibody, 2G4, has been obtained by traditional hybridoma techniques [Horimoto 1992]. To date there are 80 HC sequences and 87 LC sequences of TPO autoantibodies, these include 57 antibodies of which both the LC and the HC have been sequenced. The CDR sequences of all published TPO autoantibodies and their predicted canonical class are listed in Appendix B. The subset of autoantibodies which have been mapped to the A or B domain on TPO are also given. The autoantibodies have different CDR sequences and predicted canonical forms of different classes. They share some sequence homologies which is in part due to the restricted gene usage and may reflect the antigenic domain bound [Guo 1998a]. However, there is no overall relationship between antibody

sequences and epitopes bound, and studies have shown that this is often the case. For example, the structures of two overlapping antibodies in complex with neuraminidase has shown that there is not necessarily a correlation between the antigen bound and the sequence of the antibody [Malby 1994]. In addition, study of 18 antibodies to $\alpha(1-6)$ dextran showed that different V region genes can form similar combining sites to a single site-filling epitope [Wang 1991].

The V genes found in TPO autoantibodies are often mutated, and share the same germline genes as other autoantibodies [Portolano 1993b]. In common with other organ-specific autoantibodies, the use of the λ LC is rare [Portolano 1995]. Studies have shown a lack of promiscuity between HCs and LCs [Portolano 1993a], with promiscuity defined as the pairing of HC with a number of different LC with no alteration of antigen specificity. It has been suggested that the LCs are the dominant partners in determining specificity [Constante 1994] and a correlation between the V-regions used and the antigenic domain bound has been found [Jaume 1996b]. The $V_{\kappa}I$ (O12) LC is commonly found in TPO autoantibodies and it appears to be associated with reactivity to the B domain [Guo 1998a]. However, only 1 in 5000 phage containing this LC in combination with random HCs bound to TPO, showing that its presence is not sufficient to cause TPO reactivity [Portolano 1993a]. An analysis of the CDR3 regions of TPO antibodies indicates that D-D fusion events are common [McIntosh & Weetman 1997], but the significance of this observation is not yet understood.

The structure of a human autoantibody to TPO, TR1.9, obtained from a phage display library, has been solved by X-ray crystallography [Chacko 1996]. The antibody combining site is relatively flat and somatically mutated residues are in positions that could interact with antigen. It was proposed that the increased deformability of H2 and H3 caused by replacement of two germline asparagines and the presence of several glycines may allow induced fit.

1.5 Experimental Aims

The crystal structures of antibody-protein complexes have confirmed that antibodies bind to conformational epitopes on the surface of fully folded proteins [Davies & Cohen 1996]. However, with the exception of lysozyme [Bentley 1996], there are insufficient structures of a single proteins bound to multiple antibodies to determine which surfaces on a protein are antigenic. The epitopes may be determined solely by the accessible surface area of the antigen, or by factors extrinsic to the antigen, for example by mechanisms of immune regulation [Benjamin 1984].

Autoantibodies are a common feature of autoimmune disease, and there is evidence that they may bind to a restricted region, or autoimmunodominant domain, on the antigen [Mackay & Gershwin 1989; Tzartos 1998; Guo 1999]. Determination of the location of such autoepitopes is important in order to help understand the mechanisms of autoimmunity [Song 1996].

Autoantibodies to TPO are found in the majority of patients with AITD and they have the ability to fix complement and mediate ADCC [McIntosh & Weetman 1997]. There are many lines of evidence that suggest that TPO autoantibodies bind to an immunodominant region of TPO that is made up of two overlapping domains, A and B [Ruf 1989; Nishikawa 1995; Arscott 1996; Czarnocka 1996]. It has been suggested that the restricted B-cell epitopic recognition may result from the ability of B-cells to capture and influence peptide presentation to T-cells [Guo 1996]. In addition, there may be features of the TPO accessible surface that are responsible for the presence of an immunodominant domain.

To date, over 50 anti-TPO autoantibodies have been sequenced (see Appendix B) and many have been mapped to immunodominant domain A or B using competition studies [McIntosh 1997]. Implicit in the definition of an epitope [Benjamin 1995] is that each autoantibody will bind to a separate epitope on TPO. Therefore, it is

is that each autoantibody will bind to a separate epitope on TPO. Therefore, it is likely, due to the large number of different autoantibody sequences, that the immunodominant domain on TPO is made up of a continuum of overlapping epitopes.

The location of the immunodominant domain on TPO, and the factors responsible for the restricted epitope recognition are at present unknown. Currently the only way to unequivocally map the antigenic sites on a protein is to solve the structure of antibody-antigen complexes [Laver 1990]. The crystal structures of multiple complexes of autoantibodies with TPO would determine the location of the immunodominant domain. Visualisation of the immunodominant domain would then allow study of mechanisms responsible for the generation of autoantibodies.

The original objective of the project was to solve the structure of TPO, and the structures of TPO in complex with two autoantibodies. The structures would then be used to study the nature of the autoantigenic sites and the structure of the enzyme active site. Prior to the start of the project, recombinant human TPO expressed in insect cells had been crystallised and data had been collected to 7 Å resolution. The aim was to characterise the protein and to grow more ordered crystals to extend this resolution to 3 Å or better. Two TPO antibodies were available which together defined the major autoimmune domains on TPO: 2G4, a human autoantibody isolated from patients with Hashimoto's disease and 4F5, a mouse monoclonal antibody. It was hoped to purify the Fab fragments of 2G4 and 4F5 and crystallise the complexes with TPO.

Despite attempts to reduce TPO heterogeneity and optimise crystallisation conditions, it was not possible to grow crystals that diffracted to higher resolution. Characterisation of TPO showed that there was heterogeneity at the N-terminus of the protein, which would explain the low resolution limits of diffraction. However, it was not feasible in the timescale of the project to make a new DNA construct in order to produce more homogeneous protein. The Fab fragments of 4F5 and 2G4

complexes could be grown. Crystals of uncomplexed 2G4 were obtained, but no diffraction data could be collected.

Therefore, the revised aims of the project were:

1. To process the 7 Å TPO dataset and attempt molecular replacement.
2. To build a molecular model of TPO, and to use it to study structural features that may be responsible for restricting autoantibody binding to an immunodominant domain.
3. To crystallise the Fab fragment of 4F5 and to solve the structure by molecular replacement. To analyse the region likely to be involved in binding TPO and to compare the structure of the antibody combining site with that of the TPO autoantibody TR1.9.

Chapter 2

Thyroid Peroxidase Characterisation

2.1 Introduction

Prior to the start of the project, crystals of recombinant human TPO had been grown and diffraction data to 7 Å resolution collected. In order to solve the structure of TPO it is necessary to grow crystals that diffract to higher resolution. Homogeneous protein is a critical factor in growing highly ordered crystals [Ducruix & Giege 1992], the protein must contain no heterogeneity (contamination with other proteins) or microheterogeneity (contamination with different forms of the same protein). The term heterogeneity is commonly used to describe both forms of contamination.

To ensure the recombinant TPO had the same properties as native TPO, the protein was assayed for protein and heme content, enzyme activity and antibody binding. The protein was further characterised in order to determine the reasons for the low resolution diffraction limit of the crystals. As TPO is a glycoprotein there are two possible sources of heterogeneity – protein and carbohydrate. Recombinant and native TPO run as a doublet on SDS-PAGE gels, and it has been suggested that this is due to proteolysis or the presence of different glycoforms [Cetani 1995]. The degree of protein and carbohydrate heterogeneity can be studied by analysis on SDS-PAGE and isoelectric focusing (IEF) gels, by N-terminal sequencing and mass spectroscopy and by using deglycosylation.

2.2 Production and Purification

Recombinant human TPO was expressed and purified by RSR Ltd, using the baculovirus-insect cell system [Grennan Jones 1996; Gut 2000]. The construct contained the soluble fragment (residues 1-839) and was expressed in *Tricoplusia ni* (High Five) insects cells supplemented with the heme precursor δ -aminolevulinic acid. The culture supernatants were harvested by centrifugation and protease inhibitors ('Complete' inhibitor cocktail, Roche) were added.

The preparations available for crystallographic studies and the purification methods used to produce each one are listed in Table 2.1. Two purification protocols were developed. The first involved ion exchange chromatography, followed by gel filtration on a S-300 column, a second ion exchange step and finally a hydroxyapatite column. This method was used to produce TPO 18. Preparations 21, 23 and 26 were purified by a single ion exchange step and gel filtration [Grennan Jones 1996], and were partially pure. More recently, a combination of hydrophobic chromatography on a butyl-sepharose column, affinity chromatography with the anti-TPO antibody 4F5 and gel filtration on a S-300 column has been used [Gut 2000]. This purification protocol was used to produce TPO 1FP, 2FP and 10FP.

Name	Purification steps	Amount (mg)
21, 23, 26	Anion exchange, gel filtration	16.6
18	Anion exchange, gel filtration, anion exchange, hydroxyapatite column	3.8
1FP, 2FP	Hydrophobic column, affinity column, gel filtration	6.6
10FP	Hydrophobic column, affinity column, gel filtration	5.3

Table 2.1

The recombinant TPO preparations available for characterisation and crystallisation.

2.3 Thyroid Peroxidase Assays

The TPO preparations used for crystallisation was assayed for protein concentration (absorbance and Bradford assay) [Bradford 1976], enzyme activity (guaiacol oxidation assay) [Nakashima & Taurog 1978] and antibody binding (radioimmunoassay) [Premawardhana 1992].

The Bradford assay was run and calibrated using a bovine serum albumin (BSA) standard line (Figure 2.1). 100 μ l of 0.2 to 1.4 mg/ml of protein in 100 mM NaCl, 2 M Tris pH 8.3 was added to 5 mls of Bradford assay dye reagent and incubated at room temperature for 5 to 30 minutes. The optical density at 595 nm was read and converted to concentration using the standard line.

Absorbance was measured at 280 nm (protein peak) and 412 nm (heme peak) in quartz cuvettes in a spectrophotometer. Protein concentration was calculated using the Beer-Lambert law:

$$\log_{10} (I_0/I_t) = \epsilon cl$$

where I_0 is the radiation incident on the protein, I_t is the radiation absorbed, c is the concentration and l is the length of the cell through which the radiation travels. ϵ is the extinction coefficient of the protein at a particular wavelength. The term $\log_{10} (I_0/I_t)$ is known as absorbance (A) or optical density (OD) and is dimensionless. The extinction coefficient of TPO at 280 nm, as calculated from the amino acid sequence, is $1.33 \times 10^6 \text{ cm}^2 \text{ mol}^{-1}$. The OD 412/280 of recombinant TPO 2FP was 0.5, compared to 0.36 for native TPO [Gut 2000]. This indicates that the heme group has been effectively incorporated into the recombinant TPO.

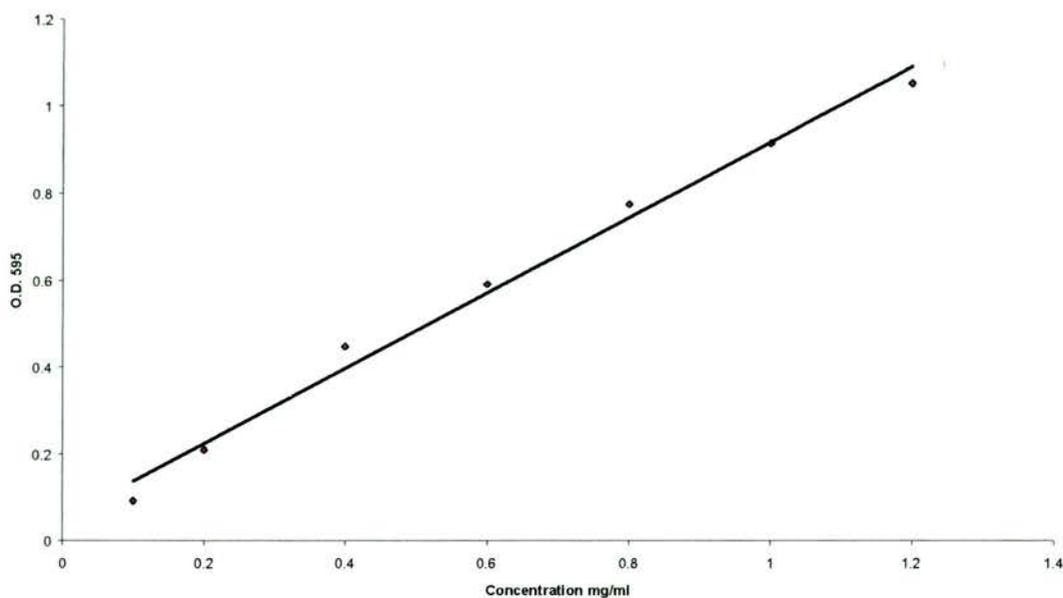
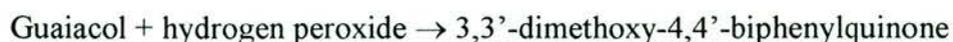


Figure 2.1
The Bradford assay calibration curve with BSA.

The guaiacol assay was carried out with 100 μ l of TPO to which 2 ml of 40 mM guaiacol in 10 mM TRIS pH 8.0 was added and incubated for 15 minutes at 37 °C [Nakashima & Taurog 1978]. 10 μ l of 60 mM H₂O₂ was added at time zero and readings were taken at 15, 30 and 60 seconds at 470 nm. The guaiacol is oxidised giving rise to coloured product, thought to be composed of dimerised guaiacol molecules [Capeillere-Blandin 1998]:



Activity was calculated in guaiacol units (Gu), defined as an increase in absorbance at 470nm over 1cm in 1 minute as described by [Nakashima & Taurog 1978]:

$$\text{Reading at 60 s} \times 10 / [\text{TPO}] \text{ mg/ml} = \text{Activity (Gu/mg)}$$

The results of the TPO guaiacol assay of 2FP culture supernatants containing 0.2 mg/ml of TPO are shown in Table 2.2. The activity is calculated to be 547 Gu/mg. On average, the recombinant TPO has an enzyme activity of about 900 Gu per mg of protein, whereas for native TPO the activity is 750 Gu per mg of protein [Gut 2000]. The comparable activity of native and recombinant TPO indicates that the heme is incorporated into recombinant TPO in the correct orientation for full activity.

Protein	OD 470 15sec	OD 470 30sec	OD 470 60sec
TPO 1:10 dilution	0.481	0.783	1.093
+ve control (TPO standard)	0.316	0.598	0.883
-ve control (TRIS only)	0.004	0.003	0.004

Table 2.2

The results of the guaiacol assay on TPO culture supernatant.

Radioimmunoassays (RIAs) carried out by RSR Ltd showed that recombinant and native TPO had identical inhibiting effects on TPO autoantibodies binding to ¹²⁵I-labelled native TPO in sera from 25 patients with AITD [Grennan Jones 1996; Gut 2000]. This provides strong evidence that the recombinant TPO is correctly folded.

2.4 SDS-PAGE Analysis

SDS-PAGE gels are used to monitor the molecular weight of a protein and of any contaminants present. The electrophoretic mobility of detergent bound proteins through an acrylamide gel is measured, and separation is based on size [Laemmli 1970]. Recombinant TPO and TPO from the dissolved crystals were run on reduced and non-reduced SDS-PAGE gels.

TPO was resolved on an 8% SDS-PAGE gel containing 7.5 mls of resolving gel (8% acrylamide, 370 mM Tris pH 8.8, 0.1% SDS, 0.04% ammonium persulphate (AMPS) and 0.1% tetramethylethylenediamine (TEMED)) and 2.5 ml of stacking gel (5% acrylamide, 120 mM Tris pH 6.8, 0.1% SDS, 0.05% AMPS and 0.2% TEMED). Approximately 2 µg of 0.2 mg/ml protein was added to 10 µl of loading buffer (containing β-mercaptoethanol for reducing gels) and run at 20 mA in tank buffer (2.5 mM Tris, 0.01% SDS, 200 mM glycine) in a Pharmacia Mighty Small gel caster and tank. The gel was stained in Coomassie Blue for 20 minutes.

Figure 2.2 shows a non-reduced and reduced 8% SDS-PAGE for native and recombinant TPO. The reduced gel of native TPO shows a doublet of 94/91 kDa with a contaminant at 37 kDa, whereas the recombinant TPO is a broad band at approximately 98 kDa. The non-reduced gel shows a doublet of native TPO 98/95 kDa and a contaminating band at 45 kDa, and a doublet of recombinant TPO at 104/100 kDa.

The difference in relative mobility between native and recombinant TPO can be explained by the sizes of the TPO preparations used. The native TPO was released from the membrane by trypsin cleavage near the transmembrane section, at a site estimated to be between amino acids 793-847 [Gut 2000]. Trypsin cleavage may also have occurred after residue 109 [Taurog 1990]. Therefore, the molecular mass of native TPO, calculated from amino acid sequence, ranges from 76 kDa (109-793) to 92 kDa (15-847).^{12.6 kDa addition} The native TPO preparation has been estimated to contain 12.6 kDa of N-linked carbohydrate [Grennan Jones 1996]. The recombinant protein is believed to consist of amino acids 1-839, with the first 14 residues acting as a signal sequence. The molecular mass of the recombinant protein (residues 15 to 840) is calculated from the primary sequence to be 91.7 kDa and the contribution of N-linked carbohydrate has been estimated to be 2.3 kDa [Grennan Jones 1996].

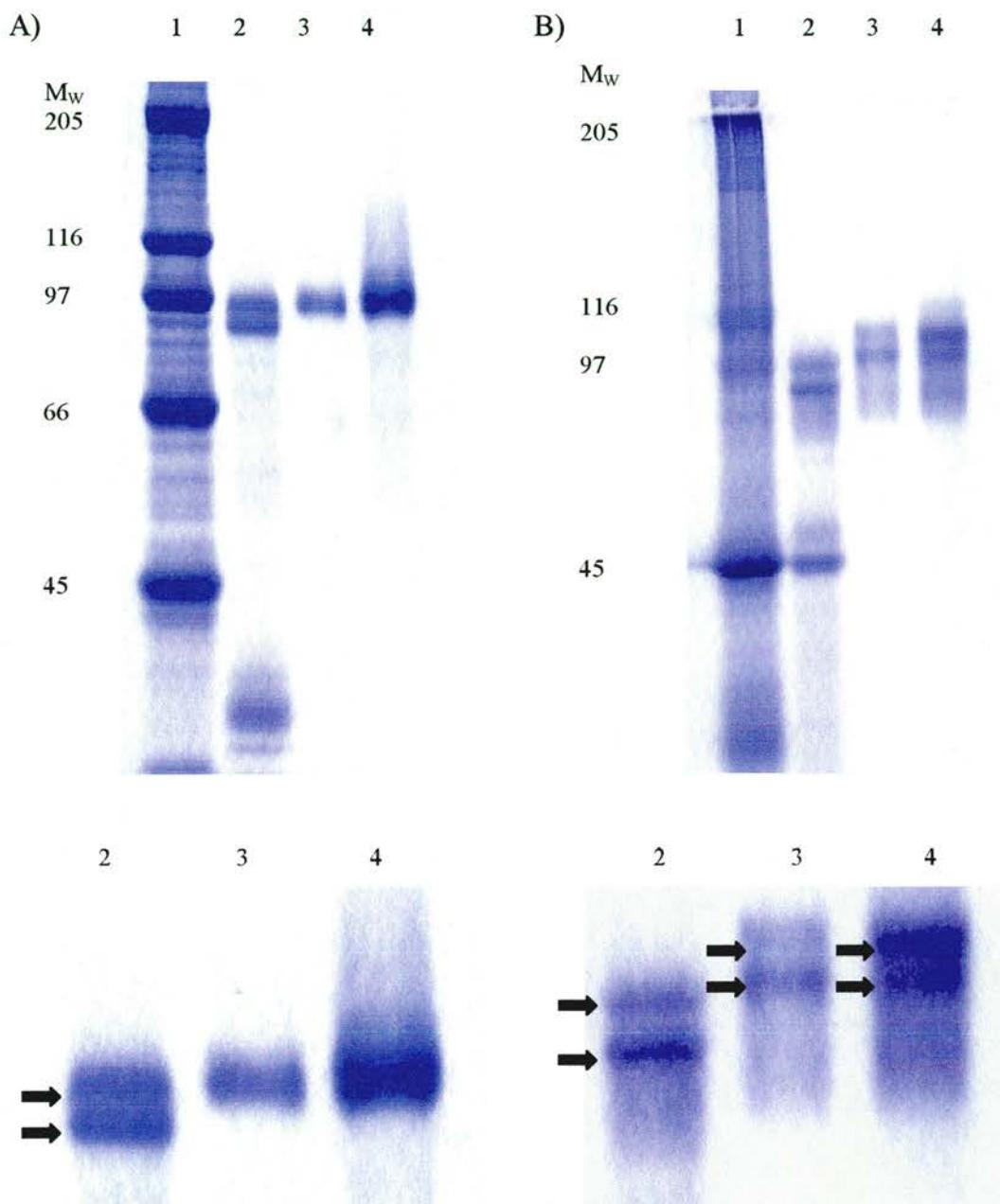


Figure 2.2

SDS-PAGE gels of TPO: A) Reduced gel B) Non reduced gel. A close-up of the doublet is shown below each gel and the doublet bands are indicated by arrows.

- | | | | |
|---|-------------------------------|---|---------------------|
| 1 | High molecular weight markers | 3 | Recombinant TPO 18 |
| 2 | Native TPO 9 | 4 | Recombinant TPO 1FP |

The recombinant preparations 18 and 1FP are produced by different purification schemes but have the same mobility on the gels. There is no evidence for native or recombinant TPO running as a dimer.

The recombinant TPO doublet is more evident on non-reduced gels than on reduced gels. The TPO band on reduced gels has been interpreted as a broad band, rather than a doublet. It is possible that a doublet is present on reduced gels, but that the individual bands are very close together and so not distinct. Alternatively, there may be more than two different TPO species present, which run as a broad band when reduced and as a doublet when not non-reduced. Proteins which contain disulphide bonds often have different mobilities, and hence different apparent molecular masses, in the presence and absence of reducing agents [Hames & Rickwood 1990]. The different behaviour of TPO when reduced suggests that the different forms of recombinant TPO may contain a different number of disulphide bridges.

Figure 2.3 shows a reduced SDS-PAGE gel of single crystal of TPO 18 that was removed from the crystallisation drop. The gel shows that the protein incorporated into the crystals is the same as that used for crystallisation. In particular, there is a doublet visible in the crystallised protein, confirming that the crystals contain the same heterogeneity as is seen in the purified protein.

2.5 IEF Analysis

IEF gels are used to measure the isoelectric point of a protein and any contaminants present. The gels separate proteins on a pH gradient imposed by applying an electric field to a solution of ampholytes [Wallis 1973]. When the proteins reach their isoelectric point they stop migrating and their position can be compared to a lane of standards.

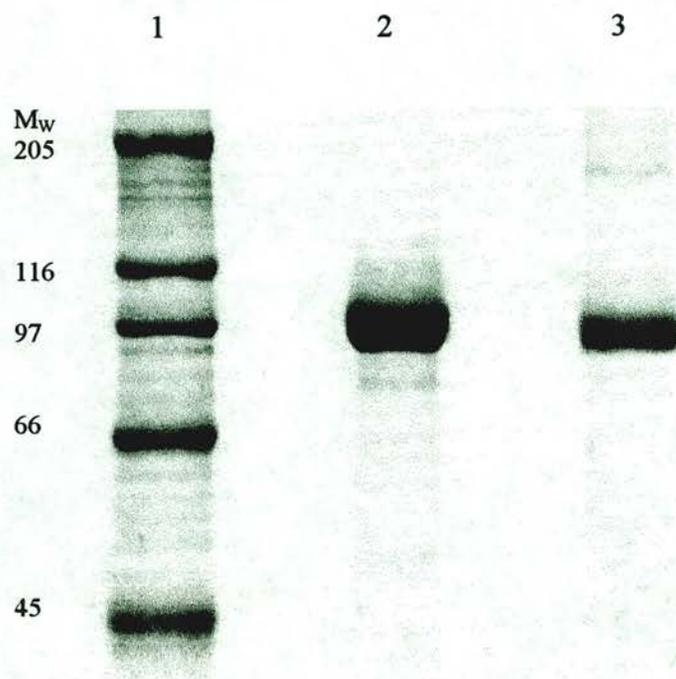


Figure 2.3

An 8% reduced SDS-PAGE of the TPO crystals.

1. **Markers**
2. **TPO 18 used for crystallisation**
3. **TPO 18 from a dissolved crystal**

IEF gels were assembled and run using a Bio-Rad mini IEF cell. Ready made gels (pH 3-10) and premixed buffers from Bio-Rad were used, approximately 10 µg of a 0.8 mg/ml protein solution was loaded onto the gel and mixed with four parts 50% glycerol to one part protein. The gel was run at 200 V for one hour and stained with Crocein Scarlet for 45 minutes.

The IEF gel of TPO is shown in lane 3 of Figure 2.4. The gel resolves multiple species with isoelectric points ranging from pH 5.0 to pH 6.2, with a predominant band at pH 5.9. The observed isoelectric points are similar to the value calculated from the TPO sequence (14-839) of pH 6.0, which does not include contributions from carbohydrate.

2.6 N-terminal Sequencing

N-terminal sequencing is a technique to identify the first amino acids in a protein [Edman 1970]. The N-terminal residue of the sample to be sequenced is chemically labelled, released and then identified using chromatography, and the process is repeated for the next amino acid. The N-terminal sequencing was carried out by the University of St. Andrews Protein Sequencing Service on an Applied Biosystems Procise 491 using the Edman degradation method [Edman 1970]. The sample (10 µl of 0.75 mg/ml TPO 10FP) was run on a non-reduced 8% SDS-PAGE gel and blotted to hyperbond membrane using the semi-dry method. The membrane was stained with amido black and the two TPO bands making up the doublet were cut out separately.

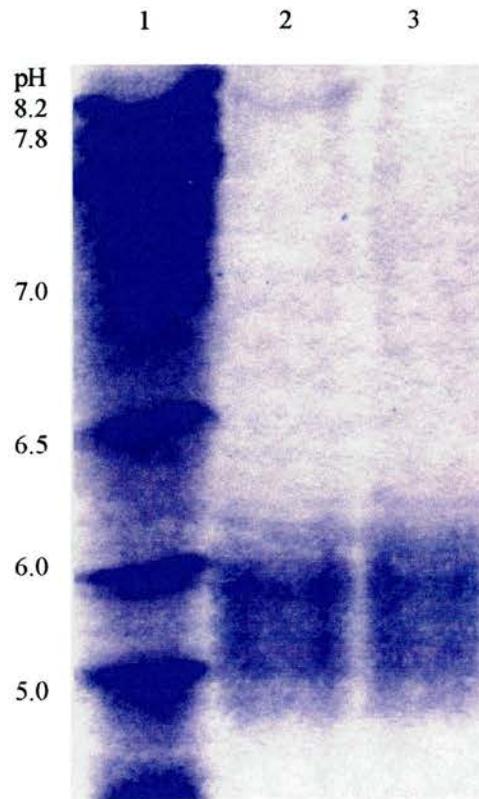


Figure 2.4

An IEF gel of glycosylated and deglycosylated 2FP TPO with PNGase F and Endo H.

1. Markers
2. TPO + PNGase/Endo F
3. TPO - PNGase/Endo F

The N-terminal sequencing results showed the upper band in the TPO doublet contains a mixture of species, with the predominant one being **XXKLPEXXSQ/G** (where X is unknown residue). The lower band also contains more than one sequence, with **TQQSQHPTDA** being the strongest. Both sequences were found to be part of the N-terminal domain of TPO and are shown in Figure 2.5 marked A and B. It can be seen that the two predominant species of the doublet begin at residues 76 and 109. For comparison the beginning of the peroxidase domain is marked on the figure at residue 142.

An analysis of the sites of cleavage by commonly known proteases [Wilkins 1997] in the N-terminal domain was carried out. The proteases included are listed in Table 2.3. Residue 109 is a trypsin cleavage site, but the cleavage site at residue 76 could not be identified. The TPO sequence between residues 77 to 108 has a calculated molecular mass of 3.6 kDa, which is similar to that observed as the difference between the doublets on non-reduced SDS-PAGE.

Enzyme	Residue before cleavage site	Residue after cleavage site
Arg-C	Arg	Any
Asp-N	Any	Asp
Chymotrypsin-1	Phe, Tyr, Trp	Any
Chymotrypsin-2	Phe, Tyr, Trp, Ala, Leu, Met	Any
Glu-C1	Glu	Any
Glu-C2	Gly, Asp	Any
Lys-C	Lys	Any
Trypsin	Arg, Lys	Any not Pro

Table 2.3

Commonly known proteases and their cleavage sites.

Assuming that the upper and lower band of the doublet are in equal ratio to each other and that there are no other species present, the molecular mass can be calculated from the protein sequence. Assuming the heme is the same mass as that of MPO (616 Da), the molecular mass of TPO is 83.6 kDa (Table 2.4). Although not including contributions from carbohydrate, this is considerable lower than that calculated from SDS-PAGE gels. The calculated isoelectric point of the TPO preparation starting at residue 76 is pH 5.8, and that at residue 109 is pH 5.6.

TPO	Residues	M _w (kDa)
Full length construct	5-840	92.3
Lower band of doublet	109-840	81.8
Upper band of doublet	76-840	85.4
Mixture of upper and lower bands	76/109-840	83.6

Table 2.4

The molecular weight of TPO as calculated from the amino acid sequence.

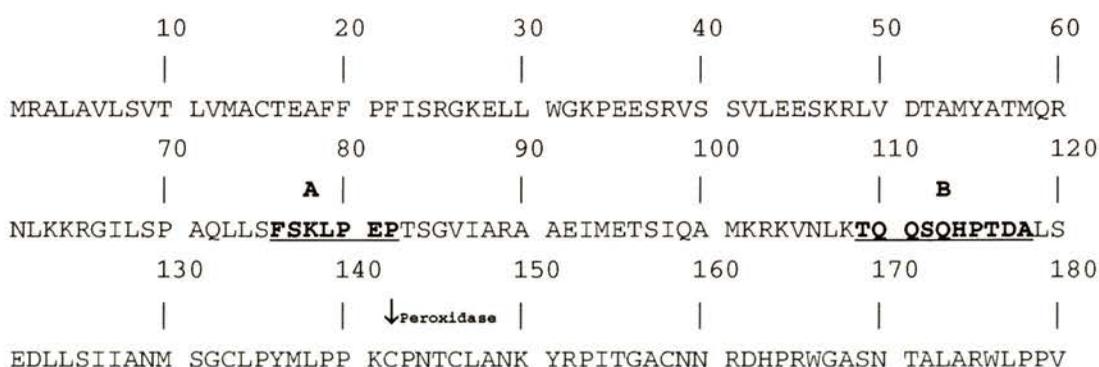


Figure 2.5

The first 180 residues of the TPO with the results of the N-terminal sequence analysis underlined (A and B).

2.7 Mass Spectrometry

Matrix-assisted laser desorption/ionization-time of flight mass spectroscopy (MALDI-TOF MS) is used to obtain an accurate molecular weight for a protein [Hillenkamp 1991]. The MALDI-TOF MS sample is mixed with a UV absorbing matrix and then hit with a UV laser which causes ionisation. The time the ions take to travel to the detector are measured and converted to a mass measurement, with larger ions travelling more slowly than smaller ones.

MS was carried out by the University of St. Andrews MALDI-TOF MS Service on a Tofspec 2E Micromass system. 0.5 μ l of 20 pmol/ μ l TPO 10FP in 0.1% trifluoroacetate (TFA) was mixed on the target plate with 0.5 μ l of matrix (a supersaturated solution of α -cyano-4-hydroxycinnamic acid in 35% acetonitrile and 65% of a 0.1% TFA solution). BSA was used to calibrate the TPO peak.

The results of the MALDI-TOF experiment are shown in Figure 2.6. TPO gives a broad peak with a molecular mass of 89.7 kDa. The lower trace shows rabbit β -galactosidase (116.4 kDa) for comparison, which gives a much narrower signal. The broadness of the TPO peak indicates that there is heterogeneity present in the TPO sample. However, the presence of individual peaks cannot be excluded due to the low resolving power of MALDI-TOF MS with protein of high molecular mass [Hillenkamp 1991].

The mass of TPO derived from MS (90 kDa) is considerable lower than that predicted from SDS-PAGE (97 kDa). The slow running of TPO on SDS-PAGE gels may be caused by the presence of carbohydrate, and anomalous behaviour of glycoproteins on SDS-PAGE has previously been observed [Segrest 1971]. It is thought that as carbohydrates do not bind SDS they lead to a lower net charge and therefore artefactually high molecular mass estimates [Hames & Rickwood 1990].

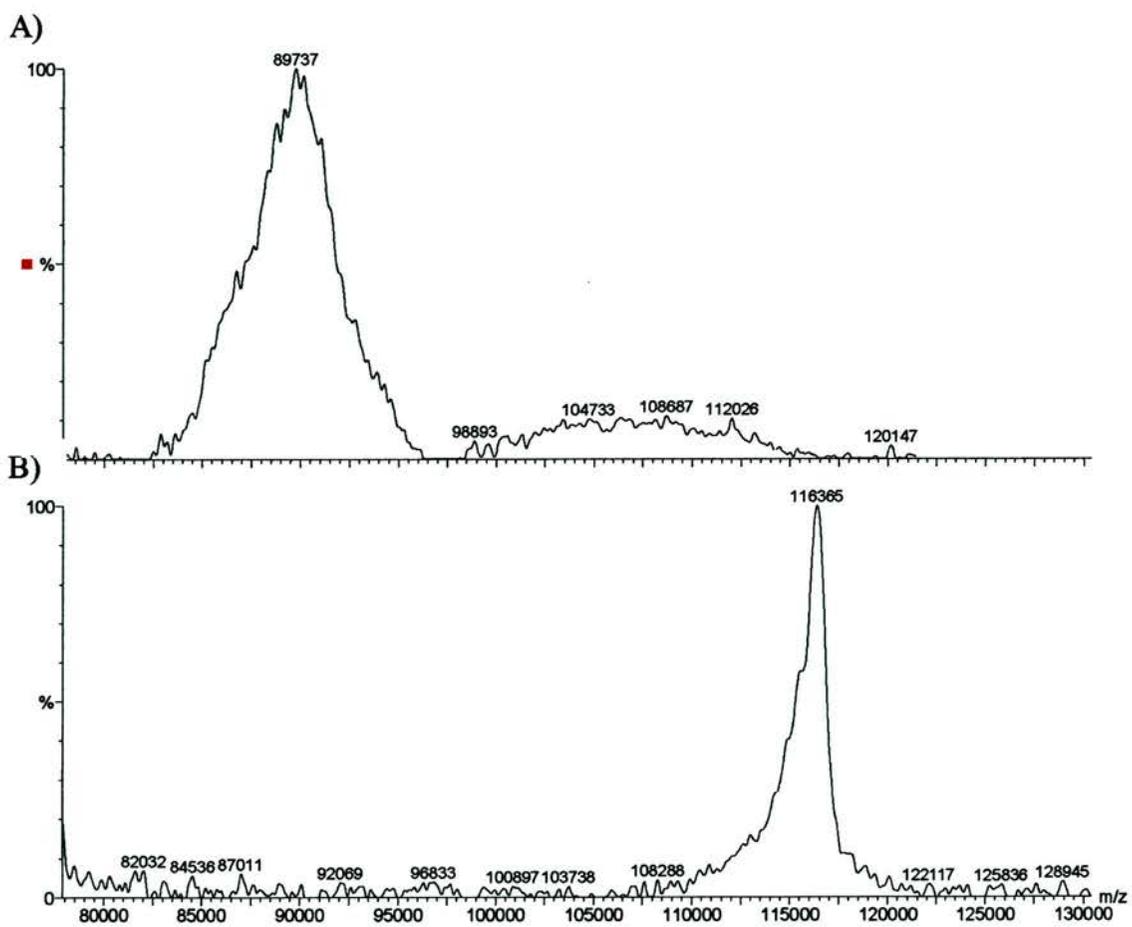


Figure 2.6

The MS spectra of A) TPO and B) β -galactosidase.

The MS derived mass supports the N-terminal sequencing results by showing that the observed mass of the TPO is too low to be that of the uncleaved full-length construct. Given that the total mass predicted from the TPO sequence (assuming the two species are in equal proportion) is 83.6 kDa the difference between this figure and the MS result is likely to represent the carbohydrate content. It can therefore be calculated that there is 6.1 kDa of carbohydrate present.

2.8 Deglycosylation Studies

To determine if the doublet and multiple bands on the SDS-PAGE gel and IEF gel were caused by different glycoforms, and to predict the amount of carbohydrate present, the mobility of glycosylated and deglycosylated TPO on SDS-PAGE and IEF gels were compared. The carbohydrates of glycosylated proteins can cause the anomalous running of SDS-PAGE gels, so only tentative conclusions can be drawn from these experiments. Carbohydrate composition was also analysed more rigorously by mass spectroscopy. TPO was proteolytically digested and the glycans sequentially released and identified. This technique gives the composition of the carbohydrates but not their position or occupancy.

Deglycosylation of TPO was performed with the enzyme peptide N-glycosidase A (PNGase A), which hydrolyses the β -aspartylglycosylamine linkages of N-linked carbohydrates, including those that contain a fucose 3-linked to the core, a common constituent of insect proteins [Altman 1993]. Reactions were set up in 0.1 M citrate phosphate buffer at pH 4, 5 and 6 using 20 μ l of 2 mg/ml of TPO 23, 20 μ l of buffer and 15 μ l of 40 ng/ml enzyme. The reactions were incubated at 37°C for 5, 8, 11 and 13 hours. A mixture of the enzymes endoglycosidase F (Endo F) and PNGase F which cleaves N-linked carbohydrates of high mannose and complex types was also used for deglycosylation. 16 μ g of TPO 18 were added to 3.75 μ l of enzyme in 40 mM sodium phosphate buffer, 100 mM EDTA at pH 7.2 and

incubated for 6 hours at 37°C. The protein was then run on a SDS-PAGE gel and an IEF gel.

The deglycosylation with PNGase A was complete after 5 hours and was successful at pH 5 and 6. pH 6 was chosen to run the preparative digest for crystallisation studies to minimize exposure of TPO to acidic conditions. Deglycosylation led to an observed decrease in molecular mass of the broad band on a reduced SDS-PAGE of about 4 kDa (Figure 2.7). Deglycosylation with PNGase F and Endo H also produced a shift of about 4 kDa on a reduced SDS-PAGE gel. The mass estimate from SDS-PAGE compares favourably with the amount of carbohydrate calculated from the difference between the MS mass and the mass calculated from TPO sequence (6.1 kDa). The difference may be due to N-linked glycans inaccessible to enzymatic cleavage, and the small amount of O-linked glycans predicted to be present by MS.

The IEF gel of TPO deglycosylated with PNGase F and Endo H showed no difference in the isoelectric point of the protein or the banding pattern (Figure 2.4). This suggests that the presence of multiple species of different charge is not due to varying amounts of accessible carbohydrate. The results also show that the sugar residues on TPO that are removed by deglycosylation are uncharged and do not contribute to the pI.

Mass spectroscopy was carried out at Imperial College, London. 50 µg of TPO 18 was digested with PNGase F and then PNGase A to remove the N-linked glycans, and reductive elimination was used to remove the O-linked glycans. Released sugars were permethylated, purified by reverse phase chromatography on a C18 Sep-Pak cartridge, and analysed by fast atom bombardment MS.

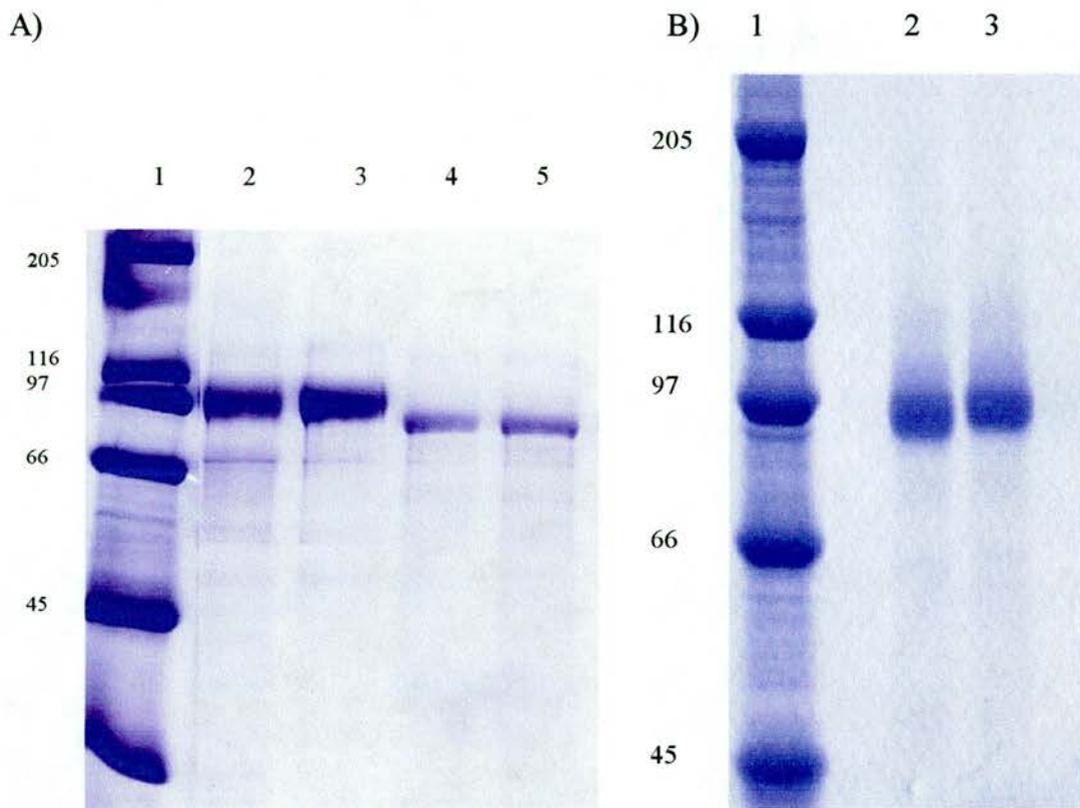


Figure 2.7

SDS-PAGE gels of deglycosylated TPO.

A) A 12% reduced SDS-PAGE of TPO 23 deglycosylated with PNGase A, after 5 hours.

- | | | | |
|----|-----------------------------|----|---------------------|
| 1. | High molecular mass markers | 4. | TPO + PNGase A pH 5 |
| 2. | TPO - PNGase A | 5. | TPO + PNGase A pH 6 |
| 3. | TPO + PNGase A pH 4 | | |

B) A 7% reduced SDS-PAGE of TPO 2FP.

- | | | | |
|----|---------------------|----|---------------------|
| 1. | Markers | 3. | TPO - PNGase/Endo F |
| 2. | TPO + PNGase/Endo F | | |

The MS spectra are shown in Figure 2.8. The PNGase F digest shows evidence for two main families of N-linked glycans (fuc = fucose, hex = hexose, hexnac = N-acetyl hexose):

Core fucosylated truncated glycans	$\text{Fuc}_{0-1}\text{Hex}_{2-4}\text{HexNac}_2$
High mannose glycans	$\text{Hex}_{5-9}\text{HexNac}_2$

There is a single peak resulting from reductive elimination that corresponds to an O-linked sugar at 534.4, the mass for HexHecNacitol. The peaks from the PNGase digest are assigned to specific glycans from their molecular weight (Figure 2.9). There are three different core fucosylated glycans, two of which are high mannose, the other complex, and there are seven different high mannose glycans that are not fucosylated. The PNGase A digest gave no meaningful signals suggesting that the fucose 3-linked core is absent. The reductive elimination gave one minor signal consistent with HexHecNacitol implying a small amount of O-linked glycosylation is present. Therefore, a total of 11 different glycans were identified. None of the glycans are charged, confirming the IEF analysis.

2.9 Summary

The recombinant TPO preparation used for crystallisation has all the characteristics of native TPO. It has similar enzyme activity, autoantibody binding behaviour and heme to protein absorbance ratio. A typical doublet is present on the SDS-PAGE gels of TPO, and is accentuated on the non-reduced gels. It is likely that this heterogeneity may be responsible for the low resolution diffraction of the TPO crystals.

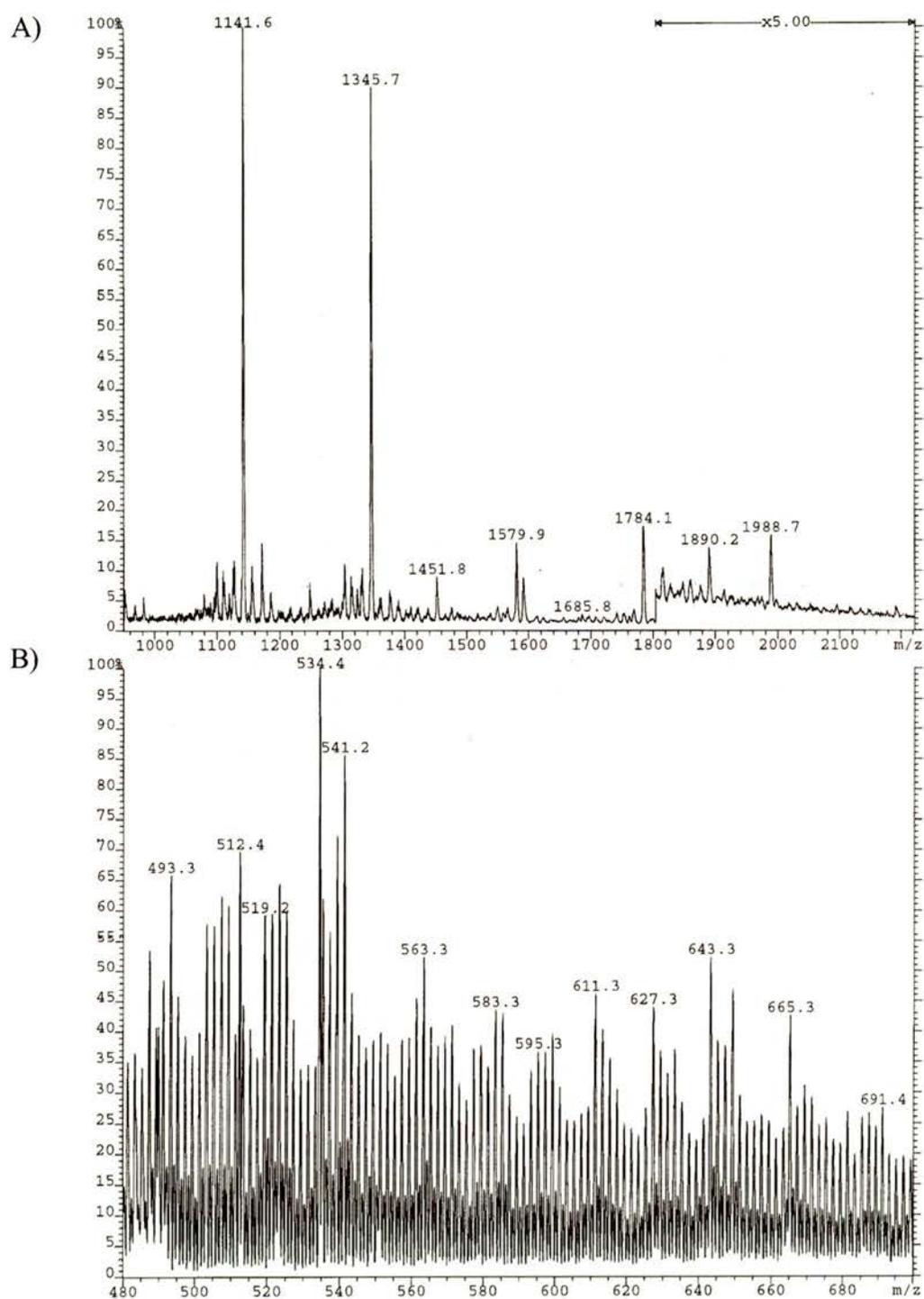


Figure 2.8

The MS spectra of the glycans released from TPO, PNGase digest (A), reductive elimination (B). The x axis corresponds to the mass in Da.

spectrum of proteins of different lengths. This is shown by MS analysis which gives a broad peak, and IEF analysis which shows multiple bands of different isoelectric points. In addition, the N-terminal sequencing identified other contaminating N-termini at concentrations too low to be sequenced, and reduced SDS-PAGE gels show a broad band rather than a doublet.

It would be interesting to see if the same N-terminal heterogeneity is responsible for the native TPO doublet. It is likely that the native TPO is cleaved at 109 as it was purified by trypsin treatment. Cleavage may be incomplete, and additional cleavage may occur at other sites, including residue 76.

The SDS-PAGE deglycosylation studies have predicted that there is 4 kDa of carbohydrate present on the recombinant TPO. The difference between the mass calculated from the protein sequence and that derived by MS indicates that there is 6 kDa of carbohydrate. MS studies of released glycans showed that they were predominantly N-linked high mannose sugars, with a small amount of N-linked complex and O-linked carbohydrates. The glycans on TPO are uncharged and so do not contribute to the multiple species present on IEF gels.

Chapter 3

Thyroid Peroxidase Crystallography

3.1 Introduction

The structures of complexes of TPO with autoantibodies will enable the autoantigenic sites on TPO to be mapped. The structure of uncomplexed TPO and autoantibodies will show if there are any conformational changes on complex formation. The structure of uncomplexed TPO will also shed light on the mechanism of action of the enzyme. Birefringent crystals of native TPO have previously been reported [Gardas 1997] but they were not of sufficient quality to diffract X-rays.

Before the start of the project, TPO crystals had been grown and a dataset to 7 Å resolution had been collected. The aim was to reproduce these crystals, and to increase the resolution limits of diffraction. It was hoped to grow crystals of the complexes of the antibodies 4F5 and 2G4 with TPO. However, the TPO preparation was shown to have heterogeneity at the N-terminus, which is likely to be responsible for the low resolution diffraction of the crystals. It was not possible to attempt to alter the TPO expression conditions in order to reduce cleavage at the N-terminus, and the TPO clone was not available to allow the generation of a new TPO construct.

Therefore, other methods were used in an attempt to increase resolution. Resolution can be extended by inducing the formation of a new set of crystal contacts, which may form crystals of higher order [Ducruix & Giege 1992]. This can be attempted by growing the crystals in different conditions, or by altering the nature of the protein surface. New crystallisation conditions for TPO were screened

and those used previously were optimised. TPO was deglycosylated in order to remove possible heterogeneity due to multiple glycoforms and to introduce new contacts in the crystal [Baker 1994a]. Co-crystallisation of proteins with Fab fragments has been shown to increase the resolution of diffraction by changing the crystal contacts [Kovari 1995], therefore crystallisation of complexes of TPO with 4F5 and 2G4 was attempted.

3.2 TPO-Fab Complex Purification

The purification of the Fab fragments of 2G4 and 4F5 is described in Chapter 5. Two methods for growing crystals of complexes were used – purification of the complexes prior to crystallisation, and the mixing of the individual proteins in the crystallisation drop.

The complexes of 2G4 and 4F5 Fab with TPO were purified from the uncomplexed molecules prior to crystallisation using different methods. The 2G4-TPO complex was co-purified on gel filtration. 2 mls of a 0.5 mg/ml TPO 23 was mixed with 2 mls of 0.25 mg/ml 2G4 Fab (1:2 molar ratio), both proteins were buffered in 50 mM sodium chloride, 20 mM Tris pH 7.6, 0.1 mM potassium iodide. The mixture was incubated for 3 hours at 4°C and then loaded onto a S-200 gel filtration column.

The gel filtration of TPO and 2G4 Fab produced three peaks, shown in Figure 3.1. These were calibrated by using a standard curve ($\log M_r$ plotted against V_e/V_o) of proteins of known molecular weight and the results are shown in Table 3.1. TPO ran at 109 kDa, the Fab at 49 kDa and the TPO-Fab complex at 151 kDa corresponding to a 1:1 ratio .

There was insufficient TPO to purify the 4F5-TPO complex by gel filtration, so a smaller scale method was used. 66 μ l of 12.7 mg/ml of TPO 1FP and 110 μ l of 8.3 mg/ml 4F5 Fab (1:2 molar ratio) in 50 mM sodium chloride, 20 mM Tris pH 7.6,

0.1 mM potassium iodide, were mixed and incubated for one hour at room temperature. The complexes were spun at 3000 rpm in a Vivaspin concentrator with molecular mass cut-off of 100 kDa for one hour and samples removed at 10 minute intervals and run on an SDS-PAGE gel. After 50 minutes, all uncomplexed Fab had passed through the membrane.

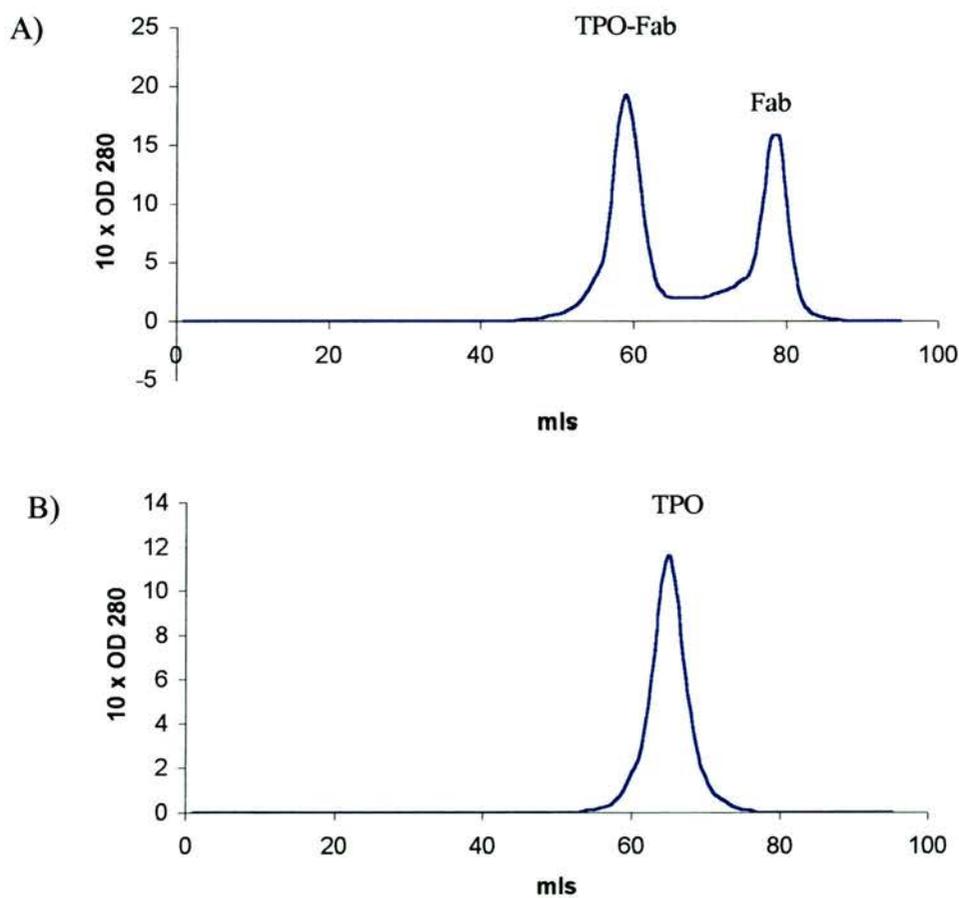


Figure 3.1
Gel filtration traces of A) TPO and 2G4, B) TPO.

	TPO-Fab	TPO	Fab
Ve (peak) mls	59	65	79
Vo (void) mls	46	46	46
Ve/Vo	1.30	1.41	1.71
Log Mr	2.2	2.0	1.7
Mr kDa	151	109	49

Table 3.1

The results of the gel filtration of TPO and 2G4 Fab complexes.

3.3 Crystallisation

An analysis of the crystallisation conditions used previously indicated that the most ordered TPO crystals grew at room temperature, with a protein concentration of 15 mg/ml in a precipitant solution of 12.5% to 17.5% PEG 4K, 0.1 M Tris pH 6.5-7.5. These conditions were repeated with TPO 18 and with new TPO preparations 2FP and 10FP, using vapour diffusion and microbatch methods. TPO was concentrated to approximately 15 mg/ml in 100 mM sodium chloride, 5 mM sodium phosphate and 0.1 mM potassium iodide at pH 7.3 using a Centricon concentrator with a molecular weight cut-off of 10 kDa. The concentration was monitored during concentration by measuring the OD₂₈₀, as described in section 2.3. Vapour diffusion was carried out using the hanging drop method and was set up with Linbro plates and 22 mm silicated coverslips. 0.5 ml of reservoir solution and 1 to 2 µl of protein mixed with an equal volume of reservoir in the drop, the coverslips were sealed with grease. For the microbatch method 1 µl of protein was mixed with 1 µl of precipitant and overlaid with 6 µl of Al's Oil (50% silicon, 50% paraffin oil) in 96 well Nunc plates.

The crystals were reproduced with TPO preparations 18 and 2FP using the hanging drop method. TPO preparation 10FP did not crystallize, probably due to low purity.

Preparations 21, 23 and 26 were too impure to use for crystallisation trials. The crystallisation conditions used are given in Table 3.2. TPO 18 crystals appeared after five days, were trigonal trapezohedral, 0.5 mm in the longest dimension and yellow in colour due to the heme. The crystals are shown in Figure 3.2.

Protein	Reservoir solution	Results
TPO 18 16.9 mg/ml	15% PEG 4K, pH 7.0-7.5	Crystals : pH 7.13 & 7.25
TPO 2FP 10.5 mg/ml	15% PEG 4K, pH 7.0-7.5	Crystals : pH 7.0 & 7.3 2µl:1µl (protein to reservoir)
TPO 10FP 15 mg/ml	18% PEG 4K, pH 7.0-7.5	No crystals

Table 3.2

The crystallization conditions used to reproduce TPO crystals with different TPO preparations. Unless stated otherwise, all experiments are hanging drop, at ~ 21°C, with drops of 2 µl protein and 2 µl reservoir.

The successful crystallisation conditions were optimised by varying protein concentration, pH, precipitants and temperature, and new crystallisation conditions were screened using grid screens from Hampton Research. The trials are summarised in Table 3.3. Crystallisation trials were set up at 4°C, 21°C and 28°C. Crystals were grown in conditions containing Tris pH 7-8, PEG 4K, 8K and 10K at 21°C. The only crystals that grew in a condition with a precipitant other than PEG were in 50% MPD, 0.1 M Tris pH 8.5, 0.2 M ammonium phosphate.



500 μm



Figure 3.2

Crystals of 18 mg/ml TPO 18 in 12% PEG 4K 0.1 M Tris pH 6.75, with drops of 1 μl of TPO to 1 μl of reservoir, at room temperature.

Protein	Reservoir solution	Results
TPO 2FP 10.5 mg/ml	15% PEG 4K, pH 7.0-7.5	Crystals : pH 7.0 & 7.3, 2µl:1µl (protein to reservoir)
TPO 2FP 10.5 mg/ml	4-12% PEG 8K, pH 7-9	Crystals: pH 8.0, 8% PEG 8K
TPO 2FP 10.5 mg/ml	8-10% PEG 8K, pH 7.75-8.25	Crystalline precipitate
TPO 2FP 10.5 mg/ml	5-15% PEG 2K, 6K, 10K, pH 6.5-8.5	Crystals: 10% PEG 10K pH 7.5
TPO 1FP 12.7 mg/ml	Hampton Screen I, hanging drop and microbatch	Crystalline precipitate – PEG 6K, 8K, 20K crystals – 30 % 2 methyl-2, 4-pentadiol (MPD), 0.1M Tris pH8.5, 0.2 M ammonium phosphate
TPO 1FP 12.7 mg/ml	Optimised 6K, 8K, 10K	Crystalline precipitate
TPO 10FP 30 mg/ml	18% PEG 4K, pH 7.0-7.5	No crystals
TPO 18 16.9 mg/ml	15% PEG 4K, pH 7.0-7.5	Crystals : pH 7.13, 7.25
TPO 18 deglycosylated 8mg/ml	Hampton crystal screen I and II	No crystals

Table 3.3

Crystallisation trials to increase the resolution of diffraction of TPO crystals. Unless stated otherwise, all experiments are hanging drop, at ~ 21° C, with drops of 2 µl protein and 2 µl reservoir.

In order to reduce the possible heterogeneity from mobile glycans and multiple glycoforms, deglycosylated TPO was used for crystallisation trials. TPO 18 was deglycosylated with PNGase A as described in section 2.6, and set down for trials using Hampton screens I and II. No crystals were observed, however there was not enough TPO to run gel filtration to remove the PNGase A protein and this may have interfered with crystallisation.

TPO-Fab complex trials were set up using co-crystallisation (Table 3.4), and with complexes purified as described in section 3.2 (Table 3.5). Co-crystallisation was carried out by adding 4F5 or 2G4 in a 1:1 molar ratio to TPO 1FP and 10FP, and to deglycosylated TPO18 and incubating for one hour at room temperature prior to setting up the drops. The TPO-2G4 gel filtration peak was concentrated to 7 mg/ml and used to set up crystallisation trials using Hampton screen I and a PEG screen. The TPO-4F5 complex purified on a size exclusion concentrator was concentrated to 13 mg/ml and screened with Hampton screen I and II.

Crystals of the TPO-4F5 complex grew with and without prior purification of the complex. However, analysis of the diffraction pattern of these crystals showed them to be 4F5 Fab only. Yellow crystalline precipitate of 2G4-TPO 23 complexes purified on gel filtration was observed in the following conditions:

1.5 M lithium sulphate, 0.1 M Na HEPES pH 7.5

20% iso-propanol, 0.1 M Na acetate pH 4.6, 0.2M calcium chloride

10% iso-propanol, 0.1 M Na HEPES pH 7.5, 20 % PEG 4000

However, there was not enough TPO to optimise the conditions.

3.4 Data Collection

Data were collected from crystals grown from TPO 18 and TPO 2FP, but the diffraction was no better than 8 Å. Therefore, the collection of data from crystals of TPO 18 grown prior to the start of the project is described.

Protein	Reservoir solution	Results
TPO 10FP 2G4 8 mg/ml	<ul style="list-style-type: none"> • 1.5 M lithium sulphate, 0.1 M Hepes pH 7.5 • 20% isopropanol, 0.1 M Na acetate pH 4.6, 0.2 M Ca chloride • 10% isopropanol, 0.1 M Na Hepes pH 7.5, 20% PEG 4K 	No crystals
TPO 1FP 4F5 12.7 mg/ml	Hampton crystal screen I and II	Crystalline precipitate – 4F5 only: <ul style="list-style-type: none"> • 0.1 M Hepes pH 7.5, 10% PEG 6K, 5% MPD • 0.1 M Hepes pH 7.5, 10% PEG 8K, 8% ethylene glycol
TPO 1FP 4F5 12.7 mg/ml	Grid around conditions: <ul style="list-style-type: none"> • 0.1 M Hepes pH 7.5, 10% PEG 6K, 5% MPD • 0.1 M Hepes pH 7.5, 10% PEG 8K, 8% ethylene glycol 	No crystals
TPO 10FP 4F5 8.5 mg/ml	<ul style="list-style-type: none"> • 0.1 M Hepes pH 7.5, 10% PEG 6K, 5% MPD • 0.1 M Hepes pH 7.5, 10% PEG 8K, 8% ethylene glycol 	No crystals
TPO 18 deglycosylated 2G4 8 mg/ml	Hampton crystal screen I and II	No crystals

Table 3.4

Summary of the crystallisation trials of TPO/Fab complexes, the proteins were mixed in the crystallisation drop at a 1:1 molar ratio. All experiments are hanging drop, at ~ 21° C, with drops of 2 µl protein and 2 µl reservoir.

Protein	Reservoir solution	Results
TPO 23 2G4 Gel Filtration 7 mg/ml	Hampton screen I PEG screen	Crystalline precipitate : <ul style="list-style-type: none"> • 1.5 M lithium sulphate, 0.1 M Hepes pH 7.5 • 20% isopropanol, 0.1 M Na acetate pH 4.6, 0.2 M Ca chloride • 10% isopropanol, 0.1 M Na Hepes pH 7.5, 20% PEG 4K
TPO 1FP 4F5 Vivaspin 12.7 mg/ml	Hampton screen I Hampton screen II	Crystals – 4F5 only: <ul style="list-style-type: none"> • 30% PEG 4K, 0.1 M tri-sodium citrate dihydrate pH 5.6, 0.2 M ammonium acetate • 0.1 M Tris pH 8.5, 30% PEG 4K, 0.2 M Mg chloride • 0.1 M Tris pH 8.5, 30% PEG 4K, 0.2 M Na acetate trihydrate • 0.2 M ammonium sulphate, 0.1 M Na citrate pH 5.6, 0.2 M Na tartrate

Table 3.5

Summary of the crystallisation trials of TPO/Fab complexes purified prior to crystallisation, all experiments are hanging drop, at ~ 21° C, with drops of 2 µl protein and 2 µl reservoir.

TPO crystals were mounted in quartz capillaries on a rotating anode X-ray generator operating at 45 kV 80 mA with a 0.3 mm collimator. Diffraction data were collected on a Mar Research image plate system. Crystals were also flash-frozen to examine the effect of freezing on resolution. The crystals were placed for two minutes in the mother liquor containing 5% v/v glycerol and were then fished into a 0.5 mm diameter loop and flash-frozen at 100K using an Oxford Cryostream cooling device. The technique of flash-annealing was used in an attempt to improve diffraction limits [Yeh & Hol 1998].

The flash-frozen crystals diffracted to approximately 7.0 Å and data were collected over 180 frames with an oscillation angle of 0.5°, an exposure time of 10 minutes per image and a crystal to detector distance of 270 mm. A diffraction image is shown in Figure 3.3. Neither flash-annealing nor subsection to synchrotron radiation extended the resolution limit of diffraction. The cell indexed in crystal class P32 and the data collection statistics are shown in Table 3.6. Rotation was around the c-axis therefore no 00l reflections were measured, consequently the space group was narrowed down to P321, P3₁21 or P3₂21. The unit cell parameters were a = b = 100 Å, c = 215 Å, $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$. The space group diagrams for all possible space groups are shown in Figure 3.4.

	All Data	Highest Resolution Shell (7.54-7.10 Å)
Space group	P3 _x 21	-
Unit cell	a=b=100 Å, c=215 Å, $\alpha = \beta = 90^\circ, \gamma=120^\circ$	-
Measured reflections	33 216	-
Unique reflections	5735	567
R _{merge} † (%)	12.7	36.0
I/sigI	6.6	2.6
Completeness (%)	96	96.8

Table 3.6

The data collection statistics.

† $R_{\text{merge}} = \frac{\sum |I(k) - \langle I \rangle|}{\sum I(k)}$, where I(k) is the value of the kth measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

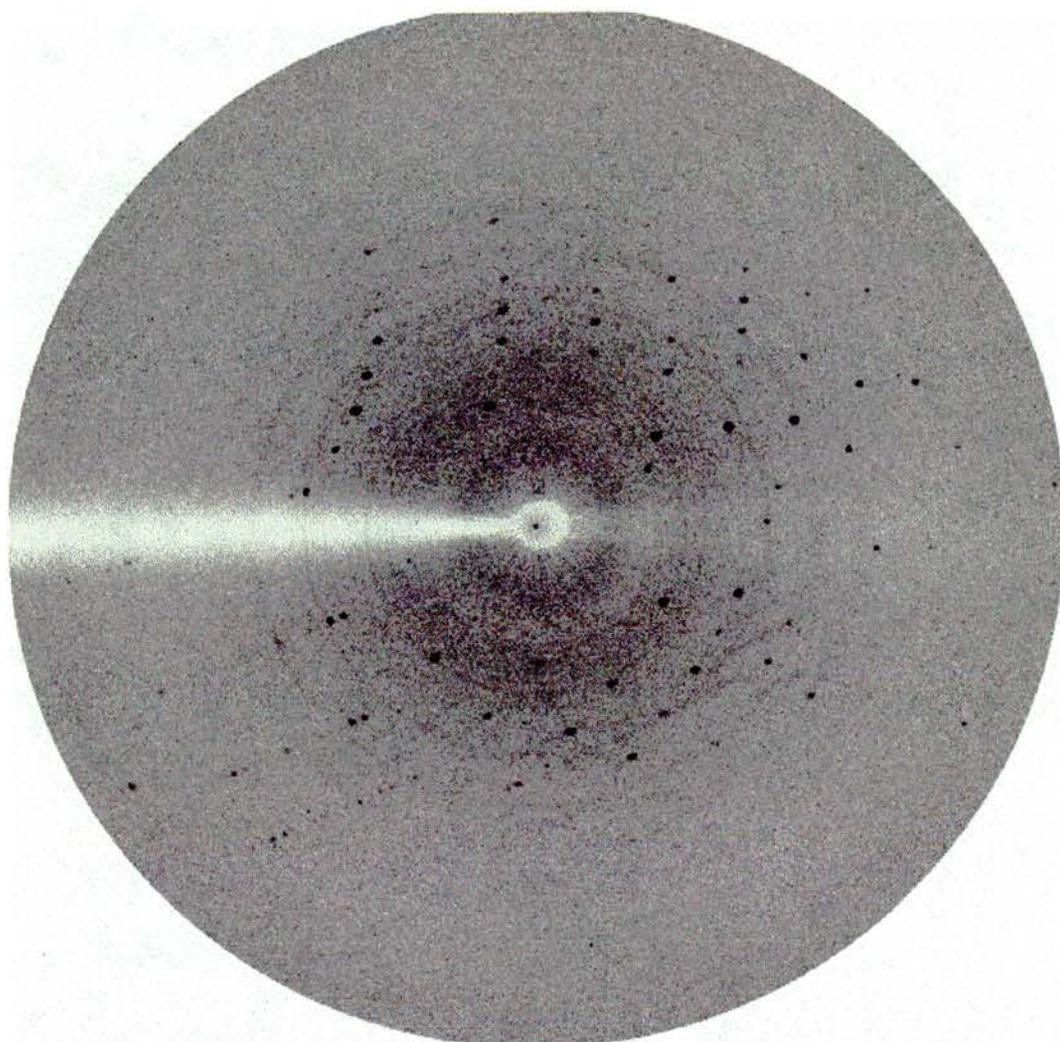


Figure 3.3

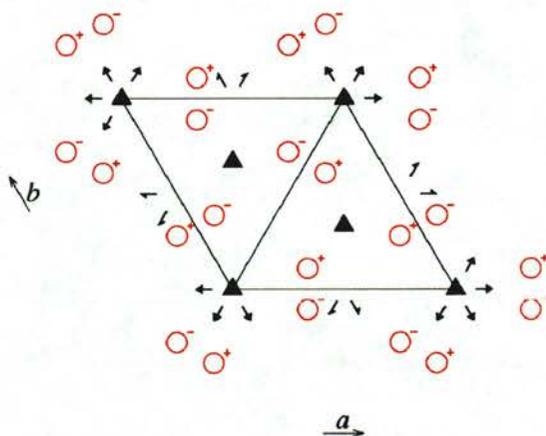
A diffraction image of the TPO crystals, the resolution at the edge of the plate is 4.86 Å.

P321

P 3 2 1

321

No. 150



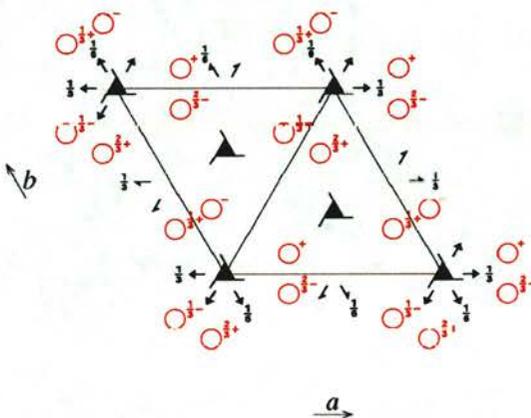
- 1 x, y, z
- 2 $\bar{y}, x - y, z$
- 3 $\bar{x} + y, \bar{x}, z$
- 4 y, x, \bar{z}
- 5 $x - y, \bar{y}, \bar{z}$
- 6 $\bar{x}, \bar{x} + y, \bar{z}$

P₃21

P 3₁ 2 1

321

No. 152



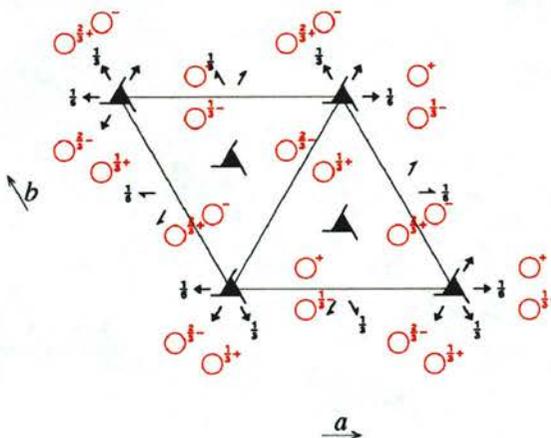
- 1 x, y, z
- 2 $\bar{y}, x - y, \frac{1}{3} + z$
- 3 $\bar{x} + y, \bar{x}, \frac{2}{3} + z$
- 4 y, x, \bar{z}
- 5 $x - y, \bar{y}, \frac{2}{3} - z$
- 6 $\bar{x}, \bar{x} + y, \frac{1}{3} - z$

P₃21

P 3₂ 2 1

321

No. 154



- 1 x, y, z
- 2 $\bar{y}, x - y, \frac{2}{3} + z$
- 3 $\bar{x} + y, \bar{x}, \frac{1}{3} + z$
- 4 y, x, \bar{z}
- 5 $x - y, \bar{y}, \frac{1}{3} - z$
- 6 $\bar{x}, \bar{x} + y, \frac{2}{3} - z$

Figure 3.4

The space group diagrams for P321, P₃21 and P₃21.

Solvent content was determined by calculating the Matthews' number [Matthews 1968] (V_m):

$$V_m (\text{\AA}^3/\text{Da}) = \text{Volume of asymmetric unit} (\text{\AA}^3) / M_r (\text{Da})$$

The specific volume of a protein molecule is approximately $0.74 \text{ cm}^3/\text{g}$, which gives:

$$\text{Volume of solvent} = 1 - 1.23 / V_m$$

The volume of the TPO unit cell is $1.87 \times 10^6 \text{ \AA}^3$, there are six asymmetric units in the unit cell and the molecular weight of the TPO monomer is 90 kDa (as calculated by mass spectroscopy). This gives a V_m of 3.4 and a solvent content of 64% for one molecule in the asymmetric unit and 27% for two molecules in the asymmetric unit. These values lie at the extremes of the typical solvent content for protein crystals as calculated by Matthews (27% to 65%).

The soluble recombinant TPO fragment used for crystallisation is monomeric on a non-reduced SDS-PAGE and gel filtration, but believed to be dimeric when membrane bound [Baker 1994b]. There is also evidence that recombinant TPO can form dimers in solution [Kaufmann 1989]. The crystal structure of MPO shows a dimer generated by two-fold rotational symmetry [Zeng & Fenna 1992], which, due to the high sequence homology, may also be present in TPO. It is therefore possible that the crystallisation of a solution of monomers of TPO may lead to the formation of a physiological dimer in the crystal.

If there are two molecules in the asymmetric unit they may form a dimer related by a non-crystallographic two-fold symmetry axis. This possibility was tested by calculating a self-rotation function in POLARRFN [Collaborative Computer Project Number 4 1994] using all data and a sphere of 20 \AA . The $\kappa = 180^\circ$ section is shown in Figure 3.5 and shows six peaks at $\phi = 0^\circ, 60^\circ, 120^\circ, 180^\circ, 240^\circ$ and 300° ,

marked A to F. These are a result of the crystallographic two-folds, but there is no evidence for a non-crystallographically related dimer. It is possible that a non-crystallographic two-fold is present, but that it is parallel to a crystallographic two-fold. However, calculation of a Patterson map showed no significant peaks (other than the origin) indicative of such an orientation.

3.5 Molecular Replacement

The data were indexed and scaled using the programmes DENZO and SCALEPACK from the HKL suite [Otwinowski & Minor 1996]. The reflections were converted to mtz format using the CCP [Collaborative Computer Project Number 4 1994] programme SCALEPACK2MTZ.

Molecular replacement was carried out using C α atoms of MPO as a phasing model with the programme AMORE [Navaza 1994] from the CCP4 suite. First, the reflection file was prepared by SORTING, and the MPO pdb file coordinates translated so their centre of gravity was at the origin using TABLING. The rotation function was then calculated in ROTING using MPO structure factors rotated with respect to structure factors generated from the TPO reflection data. The integration radius was 30 Å and data from 25 Å to 7 Å were used in the calculation. The output was given in Eulerian angles [Rossmann & Blow 1962], with each solution scored with a correlation coefficient and R-factor. Translation was then carried out in TRAIING, using the top 20 rotation solutions, and the output was a translation matrix in fractional coordinates. A separate translation function was run for each of the possible space groups. The solution underwent rigid body refinement in FITING, and the rotation and translation matrices were applied to the MPO pdb file using the CCP4 programme PDBSET. The procedure is shown in Figure 3.6. Symmetry related molecules were generated in O [Jones 1991]. Figures were created with BOBSCRIPT [Esnouf 1999] and MOLSCRIPT [Kraulis 1991].

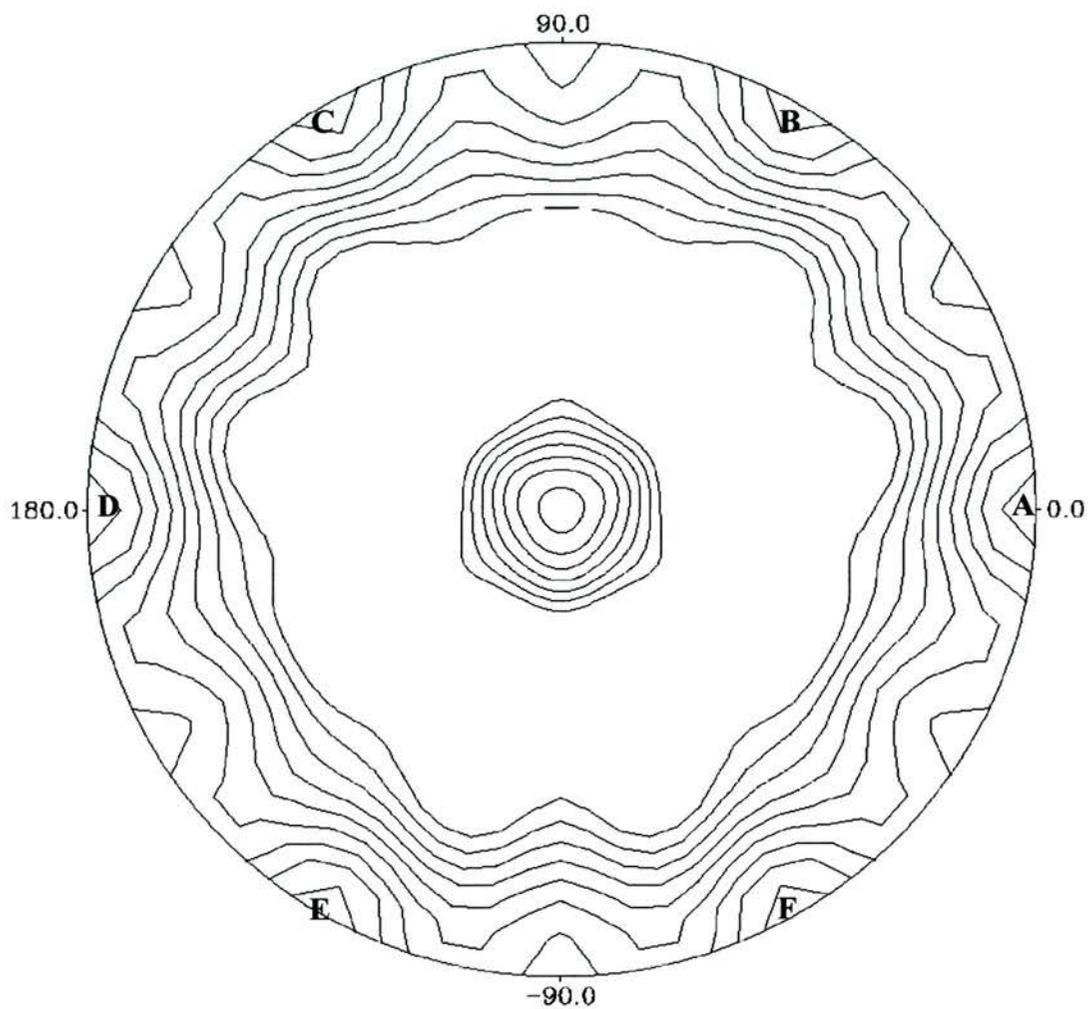


Figure 3.5

The $\kappa = 180^\circ$ section of the self-rotation function for TPO, using data from 15-7 Å and a 20 Å sphere. Axis directions are: ω (angle from pole), ϕ (angle around equator). The six crystallographic-two fold axes are labelled A to F.

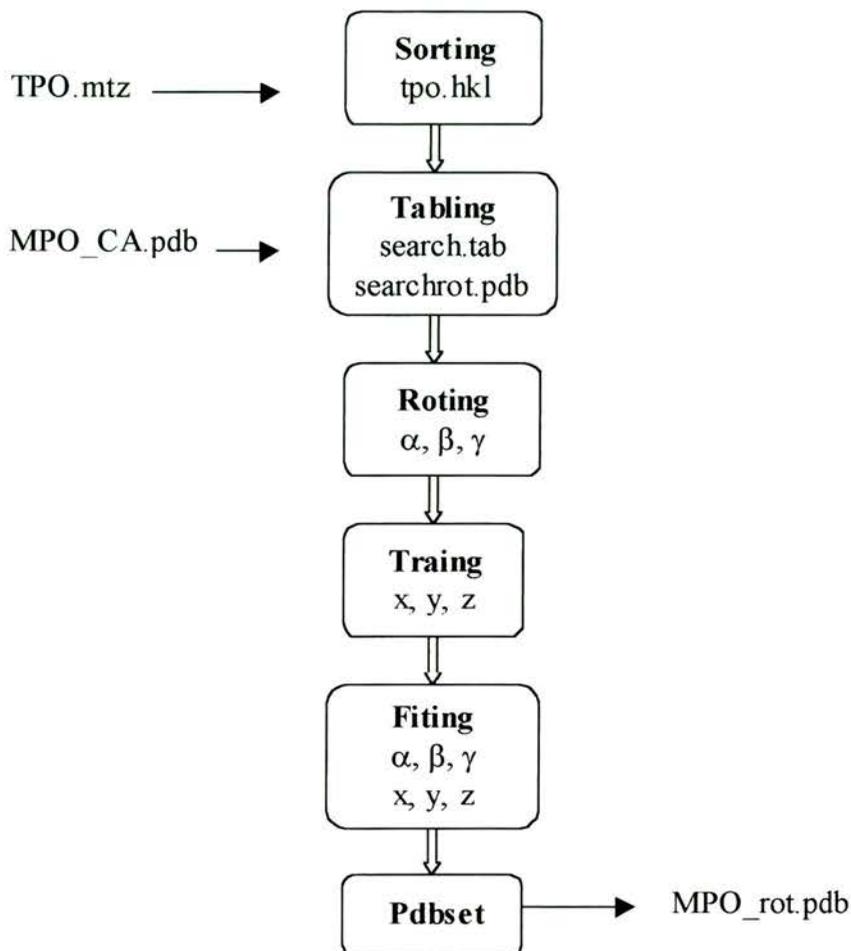


Figure 3.6

The molecular replacement procedure in AMORE.

A molecular replacement solution was obtained for space group $P3_121$, the top ten rotation and translation solutions are shown in Table 3.7. The correlation coefficient after rotation and translation was 33.2% compared to a next highest coefficient of 23.3%. Following one round of rigid-body refinement, the correlation coefficient

was 47.8% and the R factor was 46.8%. Due to the possibility of two molecules in the asymmetric unit, the first solution was fixed and a second molecule searched for in AMORE. No significant solutions were found, indicating there is likely to be one molecule in the asymmetric unit. In addition, the solution was fixed and a CCP domain (1vvc) [Wiles 1997] and an EGF domain (1emn) [Knott 1996] used as search models to locate the C-terminal regions of TPO. However, no solutions were found.

The packing of MPO monomers in the TPO unit cell is shown in Figure 3.7. There are two different dimers formed by different crystallographic two-folds, either of which could be a physiological dimer. One dimer involves the interchain disulphide (cysteine C153) loop in MPO (cysteine 296 in TPO) and is shown by molecules A and B in the figure. The other dimer, formed by molecules A and C, involves residues C447-460 (605-618 in TPO), which make up helix 14 and the N-terminus of helix 15 of MPO.

The positions of the two sets of MPO dimers involving the interchain disulphide bridge (the MPO dimers generated by the TPO crystallographic symmetry and the dimers in the MPO structure) are compared in Figure 3.8. The two dimers are formed by monomers at different relative orientations, and the potential TPO dimers are less closely packed than the MPO dimers. The closest residues in the potential TPO dimer are the C157 pair, the C α atoms of which are 4.2 Å apart. The cysteines involved in the interchain disulphide are 18.4 Å apart. However, there is a ten residue insertion in this interchain disulphide loop in TPO and the extra residues may allow formation of the disulphide bond. The dimer interface of the TPO peroxidase domains may not be as closely packed as that of MPO as the TPO dimer may be stabilised by attachment of the C-terminus to the membrane. In addition the TPO dimer interface may involve interactions between the C-terminal and N-terminal domains.

		α	β	γ	tx	ty	tz	Corr	R	
Rotation										
SOLUTIONRC	1	75.17	45.10	329.17	0.0000	0.0000	0.0000	24.3	0.0	1
SOLUTIONRC	1	10.70	81.80	135.80	0.0000	0.0000	0.0000	21.2	0.0	3
SOLUTIONRC	1	72.79	89.25	140.94	0.0000	0.0000	0.0000	21.1	0.0	5
SOLUTIONRC	1	9.79	53.40	341.66	0.0000	0.0000	0.0000	18.4	0.0	7
SOLUTIONRC	1	34.59	126.83	194.68	0.0000	0.0000	0.0000	18.4	0.0	9
SOLUTIONRC	1	68.91	85.26	79.44	0.0000	0.0000	0.0000	17.9	0.0	11
SOLUTIONRC	1	16.98	37.21	24.46	0.0000	0.0000	0.0000	16.0	0.0	13
SOLUTIONRC	1	76.09	27.48	210.16	0.0000	0.0000	0.0000	15.3	0.0	15
SOLUTIONRC	1	5.37	95.10	15.42	0.0000	0.0000	0.0000	15.2	0.0	17
SOLUTIONRC	1	51.53	103.93	11.89	0.0000	0.0000	0.0000	14.8	0.0	19
Translation										
SOLUTIONTF1	1	75.17	45.10	329.17	0.6675	0.5538	0.1500	33.2	52.1	1
SOLUTIONTF1	1	10.70	81.80	135.80	0.3140	0.9589	0.3088	19.5	56.3	2
SOLUTIONTF1	1	72.79	89.25	140.94	0.7143	0.7718	0.4575	18.9	57.0	1
SOLUTIONTF1	1	9.79	53.40	341.66	0.1868	0.7938	0.4812	17.9	56.9	2
SOLUTIONTF1	1	34.59	126.83	194.68	0.8485	0.8243	0.0414	21.4	57.3	1
SOLUTIONTF1	1	68.91	85.26	79.44	0.3057	0.8243	0.1984	16.7	56.0	6
SOLUTIONTF1	1	16.98	37.21	24.46	0.0033	0.6315	0.1395	19.6	57.7	3
SOLUTIONTF1	1	76.09	27.48	210.16	0.3243	0.2364	0.1828	16.8	57.6	4
SOLUTIONTF1	1	5.37	95.10	15.42	0.7596	0.4041	0.1551	15.7	58.0	6
SOLUTIONTF1	1	51.53	103.93	11.89	0.2577	0.9549	0.3426	16.6	55.9	10
Fiting										
SOLUTIONF	1	66.53	42.13	335.52	0.6601	0.5469	0.1500	47.8	46.8	1

Table 3.7

The AMORE output, the identified solution is shown in bold.

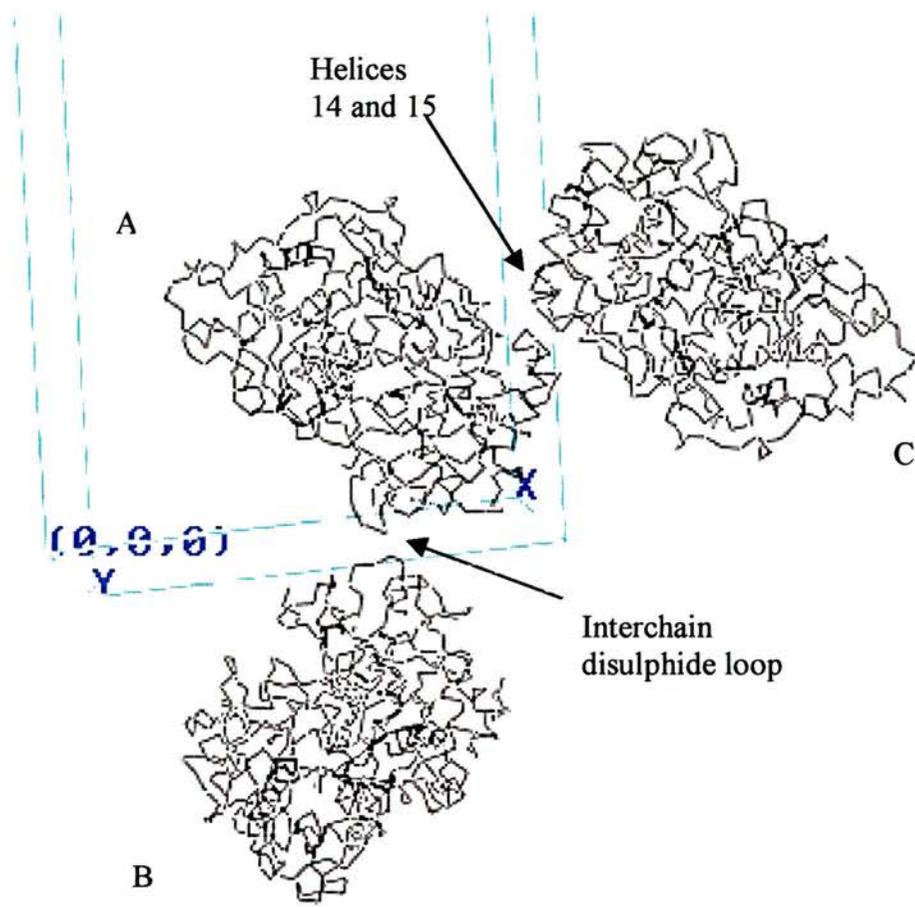


Figure 3.7

The packing of MPO monomers A, B and C in the TPO unit cell.

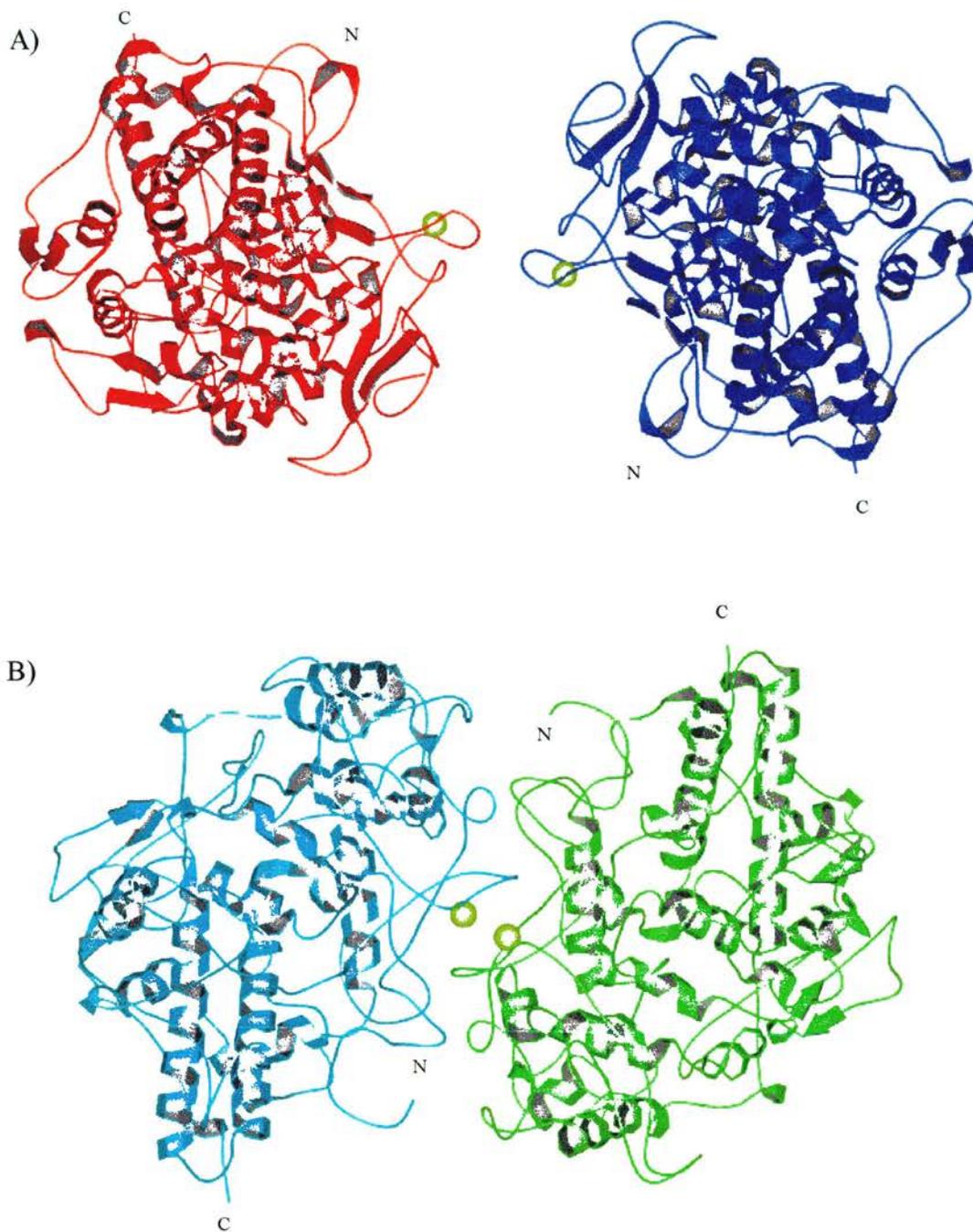


Figure 3.8

A comparison of the MPO dimers and the potential TPO dimers, the cysteines involved in the interchain disulfide are shown as yellow spheres. A) TPO, B) MPO.

3.6 Refinement

Density modification was carried out with the programme DM [Collaborative Computer Project Number 4 1994], involving 30 cycles of solvent flattening and histogram matching. The resulting R-factor was 38.9%. A $2F_{\text{obs}} - F_{\text{calc}}$ map was calculated to 7.1 Å resolution using the CCP4 programme FFT. Figure 3.9 shows a section of the map around the central helix hairpin made up of helices 6 and 7 in MPO (residues 401 to 459 in TPO). The TPO electron density map did not show any additional features beyond the MPO model. In particular, there was no electron density for atoms in the heme prosthetic group, which was not included in the MPO model used in molecular replacement. Assuming the heme is in the same position in MPO and TPO, then a peak for the strongly diffracting iron atom would be expected in the TPO map. It is possible that the heme is in a different position in the TPO model, but this is unlikely considering the conservation of heme ligands in TPO. Attempts to further refine the TPO solution using energy minimisation and simulated annealing did not lead to a decrease in R-factor.

3.7 Summary

It was not possible to grow crystals of TPO that diffracted to resolution higher than 7 Å. In order to increase the resolution it is likely that the N-terminal heterogeneity must be eliminated. However, due to restricted TPO supply, studies of deglycosylated TPO were not exhaustive, and may still provide a means of increasing resolution. In addition, promising crystalline precipitate was obtained using complexes of TPO and 2G4 purified on gel filtration and it may be possible to optimise these conditions in order to grow crystals.

A potential molecular replacement solution for the 7 Å TPO dataset was obtained and indicated that physiological dimers may be formed using a crystallographic two-fold. Analysis of the dimer formation showed that the same interchain

disulphide as MPO may be used, but that the dimers are less closely associated and at a different orientation to each other. However, due to the low resolution of the data used to calculate the molecular replacement solution, the lack of heme density, and the high R-factor, the solution should be treated with caution.

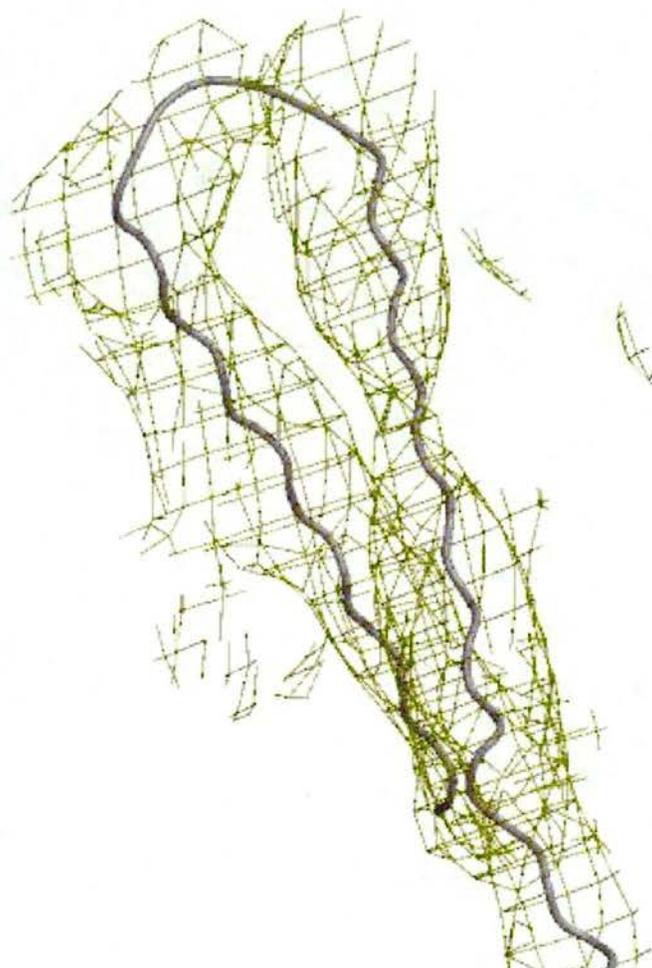


Figure 3.9

A sample region of a $2F_{\text{obs}} - F_{\text{calc}}$ TPO electron density map, contoured at 1 sigma.

Chapter 4

Thyroid Peroxidase Modelling

4.1 Introduction

In the absence of a high resolution X-ray or NMR structure, a model of a protein can be built using information from the amino acid sequence and other known structures [Morea 1998a]. The first steps in modelling are to carry out a database search and perform a multiple sequence alignment in order to identify homologous sequences. If any homologues have a known structure then this structure can be used in comparative modelling. If there are no homologous proteins of known structure then the method of *de novo* structure prediction is used. TPO is a multidomain protein, three of the extracellular domains can be modelled by homology methods whilst the fourth domain must be modelled *de novo*.

A model of TPO can be used to study many features of the protein. The differences in structure between MPO and TPO may provide information on how the catalysis of similar reactions using different substrates is achieved. Given the molecular replacement solution indicating that TPO may have a different dimer interface from MPO, the residues involved in the contact between dimers in MPO and their equivalents in TPO can be compared. The physiological role of the N-terminal domain may be better understood if it can be modelled, and the position of the N-terminal proteolysis sites can be visualised.

Features responsible for the immunodominant domain on TPO can be also be investigated on a model. The position of conformational epitopes that have been mapped to date can be located. A model of the quaternary structure of TPO can be used

to identify which regions of the surface are likely to be accessible to autoantibody binding.

There are many sources of errors and inaccuracies in protein modelling. An incorrect sequence alignment can lead to a wrong model, and the likelihood of making a mistake goes up as the sequence homology decreases. Non-conserved loops normally vary in position between homologous structures and may be difficult to model accurately, whereas the core regions are normally more conserved and reliable. *De novo* structure prediction is generally less accurate than homology modelling, with accuracy decreasing as protein size increases. Protein models should always be treated with caution and analysed in combination with experimental data.

4.2 Protein Sequence Alignments

The amino acid sequence for human TPO was downloaded from the SWISS-PROT database [SWISS-PROT 2000], accession code P07202 [Kimura 1989]. The protein structure database (pdb) was searched using TPO as a query sequence on the NCBI server [NCBI 2000] using BLASTP. The BLOSUM62 matrix was used, which assigns a probability score for each position in the alignment based on the frequency with which that substitution is known to occur among consensus blocks within related proteins. Sequence similarity is defined as the sum of the identical amino acids and amino acids with conserved physico-chemical properties. Sequence alignments were carried out using ClustalW [Thompson 1994], visualised with ESPript [Courcelle 2000] and secondary structure calculated with DSSP [Kabsch & Sander 1983].

The results of the Blast search are shown in Table 4.1 and Figure 4.1. The peroxidase domain is best modelled with human MPO (1mhl) [Fenna 1994], with residues 142 to 734 showing 46% identity with TPO. Cytochrome c oxidase (CCO) and prostaglandin H2 synthase (PGHS) also show some sequence identity, but are not as useful for modelling purposes as they are homologous to shorter sections of

the peroxidase domain. The best model of the CCP domain is the NMR structure of vaccinia control protein (1vvc) which has 35% identity to TPO over residues 741 to 797 [Wiles 1997]. The NMR structure of the second EGF domain of a fibrillin pair (1emn) shows the highest identity (42%) with TPO extending over residues 794 to 838 [Knott 1996]. The domain boundaries, based on sequence alignments, are listed in Table 4.2.

The sequence alignment of TPO with MPO is shown in Figure 4.2, and the secondary structure of MPO is listed in Table 4.3. MPO is a disulphide linked dimer, with each monomer containing 19 α -helices and 10 β -sheets forming five antiparallel strands (see Figure 1.10). The peroxidase domain of TPO has the key MPO structural residues conserved - all six intrachain disulphides, the interchain disulphide, the key active site residues, and all calcium ligands. There are three single amino acid deletions/insertions and three larger insertions of three, eight and 10 residues. The eight residue insertion corresponds to the residues that are post-translationally cleaved in MPO, the 10 residue insertion begins three residues C-terminal to the interchain disulphide cysteine.

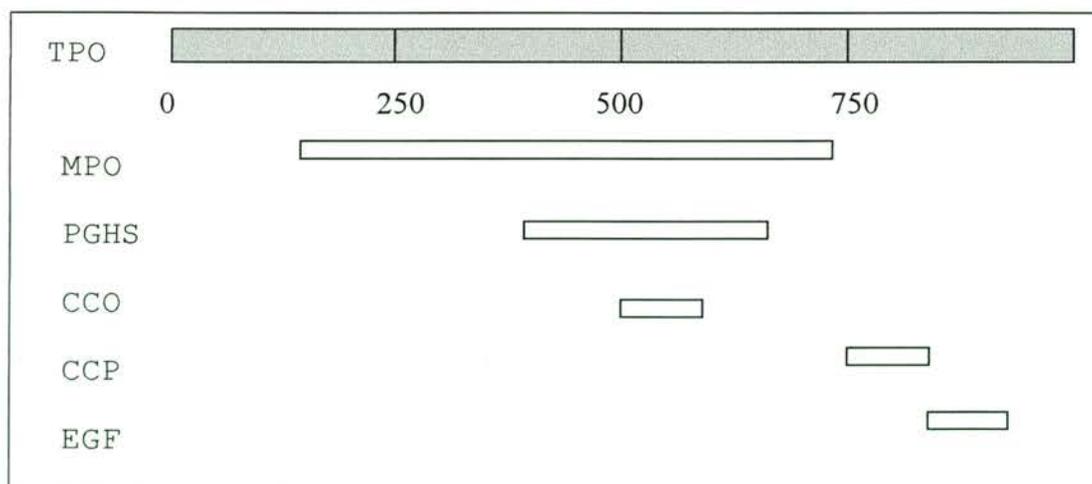


Figure 4.1

The alignment of proteins of known structure with TPO, see text for abbreviations.

Structure	Homologous TPO sequence	Similarity (%)	Identity (%)
Myeloperoxidase	142-734	61	46
Fibrillin EGF-like domains	794-838	57	39
Prostaglandin H2 Synthase-1	400-650	37	20
Vaccinia Virus CCP	741-797	46	35
Bovine Factor Xa	814-845	58	43
Apolipoprotein-H	742-795	49	32
Thrombin-Thrombomodulin	793-838	46	34
Cytochrome C oxidase	511-567	47	35
Clr EGF like domain	796-838	45	33
Cd46 (Mcp)	766-797	51	38
Factor H CCP domain	742-794	39	27
Activated Protein C	768-845	41	25

Table 4.1

The results of a Blast search for proteins of known structure with sequence similarity to TPO.

Domain/region	Amino acid sequence
Signal Sequence	1-14
N-terminal domain	15-141
Peroxidase domain	142-734
CCP-like domain	741-795
EGF-like domain	796-838
Transmembrane helix	847-871
C-terminal tail	871-933

Table 4.2

The domain boundaries of TPO.

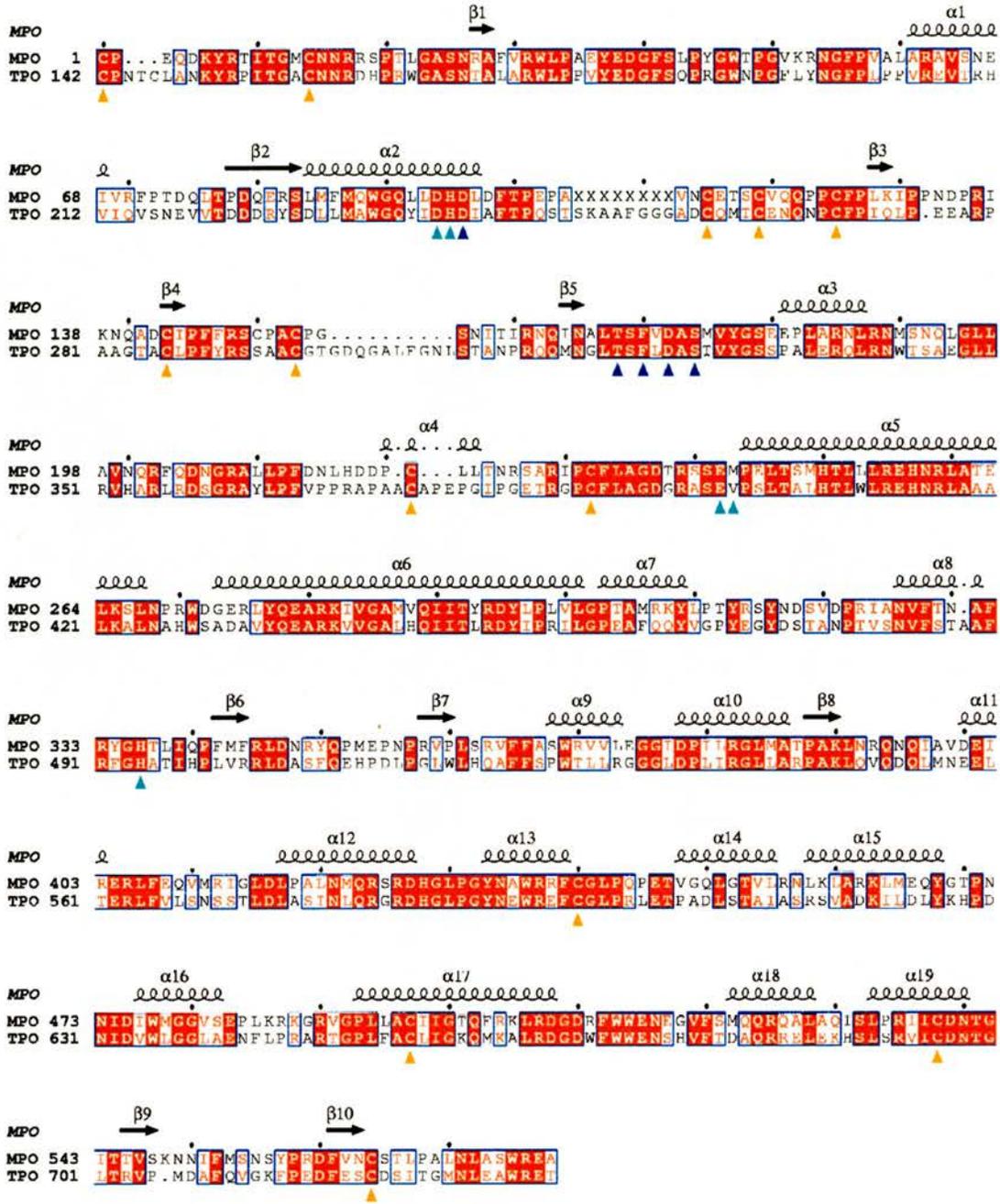


Figure 4.2

A protein sequence alignment of TPO with MPO. The secondary structure of MPO is shown above the alignment (X = residue posttranslationally cleaved)

-  Heme/active site ligands
-  Ca ligands
-  Cysteines in disulphides
-  identity
-  similarity

α -helices	Residues		β -sheets	Residues	Sheet formation
1	61-69		1	28-29	Antiparallel with β 5
2	84-97		2	78-83	Antiparallel with β 8
3	181-187		3	128-129	Antiparallel with β 4
4	220-224		4	143-144	Antiparallel with β 3
5	244-267		5	164-165	Antiparallel with β 1
6	273-301		6	342-344	Antiparallel with β 7
7	303-309		7	358-360	Antiparallel with β 6
8	326-338		8	388-390	Antiparallel with β 2
9	368-374		9	545-547	Antiparallel with β 10
10	377-385		10	561-563	Antiparallel with β 9
11	400-405				
12	417-428				
13	433-439		Disulphides	Residues	Position
14	448-455		1	1-14	N-terminus
15	458-467		2	115-125	Between α 2 and β 3
16	476-482		3	119-143	Joins N-term of β 3 to β 4
17	493-509		4	221-232	Joins α 4 to N-term of α 5
18	522-529		5	440-497	Joins α 13 to α 17
19	533-540		6	538-564	Joins α 19 to β 10

Table 4.3

The secondary structure of MPO, as identified by DSSP.

The sequence alignments of TPO with vaccinia virus CCP and the second of the pair of fibrillin EGF domains are shown in Figure 4.3. The fibrillin EGF domain contains a two stranded β -sheet, three disulphide bonds (1-3, 2-4, 5-6) and a bound calcium ion (Figure 4.4A). All cysteines are conserved in TPO. Five residues in calcium binding EGF domains have been identified to bind the ion, three of these residues interact via their side chain atoms and are identical in TPO and fibrillin, the remaining two interact via their carbonyl groups and show conservative substitution. The vaccinia virus CCP domain is shown in Figure 4.4B. It contains six β -strands and two disulphide bonds (1-3, 2-4), all corresponding cysteines are conserved in TPO.

4.3 Peroxidase Domain

The MPO structure was read into the interactive molecular graphics programme O [Jones 1991] and the sequence mutated to that of TPO. Loops were built using sections of known loop structures in lego-loop. The resulting pdb file underwent 200 cycles of energy minimisation in CNS [Brunger 1998], in which the C α atoms were harmonically restrained. The programmes PROCHECK [Laskowski 1993] and DSSP [Kabsch & Sander 1983] were used to analyse the model, figures were drawn with MOLSCRIPT [Kraulis 1991] and GRASP [Nicholls 1991].

The model of the peroxidase domain is shown in Figure 4.5, and the validation statistics in Table 4.4. After energy minimisation all bond lengths and 96% of bond angles were within limits and there were no disallowed dihedral angles. The model of the peroxidase domain of TPO fits closely to that of MPO, with a root mean squared deviation (rmsd) between MPO and the TPO model of 0.65 Å. All disulphide bonds and secondary structure are maintained, helix 4 was remodelled in TPO to preserve the disulphide bond on insertion of four residues. The eight residue insertion forms a bridge-like structure on the surface of TPO. The position of all active site residues and calcium ligands are very similar in MPO and the TPO

model and are compared in Figure 4.6. The distance between the C α atoms of seven key residues in the active site of the TPO model and MPO are listed in Table 4.5, all but one have the same amino acid type in both proteins.

TPO has four N-linked glycosylation sites (sequence N X S/T where X is not proline), at residues 129 (N-terminal domain), 307, 342 and 569. All of these residues in the TPO peroxidase domain model are on the surface of the structure. The MPO sequence contains four N-linked sites 157, 189, 255 and 317, all of which, apart from 157, have attached carbohydrate in the crystal structure. The glycosylation site at asparagine 317 is the most extensively glycosylated in the structure, and the sugars are important in making dimer contacts. Interestingly, this glycosylation site is not present in TPO. The glycosylation site at asparagine 255 is also missing in TPO. The site unoccupied in MPO (asparagine 157) is present in TPO, and TPO has one N-linked glycosylation site (asparagine 571) that is absent in MPO. The glycosylation sites on MPO and TPO are compared in Figure 4.7 and Table 4.6.

The peroxidase domain model is of a monomer, however, predictions can be made about the nature of the dimer interface by making comparisons between the TPO model and regions of the MPO structure involved in the interface. The dimer interface in MPO is stabilised by an interchain disulphide, carbohydrate-carbohydrate and carbohydrate-protein interactions, and hydrogen bonds [Fenna 1994]. The position of the disulphide bonded cysteine is conserved in TPO, but the loop of which is part is 10 residues longer in TPO than MPO. As discussed above, the glycosylation site occupied by sugars with an important role in the dimer interactions is missing in TPO.

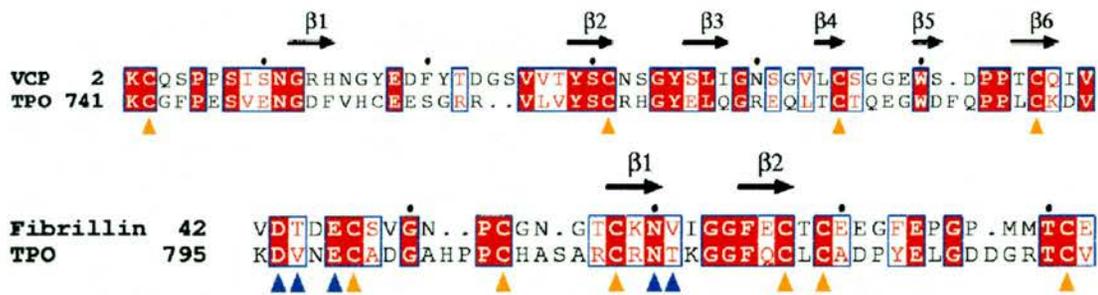


Figure 4.3

A protein sequence alignment of TPO vaccinia control protein and fibrillin. The secondary structure of VCP and fibrillin are shown above the alignments.

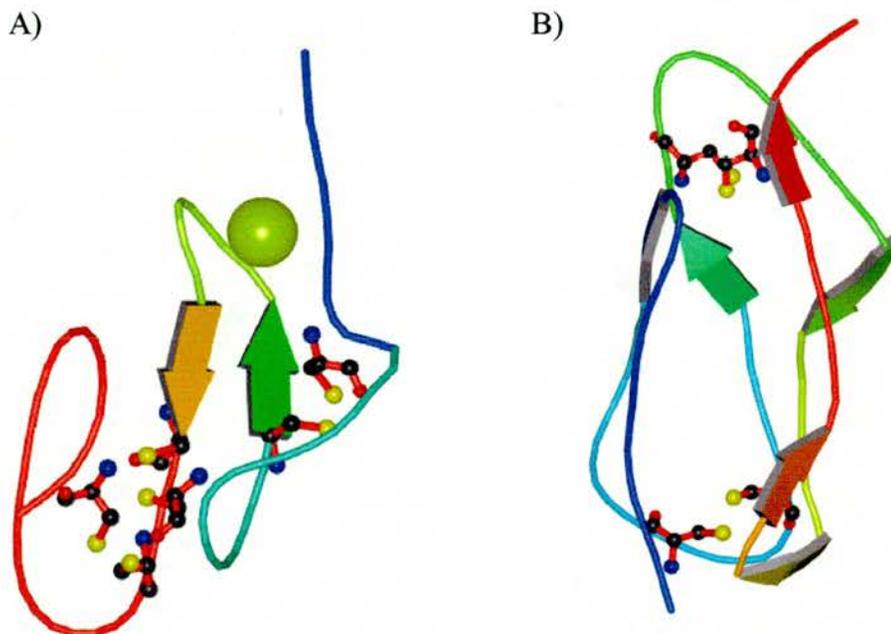
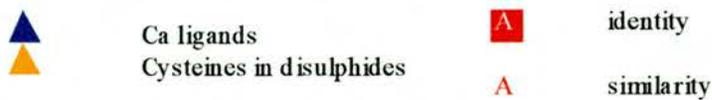


Figure 4.4

A) EGF domain of fibrillin and B) CCP domain of vaccinia virus.

The proteins are coloured blue to red from N- to C-termini, disulphide bonded cysteines are shown as ball and stick, and the calcium ion as a yellow sphere.

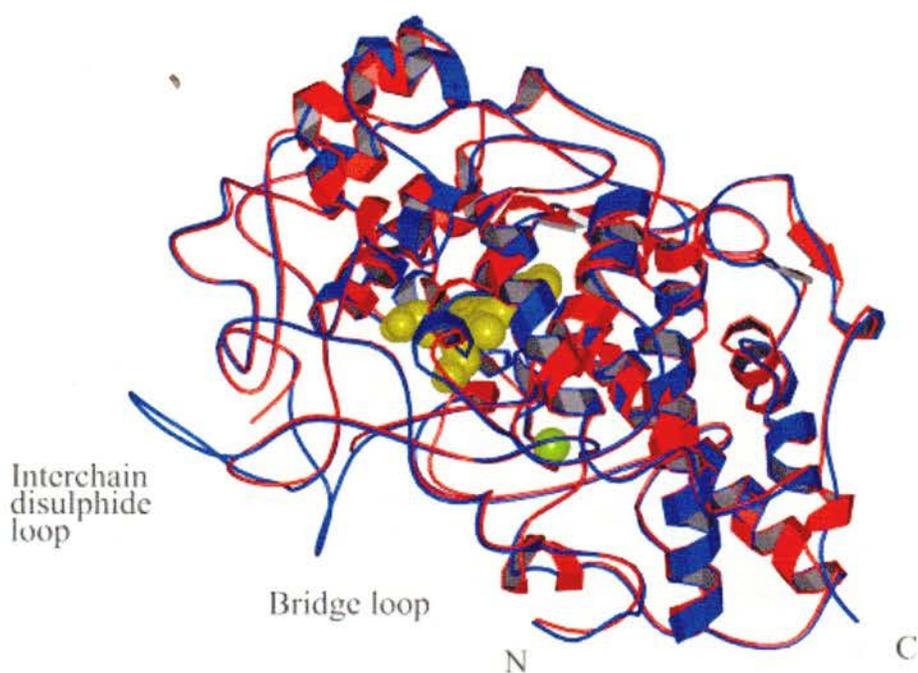


Figure 4.5

Superimposition of MPO (red) and the TPO model (blue). The heme is shown as space filling in yellow and the calcium ion in green.

Parameter	Before energy minimisation	After energy minimisation
Disallowed dihedral angles	3	0
% of bond lengths within limits	70%	100%
% of bond angles within limits	57%	96%
rmsd Å ² from MPO	0.39	0.65

Table 4.4

Summary of the validation statistics of the peroxidase domain model before and after energy minimisation.

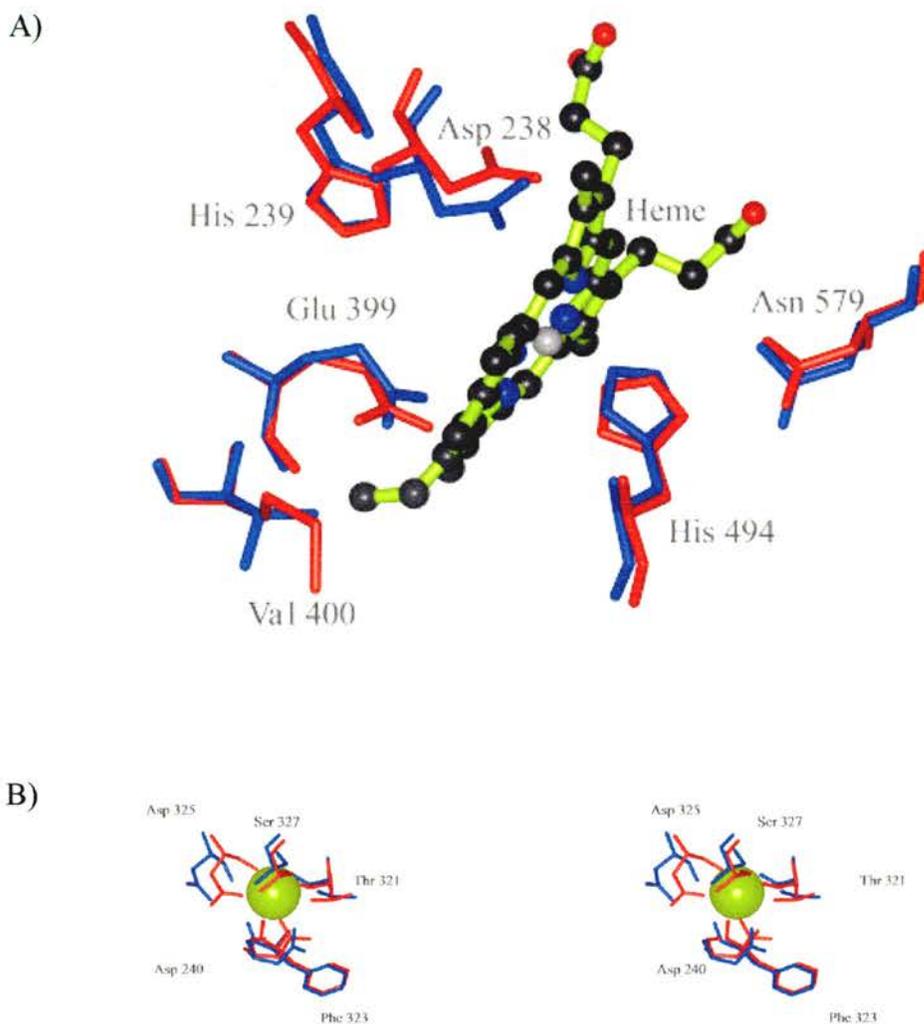


Figure 4.6

The ligation of the heme and Ca ions in the peroxidase domain model.

A) A close up of the active site of MPO (red) superimposed with the TPO model (blue), residue types and numbers correspond to TPO. The heme is shown in yellow, the active site entrance is at the back of the picture.

B) A stereo diagram of the Ca binding site of MPO (red) superimposed with the TPO model (blue), residue types and numbers correspond to TPO. The calcium is shown in yellow.

MPO	TPO model	Distance of Cα (Å)	Role
Asp 94	Asp 238	0.57	Ester linkage to heme
His 95	His 239	0.48	Distal heme ligand
Arg 239	Arg 396	0.57	Stabilised –ve charge on leaving group
Glu 242	Glu 399	0.26	Ester linkage to heme
Met 243	Val 400	0.21	Sulphonium ion linkage to heme
His 336	His 494	0.25	Proximal heme ligand
Asn 421	Asn 579	0.36	H-bond donor to proximal his

Table 4.5

The residues involved in the reaction mechanism of MPO and their homologues in the model of TPO.

Residues 19 and 21 from one MPO monomer are hydrogen bonded to residues 35 and 40 from the other. Residues 19, 21 and 35 are mutated in TPO. The sidechain hydroxyls of serine 19 and threonine 21 involved in hydrogen bonds in MPO and are replaced by a histidine and arginine in TPO. Their hydrogen bonding partners in residues 35 and 40 are mainchain atoms. Although the mutations in TPO could accommodate the interchain hydrogen bonds, rearrangements would be needed in the structure of the surrounding protein. The residues in the loops mediating the interactions (14-22 and 37-42) are highly conserved between TPO and MPO (see the alignment in Figure 4.2), suggesting that they have not been under selective pressure to mutate. However, the fact that those involved in the hydrogen bonds have mutated, indicates that they do not play an important role in TPO. The interactions involved in the dimer interface of MPO are shown in Figure 4.8.

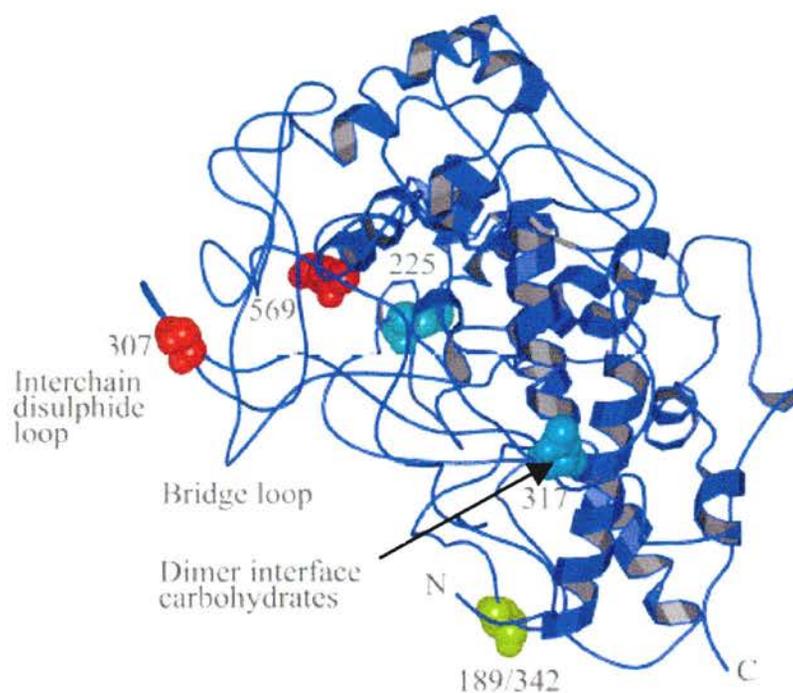


Figure 4.7

The TPO model showing the N-linked glycosylation sites. Blue – occupied site in MPO structure, no site in TPO model, yellow – occupied site in MPO structure, site in TPO model, red - site in TPO model, not in MPO structure.

MPO (MPO numbering)	TPO (TPO numbering)	Present in MPO structure	Position
157	307	No	TPO site is 3 residues N-terminal to MPO site
189	342	1 N-acetylglucosamine	Identical in MPO/TPO
255	-	1 N-acetylglucosamine	Not present in TPO
317	-	2 N-acetylglucosamines 3 mannose, 1 fucose	Not present in TPO At dimer interface
-	569	-	Not present in MPO

Table 4.6

Conservation of glycosylation sites in TPO and MPO.

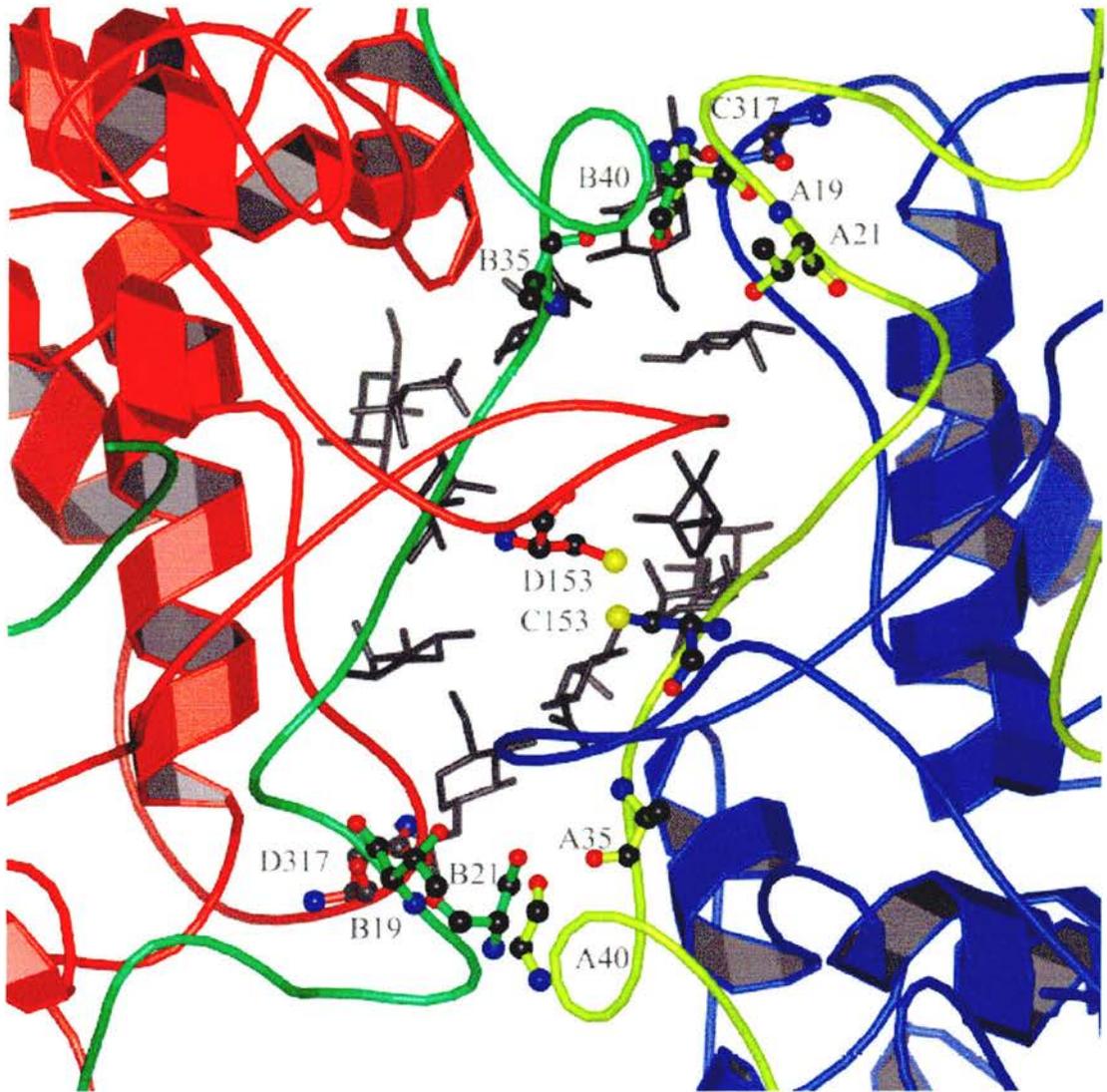


Figure 4.8

The structure of the MPO dimer interface. The light chains are shown in yellow (A) and green (B) and the heavy chains in blue (C) and red (D). Residues that form interchain hydrogen bonds (A/B 19, 21, 35, 40) and the interchain disulphide bond (C/D 153) are shown as ball and stick and are labelled. Carbohydrate residues are shown as grey lines and the asparagine residues to which they are bonded (C/D 317) are shown in ball and stick and are labelled.

Figure 4.9 shows a Grasp representation of the surface of the TPO model compared with that of MPO. The molecules have the same overall shape, but a different surface charge distribution. The most striking difference is in the active site and surrounds. In MPO there is a concentration of negative charge, whereas in the TPO the charge is less pronounced. This can be explained by the absence of negatively charged amino acids in the TPO sequence which are present in MPO. In particular, those corresponding to glutamates 102, 116 and 180, and aspartates 218 and 214 in MPO are replaced with neutral amino acids in TPO.

The position of the charged residues lining the entrance to the active site in MPO which are replaced by neutral amino acids in the TPO model are compared in Figure 4.10. TPO and MPO catalyse similar reactions, but the MPO substrate is chloride while TPO uses iodide. In addition, TPO uses the resulting hypoiodate to iodinate and join tyrosine residues on the large protein thyroglobulin, whereas hypochloride is the end point of the MPO catalysed reaction. The decrease in negative charge around the TPO active site is unlikely to be due to the interaction with iodide instead of chloride, as iodine is less electronegative than chlorine. Therefore, it can be postulated that the low charge around the TPO active site may have arisen in order to mediate interactions with thyroglobulin.

In the TPO model the active site is partially occluded by a bridge made up from the eight-residue insertion joining the two MPO chains. This is a prominent structure and it is clear why it is accessible to proteolytic cleavage in MPO. In TPO it may play a role of defining access to the active site and perhaps interacting with thyroglobulin.

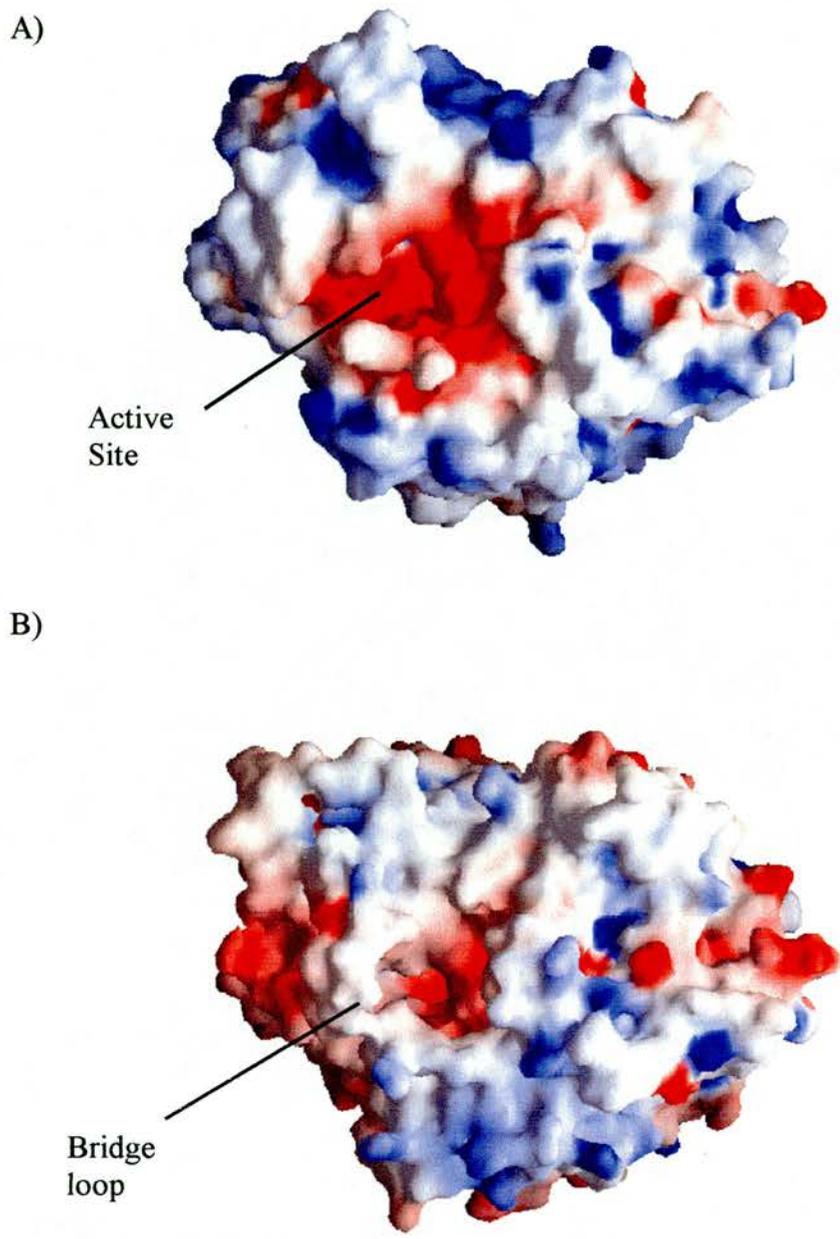


Figure 4.9

A grasp representation of the surfaces of A) MPO and B) TPO model. Red indicates negative charge, blue indicates positive charge, and white shows neutral residues.

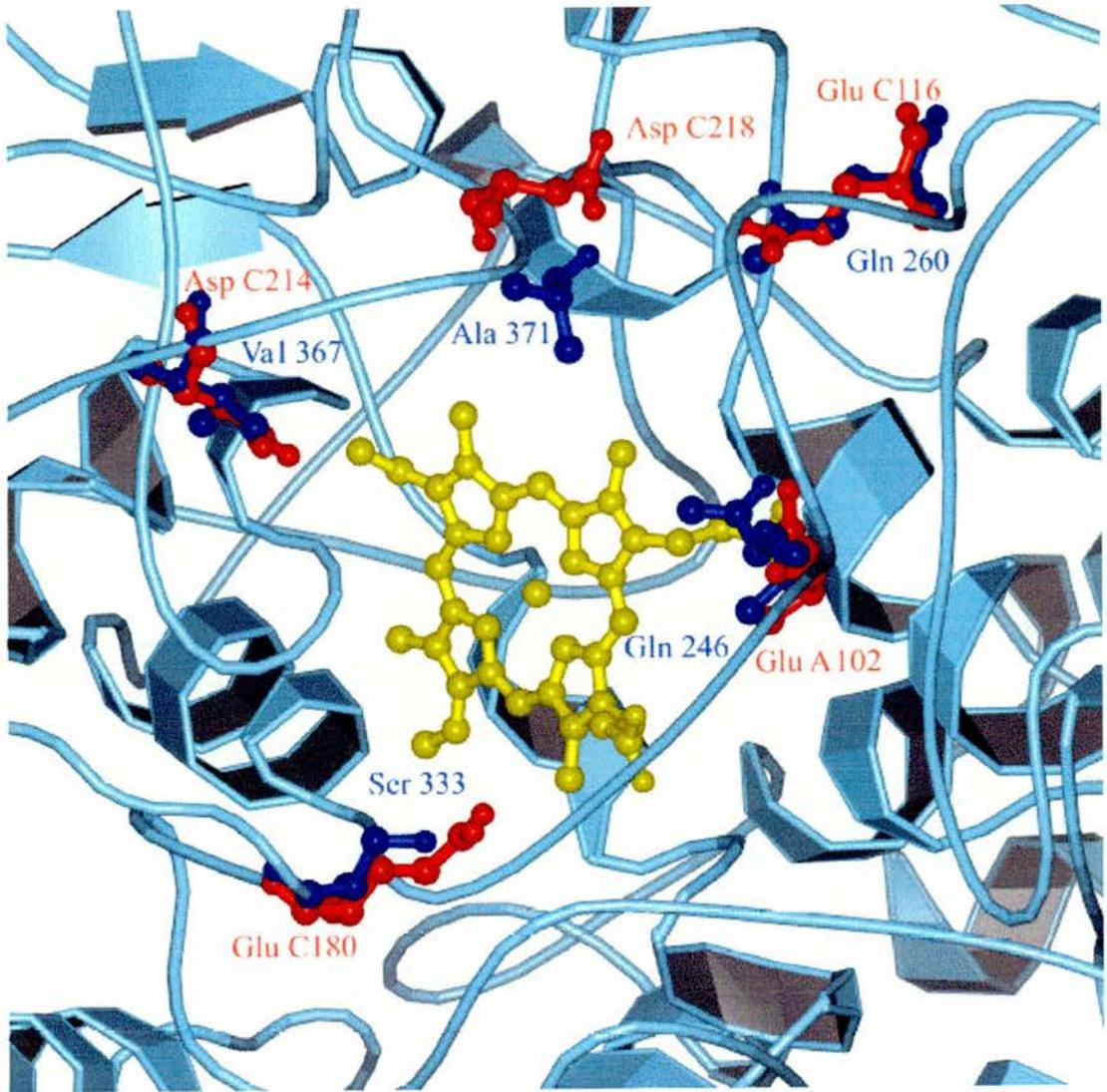


Figure 4.10

The entrance to the active site of the TPO model (light blue) highlighting negatively charged residues in MPO (red) that have neutral homologous residues (mid-blue) in the TPO model. The heme is shown in yellow.

MPO has been shown to interact with integrins [Johansson 1997] and it has been proposed that the animal peroxidases evolved from adhesion proteins [Taurog 1999]. Peroxinectin, an invertebrate peroxidase, has both adhesion and peroxidase activities and the site of adhesion has been localised to a KGD peptide [Johansson 1995]. A protein sequence alignment (Figure 4.11) shows the homologous sequence in TPO is RGD (residues 665-668), a motif related to KGD and known to bind integrins [Ruoslahti 1987]. The RGD motif is also present in the same position in MPO. However, as shown in red in Figure 4.12, this motif is not positioned on the surface of the TPO model. This implies that either the KGD motif is not involved with adhesion, or that this region of TPO has been modelled incorrectly.

The C21 epitope (713-721) maps to the C-terminus of the peroxidase domain and is shown in green in Figure 4.12. As expected from a sequence at a domain boundary, this region is on the surface of the structure and so accessible to antibodies. An immunodominant peptide previously predicted to be on the surface of a model of TPO [Hobby 2000], is shown in yellow, and is shown to occupy a surface region on this model. The TPO model shows that the peptide cleaved in TPO-2 (533-589) is integral to the TPO structure, and is highlighted in Figure 4.12. Without this peptide it is unlikely that the domain folds correctly, explaining the reported inactivity of TPO-2.

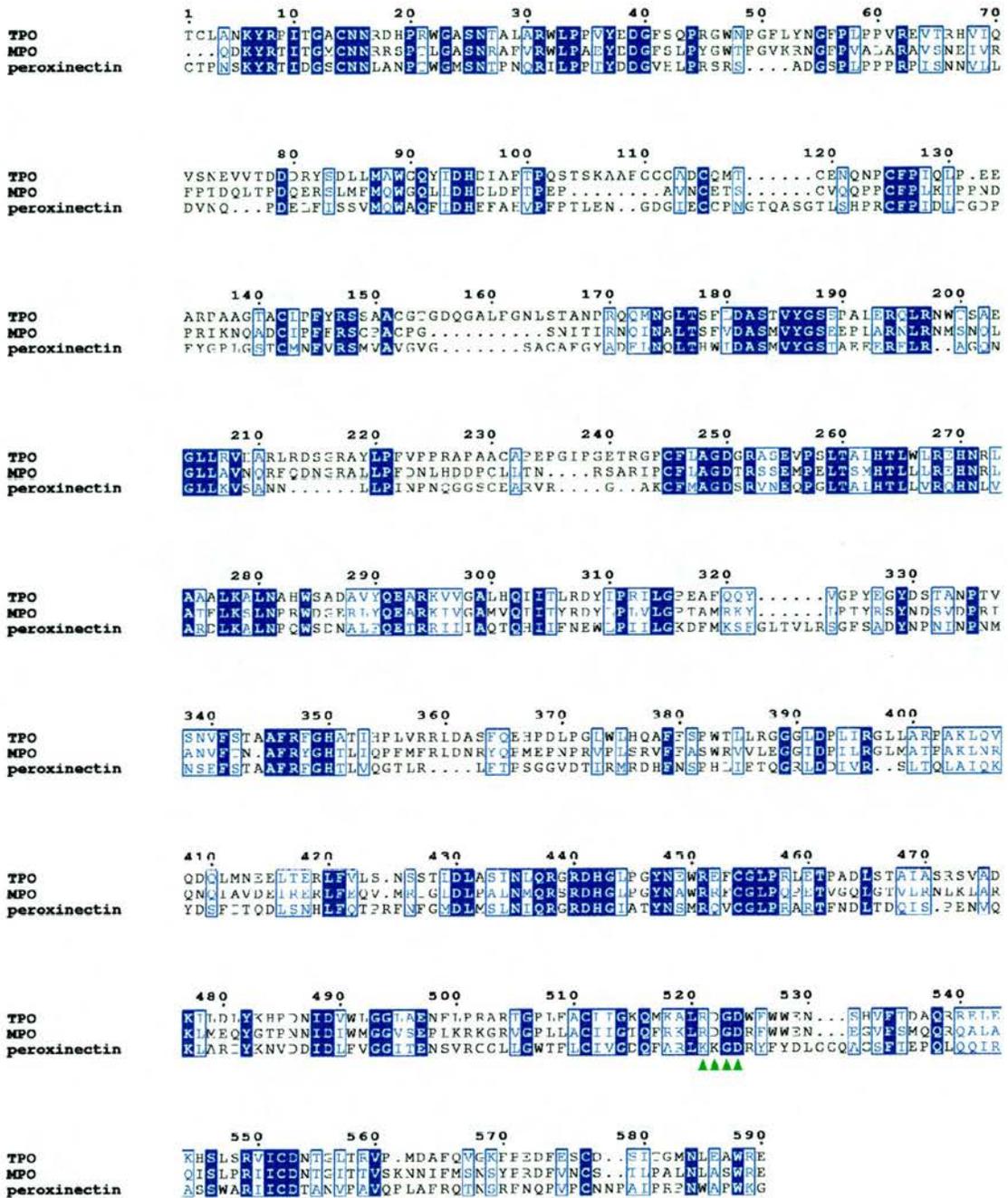


Figure 4.11

A sequence alignment of the peroxidase domains of TPO, MPO and peroxinectin. The potential adhesion motifs are indicated by green arrows.

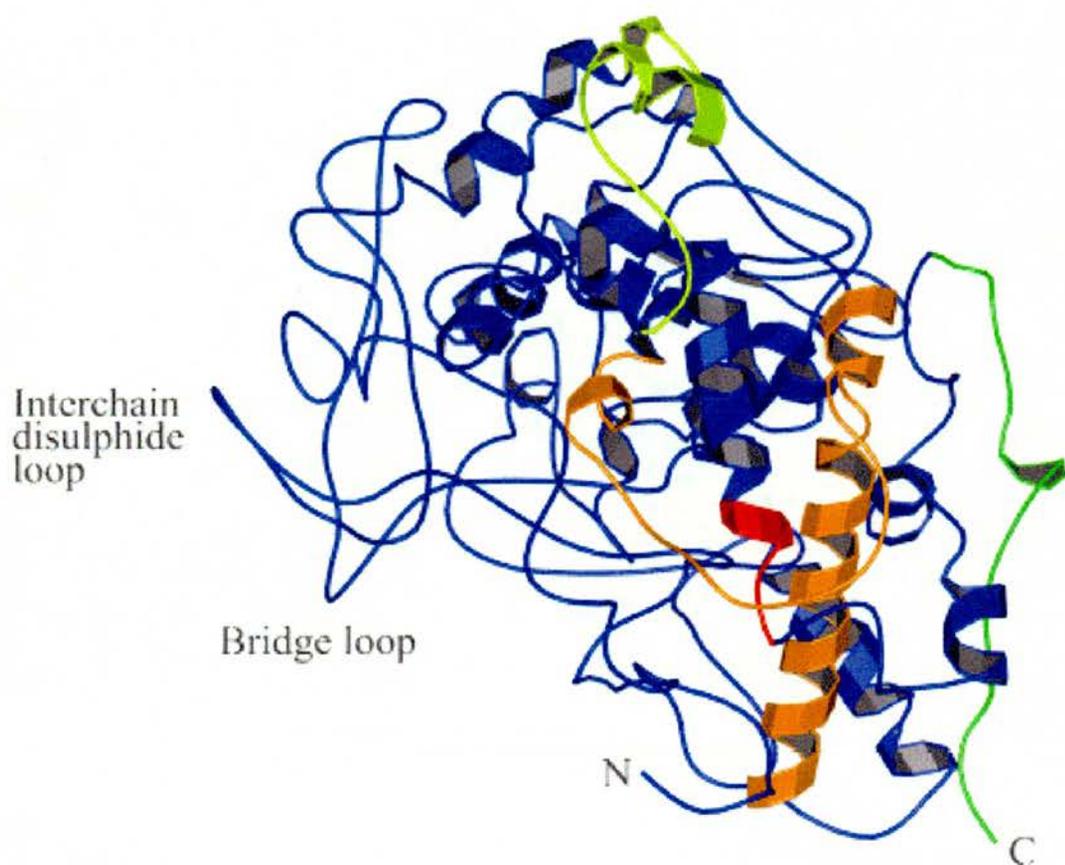


Figure 4.12

The TPO model showing the immunodominant peptide from Hobby *et al.* in yellow, the C21 epitope in green, the RGD motif in red and the cleaved peptide in TPO-2 in orange. The active site entrance is at the back of the picture.

4.4 CCP and EGF Domains

The CCP, 1vve [Wiles 1997] and EGF, 1emo [Knott 1996], structures identified by sequence alignments have both been solved by NMR. The minimised mean structure of each was used for modelling (1vvc and 1emn). The modelling procedure was the same as that used for the peroxidase domain, described in 4.3.

Diagrams of the EGF and CCP models of TPO are shown in Figure 4.13. The validation statistics of the models are shown in Table 4.7. The models of the EGF and CCP domains have all disulphide bonds conserved, (1-3 and 2-4) in CCP and (1-3, 2-4 and 5-6) in EGF. The position of all calcium ligands is conserved in the EGF domain model and are shown in Figure 4.14.

Parameter	Before energy minimisation		After energy minimisation	
	CCP	EGF	CCP	EGF
Disallowed dihedral angles	3	0	2	0
% of bond lengths within limits	70%	76%	100%	100%
% of bond angles within limits	69%	75%	94%	98%
rmsd Å ²	1.3	0.83	1.1	0.98

Table 4.7

Summary of the validation statistics of the EGF model before and after energy minimisation.

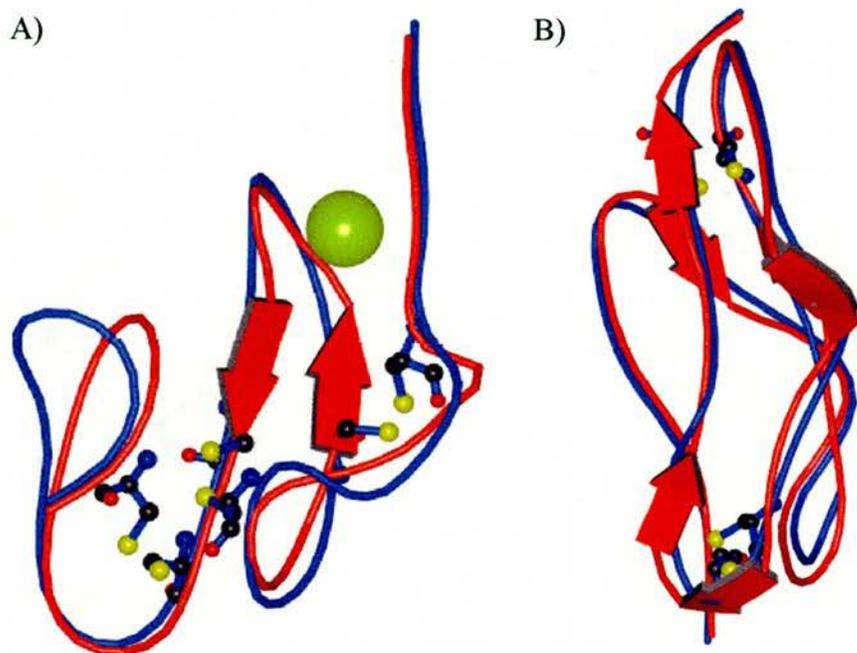


Figure 4.13

A superimposition of the template structures (red) and the TPO models (blue) of the EGF (A) and CCP (B) domain. The calcium is shown in yellow and the intrachain disulphide bonded cysteines are shown as ball and stick.

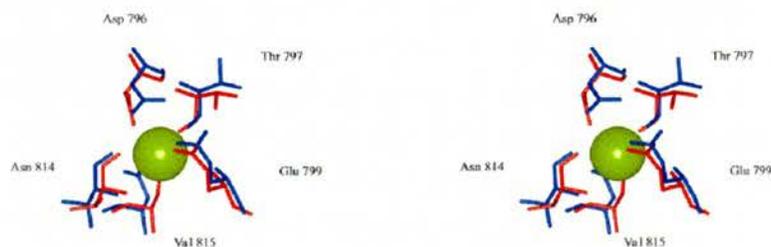


Figure 4.14

A stereo diagram of Ca binding site in EGF domain (red) compared with the TPO model (blue). Residues in the TPO model are labelled.

4.5 N-terminal Domain

A Blast search of the database of known structures failed to find a match for the N-terminal domain of TPO. MPO contains the N-terminal domain but it is post-translationally cleaved. The programme JPRED, which uses a consensus of secondary structure prediction programmes [Cuff 1998], was used to predict secondary structure. A multiple alignment of all known TPO N-terminal domain sequences (human, rat, mouse, pig) was used as this is believed to give more accurate results than single sequences. JPRED uses nine independent prediction methods (PHD, DSC, PREDATOR, NNSSP, MULPRED, ZPRED, JNET, COILS AND MULTICOIL) to produce a consensus which has been shown to be over 70% accurate.

Secondary structure is predicted by recognising sequence patterns that are indicative of α -helices or β -sheets. α -helices have a periodicity of 3.6, so positions at residues i , $i + 3$, $i + 4$ and $i + 7$ will lie on one face of the helix. For β -sheets, the side chains of adjacent residues point in opposite residues, so the pattern will be $i + 2$, $i + 4$ and $i + 6$. Amphipathic structures have one face buried in the protein core that is composed of hydrophobic residues, and the other exposed to solvent containing polar residues. Totally buried structures are composed of runs of hydrophobic residues.

The results of the consensus secondary structure prediction are shown in Figure 4.15. The prediction is for a four-helix structure with three, short β -sheets. Cleavage sites were identified in the N-terminal domain of the recombinant TPO used for crystallisation trials after residues 76 and 109. Residue 76 occurs in strand 3, and residue 109 occurs between helices 3 and 4, these sites are marked on the alignment. Helical wheel diagrams of the predicted helices, with residues coloured using the hydrophobicity scale of Brenner *et al.* [Brenner 1994], were calculated using a Java applet [Turcotte 2000] and are shown in Figure 4.16. Helix 1 is hydrophobic, so is likely to be buried in the structure, helix 3 is amphipathic so

likely to have one face buried and the other exposed. Helices 2 and 4 show no clear patterns of hydrophobicity, so their packing may vary across their length.

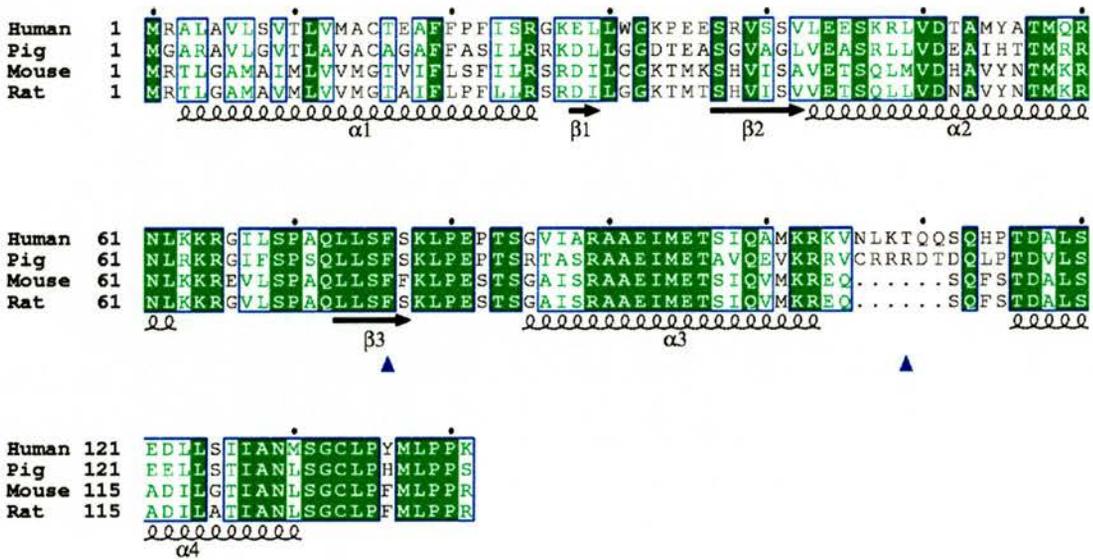


Figure 4.15

The results of the consensus secondary structure prediction for the N-terminal domain of TPO. The predicted secondary structure is indicated at the bottom of the alignment of the N-terminus of TPO sequences from different species, arrows indicate the residues found to be at the N-terminus of the recombinant TPO used for crystallisation.

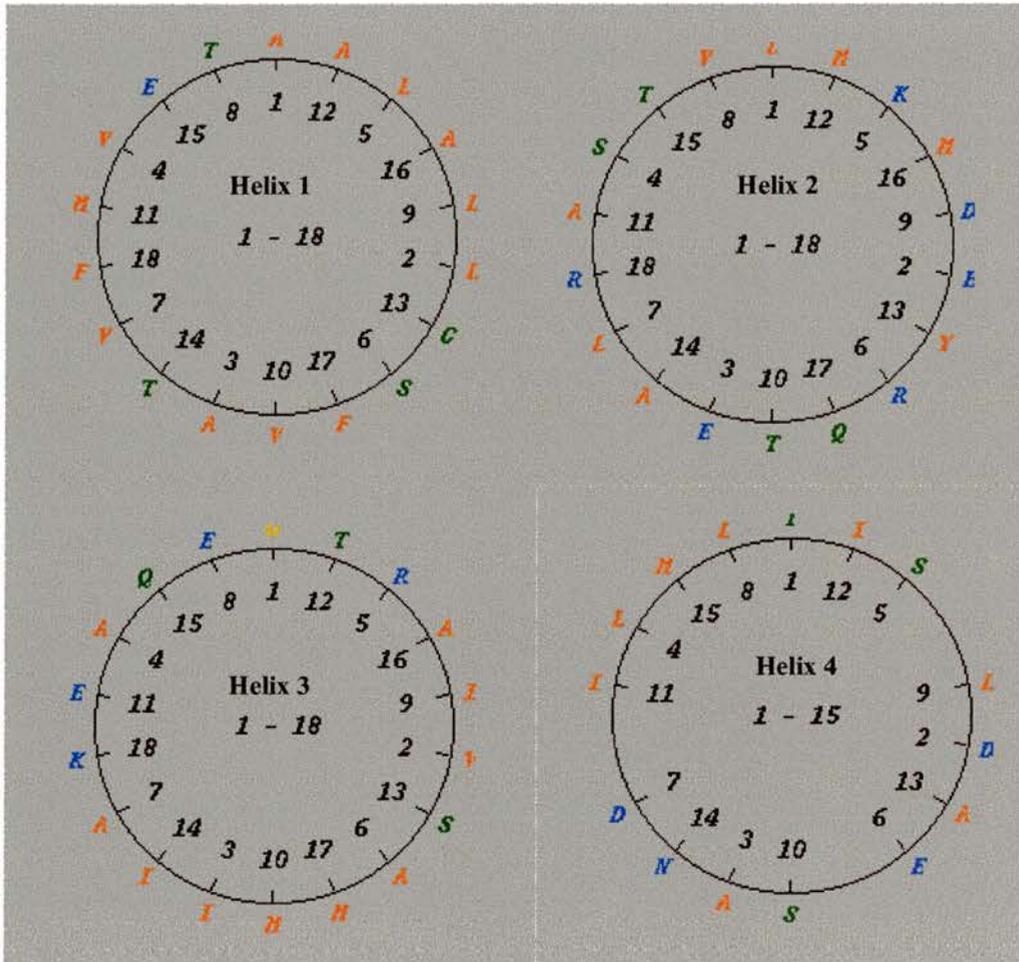


Figure 4.16

Helical wheels calculated for the four predicted helices in the N-terminal domain of TPO. Residues are coloured according to hydrophobicity: Red (FAMILYVW) > Orange (PG) > Green (CHQST) > Blue (KREND).

The programme 3D-PSSM [Kelley 2000] was used for fold recognition, using the protein sequence of the N-terminal domain of TPO. The algorithm ‘threads’ the sequence onto the members of a library of protein structures representative of each structural superfamily, and scores for compatibility. Scoring is based on sequence alignments with sequence profiles and structural profiles, and secondary structure and solvation potential matching.

The three highest scoring hits from threading were a ferritin homologue (1qgh), dehaloperoxidase (1ew6) and sperm lysin (3lyn). The first hit was rejected as the protein is dodecameric, binds DNA, and has an iron binding site, all properties which are not likely to be shared by the N-terminal domain of TPO. The second hit was also discarded as the protein binds a heme prosthetic group, also unlikely for the N-terminal domain. Lysin is an extracellular molecule on green abalone sperm cells that is responsible to adhesion to the egg receptor. An ungapped alignment between residues 1-121 of TPO and lysin is shown in Figure 4.17. Lysin is a five helix bundle, the structure has been solve by crystallography to 1.8 Å and is shown in Figure 4.18. There is very low sequence identity between the two proteins, but the predicted secondary structure of the TPO N-terminal domain and the secondary structure lysin align well. Helices 1 and 2 in lysin and the N-terminal domain are equivalent, helices 3 and 4 in lysin correspond to helix 3 in the N-terminal domain, and helix 5 in lysin is equivalent to helix 4 in the N-terminal domain.

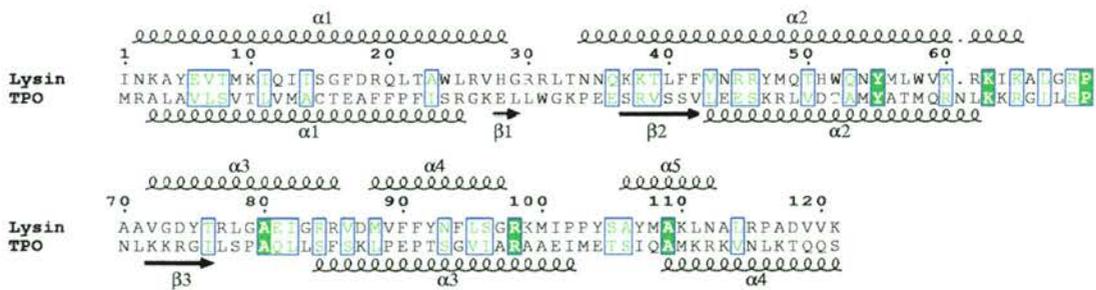


Figure 4.17

An ungapped alignment of lysin with the N-terminal domain of TPO. The secondary structure of lysine is at the top of the alignment and the predicted secondary structure (from JPRED) of TPO is indicated at the bottom.

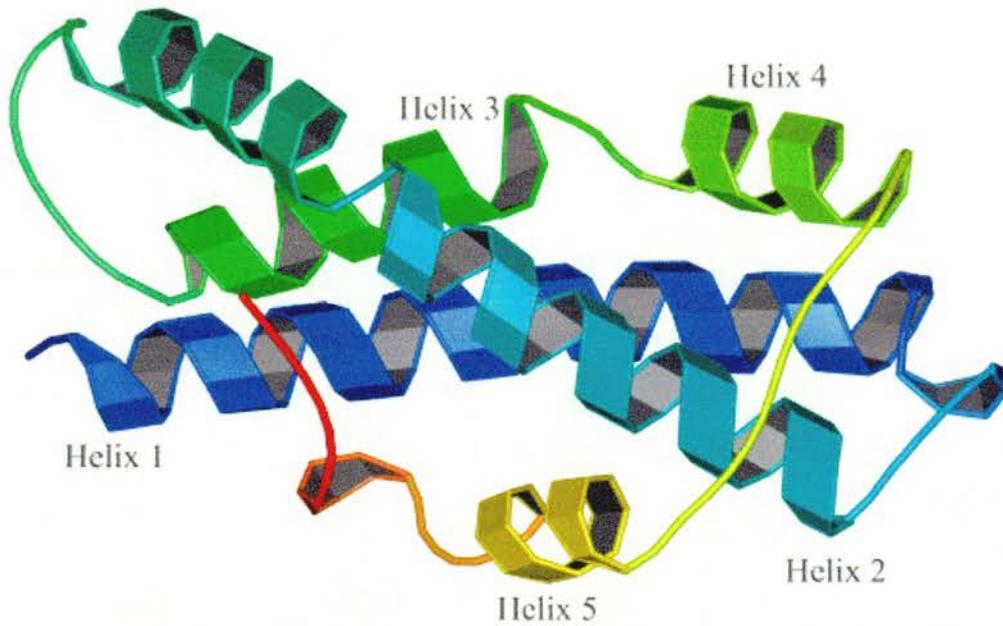


Figure 4.18

The structure of a monomer of lysin, a model for the N-terminal domain of TPO. The protein is coloured blue to red from N- to C-termini and the helices are numbered.

Sequence analysis also shows that there is a potential N-glycosylation site at residue 129 in the N-terminal domain, 13 residues before the start of the peroxidase domain. This amino acid is eight residues after the C-terminus of the lysin model, so is predicted to be on a loop linking the two domains on the surface of the protein.

4.6 Quaternary Structure

Models of the individual domains of TPO do not give information about how the domains may pack together in the membrane bound dimer. However, the quaternary structure of TPO is constrained by the dimer and membrane interactions, the length of the linkers between domains, and the position of the N- and C-termini of the domains.

The position of the termini and the length of the linkers between domains in the TPO model are shown in Figure 4.19. The linkage between all domains is less than ten amino acids, apart from that between the peroxidase domain and the N-terminal domain which is 21 residues. This implies that there is little flexibility in the C-terminal domains of TPO. The N- and C-termini of the CCP and EGF domains are at either end of the structure, implying the domains could form an extended domain pair. The N- and C-termini of the peroxidase and N-terminal domains are on the same face of each domain. The proximity of the termini of the peroxidase domain means that the N-terminal domain and the CCP domain must be relatively close together. These considerations can be used to propose a model of the potential orientation of domains, which is shown in Figure 4.20. It is assumed that TPO is a disulphide linked dimer, but the orientation of the dimer interface is not taken into account.

4.7 Summary

Models of the individual domains of TPO have been built, and the factors restricting the orientation of the domains with respect to each other have been studied. A comparison of the structure of MPO with the model of the peroxidase domains shows that the key active site residues have a similar position but that the entrance to the active site has different charge properties. Evidence has been

provided, based on the position of glycosylation sites, interface hydrogen bonds and the interchain disulphide, that the dimer interface is different in TPO than MPO. The conformational epitopes so far reported appear on the surface on the TPO model. An alignment of TPO with MPO and peroxinectin has identified an RGD motif in TPO that suggests it may have evolved from an adhesion protein. A model of the quaternary structure of membrane bound TPO shows that many regions of the structure will be inaccessible to autoantibody binding due to steric considerations.

The structure of sperm lysin has been proposed as a model for the N-terminal domain. Lysin is an adhesion protein, suggesting that the N-terminal domain may have a similar role, perhaps in mediating interactions with thyroglobulin. The model of the N-terminal domain shows that the proteolytic cleavage sites identified in TPO are in accessible sites in loops between α -helices.

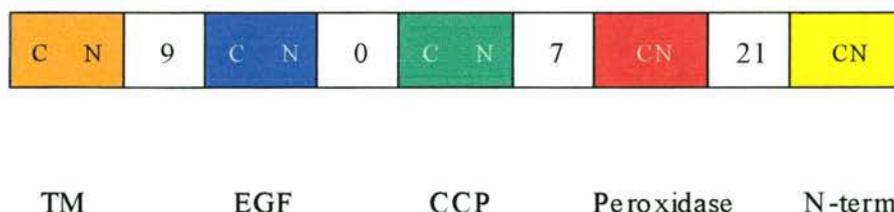
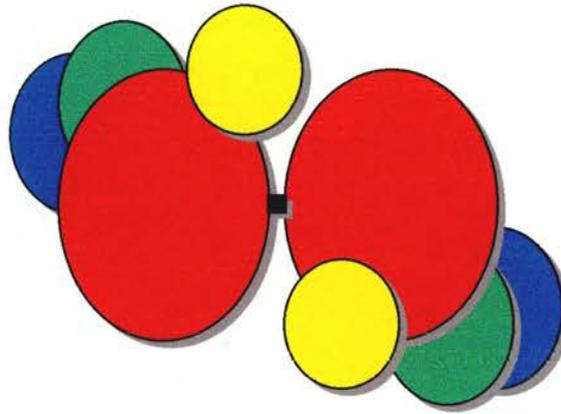


Figure 4.19

The length of the linker and position of the termini of the domains in the TPO model. The linker lengths are shown in white boxes. C- and N-termini are marked on opposite ends of the boxes if they are at opposite end of the domain, and close together in the boxes if they are on the same face in the domain.

A)



B)

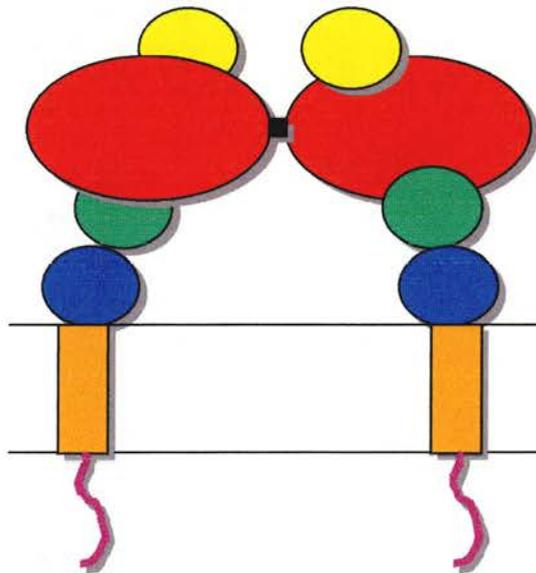


Figure 4.20

A model of the possible relative dispositions of the domains of TPO, A) is looking from above the membrane, B) is rotated by 90° . Yellow, N-terminal domain; red, peroxidase domain; green, CCP domain; blue, EGF domain; orange, transmembrane helix; purple, C-terminal domain; black, disulphide bridge.

Chapter 5

4F5 and 2G4 Characterisation and Fab Production

5.1 Introduction

4F5 and 2G4 are antibodies with high affinity for TPO that are produced by RSR Ltd. 4F5 is a mouse IgG1, kappa monoclonal antibody. 2G4 is a human IgG1, kappa monoclonal antibody isolated from Hashimoto thyroid lymphocytes. Both antibodies compete with sera from AITD patients for the binding of TPO. Structures of 4F5 and 2G4 in complex with TPO would allow identification of the autoimmune epitopes on TPO.

2G4 has already been studied in detail. The DNA sequence of the V genes has been determined [Hexham 1992] and competition studies with patient autoantibodies have been reported [Horimoto 1992]. The DNA sequencing of 4F5 was carried out in order to aid model building of the 4F5 structure. The sequence was analysed to identify which germline genes had been used, and to locate the position of somatic mutations. Characterisation of the binding of 4F5 and 2G4 to TPO was undertaken in order to determine if the antibodies bound the same region on TPO.

IgG molecules are composed of two Fab domains that bind antigen and a Fc domain that mediates effector functions. The intersegmental flexibility of antibody molecules hinders crystallisation and can be overcome by crystallising individual antibody fragments [Boulet 1988]. The structure of the combining site of the Fab domain provides insights into antigen binding. The Fab fragments of 4F5 and 2G4

were produced and purified. The binding of the Fab fragments to TPO was assayed to ensure the ability to bind TPO was still present.

The production of 4F5, the DNA sequencing of the V genes of 4F5, and the competition assays with AITD patient sera were carried out by RSR Ltd. Antibody fragment purification, Fab fragment assays and analysis of the 4F5 sequence were carried out by myself.

5.2 4F5 Production and Sequencing

4F5 was produced in mice in response to immunisation with native human TPO and Freund's complete adjuvant. The spleen was removed and fused with a murine myeloma cell line to yield a stable antibody-producing hybridoma. The hybridoma culture supernatant containing 4F5 was then harvested.

4F5 RNA was isolated by using the single-step method with guanidinium thiocyanate-phenol-chloroform extraction [Chomczynski & Sacchi 1987]. A one step reverse transcriptase polymerase chain reaction was carried out using SUPERSRIPT and GIBCOBRL kits from Life Technologies, with primers optimised for cloning mouse immunoglobulin genes [Kettleborough 1993]. The primers are shown in Table 5.1 and the PCR cycle is shown in Table 5.2. The PCR products were phenol extracted, ethanol precipitated and digested with *AntII/SalI* (HC) and *NcoI/XhoI* (LC). The digests were gel purified and cleaned with GeneClean II kit from Anachem. The HC was ligated to *AntII/SalI* cut pUC18, the LC was ligated to a modified pUC18 in which the standard polylinker had been replaced with a polylinker containing *NcoI/XhoI* sites. DNA sequencing was carried out using the Sanger dideoxy method [Sanger 1977] with sequence version II kit from Amersham using ³⁵S dATP. Synthetic primers were made 140 bp apart and both strands were sequenced with long overlaps.

HC

5'-primer

5' -GTA AGA CGT CMA GCT TCA GGA GTC RGG ACC-3'

3'-primer

3' -ATT AAG TCG ACC KYG GTS YTG CTG GCY GGG TG-3'

LC

5'-primer

5' -ATA TCC ATG GCA RAM ATT KTG CTG ACY CAR TYT CC-3'

3'-primer

5' - CTA CCT CGA GTT AAC ACT CAT TCC TGT TGA AGC-3'

K = T/G, M = C/A, R = A/G, S = C/G, Y = C/T

Table 5.1

The PCR primers used to clone 4F5.

1. cDNA synthesis and pre-denaturation, 1 cycle		
	50°C	30 mins
	94°C	2 mins
2. PCR amplification, 40 cycles		
denature	94°C	30 secs
anneal	48°C	1 min
extend	72°C	2 mins
3. Final extension, 1 cycle		
	72°C	10 mins

Table 5.2

The PCR cycle used to clone 4F5.

5.3 4F5 Sequence Analysis

The DNA and protein sequence of 4F5 were analysed using IGBLAST [Igblast 2000] and ABCHECK [Martin 1996], and DNA alignments were visualised in CINEMA [Parry-Smith 1998]. The Kabat numbering scheme [Wu & Kabat 1970] is used throughout. A recent report estimates that all but 2-5 mouse Ig kappa germline V genes have been isolated [Thiebe 1999]. However, the mouse IgH V gene locus has not been systematically studied in detail and the exact number of germline V genes is unknown.

The DNA sequence of the sense strand of the HC and LC of 4F5, and the deduced amino acid sequence are shown in Figures 5.1 and 5.2. Alignments of the HC sequence with that of known mouse Ig germline genes identified homology with MVARG2 of the V_H1 family [Blankenstein 1984], and the J_H2 gene [Sakano 1980]. No D-segment gene was unequivocally identified in the HC gene so it is not clear how many N region insertions make up the 10 bases between the V and J genes. The LC sequence is most similar to ap4 [Thiebe 1999] of the V_κ4/5 family and J_κ2 [Max 1981]. There is a single base N region insertion in the LC. The alignments of 4F5 DNA with the germline genes are shown in Figure 5.3

In the framework regions there are 18 DNA mutations, of which nine lead to a change in amino acid type. In the CDRs there are 12 mutations, of which 11 are replacement mutations. The somatic mutations are listed in Table 5.3. Assuming the germline genes have been correctly identified, this gives a ratio of replacement to silent mutations (R:S) of 1.0 for the framework regions (FWRs) and 11.0 for the CDR 1 and 2. This compares to the R:S ratio of 2.925 for a random sequence [Jukes & King 1979]. The high R:S ratio in the CDRs compared to the FWRs indicates that 4F5 has undergone antigen driven selection. The Kabat numbering of 4F5 is shown in Figure 5.4, the somatic mutations are highlighted.

1 - G A U G T C A A G C T T C A G G A G T C G G G A C C T G A G C T G G T G A A G C C T G G G G C T T C A G T G A G G A T A - 60
 D V K L Q E S G P E L V K P G A S V R I

61 - T C C T G C A A G G C T T C T G G C T A C A C C T T C A C A A G C T A C T A T A T A C A C T G G G T G A A A C A G A G G - 120
 S C K A S G Y T F T S Y Y I H W V K Q R

121 - C C T G G A C A G G G A C T T G A G T G G A T T G G A T T G A T T T A T C C T G G A A A T G T T T A T A C T A A G T A G - 180
 P G Q G L E W I G W I Y P G N V Y T K Y

181 - A G T G A G A A G T T C A A G G A C A A G G C C A C A C T G A C T G C A G A C A A A T C C T C C A G C A C G G C C T A G - 240
 S E K F K D K A T L T A D K S S S T A Y

241 - A T G C A G C T C A G C A G C C T G A C C T C T G A G G A C T C T G C G G T C T A T T T C T G T G G A A G A G A T G C T - 300
 M Q L S S L T S E D S A V Y F C G R D A

301 - T A C C T T G A G T A C T G G G G C C A A G G C A C C A C T C T C A C A G T T T C C T C A G C C A A A C G A C A C C C - 360
 H L E Y W G Q G T T L T V S S A K T T P

361 - C C A T C T G T C T A T C C A C T G G C C C T G G A T C T G C T G C C C A A A C T A A C T C C A T G G T G A C C C T G - 420
 P S V Y P L A P G S A A Q T N S M V T L

421 - G G A T G C C T G G T C A A G G G C T A T T T C C C T G A G C C A G T G A C A G T G A C C T G G A A C T C T G G A T C C - 480
 G C L V K G Y F P E P V T V T W N S G S

481 - C T G T C C A G C G G T G T G C A C A C C T T C C A G C T G T C C T G C A G T C T G A C C T C T A C A C T C T G A G C - 540
 L S S G V H T F P A V L Q S D L Y T L S

541 - A G C T C A G T G A C T G T C C C C T C C A G C A C C T G G C C C A G C G A G A C C G T C A C C T G C A A C G T T G C C - 600
 S S V T V P S S T W P S E T V T C N V A

601 - C A C C C A G C C A G C A A G A C C A A G G T C G A C - 627
 H P A S K T K V D

Figure 5.1

The DNA sequence of the 4F5 HC and the deduced protein sequence shown below. Boxed sequence is derived from the PCR primer, green sequence corresponds to the MVARG2 gene, yellow sequence to the JH2 gene and purple sequence to the D_H gene and N-region insertion. The CDRs are shown in grey.

```

1 - ATGGCAGAAATTTTGCTGACCCAATTTCCAGCAGTCATGTCTGCATCTCTCTGGGGTGAAG - 60
    M A E I L L T Q F P A V M S A S P G V K

61 - ETCACCATAACCTGCAGTGCCAGCTCAAGTGTAATTCATTCACTGGTTCAGCAGAAG - 120
    V T I T C S A S S S V N F I H W F Q Q K

121 - CCAGGCACTTCTCCAAACTCTGGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCT - 180
    P G T S P K L W I Y S T S N L A S G V P

181 - GCTCGCTTCAGTGGCAGTGGATCTGGGACCTCTTACTCTCTCACAATCAGCCGAATGGAG - 240
    A R F S G S G S G T S Y S L T I S R M E

241 - GCTGAGGATGCTGCCACTTATTACTGCCTGCAAAGGAGTAGTTACCCGTACACGTTCCGGA - 300
    A E D A A T Y Y C L Q R S S Y P Y T F G

301 - GGGGGGACCAAGCTGGAAATAAAACGGGCTGATGCTGCACCAACTGTATCCATCTTCCCA - 360
    G G T K L E I K R A D A A P T V S I F P

361 - CCATCCAGTGAGCAGTTAACATCTGGAGGTGCCTCAGTCGTGTGCTTCTTGAACAACTTC - 420
    P S S E Q L T S G G A S V V C F L N N F

421 - TACCCCAAAGACATCAATGTCAAGTGGAAGATTGATGGCAGTGAACGACAAAATGGCGTC - 480
    Y P K D I N V K W K I D G S E R Q N G V

481 - CTGAACAGTTGGACTGATCAGGACAGCAAAGACAGCACCTACAGCATGAGCAGCACCCCTC - 540
    L N S W T D Q D S K D S T Y S M S S T L
    *

541 - ACGCTGACCAAGGACGAGTATGAACGACATAACAGCTATACCTGTGAGGCCACTCACAAG - 600
    T L T K D E Y E R H N S Y T C E A T H K

601 - ACATCAACTTCACCCATTGTCAAGAGCTTCAACAGGAATGAGTGTTAACTCGAG - 654
    T S T S P I V K S F N R N E C * L E

```

Figure 5.2

The DNA sequence of the 4F5 LC and the deduced protein sequence shown below. Boxed sequence is derived from the PCR primer, red sequence corresponds to the ap4 gene, pink sequence the J κ 2 gene, and blue sequence the N-region insertion. The CDRs are shown in grey.

```

ap4      A G A G G A C A A A T T G T T C T C A C C C A G T C T C C A G C A A T C A T G T C T G C A T C T C C
4F5 LC   A T G G C A G A A A T T T T G C T G A C C C A A T T T C C A G C A G T C A T G T C T G C A T C T C C
.....10.....20.....30.....40.....50

ap4      A G G G G A G A A G G T C A C C A T A A C C T G C A G T G C C A G C T C A A G T G T A A G T T A C A
4F5 LC   T G G G T G A A G G T C A C C A T A A C C T G C A G T G C C A G C T C A A G T G T A A A T T T C A
.....60.....70.....80.....90.....100

ap4      T G C A C T G G T T C C A G C A G A A G C C A G G C A C T T C T C C C A A A C T C T G G A T T T A T
4F5 LC   T T C A C T G G T T C C A G C A G A A G C C A G G C A C T T C T C C C A A A C T C T G G A T T T A T
.....110.....120.....130.....140.....150

ap4      A G C A C A T C C A A C C T G G C T T C T G G A G T C C C T G C T C G C T T C A G T G G C A G T G G
4F5 LC   A G C A C A T C C A A C C T G G C T T C T G G A G T C C C T G C T C G C T T C A G T G G C A G T G G
.....160.....170.....180.....190.....200

ap4      A T C T G G G A C C T C T T A C T C T C T C A C A A T C A G C C G A A T G G A G G C T G A A G A T G
4F5 LC   A T C T G G G A C C T C T T A C T C T C T C A C A A T C A G C C G A A T G G A G G C T G A G G A T G
.....210.....220.....230.....240.....250

ap4      C T G C C A C T T A T T A C T G C C A G C A A A G G A G T A G T T A C C C
4F5 LC   C T G C C A C T T A T T A C T G C C T G C A A A G G A G T A G T T A C C C
.....260.....270.....280.....290.....300

4F5 LC   T A C A C G T T C G G A G G G G G A C C A A G C T G G A A A T A A A A C G
JK2      T A C A C G T T C G G A G G G G G A C C A A G C T G G A A A T A A A A C G
.....10.....20.....30.....40.....50

MVAR2    C T G C A G C A G T C T G G A C C T G A G C T G G T G A A G C C T G G G G C T T C A G T G A A G T T
4F5 HC   C T T C A G G A G T C G G G A C C T G A G C T G G T G A A G C C T G G G G C T T C A G T G A G G A T
.....10.....20.....30.....40.....50

MVAR2    G T C C T G C A A G G C T T C T G G C T A C A C C T T C A C A A G C T A C G A T A T A A A C T G G G
4F5 HC   A T C C T G C A A G G C T T C T G G C T A C A C C T T C A C A A G C T A C T A T A T A C A C T G G G
.....60.....70.....80.....90.....100

MVAR2    T G A A G C A G A G G C C T G G A C A G G G A C T T G A G T G G A T T G G A T G G A T T T A T C C T
4F5 HC   T G A A A C A G A G G C C T G G A C A G G G A C T T G A G T G G A T T G G A T G G A T T T A T C C T
.....110.....120.....130.....140.....150

MVAR2    A G A G A T G G T A G T A C T A A G T A C A A T G A G A A G T T C A A G G G C A A G G C C A C A T T
4F5 HC   G G A A A T G T T T A T A C T A A G T A C A G T G A G A A G T T C A A G G A C A A G G C C A C A C T
.....160.....170.....180.....190.....200

MVAR2    G A C T G T A G A C A C A T C C T C C A G C A C A G C G T A C A T G G A G C T C C A C A G C C T G A
4F5 HC   G A C T G C A G A C A A A T C C T C C A G C A C G G C C T A C A T G C A G C T C A G C A G C C T G A
.....210.....220.....230.....240.....250

MVAR2    C A T C T G A G G A C T C T G C G G T C T A T T T C T G T G C A A G A
4F5 HC   C C T C T G A G G A C T C T G C G G T C T A T T T C T G T G G A A G A
.....260.....270.....280.....290.....300

JH2      T T G A C T A C T G G G G C C A A G G C A C C A C T C T C A C A G T C T C C T C A
4F5 HC   T T G A G T A C T G G G G C C A A G G C A C C A C T C T C A C A G T T T C C T C A
.....10.....20.....30.....40.....50

```

Figure 5.3

An alignment of 4F5 with the germline genes from which the sequences are likely to be derived.

The alignment identified differences between the primer derived DNA sequence of 4F5 and the germline genes. At the 5' end there are seven differences in the LC and three in the HC. The HC germline sequence begins nine bases upstream from the start site, so the first three amino acids are unverifiable. In the primer regions the germline sequence was taken to be correct. The sequences of the C_H and C_L domains of 4F5 were compared to the sequence of murine immunoglobulin constant domain gene. A single mutation was found in the LC (546 T→C) and three differences were found in the 3' primer region (606 G→A, 614 G→A, 615 C→G).

Region (HC & LC)	R:S (R:S)	DNA mut'ns		Protein mutations	
		LC	HC	LC	HC
FWR1	2 (4:2)	3	3	I→V, E→V	K→R, L→I
CDR1	∞ (5:0)	3	2	S→N, Y→F, M→I	D→Y, N→H
FWR2	0 (0:1)	0	1	none	none
CDR2	6 (6:1)	0	7	none	R→G, D→N, G→V, S→Y, N→S, G→D
FWR3	0.83 (5:6)	1	10	none	V→A, T→K, E→Q, H→S, A→G
J	1 (1:1)	0	2	none	D→E

Table 5.3

The distribution of mutations in the DNA and protein sequence of 4F5 compared with the identified germline sequences.

L1 Q	L41 G	L80 A	H8 G	H47 W	H82C L
L2 I	L42 T	L81 E	H9 P	H48 I	H83 T
L3 V	L43 S	L82 D	H10 E	H49 G	H84 S
L4 L	L44 P	L83 A	H11 L	H50 W	H85 E
L5 T	L45 K	L84 A	H12 V	H51 I	H86 D
L6 Q	L46 L	L85 T	H13 K	H52 Y	H87 S
L7 S	L47 W	L86 Y	H14 P	H52A P	H88 A
L8 P	L48 I	L87 Y	H15 G	H53 G	H89 V
L9 A	L49 Y	L88 C	H16 A	H54 N	H90 Y
L10 I	L50 S	L89 L	H17 S	H55 V	H91 F
L11 M	L51 T	L90 Q	H18 V	H56 Y	H92 C
L12 S	L52 S	L91 R	H19 R	H57 T	H93 G
L13 A	L53 N	L92 S	H20 I	H58 K	H94 R
L14 S	L54 L	L93 S	H21 S	H59 Y	H95 D
L15 P	L55 A	L94 Y	H22 C	H60 S	H96 A
L16 G	L56 S	L95 P	H23 K	H61 E	H97 H
L17 V	L57 G	L96 Y	H24 A	H62 K	H98 L
L18 K	L58 V	L97 T	H25 S	H63 F	H101 E
L19 V	L59 P	L98 F	H26 G	H64 K	H102 Y
L20 T	L60 A	L99 G	H27 Y	H65 D	H103 W
L21 I	L61 R	L100 G	H28 T	H66 K	H104 G
L22 T	L62 F	L101 G	H29 F	H67 A	H105 Q
L23 C	L63 S	L102 T	H30 T	H68 T	H106 G
L24 S	L64 G	L103 K	H31 S	H69 L	H107 T
L25 A	L65 S	L104 L	H32 Y	H70 T	H108 T
L26 S	L66 G	L105 E	H33 Y	H71 A	H109 L
L27 S	L67 S	L106 I	H34 I	H72 D	H110 T
L29 S	L68 G	L106A -	H35 H	H73 K	H111 V
L30 V	L69 T	L107 K	H36 W	H74 S	H112 S
L31 N	L70 S	L108 R	H37 V	H75 S	
L32 F	L71 Y	L109 A	H38 K	H76 S	
L33 I	L72 S		H39 Q	H77 T	
L34 H	L73 L	H1 D	H40 R	H78 A	
L35 W	L74 T	H2 V	H41 P	H79 Y	
L36 F	L75 I	H3 K	H42 G	H80 M	
L37 Q	L76 S	H4 L	H43 Q	H81 Q	
L38 Q	L77 R	H5 Q	H44 G	H82 L	
L39 K	L78 M	H6 Q	H45 L	H82A S	
L40 P	L79 E	H7 S	H46 E	H82B S	

Figure 5.4

The 4F5 sequence numbered according to Kabat, with CDRs boxed. Residues highlighted in green are those that differ from the germline sequence. The sequences in the primer region have been changed to those of the germline and are highlighted in yellow.

5.4 4F5 and 2G4 IgG Characterisation

4F5 and 2G4 are bind preferentially to non-reduced rather than reduced TPO (personal communication RSR Ltd.) and do not inhibit enzyme TPO activity [Gut 2000]. Enzyme-linked immunosorbant assay (ELISA) experiments were carried out in order to determine the relationship between the epitopes on TPO recognised by 4F5 and 2G4. Further experiments were undertaken to see whether these epitopes are the same as those bound by autoantibodies from patients with AITD.

100 μ l of 0.025 μ g/ml of recombinant TPO was added to a Nunc Immuno Module MaxiSorp plate coated with 2G4. 50 μ l of 0.2 μ g/ml biotin labelled 4F5 was added and the reaction incubated for 30 minutes at room temperature with shaking. The plates were washed three times and 100 μ l of a streptavidin-peroxidase (SA-POD) conjugate and 100 μ l of tetramethyl benzidine (TMB) substrate were added. The reaction was incubated for 15 minutes in the dark and quenched with 50 μ l of 2M H₂SO₄. The absorbance at 450 nm was then measured. The assay is illustrated in Figure 5.5A. The OD₄₅₀ was 0.14 with no TPO added, and 2.25 with TPO added. As 4F5 can only bind to TPO if it reacts with a different epitope from 2G4, this shows that the 4F5 and 2G4 recognise different epitopes on TPO.

In a second experiment (Figure 5.5B), a NIBSC (National Institute for Biological Standards and Control, Potters Bar, UK) polyclonal, human TPO autoantibody standard was used with 4F5 or 2G4 coated plates. 100 μ l of NIBSC standard, diluted 1:10 in 0.025 μ g/ml of TPO in assay buffer and 10% normal human serum, and 50 μ l of 2G4-Bi in assay buffer (0.2 μ g/ml) were added to the plate and incubated for 30 minutes shaking at room temperature. The plate was washed three times, 100 μ l of SA-POD conjugate was added, and the reaction incubated for 15 minutes in the dark. 50 μ l of 2M H₂SO₄ was used to quench the reaction and the OD₄₅₀ read. The results (Figure 5.6A) show that the TPO autoantibody standard inhibited the binding of 2G4 and 4F5 to TPO. This indicates that the epitopes

recognised by 2G4 and 4F5 antibodies are similar to those recognised by TPO autoantibodies from patients with AITD.

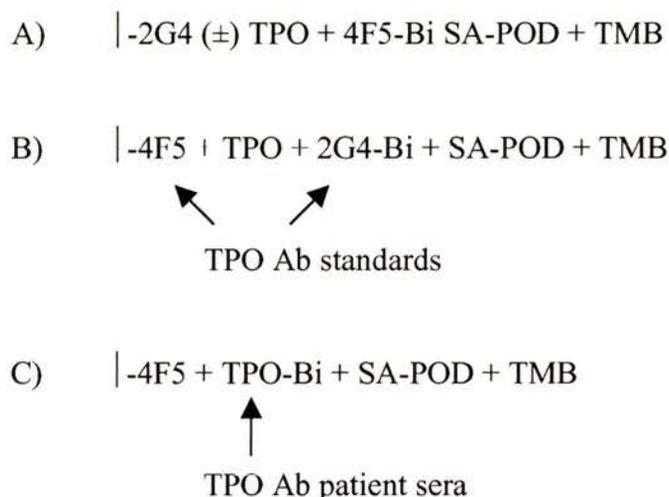


Figure 5.5

The ELISA experiments.

A) Comparison of the epitopes of 4F5 and 2G4 on TPO, B) Competition for TPO binding with antibody standards, C) Competition for TPO binding with patient sera. Bi (biotin), (SA-POD (streptavidin-peroxidase), TMB (tetramethyl benzidine).

A third experiment (Figure 5.5C) used 4F5 or 2G4 coated plates to test the effect of TPO antibody positive sera of TPO-Bi binding. 50 µl of a patient serum sample and 100 µl of 0.01 µg/ml TPO-Bi were incubated for 60 minutes with shaking at room temperature and then washed three times. 50 µl of a 1 µg/ml SA-POD conjugate was added and the reaction incubated for 30 minutes at room temperature with shaking. After three washes, 100 µl of TMB substrate was added. The reaction was incubated for 20 minutes in the dark and 50 µl of 2 M H₂SO₄ was added to quench the reaction. The absorbance at 450 nm was then measured. For comparison, the total amount of TPO autoantibodies in each sera were estimated using a radioimmunoassay. The results (Figure 5.6B) show that all the TPO

antibody positive sera inhibit binding of TPO to 4F5 or 2G4 indicating that these sera contain TPO Ab reactive with 4F5 and 2G4 epitopes. The degree of inhibition was related to the total amount of TPO autoantibodies present.

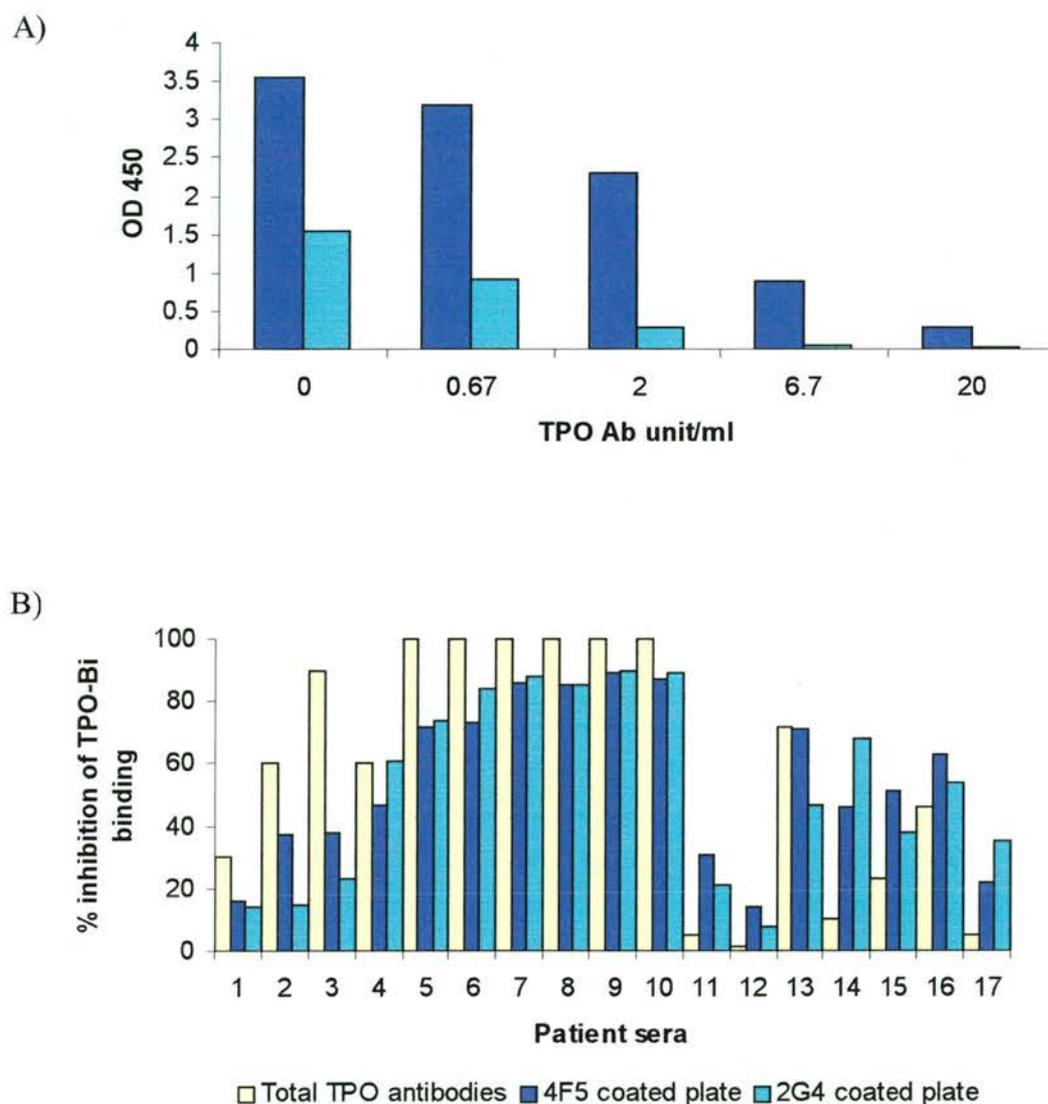


Figure 5.6

Results from the 4F5 and 2G4 ELISAs.

A) Competition for TPO binding with antibody standards.

B) Competition for TPO binding with patient sera, the total TPO antibodies is in antibody units /ml.

5.5 4F5 and 2G4 Fab Fragment Production

4F5 was received as a hybridoma culture supernatant, intact 2G4 had already been purified. 4F5 supernatant was mixed with double strength binding buffer (2 M glycine, 0.3 M sodium chloride pH 8.6) and run on an affinity column containing 2 mls of protein A bound to glass beads (Prosep A) in binding buffer. The antibody was eluted with 0.1 M citrate pH 3.0, the fractions were monitored by their OD₂₈₀, and the elution peak containing antibody was pooled. The pooled peak was dialysed overnight into phosphate buffered saline (PBS) at pH 7.4. The elution profile of the protein A column and corresponding SDS-PAGE are shown in Figure 5.7. The elution peak is 4F5 and the flow-through is contaminating proteins, mainly BSA.

Purified 4F5 and 2G4 IgGs were cleaved into Fab and Fc fragments using papain. The IgG was concentrated to 4 mg/ml and then incubated with 20 mM EDTA, 100 mM L-cysteine and 18 mg/ml papain at 0.05% v/v (2G4) or 0.1% v/v (4F5) for 6 hours at 37°C. These conditions were found to be optimal by setting up trial digests at different papain concentrations and incubation times. The maximum reaction volume that was successful was 500 µl. The reaction was quenched by adding 200 mM iodoacetamide and incubating for 30 minutes at room temperature.

The digested antibody was dialysed into binding buffer overnight and run on a protein A column, using the protocol for the intact IgG purification. The flow through was collected and dialysed overnight into TPO buffer (50 mM sodium chloride, 20 mM Tris pH 7.6, 0.1 mM potassium iodide). The purified Fab and Fc were filtered through a 0.45 µm filter and concentrated to about 8 mg/ml in a Centricon 10 M_w cut-off concentrator. The SDS-PAGE of the papain treated 4F5 and 2G4 is shown in Figure 5.8, along with the elution profile of the protein A column. The percentage yield was 11% for 4F5 and 20% for 2G4. The yield for each step is shown in Table 5.4.

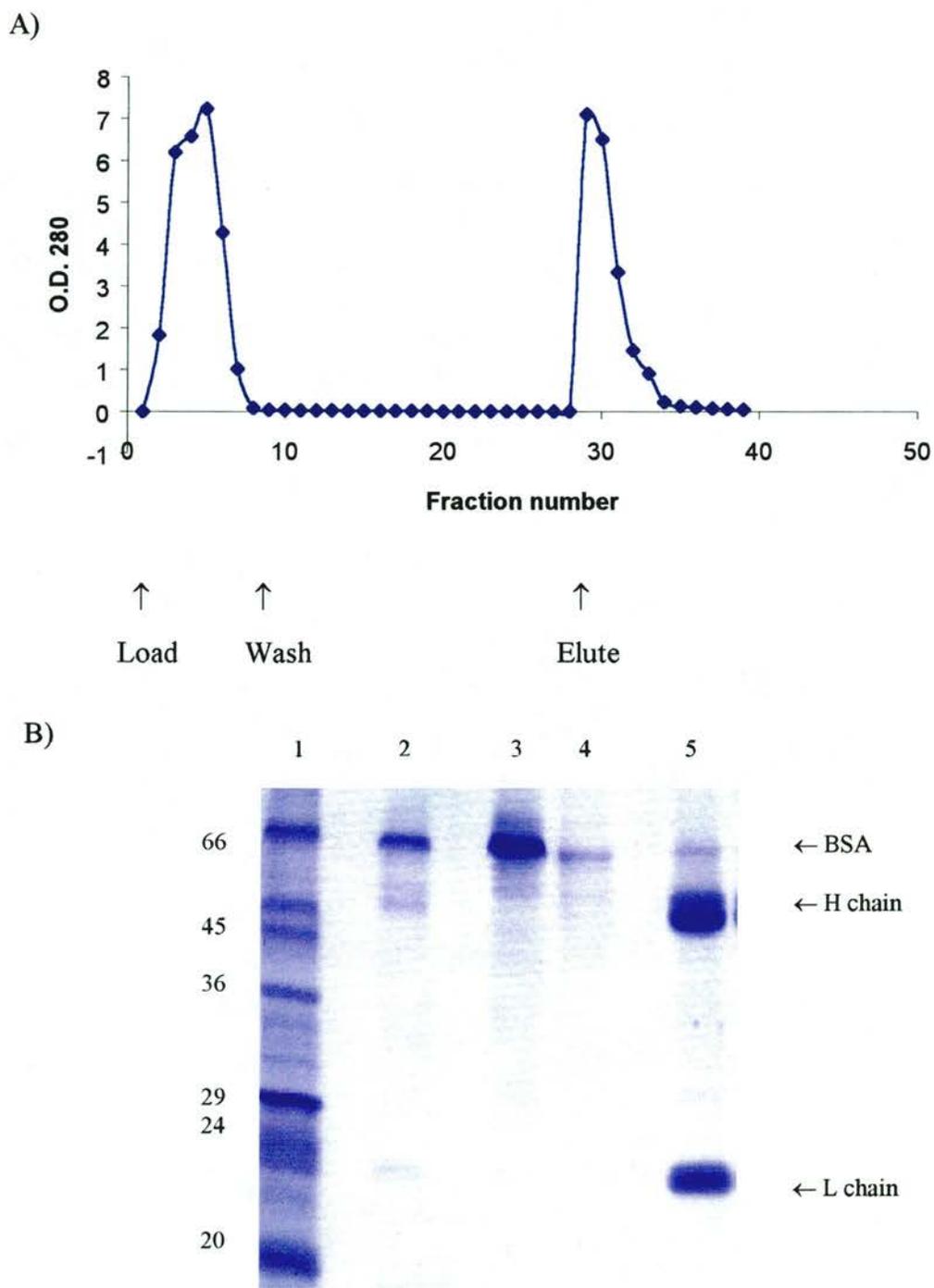


Figure 5.7

The elution profile of the Protein A column used to purify intact 4F5 (A) and a 12% reduced SDS-PAGE gel showing the 4F5 purification (B).

- | | | | | | |
|----|---------------|----|--------------|----|---------|
| 1. | Markers | 3. | Flow through | 5. | Elution |
| 2. | Load material | 4. | Wash | | |

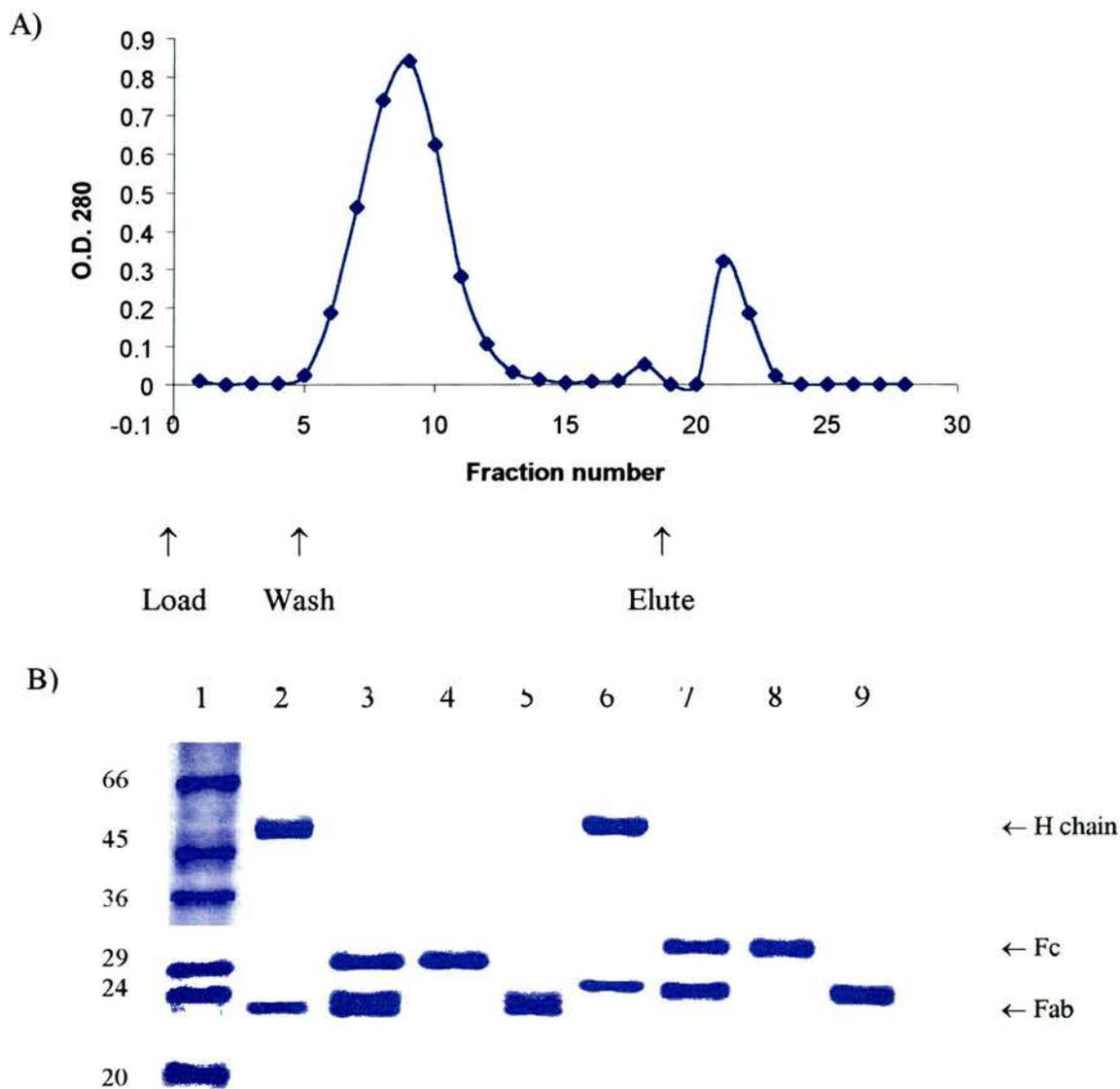


Figure 5.8

The elution profile of the Protein A column used to purify 4F5 Fab (A) and a 12% reduced SDS-PAGE gel of 4F5 and 2G4 Fab purification (B).

- | | |
|---------------------------------|----------------------|
| 1. Low molecular weight markers | |
| 2. 4F5 IgG | 6. 2G4 IgG |
| 3. 4F5 Papain digest | 7. 2G4 Papain digest |
| 4. 4F5 Fc | 8. 2G4 Fc |
| 5. 4F5 Fab | 9. 2G4 Fab |

Step	4F5		2G4	
	Amount (mg)	Yield (%)	Amount (mg)	Yield (%)
Cell supernatant	21.3	100	-	-
Protein A column I	14.6	69	22.9	100
Papain digest	9.8	46	16.8	73
Protein A column II	5.7	26	9.7	42
Concentration	2.3	11	4.5	20

Table 5.4

The yield from each step of the 4F5 and 2G4 Fab purification.

5.6 4F5 and 2G4 Fab Fragment Characterisation

During the purification procedure a radioimmunoassay (RIA) was used to monitor the concentration of 4F5 and 2G4 IgGs and Fab fragments. Following purification, the binding of the Fab fragments to TPO was characterised and the affinity constant for the interaction was estimated using Scatchard analysis. The affinity constant of the binding of full length IgGs to TPO was also carried out to allow comparison.

The full-length antibodies were assayed for TPO binding using a direct radioimmunoassay [Beever 1989]. ^{125}I -labelled TPO (RSR Ltd), anti-mouse IgG coated cellulose suspension (Sac-cel) and the antibody sample were diluted in assay diluent (10mM Tris pH 7.5, 150 mM NaCl, 0.1 mM bovine serum albumin, 10 mM sodium azide, 0.1% Tween). 50 μl of serial dilutions of a 0.2 mg/ml antibody preparation were mixed with 50 μl of ^{125}I -TPO and the reactions incubated for one hour at room temperature. 100 μl of Sac-Cel was added to each tube and incubated for a further hour at room temperature. Finally 1 ml of assay diluent was added to each tube, the tubes vortexed, and then spun at 2400 rpm for 30 minutes at 4°C.

The supernatant was aspirated and the radioactivity of the pellets counted for 1 minute.

The assay for 2G4 was identical, except that 50 μ l of Protein A instead of Sac-cel was used as the precipitating agent. All tubes were set up in duplicate and the mean count used in calculations. Percentage binding was calculated as the counts per minute for the sample divided by the total counts per minute for the 'total' tube (50 μ l of 125 I-TPO only). Percentage binding was then converted to antibody concentration using a calibration curve of antibody standards. The amount of 125 I in the pellet measures the percentage antibody binding as the radiolabelled TPO will only be precipitated if bound to Sac-Cel via the antibody. The calibration curve using 4F5 standards for the direct assay RIA is shown in Figure 5.9. The graph is sigmoidal and calibrations readings were taken from the central linear section.

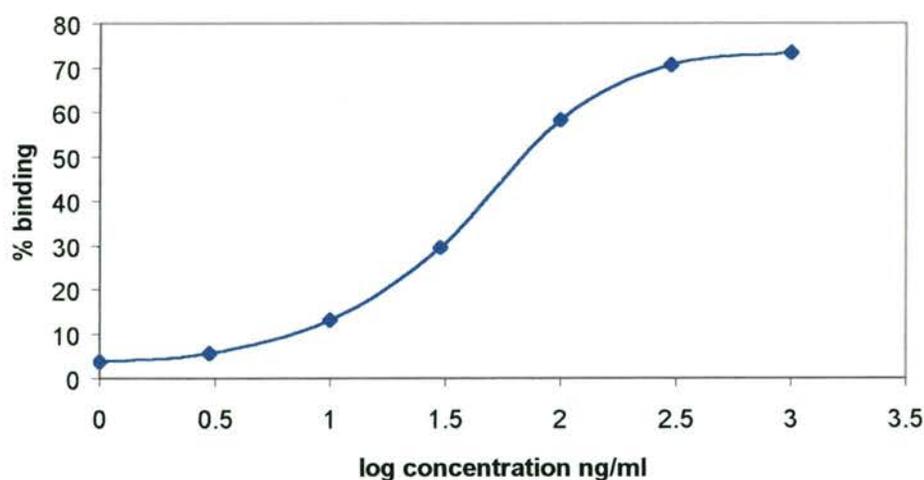


Figure 5.9

The calibration curve for the 4F5 antibody direct assay RIA.

The same assays were used to characterise the binding to TPO of Fab fragments. However, as Protein A interacts with the Fc region of IgGs, the 4F5 Fab assay was modified. A new assay was developed incorporating an anti-Fab antibody to interact with the Sac-cel antibody. 50 μl of ^{125}I -labelled TPO and 50 μl of 4F5 or 2G4 Fab (10 to 10^6 dilution of a 0.2 mg/ml solution) were incubated for one hour at room temperature. 50 μl of Fab specific antimouse IgG from goat (diluted 1:10²) was then added and incubated for a further hour at room temperature. 100 μl of Sac-cel binding goat antibodies was added and the tubes were incubated for a final hour at room temperature before spinning and counting as before. The method used to assay IgG and Fab concentration is summarised in Figure 5.10.

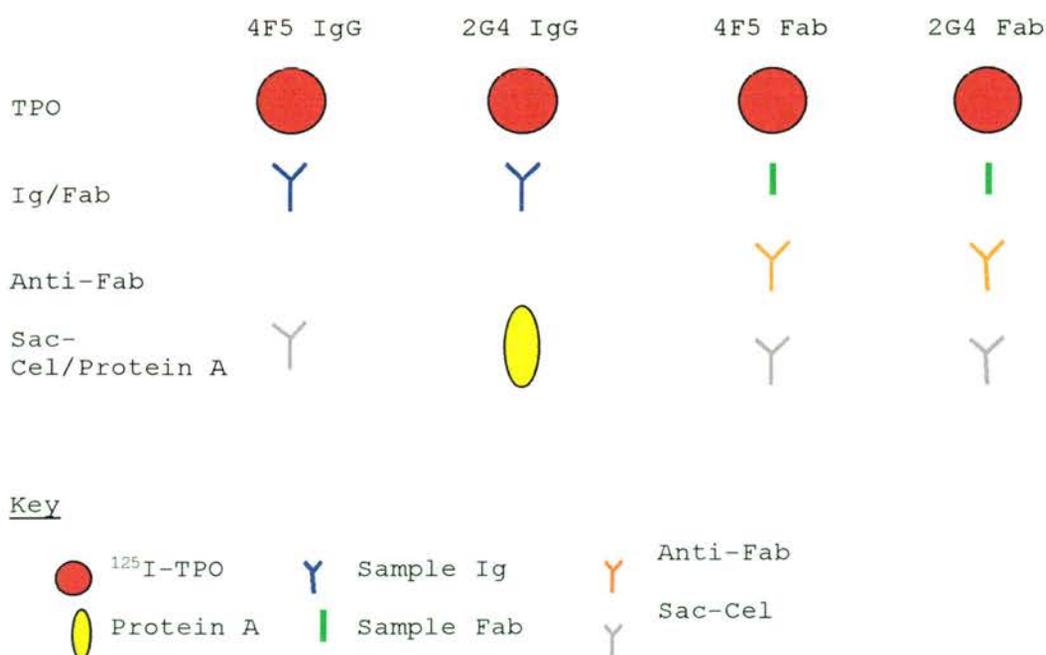


Figure 5.10

The methods used to assay the 2G4 and 4F5 IgG and Fab fragments.

In order to determine the affinity of 4F5 and 2G4 for TPO, Scatchard analysis was used [Scatchard 1949]. A similar assay to that used to calculate percentage binding was carried out, except that 50 µl of cold TPO was added to the first incubation. Each assay was repeated. 4F5 Fab was assayed with mouse Sac-Cel in addition to adding anti-Fab IgG and goat Sac-Cel, as the two methods gave similar results. The affinity constant, K_S , was calculated using the equation:

$$v = n - K_S v/[L]$$

where v is the number of moles of TPO bound per mole of antibody, n is the number of identical non-interacting sites on the antibody and $[L]$ is the concentration of free TPO. A graph of v against $v/[L]$ (Scatchard plot) has a slope of $-K_S$ with an intersect on the $v/[L]$ axis of n . The specific activity of the ^{125}I -TPO was 740 Bq/ng and the gamma counter efficiency was 0.78.

The total TPO (ng/tube) is calculated as follows:

$$\text{Total TPO} = \text{Cold TPO} + \text{Hot TPO}$$

Where:

$$\text{Hot TPO} = \text{Count of 'total' tube} / 740 \text{ (Bq/ng)} \times 0.78 \times 60 \text{ sec}$$

The bound TPO concentration (ng/tube) is calculated as follows:

$$\frac{[\% \text{ binding} - \text{background } \% \text{ binding (buffer only control)}] \times \text{total TPO ng/tube}}{0.78 \times 100}$$

The free TPO concentration (ng/tube) is:

$$\text{Total TPO} - \text{Bound TPO}$$

The bound TPO is converted to molar concentration:

$$\text{TPO (g) / Mr TPO (90 000) x vol in assay tube (l)}$$

A graph was plotted of bound/free TPO against bound TPO, and the gradient of the best-fit line through the points was calculated using Trendline in Microsoft Excel.

The percentage binding of the purified, intact 4F5 and 2G4 and Fab fragments to TPO is shown in Figure 5.11. Nonspecific binding is measured by the percentage binding in the presence of zero standard and was $2.73\% \pm 0.69$ (mean \pm SD, $n = 11$) for 2G4, and $2.92\% \pm 0.34$ ($n = 19$) for 4F5. Five out of the six curves show similar sigmoidal behaviour as the calibration curve, showing that the purified IgGs and Fabs bind to TPO. 2G4 Fab assayed with Protein A show no evidence for binding, as protein A binds to the Fc region of antibodies this result was expected. In contrast, mouse Sac-cel interacts with 4F5 Fab to the same extent as the interaction of with full length IgG. This indicates that the anti-Ig antibody is specific for the Fab region of murine antibodies.

The concentration of IgG molecules read from the calibration curve was comparable with that estimated by measuring the absorbance at 280 nm. The same concentration of IgG and Fab molecules were used in all six experiments, however the different assay procedure used in each case means that the results are not directly comparable. For example, 2G4 Fab (assayed with anti-Fab and goat Sac-cel) shows higher percentage binding than 2G4 IgG (assayed with protein A) due to the use of protein A instead of Sac-cel.

Scatchard analysis of the anti-TPO antibody interactions is shown in Figure 5.12. The analysis gave approximately linear plots, as measured by the R^2 value (a perfect straight line has an R^2 of 1). The affinity constants were calculated assuming a univalent interaction and are shown in Table 5.5. The results show that the affinity of 4F5 IgG for TPO ($2.4 \times 10^9 \text{ M}^{-1}$) is about ten-fold higher than 2G4 IgG ($4.2 \times 10^8 \text{ M}^{-1}$). Both the 4F5 Fab and 2G4 Fab have slightly lower affinities than the full

length antibodies for TPO ($1.5 \times 10^9 \text{ M}^{-1}$ and $1.4 \times 10^8 \text{ M}^{-1}$ respectively). The two runs of the 4F5 Fab assay (with and without the anti-Fab) gave similar results, showing that addition of the anti-Fab antibody does not affect the outcome of the 2G4 experiment.

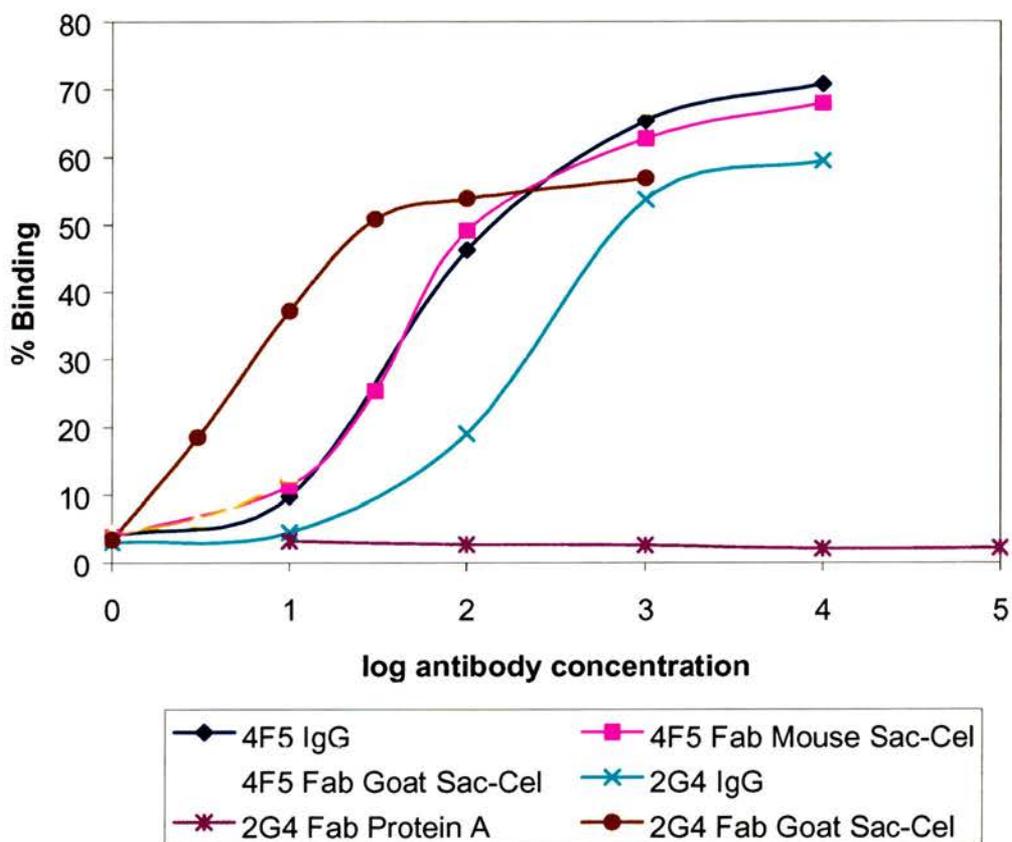


Figure 5.11

The percentage binding of 4F5 and 2G4 Igs and Fabs to TPO.

Antibody	Experiment 1		Experiment 2		Mean $K_S M^{-1}$
	$K_S M^{-1}$	R^2	$K_S M^{-1}$	R^2	
2G4 IgG	4.2×10^8	0.89	4.2×10^8	0.98	4.2×10^8
2G4 Fab (anti-Fab)	1.5×10^8	0.74	1.3×10^8	0.74	1.4×10^8
4F5 IgG	2.9×10^9	0.97	1.9×10^9	0.75	2.4×10^9
4F5 Fab	1.5×10^9	0.94	1.5×10^9	0.76	1.5×10^9
4F5 Fab (anti-Fab)	2.0×10^9	0.87	4.2×10^9	0.82	3.1×10^9

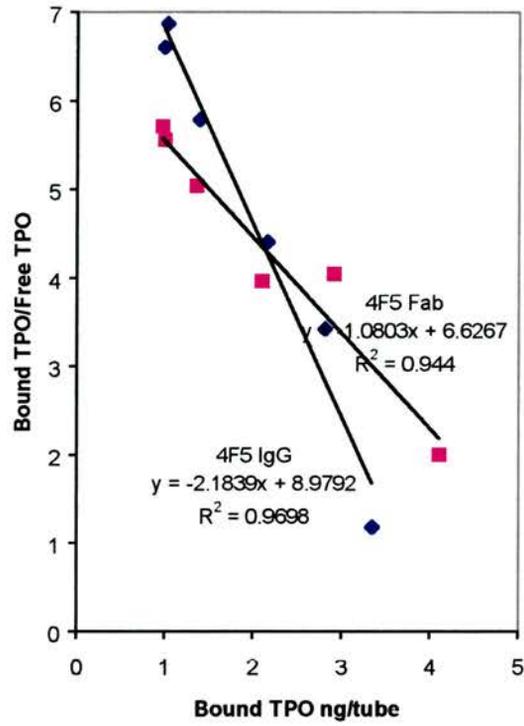
Table 5.5

The results of Scatchard analysis of 2G4 and 4F5 intact antibody and Fab fragments, each experiment was repeated and the mean calculated.

Affinity constants of cloned TPO autoantibodies range from $1.4 \times 10^8 M^{-1}$ to $1 \times 10^{11} M^{-1}$ [McIntosh 1998], and the affinity of autoantibodies to TPO in patients' sera are reported to be $1 \times 10^9 M^{-1}$ [Beever 1989]. The calculated affinities of 4F5 and 2G4 antibody and Fab fragments fall within the range of those observed in cloned autoantibodies and are similar to that of patient sera. The affinity of 2G4 for TPO is ten fold lower than that previously reported for 2G4 [Horimoto 1992] ($2.5 \times 10^9 M^{-1}$). The same method for calculating the affinity constant was used in both cases, the only difference being that 2G4 had been re-cloned for the second determination. It is unclear if re-cloning has altered the affinity constant, or if the difference was due to experimental technique.

The similar affinity constants obtained for the full length and Fab fragments of 2G4 and 4F5 confirm that the purified Fab domains have retained the ability to bind TPO. The slightly higher affinity of the full-length molecules compared with the Fab may be due to the ability of the full-length antibody to form bivalent interactions with two TPO molecules. Alternatively, some of the Fab may have lost binding affinity, for example due to denaturation during preparation.

A)



B)

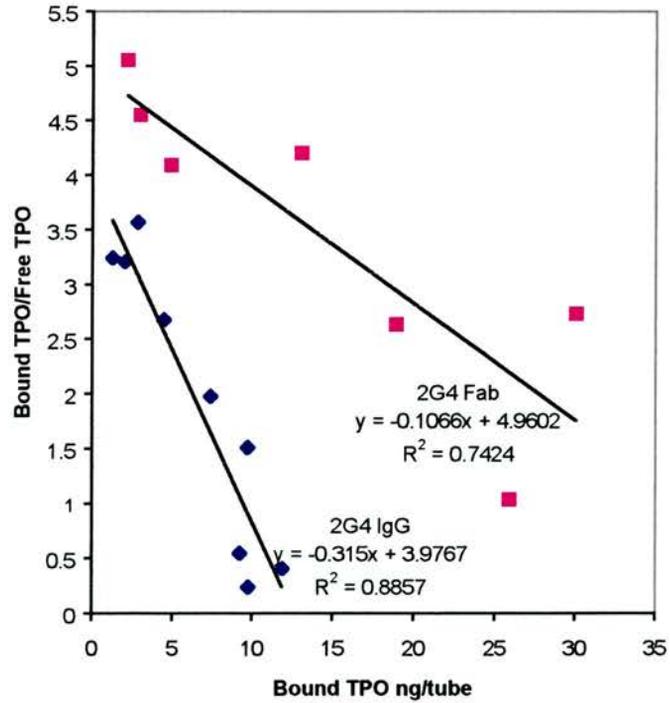


Figure 5.12

Scatchard analysis of A) 4F5 and B) 2G4 IgGs and Fab fragments.

5.7 Summary

The variable domain DNA of the LC and HC of 4F5 has been sequenced. The germline V and J genes have been identified and replacement somatic mutations found to be concentrated in the CDRs.

The binding of 4F5 and 2G4 to TPO has been studied. It has been shown that 2G4 and 4F5 recognise two different epitopes on TPO, and that these are the major epitopes recognised by autoantibodies from patients with AITD.

In preparation for crystallisation of 4F5 and 2G4, the Fab fragments of both antibodies have been produced and purified to homogeneity. Characterisation of the Fab fragments showed that they bound to TPO with high affinity. 2G4 Fab bound to TPO with an affinity of $1.4 \times 10^8 \text{ M}^{-1}$, 4F5 Fab bound with an affinity of $1.5 \times 10^9 \text{ M}^{-1}$. The affinities were comparable with those of the full-length antibodies, showing that the Fab fragments had the same TPO binding properties.

Chapter 6

4F5 Crystallography

6.1 Introduction

2G4 is a human IgG1, kappa monoclonal antibody isolated from Hashimoto thyroid lymphocytes. 4F5 is a mouse IgG1, kappa monoclonal antibody. Each antibody has been shown to bind to a different autoimmune epitope on TPO.

The structure of the Fab fragments of 2G4 and 4F5 will allow study of the nature of the antibody combining sites. The Fc fragment of TPO autoantibodies also has biological interest as the effector functions mediated by this region have been implicated in disease [Salvi 1988].

Crystallisation of the Fab fragments of 4F5 and 2G4 and the Fc fragment of 2G4 was therefore attempted. Crystals were grown of each fragment and data were collected from 4F5 Fab. The large number of solved Fab structures, all with the same fold, makes molecular replacement the obvious method for solving the phase problem.

6.2 Crystallisation

Pure 4F5 Fab (8.3 mg/ml), 2G4 Fab (9.3 mg/ml) and 2G4 Fc (6.2 mg/ml) in 50 mM sodium chloride, 20 mM Tris pH 7.6, 0.1 mM potassium iodide were used to set up Hampton Screens I and II using the hanging drop method as described in 3.2. The

drops contained 1-2 μl of protein and 1-2 μl of reservoir, and the plates were kept at room temperature.

a) 4F5 Fab

Crystals and plates of 4F5 were observed in the following conditions after one to three days:

- 20 % PEG 8K, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate tetrahydrate
- 30 % PEG 4K, 0.1 M Tris pH 8.5, 0.2 M sodium acetate trihydrate
- 30% PEG 4K, 0.1 M trisodium citrate dihydrate pH 5.6, 0.2 M ammonium acetate
- 2% PEG 400, 2.0 M ammonium sulphate, 0.1 M Hepes pH 7.5
- 2 M ammonium sulphate, 0.1 M Tris pH 8.5
- 2 M ammonium sulphate
- 0.2 M ammonium sulphate, 0.1 M MES pH 6.5, 30 % PEG MME 5K
- 20 % PEG 10 K, 0.1 M Hepes pH 7.5

The most promising looking crystals were in the first two conditions and these were optimized by varying pH from 6 to 9, using 30% PEG 4K and 20% PEG 8K and with both 0.2 M magnesium and sodium acetate. Crystals grew in many of the optimised conditions. The crystals dissolved due to temperature fluctuations during transportation but re-grew after two weeks at 20°C. The crystals were about 200 μm long and are shown in Figure 6.1.

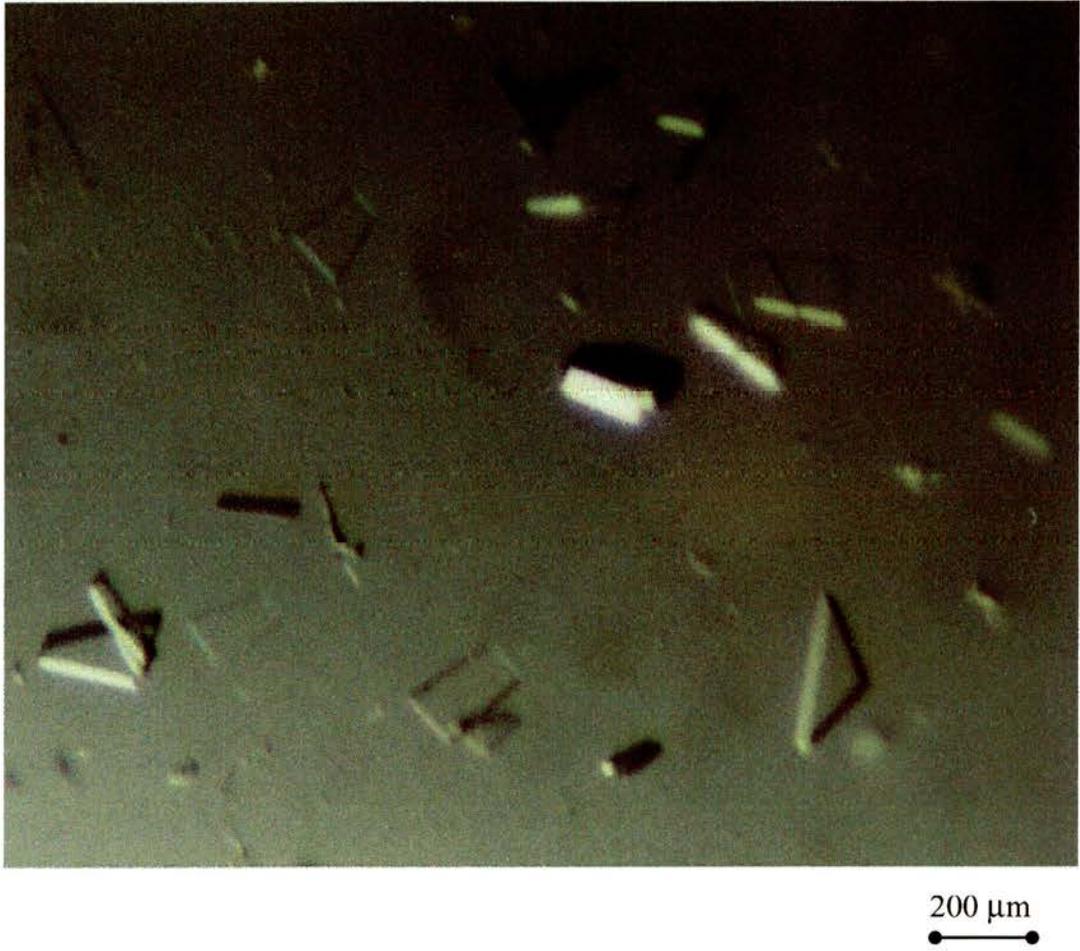


Figure 6.1

Crystal of 4F5 Fab grown in 20 % PEG 8K, 0.1 M Tris pH 8.0, 0.2 M sodium acetate.

b) 2G4 Fab

Rhombohedral crystals about 50 μm in length were observed in 0.5 M lithium sulphate monohydrate, 15% PEG 8K after two days. Attempts to optimise the crystals in 0.3 M to 0.8 M lithium sulphate and 10 % to 20 % PEG 8K failed to produce any crystals. Fishing for the original crystals with cryoloops was problematic as the drop had become biphasic and the crystals clumped together. A repeat of the successful Hampton Screen condition on the same protein preparation produced heavy precipitate rather than crystals, and this was thought to be due to a modification of the protein caused by storage at -70°C . A fresh protein preparation was carried out and the protein set down for crystallisation trials without being subjected to low temperatures, but no crystals could be grown. As no diffraction data could be collected, and the crystals could not be reproduced, it could not be confirmed that the crystals were of 2G4.

c) 2G4 Fc

After a few hours many small rods were visible in 0.2 M ammonium acetate, 0.1 M trisodium dihydrate pH 5.6, 30% PEG 4K. After one day larger rods had grown in 0.2 M ammonium sulphate, 0.1 M sodium cacodylate pH 6.5, 30% PEG 8K. Both crystals are shown in Figure 6.2. The crystals were flash frozen in 15% glycerol and also mounted in quartz capillaries at room temperature, but no diffraction was observed with in-house X-rays. This was probably due to disorder in the crystals. Optimization of the crystallisation conditions failed to produce any diffracting crystals.

A)



50 μm

B)



200 μm

Figure 6.2

2G4 Fc crystals in A) 0.2 M ammonium acetate, 0.1 M trisodium dihydrate pH 5.6, 0% PEG 4K B) 0.2 M ammonium sulphate, 0.1 M sodium cacodylate pH 6.5, 30% PEG 8K.

6.3 Data Collection

Crystals of 4F5 Fab grown in 0.2 M sodium acetate, 20 % PEG 8K 0.1 M Tris pH 8.0 were placed in a 15% solution of glycerol in mother liquor for 2 minutes and then fished into a 0.5 mm diameter loop and flash-frozen at 100 K in liquid nitrogen. Data were collected at the Deutsches Elektronen-Synchrotron (DESY), Hamburg on beamline X11, $\lambda = 0.902 \text{ \AA}$, on a MAR-CCD detector (165 mm), with a crystal to detector distance of 150 mm. 145 frames were collected with an oscillation angle of 1° and a dose dependent exposure time of approximately five minutes per image. The experimental set-up is shown in Figure 6.3. The data were indexed and scaled using the programmes DENZO and SCALEPACK from the HKL suite [Otwinowski & Minor 1996].

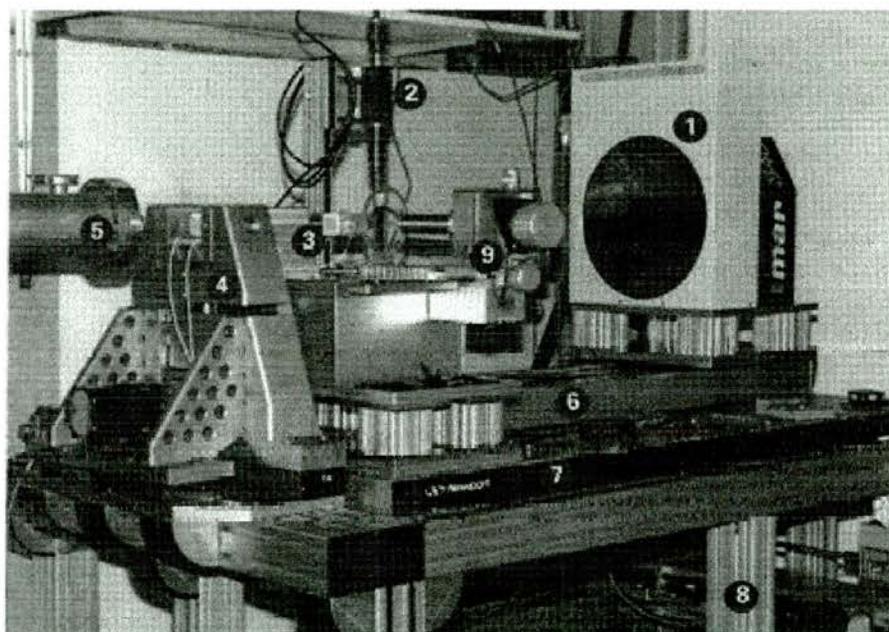


Figure 6.3

The experimental set-up at the X-11 station.

(1) Mar imaging plate, (2) video microscope, (3) collimator (4) mounting structure (5) beam exit (6) detector translation (7) optical table (8) aluminium frame structure on wheels, (9) rotation axis for crystal with an integrated angle measuring device.

The 4F5 Fab crystals diffracted to 1.86 Å, a typical diffraction image is shown in Figure 6.4. The images indexed in crystal class P2 and the presence of 0k0 systematic absences indicated a screw axis along b. Therefore the space group was assigned as P2₁. The unit cell is a = 36.7 Å, b = 83.7 Å, c = 68.7 Å, α = γ = 90°, β = 99.1°. The Matthews calculation (Section 3.2) gave 43% solvent with one molecule in the asymmetric unit. The data collection statistics are given in Table 6.1 and the space group diagram in Figure 6.5.

	All Data	Highest Resolution shell (1.93 – 1.86 Å)
Space group	P2 ₁	-
Unit Cell	a = 36.7 Å, b = 83.7 Å, c = 68.7 Å, β = 99.1°	-
Measured Reflections	153 646	-
Unique Reflections	33 420	3 138
R _{merge} † (%)	3.0	9.0
I/sigI	27.5	6.5
Completeness (%)	96.3	91.1
Mosaicity (°)	0.30	-
Resolution (Å)	1.86	-

Table 6.1

The data collection statistics for 4F5.

† R_{merge} = $\sum |I(k) - \langle I \rangle| / \sum I(k)$, where I(k) is the value of the kth measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

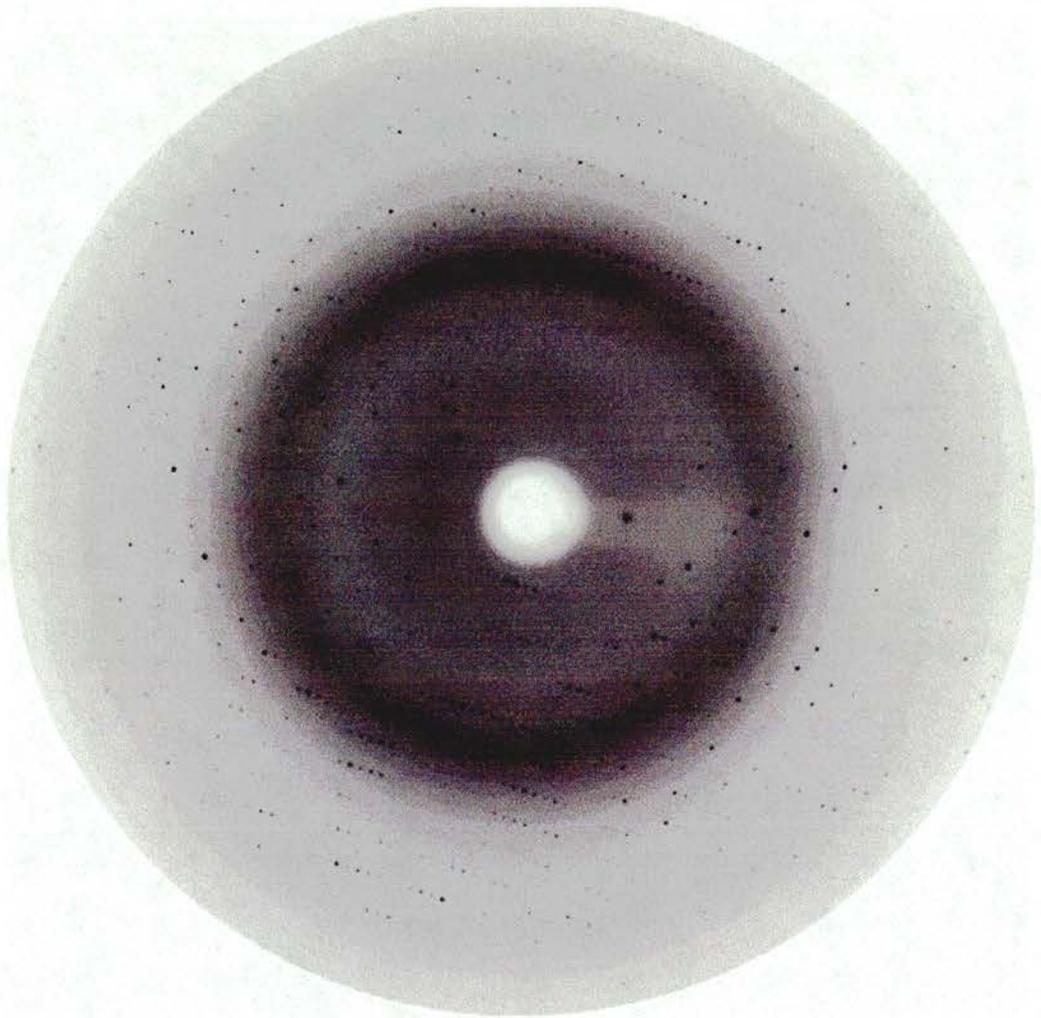


Figure 6.4

The diffraction image of a 4F5 crystal.

$P2_1$

$P 1 2_1 1$

2

No. 4

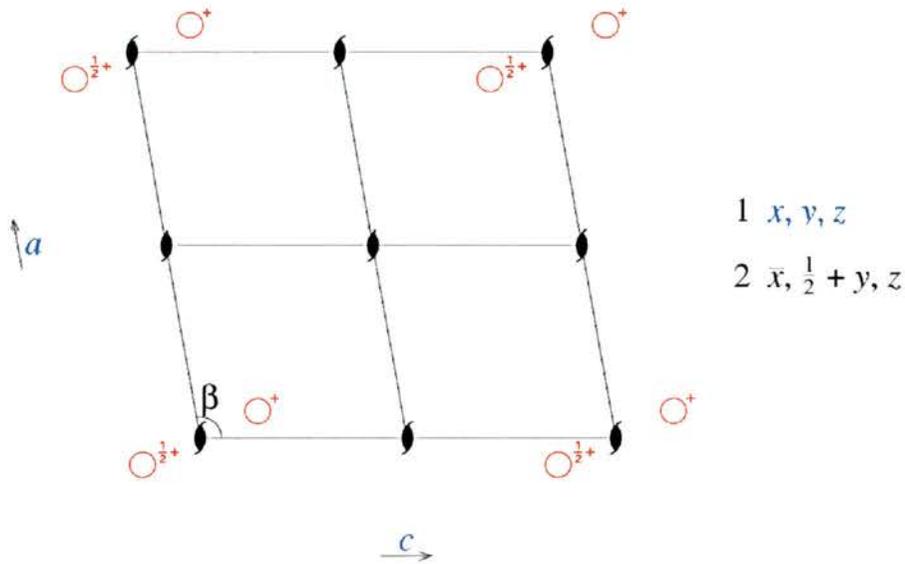


Figure 6.5

The space group diagram for $P2_1$.

6.4 Molecular Replacement

The scaled intensities must be combined with phases in order to calculate electron density. Molecular replacement is an ideal approach for solving a Fab structure given that there are over 200 known structures in the database [ABG 2000]. All known Fab structures have the same overall structure, the main difference being the elbow angle between the constant and variable domains, which can range from 127.2° to 176.2° [Padlan 1994]. Therefore, the variable and constant domains need to be treated separately in molecular replacement.

The merged reflection file from SCALEPACK was converted to mtz format using the CCP4 programme SCALEPACK2MTZ [Collaborative Computer Project

Number 4 1994], which uses TRUNCATE to convert intensities into structure factor amplitudes (F_s) and CAD to sort the data into the correct asymmetric unit of reciprocal space. The format of the reflection file was converted into X-PLOR [Brunger 1992] format (F_{obs}) using MTZ2VARIABLES. The pdb file used for molecular replacement was HyHEL-5, an antibody complexed with lysozyme (2hfl). This is a mouse IgG1 kappa antibody solved to 2.5 Å. Sequence identity could not be taken into account when choosing the molecular replacement model as the sequence of 4F5 was not known at this stage. The pdb file was converted to X-PLOR format using PDBSET [Collaborative Computer Project Number 4 1994], and a pdb and protein structure (psf) file for input into X-PLOR were created using GENERATE. A psf file contains atom, bond, dihedral and hydrogen bonding terms.

The first step is a cross-rotation function performed in Patterson space. A Patterson map of the whole 2hfl Fab molecule was calculated, and rotated with respect to that of 4F5, using data from 10 Å to 4 Å and an angular grid size of 5°. The structure factors were compared at each orientation and their correlation expressed by an R-function. The orientation was expressed in Eulerian angles, θ_1 , θ_2 , and θ_3 [Brunger 1992]. Local rotation functions using the variable and constant domains were performed 30° either side of the direct rotation function peak.

The translation search involves moving the 2hfl molecule through the asymmetric unit and calculating structure factors which are then compared to those of 4F5. Both domains were rotated to their refined position and a Patterson correlation translation search was carried out with data from 15 Å to 4 Å. A pdb file of the translated constant and variable domains was then produced and an R-factor calculated. The resulting pdb file was used in a translation search with the whole molecule. The method is summarised in Figure 6.6.

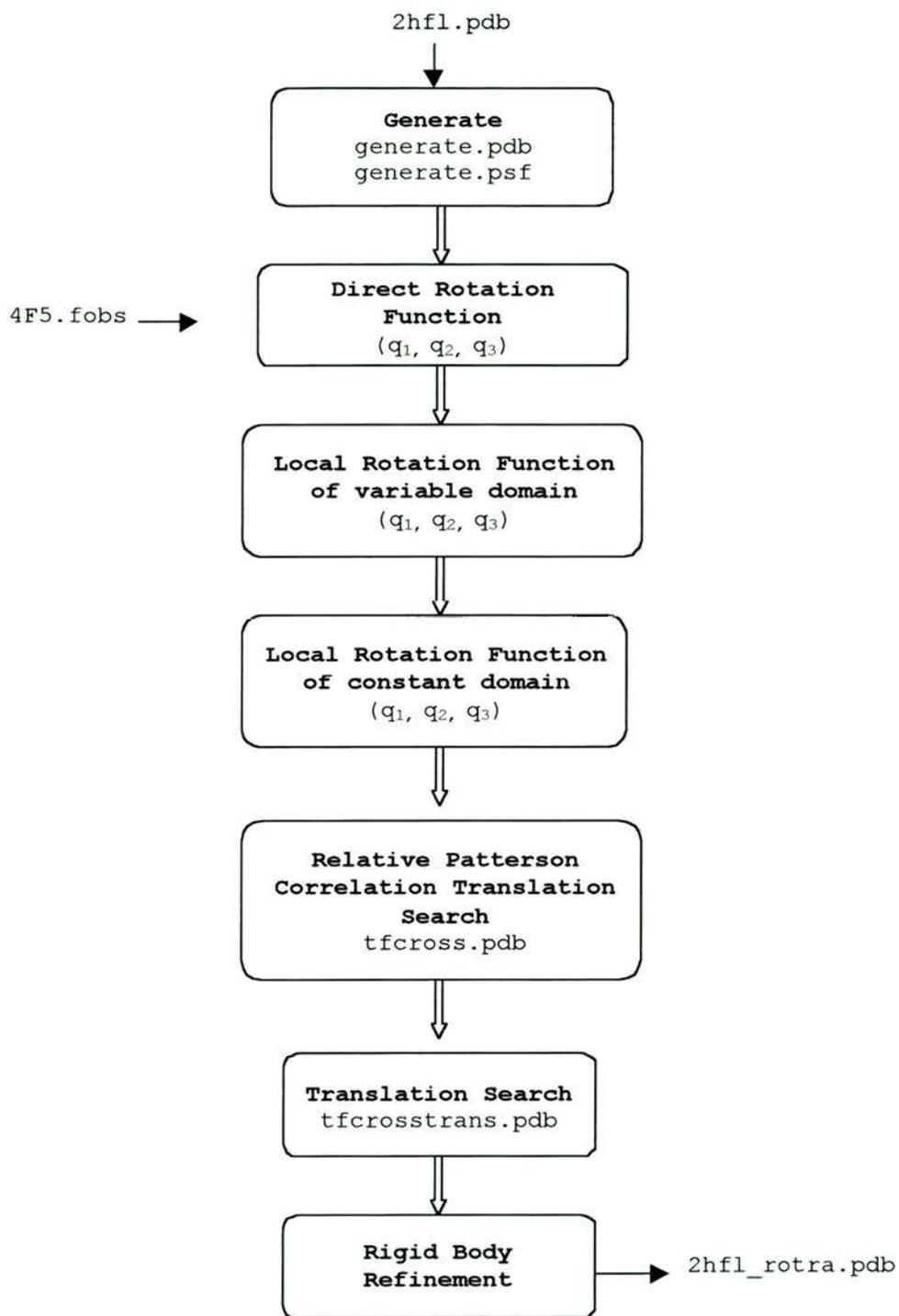


Figure 6.6

The procedure for molecular replacement of a Fab molecule in X-PLOR.

The direct rotation function gave a top peak of ($\theta_1 = 182.254$, $\theta_2 = 15.000$, $\theta_3 = 142.254$) with a R-function of 0.1165. The next highest R-function was 0.0888. The local rotation functions gave peaks of ($\theta_1 = 182.005$, $\theta_2 = 15.005$, $\theta_3 = 144.505$) and ($\theta_1 = 192.007$, $\theta_2 = 35.008$, $\theta_3 = 134.504$) for the variable and constant domains respectively, with R-functions of 0.1625 and 0.1378. The relative translation search gave an R-factor of 50.4% and the translation search of the whole molecule gave an R-factor of 43.6%.

Rigid body refinement was carried out with the whole Fab molecule using X-PLOR with data from 10 Å to 4 Å. Rigid geometry is assigned to the whole structure and the position and the orientation in the unit cell is optimised. This was followed by refinement in which the variable and constant domains were treated as separate rigid bodies using data from 8 Å to 4 Å. Rigid body refinement of the whole Fab to 4 Å gave an R-factor of 42.0%, refinement of the position of individual domains to 2 Å gave an R-factor of 47.3%. The translated and rotated pdb file was read into the molecular graphics programme O [Jones 1991], and the symmetry related molecules displayed. The molecules pack in a head-to tail fashion as is shown in Figure 6.7.

6.5 Refinement

Refinement was carried out with of cycles of reciprocal space calculations, and real space model building, as shown in Figure 6.8. Reciprocal space refinement was carried out in CNS [Brunger 1998] and progression was monitored using R-factors. A 4F5 reflection file was created with a test set of 5% of all reflections flagged, these were then omitted from refinement. The R_{work} is the R-factor corresponding to reflections in the working set, and the R_{free} is a more independent measure as it is the R-factor calculated from test set reflections only.

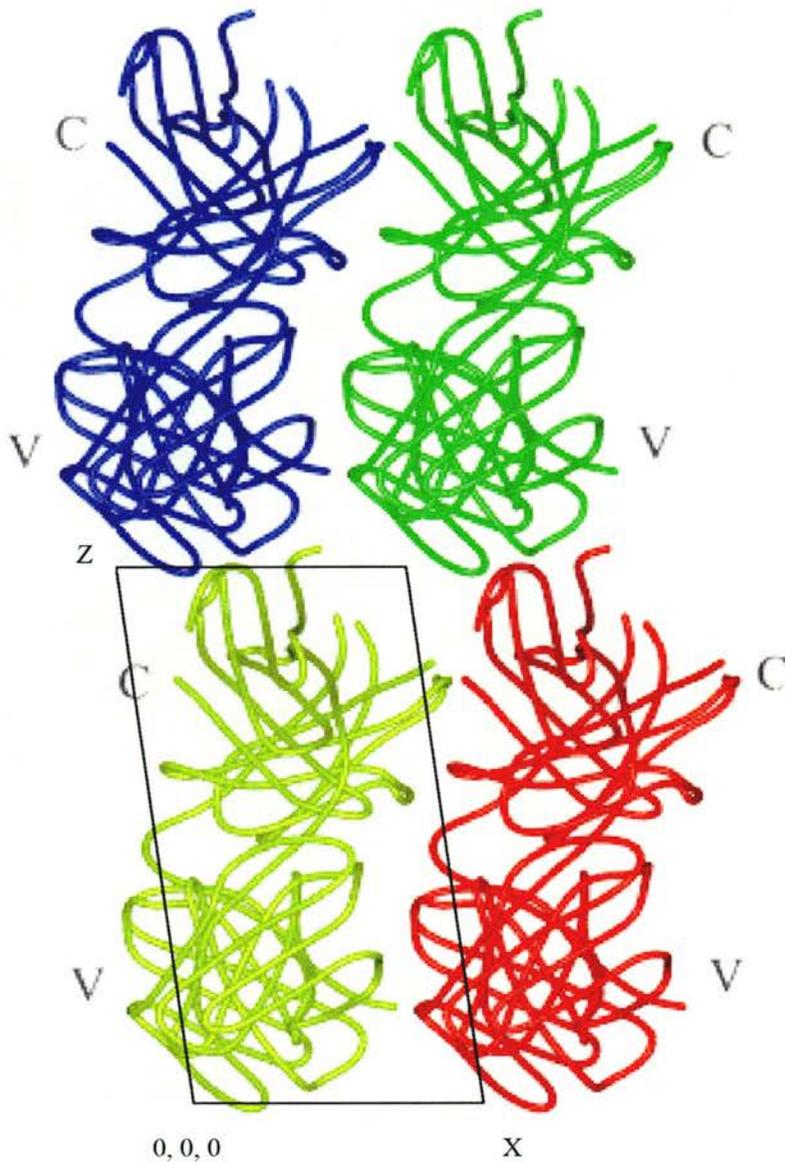


Figure 6.7

Packing of the 4F5 molecules in the unit cell, viewed down the Y-axis. Each Fab molecule is shown in a different colour, variable and constant domains are labelled.

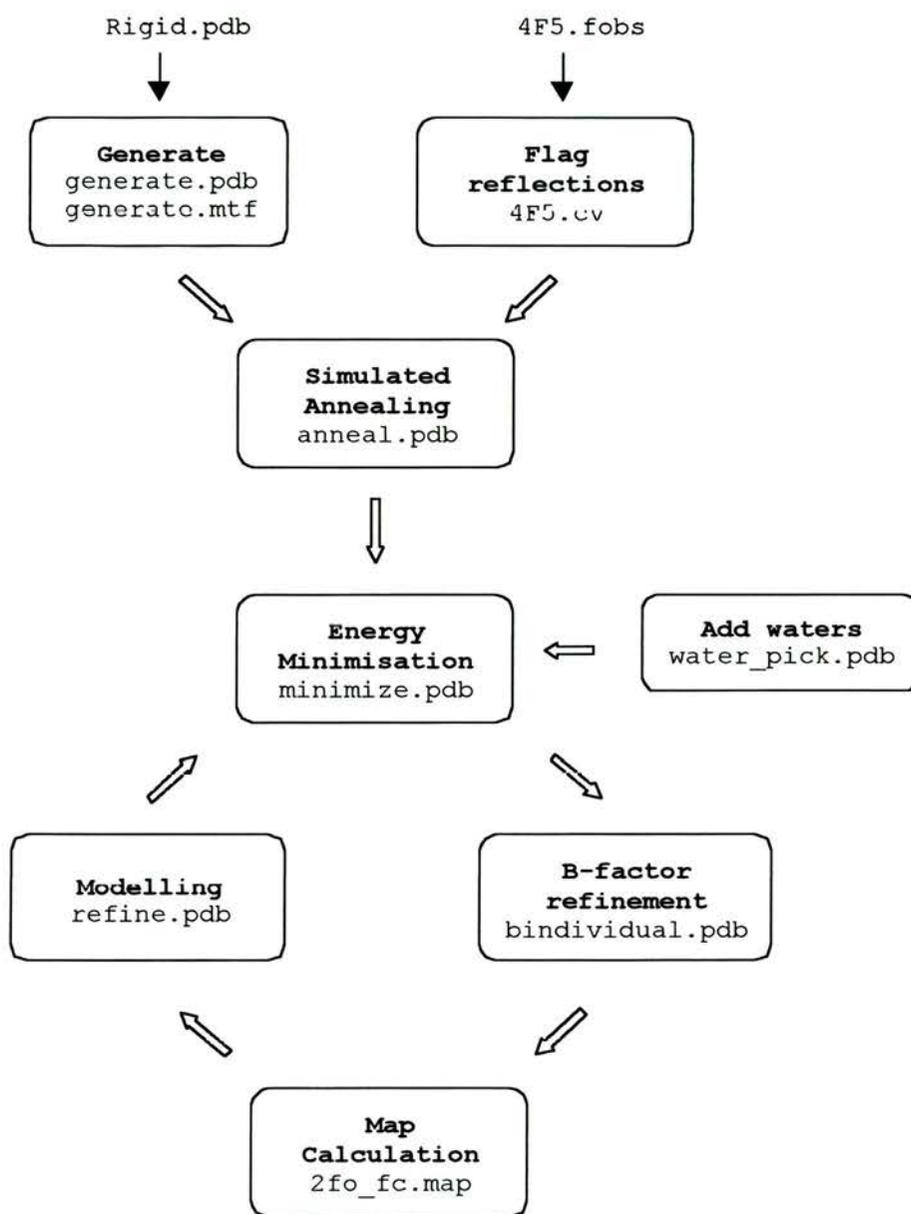


Figure 6.8
The refinement procedure in CNS.

The correctly positioned and orientated 2hfl pdb file was used to create a pdb and molecular topology file (mtf) in GENERATE. An mtf file is the equivalent of an X-PLOR psf file. B-factors were estimated from the Wilson plot in TRUNCATE to be 16.8 \AA^2 . Bulk solvent correction, in which the regions of electron density corresponding to solvent are flattened, was used throughout.

Simulated annealing was carried out using data from 20 \AA to 1.86 \AA with a starting temperature of 2500 K and 1000 cycles of molecular dynamics steps with a drop in temperature of 25 K at each step. Least-squares fitting of the position of the atoms and the B-factors was then used, initially with overall anisotropic B-factor correction. 20 steps of energy minimisation were followed by 10 steps of restrained individual isotropic B-factor minimisation. Refinement in CNS gave a R_{free} of 39.5% and a R_{work} of 34.9%.

A sigma-A weighted $2F_{\text{obs}} - F_{\text{calc}}$ map was calculated in CNS using phase information from the 2hfl model. The electron density maps were visualized in O, the initial $2F_{\text{obs}} - F_{\text{calc}}$ map was of good quality. As the amino acid sequence of 4F5 was not available, the residue type present in the phasing model was maintained, unless a mutation was obvious, in which case the residue was changed to alanine. 36 out of 212 residues of the LC and 21 out of the 215 residues of the HC were mutated to alanine. Loop regions that deviated from the electron density were remodelled using lego-loop. Loops L1, L2 and H1 appeared to have similar conformations to the phasing model and were retained. However, residues L90-93 (L3), H61-67 (H2) and H97-105 (H3) deviated from the density and were removed from the model. After a further round of refinement in CNS H2, H3 and L3 could be modelled with confidence.

Waters were added in CNS using automatic selection, with the minimum peak height set at 3σ and the distance between water and any atom in the range 2.6 \AA to 4 \AA . Rounds of manual rebuilding, refinement in CNS, and addition of water molecules led to a steady improvement of the map quality and assignment of all the

residues to specific amino acids. Ambiguous residues in the complementarity determining regions, such as aspartate/asparagine and glutamate/glutamine were assigned by studying $F_{\text{obs}} - F_{\text{calc}}$ maps, contoured at $+3\sigma$ and -3σ , of the 4F5 structure containing each of the possible amino acids. A cis-alanine was identified at residue H96 (in CDR H3), and a new parameter file to maintain cis geometry was created in CNS and used in subsequent refinement steps.

Towards the end of the refinement, the DNA sequence of 4F5 was determined. The amino acid sequence changes were incorporated into the model and refinement concluded with cycles of energy minimisation and individual B-factor refinement. The 4F5 amino acid sequence as deduced from the DNA sequence was the same as that derived from the structure except for 67 residues (16% of all residues). The single residue deletion in 4F5 L1 and the five residue deletion in 4F5 H3, as compared to 2hfl, had been correctly identified from the electron density.

The 67 residues that were incorrectly identified from the electron density map fall into two categories. Slightly over half of the changes led to small differences in the properties of the amino acids, for example from an asparagine to an aspartate, and were due to ambiguous density. The remainders were more substantial changes, for example alanine to glutamate. Inspection of the electron density map around the latter residues showed that density did not extend over the whole side chain. For example, density limited to the $C\alpha$ of a glutamate, made it appear to be an alanine. The lack of electron density around these residues was probably due to thermal or static disorder.

An alignment of the phasing model, the sequence identified from the structure, and the sequence derived from the DNA sequence is shown in Figure 6.9. The identity between the phasing model amino acid sequence and that of 4F5 is 79%. The amino acid differences between 4F5 and 2hfl are clustered in the CDRs. Both 4F5 and 2hfl share the same HC and LC constant regions.

VL

Phasing model	DIVMTQSOKF	MSISVGRVVS	ITC	KASQNVG	FAVAWYQOKP	GQSPKLMIYS
From structure	SIVLTQSPAV	MSASPGSKVT	VTCSASS.AV	VFIHWFQOKP	GIVPALWIIYS	
From DNA	QIVLTQSPAI	MSASPGVKVT	ITCSASS.SV	NFIHWFQOKP	GTSPKLWIIYS	

Phasing model	ASNRYTGVPD	RFTGSGSGTD	FTLTISNMQS	EDLADYFCQQ	YSSYPITFGA
From structure	TSNLASGVPA	RFSGTGSGTS	YSLTISAMSA	EDAATYYCLQ	RSSYPYTFGG
From DNA	TSNLASGVPA	RFSGSGSGTS	YSLTISRMEA	EDAATYYCLQ	RSSYPYTFGG

CL

Phasing model	GTKLEIKRAD	AAPTVSIFPP	SSEQLTSGGA	SVVCFLNIFY	PKDINVKWKI
From structure	GTKLSIKRAD	AAPTVIFIPP	SAEQLSSGGA	SVVCFLNIFY	PANIASAWLI
From DNA	GTKLEIKRAD	AAPTVSIFPP	SSEQLTSGGA	SVVCFLNIFY	PKDINVKWKI

Phasing model	DGSERQNGVL	NSATDQDSKD	STYSMSSTLT	LTKDEYERHN	SYTCEATHKT
From structure	DGSE...SVL	NSHSNQDGGD	STYSMTSTST	LTKDEYECHL	SYTCAATHKT
From DNA	DGSER..GVL	NSWTDQDSKD	STYSMSSTLT	LTKDEYERHN	SYTCEATHKT

VH

Phasing model	STSPIVKSFN	RNECEVQLQQ	SGAEVVRSGA	SVKLSCTASG	ENIKDYIHW
From structure	STAPIVVSFS	RA.AAVKLAE	SGPELILPGA	SVRISCAASG	YTFTSYIHW
From DNA	STSPIVKSFN	RN.EDVKLQQ	SGPELVKPGA	SVRISCKASG	YTFTSYIHW

Phasing model	VKQRPKGGLE	WIGWIDPEIG	DTEYVPKFOG	KATMTADTSS	NTAYLQLSSL
From structure	VKQRP...LE	WIGWIYPGNS	YTKYSVKFVG	AATLTADSA	STAYMQLVAL
From DNA	VKQRPQGGLE	WIGWIYPGNV	YTKYSEKFKD	KATLTADKSS	STAYMQLSSL

CH

Phasing model	TSEDTAVYIC	NAGHDYDRGR	EPYWGQGTLV	TVSSAKTTPP	SVYPLAPGSA
From structure	TSQDSAVYFC	GR.....DPF	LOYWGQGTTL	TVSSASTVPP	SVYPLAPG..
From DNA	TSEDSAVYFC	GR.....DAH	LEYWGQGTTL	TVSSAKTTPP	SVYPLAPG..

Phasing model	AQTNSMVTLG	CLVKGYFPEP	VTVTWNSGSL	SSGVHTFPAV	LQSDLYTLSS
From structureAMSTLG	CLVKGYFPEP	VVTWNSGSL	SVGVHTFPAV	LKSDLYTLSS
From DNASMVTLG	CLVKGYFPEP	VTVTWNSGSL	SSGVHTFPAV	LQSDLYTLSS

Phasing model	SVTVPSSTWP	SETVTCNVAH	PASSTKVDKK	IVPRD
From structure	SVTVPASTWP	SETVTCNVAH	PASSTKVDKK	ITP..
From DNA	SVTVPSSTWP	SETVTCNVAH	PASKTKVDKK	IVPR.

Figure 6.9

An alignment of the 2hfl sequence, the 4F5 sequence derived from the electron density map, and the 4F5 sequence deduced from the DNA sequence. The differences between the true 4F5 sequence and that derived from the structure are highlighted in blue, and between 4F5 and 2hfl in green. CDR residues are boxed.

Final refinement gave an R_{free} of 22.9% and an R_{work} of 19.3%. 327 water molecules were built into the structure. The LC is from L1 to L213, with L156 and L157 omitted from the model due to insufficient electron density. The HC model runs from H1 to H213, and is missing residues H128 to H133. Example sections of the final electron density map are shown in Figure 6.10. The progression of R_{free} and R_{work} during refinement is shown in Figure 6.11.

6.6 Validation

A Ramachandran plot was generated using Procheck [Laskowski 1993]. Further model statistics were generated in CNS.

The dihedral angles ϕ (ϕ) and ψ (ψ) are defined in Figure 6.12 and the Ramachandran plot is shown in Figure 6.13. The geometry and stereochemistry of the model are good, with 92.2% of residues in the most favoured region of conformational space. The residue in a disallowed region (threonine L51) is in a strained conformation with angles $\phi = 70.0^\circ$ $\psi = -49.6^\circ$. These are similar to the average angles that this residue adopts in 17 high resolution Ig structures [Al-Lazikani 1997]. The rmsd for bond angles is 1.39° and 0.0048 \AA for bond lengths. The average B-factor over the whole structure is 17.00 \AA^2 , with a minimum of 4.33 \AA^2 and a maximum of 66.54 \AA^2 .

A total of eighteen residues of amino acid type arginine, lysine and glutamate had poor electron density for some side chain atoms. Two of these residues were in the CDRs (glutamate H61 and lysine H64). The lack of density was probably due to thermal or static disorder of the side chains. The residues that have poor density are listed in Table 6.2 and example electron density from two of these residues shown in Figure 6.14.

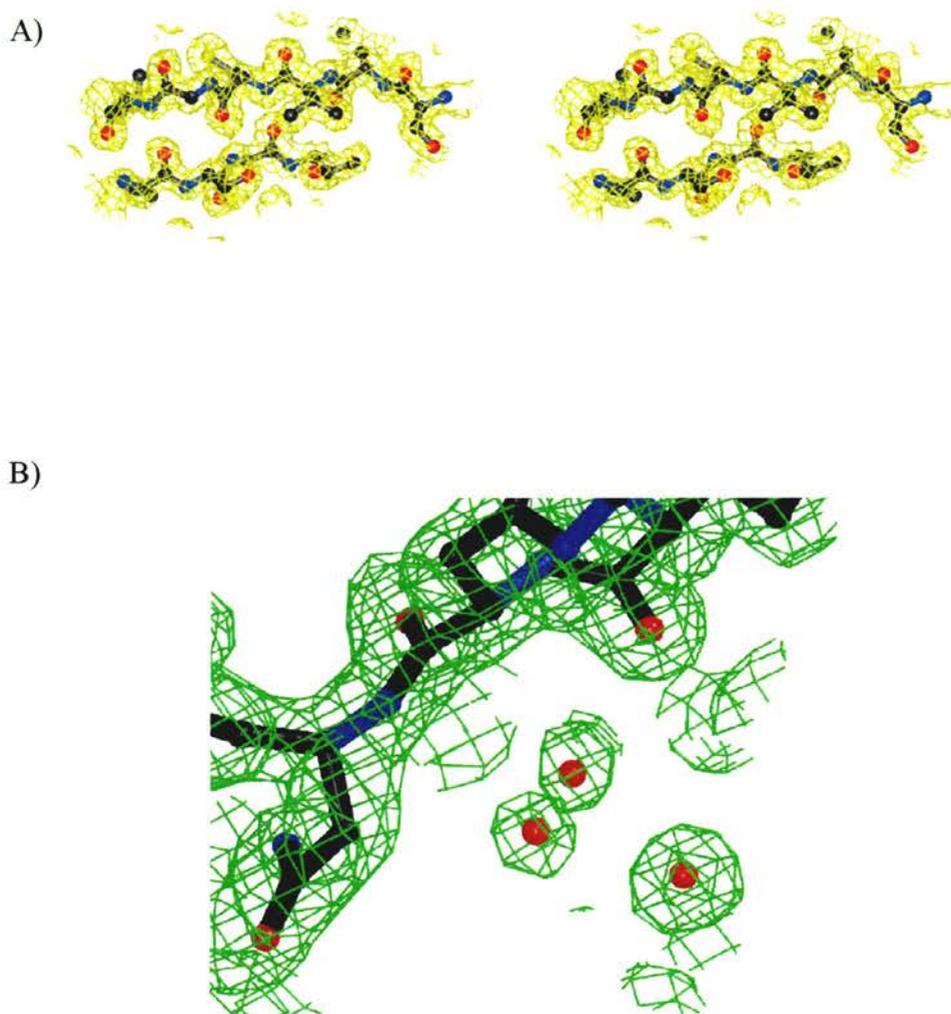


Figure 6.10

Sections of the final $2F_{\text{obs}} - F_{\text{calc}}$ electron density map.

A) Stereo diagram of β -strands (b and e) of the LC forming anti-parallel sheet C, contoured at 1 sigma.

B) An example of the electron density around water molecules, contoured at 1 sigma.

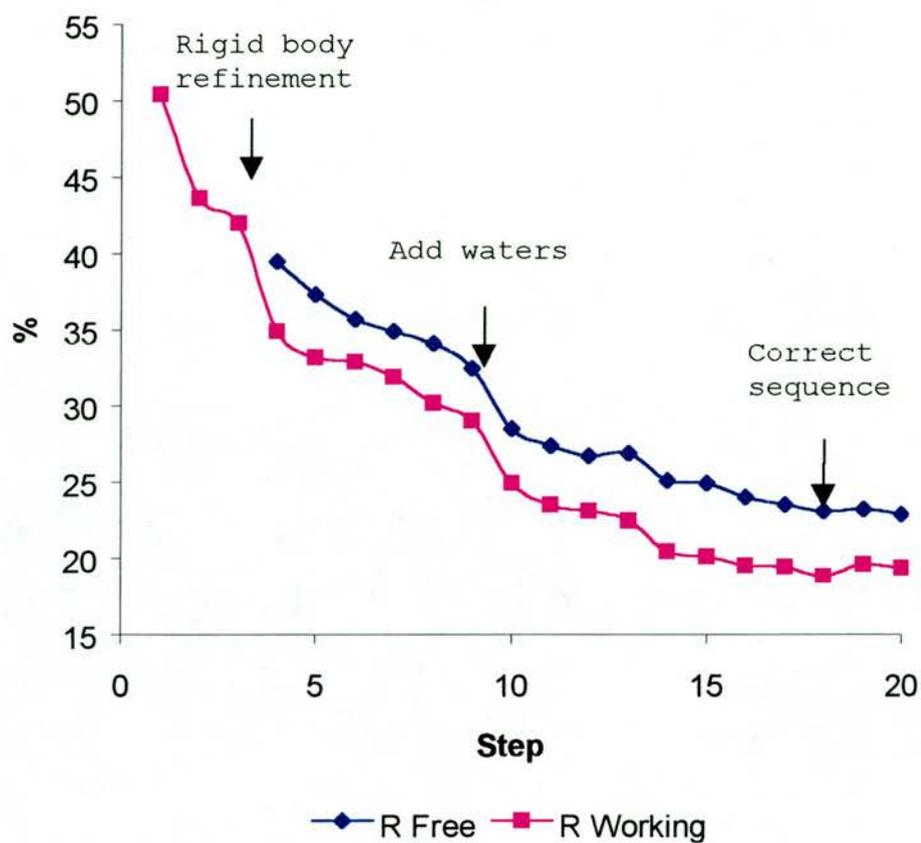


Figure 6.11

A graph showing the value of R_{work} and R_{free} during refinement.

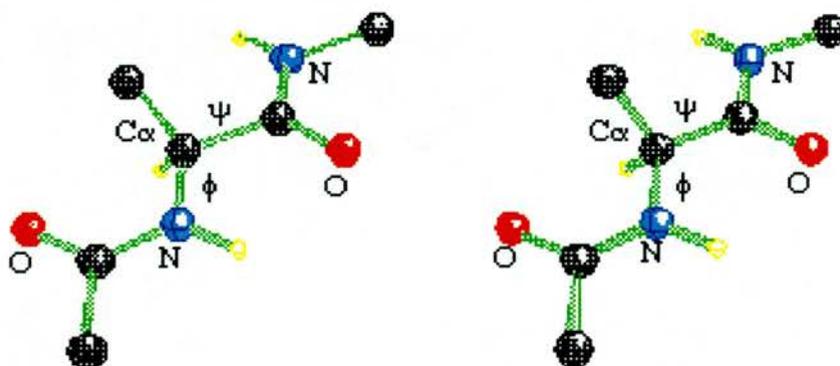


Figure 6.12

A stereo diagram of the main chain conformational angles in a polypeptide, phi (ϕ) and psi (ψ). In the diagram both angles are 180° .

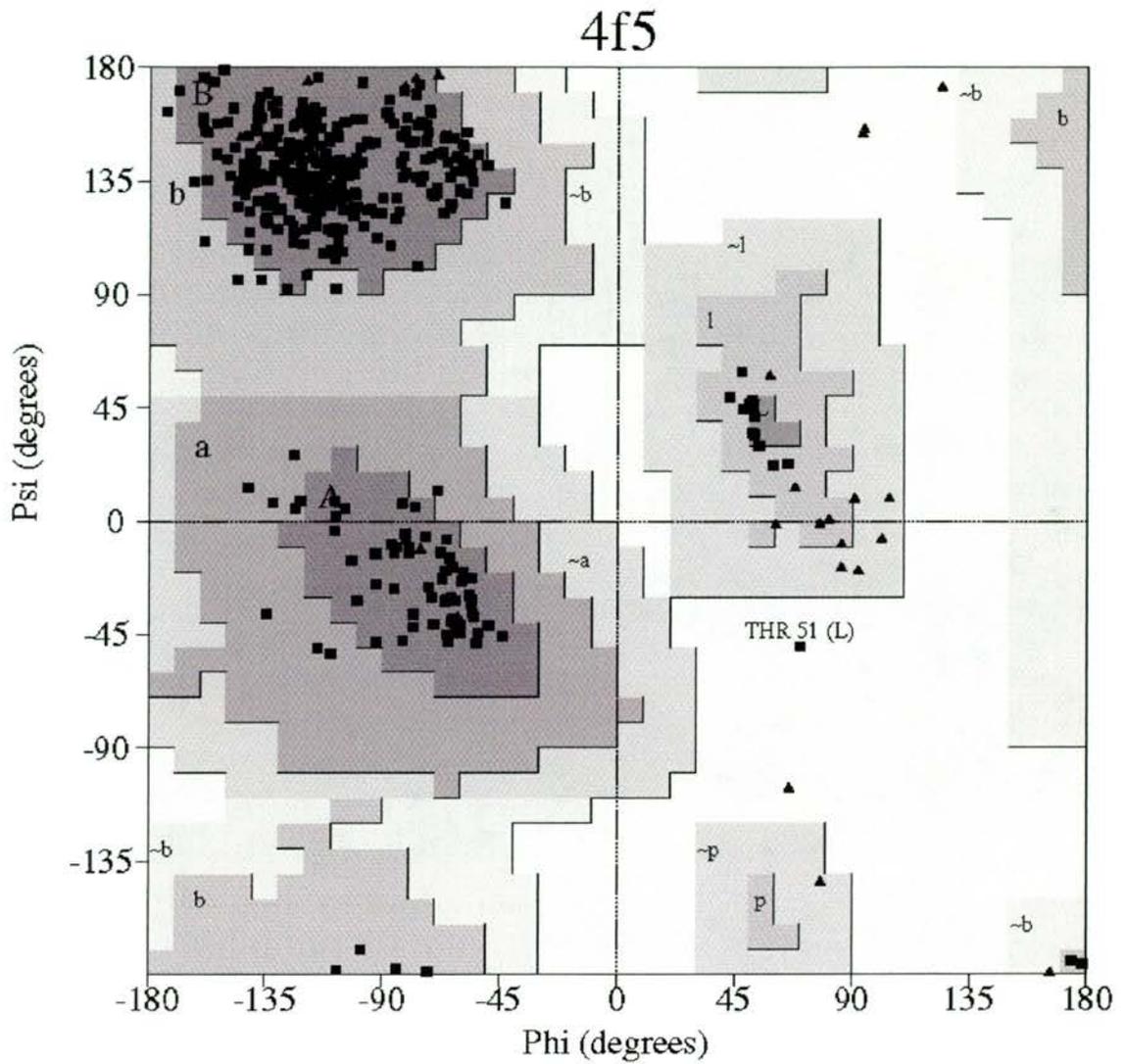


Figure 6.13

A Ramachandran plot of the 4F5 structure.

The dihedral angles ϕ and ψ are plotted. The most favourable regions of conformational space are indicated by the dark polygons, the lighter grey indicates less favourable regions and disallowed angles are found in white regions. Glycines are indicated by triangles, squares represent all other residues. The region labelled a corresponds to α -helices, b to β -sheets, l to left handed helices, and p to polyproline conformations.

LC	No density from atom	HC	No density from atom
K 77	C γ	K 3	N ζ
K 103	N ζ	K 13	N ζ
E 105	C γ	K 23	C γ
K 147	C γ	*E 61	C δ
K 149	C ϵ	*K 64	C δ
R 155	C δ	K 66	C ϵ
K 169	C β	E 85	C γ
E 195	C γ	K 115	C δ
		K 203	C δ
		K 205	C ϵ

Table 6.2

Residues that had side chain density missing, starred residues are in the CDRs.

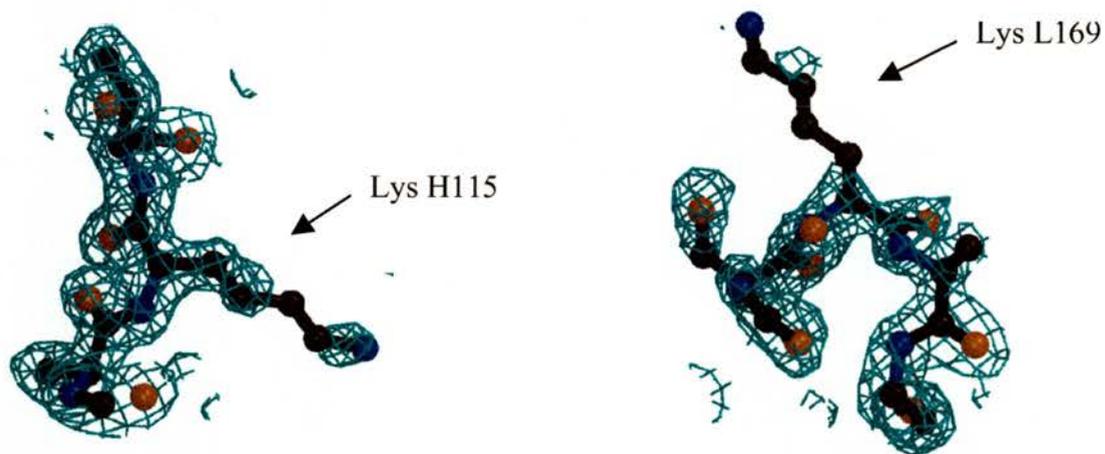


Figure 6.14

Lysine residues with side chain density missing, a $2F_{obs} - F_{calc}$ map contoured at 1 sigma.

6.7 Summary

Crystals of the Fab and Fc fragments of 2G4 were grown, but no diffraction data could be collected. It is possible that further purification and optimisation of the crystallisation conditions may yield diffracting crystals. This was attempted for 2G4 Fab but the crystals could not be reproduced using the original conditions.

The crystals of 4F5 Fab diffracted to 1.9 Å resolution and were of space group P2₁. A data set of 96% completeness was collected using synchrotron radiation. The structure was solved using molecular replacement in X-PLOR, with HyHEL-5 lysozyme as a phasing model. Refinement was carried out in CNS, and gave a final R-factor of 19.3% and an R-free of 22.9%. The stereochemistry and geometry of the final model is good. Residue L51 has a disallowed dihedral angle, a common occurrence in Fab structures. The map extends over the whole LC and HC with the exception of the loop L156-L157 and helix H128-H133.

Chapter 7

4F5 Structure Analysis

7.1 Introduction

It has been shown that a panel of murine anti-TPO antibodies recognize four major domains on TPO, two of which (A and B) are also recognised by human anti-TPO autoantibodies [Ruf 1989]. 4F5 is a mouse monoclonal antibody that competes with AITD patient sera for TPO binding, it is therefore likely to bind to the A or B domain of TPO.

The structures of antibody-antigen complexes show that the interactions are based on topological and charge complementarity [Braden 1998]. An uncomplexed antibody structure can be used to make predictions about the residues likely to be involved in binding to the antigen. The size and charge of surface accessible CDR residues determine the properties of the potential antigen combining site, assuming there is no induced fit. The contact definitions of CDRs give the residues most likely to contact antigen based on contacts observed in antibody-antigen structures [MacCallum 1996]. The specificity determining residues (SDRs) [Padlan 1995] indicate residues likely to contact antigen based on sequence hypervariability.

The structure of TR1.9, a human autoantibody to TPO has also been solved by crystallography [Chacko 1996]. This autoantibody binds to the A domain. There is no experimental information on how the TR1.9 epitope relates to the 4F5 epitope. The antibody combining site of 4F5 was studied and compared to that of TR1.9. The colour scheme used in analysis given in Table 7.1.

Colour	Amino acid	Property
Dark grey	Alanine, glycine, isoleucine, leucine, proline, valine	Aliphatic
Light grey	Asparagine, glutamine, histidine, serine threonine	Polar
Green	Phenylalanine, tyrosine, tryptophan	Aromatic
Yellow	Cysteine, methionine	Sulphur containing
Red	Glutamate, aspartate	Negative
Blue	Arginine, lysine	Positive

Table 7.1

The colour code used in analysis.

7.2 Framework Regions

B-factor distribution (all atoms) and buried and exposed surface area were calculated in CNS [Brunger 1998]. Surface accessibility was calculated using the Lee and Richards method [Lee & Richards 1971] with a probe radius of 1.4 Å. Residues were assigned as accessible if the side chains had a total accessible area of greater than 5 Å². Secondary structure and hydrogen bonding were identified using DSSP [Kabsch & Sander 1983]. Interactions were analysed using the Protein-Protein Interaction Server [Jones & Thornton 1996]. Superimpositions were carried out in O [Jones 1991] using LSQ. The elbow angle, defined as the angle between the V_L-V_H and C_L-C_{H1} pseudodiads, was calculated by superimposing the V_L domain onto V_H, and C_L onto C_{H1} in O. The resulting matrices were deconvoluted into vectors using NCSSTRICT in X-PLOR [Brunger 1992]. The dot product of the vectors was calculated, the inverse cosine of this number gives the elbow angle. Diagrams were created with MOLSCRIPT [Kraulis 1991] and GRASP [Nicholls 1991].

The structure of 4F5 is that of a standard Fab fragment and is shown in Figure 7.1. There are four Ig domains, each made up of two antiparallel β -sheets of four and five strands (variable domains) and three and four strands (constant domains). The secondary structure is listed in Table 7.2. The main chain atoms of the two β -sheets are about 10 Å apart and the sheets are inclined at an angle of -30° . Disulphide bonds are present between residues L23-L88, L134-L194, H22-H92 and H140-H195. The constant domains have a short stretch of α -helix linking strands a and b, this linkage could not be built into the HC domain model due to insufficient electron density. Surface representations of the LC and HC with their interaction surfaces exposed, and of the whole Fab are shown in Figure 7.2. The Kabat numbering scheme [Wu & Kabat 1970] has been used throughout.

The surface area of the LC is 11 140 Å² that of the HC is 11 160 Å². The surface area buried in the interface between the two chains is 1 735 Å² for the LC and 1 641 Å² for the HC. 34% of the atoms in the interface are polar and there are 10 hydrogen bonds, two salt bridges and 273 hydrogen bonds. The interface between the HC and LC is similar to that observed in other Fab structures [Padlan 1994]. The elbow angle was calculated to be 141.2° , which falls within the range of those previously observed [Padlan 1994]. The distribution of B-factors over the whole molecule is shown in Figure 7.3. It can be seen that the B-factors are high at the termini of the molecules due to increased flexibility in these regions. Areas of high B-factor also correlate with the position of the CDR loops L1, L3 and H2. The amino acids that are missing from the final model are surrounded by regions of high B-factor.

A superimposition of the variable domain of the phasing model, 2hfl, with 4F5 shows a root mean square deviation (rmsd) over the variable domain of 0.77 Å, over the constant domain of 0.75 Å, and over the whole Fab of 1.6 Å. The superimposition is shown in Figure 7.4. The difference in elbow angle between the structures is 18° .

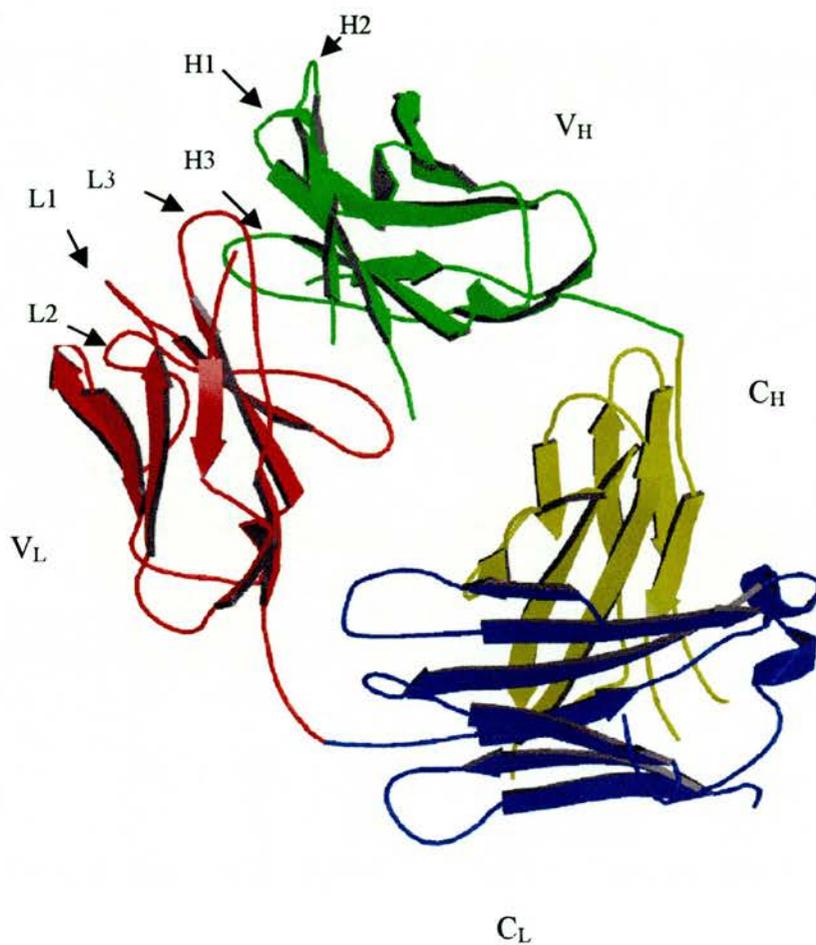


Figure 7.1

The structure of 4F5.

CDRs are labelled. Red - V_L, blue - C_L, green - V_H, yellow - C_H.

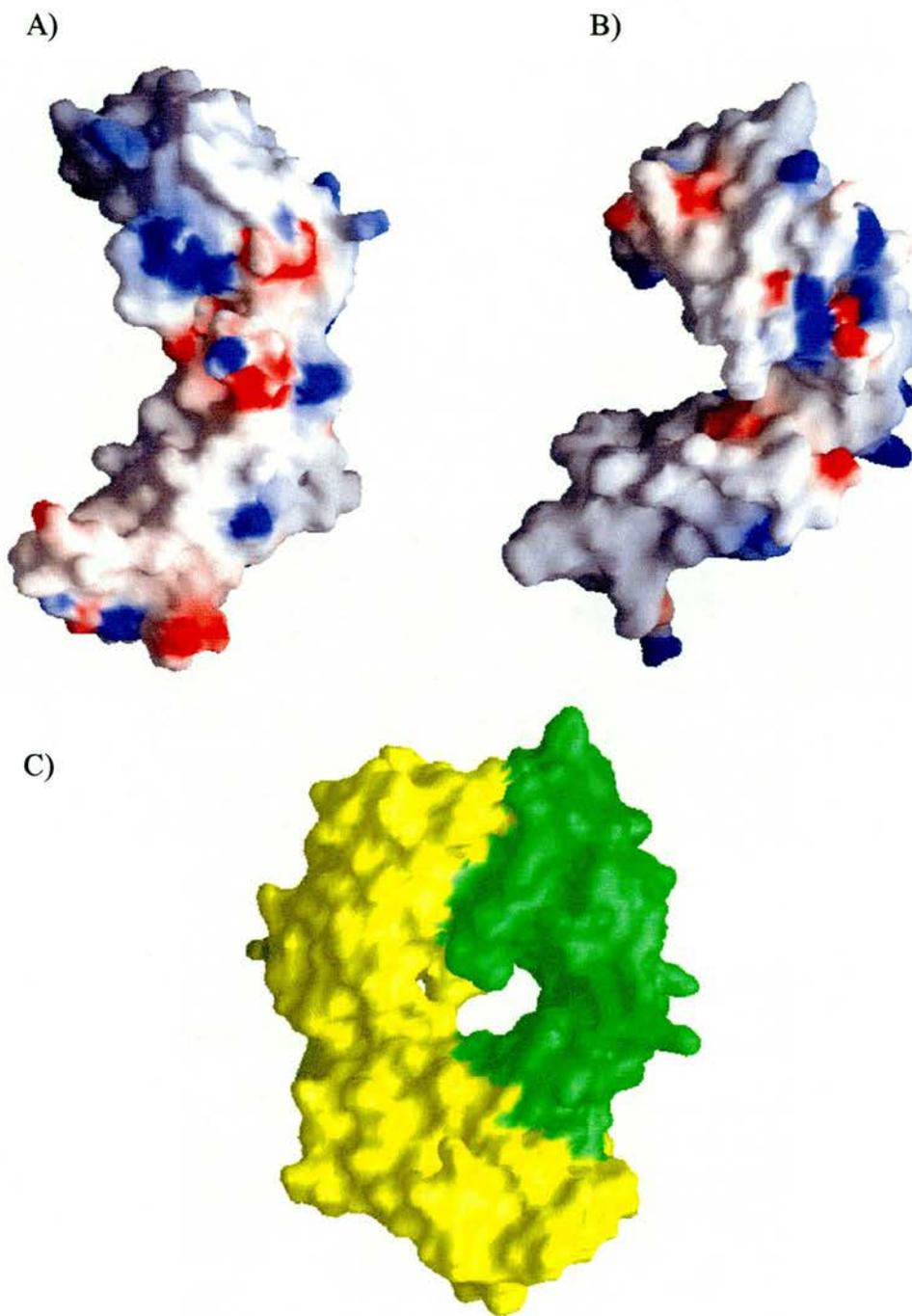


Figure 7.2

A grasp representation of the interface region of the light (A) and heavy (B) chains, red indicates negative charge and blue positive charge. The Fab dimer is shown in (C), HC in green and LC in yellow. In all pictures the variable domain is at the top.

Residues	2° Structure	Part of sheet	Residue	2° Structure	Part of sheet
L4-7, 10-13	Strand a	A	H3-6, 10-12	Strand a	E
L19-25	Strand b	A	H18-27	Strand b	E
L24-34	L1	-	H31-35	L1	-
L34-40	Strand c	B	H34-40	Strand c	F
L43-49	Strand c'	B	L44-52	Strand c'	B
L50-56	L2	-	H50-65	L2	-
L62-67	Strand d	A	H56-59	Strand c''	B
L69-72	Strand e	A	H67-72	Strand d	E
L84-90	Strand f	B	H77-82	Strand e	E
L89-97	L3	-	H88-94	Strand f	F
L101-105	Strand g	B	H95-102	L3	-
L114-118	Strand a	C	H102-103, 107-111	Strand g	F
L122-127	Helix a	-	H120-124	Strand a	G
L129-139	Strand b	C	Missing	Helix a	-
L145-150, 153-4	Strand c	D	H134-H145	Strand b	G
L159-163	Strand d	C	H151-154	Strand c	H
L171-182	Strand e	C	H163-5, 169-71	Strand d	G
L183-186	Helix b	-	H 174-184	Strand e	G
L191-197	Strand f	D	H194-199	Strand f	H
L205-210	Strand g	D	H204-209	Strand g	H

Table 7.2

The secondary structure of 4F5, calculated in DSSP.

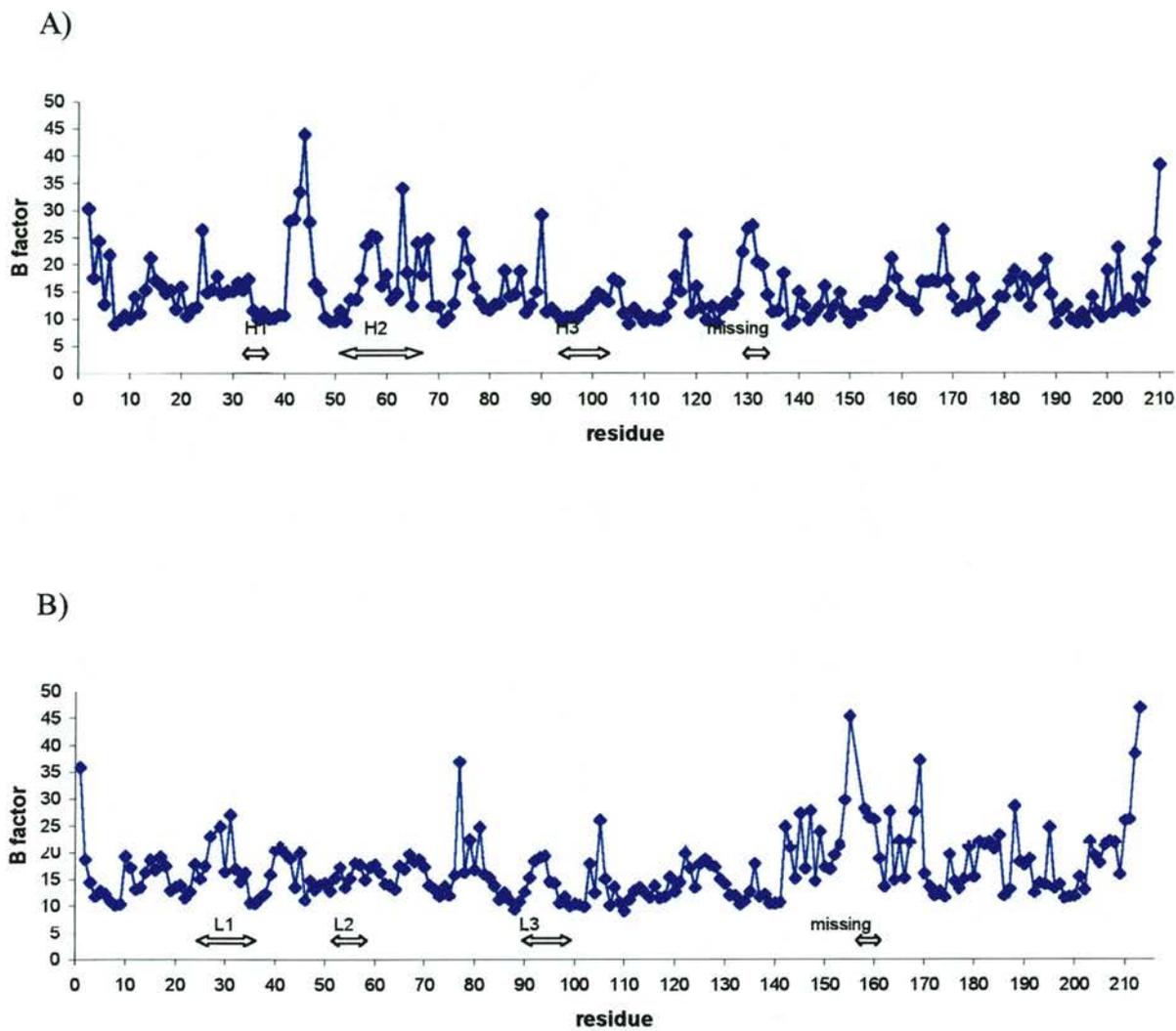


Figure 7.3

A plot of B factor against residue number for the HC (A) and LC (B).



Figure 7.4

Superimposition of the phasing model 2hfl (green) and 4F5 (red), residues from the variable domains were used to position the alignment.

7.3 Complementarity Determining Regions

The CDRs of 4F5 are listed in Table 7.3. The Kabat definition is based on sequence variability [Wu & Kabat 1970] and the contact definitions is based on the analysis of structures of antibody-antigen complexes [MacCallum 1996]. The SDRs [Padlan 1995] indicate residues likely to contact antigen based on sequence hypervariability.

CDR	Sequence
L1	Kabat A V FI
	Contact V FI WF
	SDR
L2	Kabat LA
	Contact LWIY LA
	SDR A
L3	Kabat L R YPY
	Contact L R YPY
	SDR L R Y
H1	Kabat YYI
	Contact YYI
	SDR Y
H2	Kabat WIYPG VY KY EKFKD
	Contact WIGWIYPG VY K
	SDR W Y G Y K
H3	Kabat DA LEY
	Contact GRDA LEY
	SDR DA

Table 7.3

The CDRs of 4F5 as using Kabat and contact definitions, with SDRs marked.

L1 is 10 residues long and has a contact range of seven residues, L2 is seven residues long and has a contact range of 10 residues, none of the residues are charged. L3 contains nine residues (contact range eight residues) including arginine L91. H1 is five residues long (contact range six residues) and is uncharged. H2 is the longest and most highly charged CDR, containing 17 residues (contact range 13) including lysine H58, glutamate H61, lysine H62, lysine H64 and aspartate H65. H3 is six residues long and contains aspartate H95 and glutamate H101, the contact range extends two residues to the N-terminal and includes arginine H94.

The structures of CDR loops have been used to generate a set of canonical forms for all loops apart from H3. Each canonical form is defined by a set of key residues in the CDRs and the framework region. Those so far identified are listed in Tables 7.4 and 7.5 [Al-Lazikani 1997; Martin 1998]. There are eight canonical forms for L1, one for L2 and five for L3. Four canonical forms have been identified for H1 and four for H2.

Stereo diagrams of the CDR loops of 4F5 are shown in Figures 7.5 and 7.6. L1 of 4F5 is likely to be class 1, with the exception that L33 is isoleucine and only leucine or methionine are allowed at this position. However, superimposition of 4F5 with the template loop (2fbj) gives a rmsd of 0.56 Å, indicating that L1 is class 1. There is only one canonical class for L2 identified, class 1, to which the structure of 4F5 conforms. L3 and H1 are both class 1, and have no mismatches with the key residues. H2 conforms to class 2 with the exception a mismatch of valine H55 in H2 with the allowed residues of glycine or serine. Again, superimposition with the template (1bbd) gives a rmsd of 0.52 Å, showing H2 is class 2. With only six residues, the H3 loop of 4F5 is on the short end of the range of loops lengths identified [Wu 1993]. The structure is classified as kinked, due to a hydrogen bond between the carbonyl of residue H98 and the ring N of tryptophan H103 (Figure 7.6C).

Class	LC	Length	Template	Key residues
L1-1	κ	10	2fbj	L2 I, L25 A, L29 VIL, L33 LM, L71 YF
L1-2	κ	11	1ikf	L2 I, L25 A, L29 VIL, L33 LV, L71 YF
L1-3	κ	17	1hil	L2 I, L25 S, L29 VIL, L33 L, L71 YF
L1-4	κ	16	1rmf	L2 V, L25 S, L29 L, L33 L, L71 F
L1-5	λ	13	2fb4	L25 G, L30 I, L33 V, L71 A
L1-6	λ	14	7fab	L25 G, L30 I, L33 V, L71 A
L1-7	κ	14	1gig	L25 S, L30 V, L33 A, L71 A
L1-8	κ	12	1fig	L2 N, L25 A, L29 V, L33 L, L71 Y
L2-1	κ	7	1lmk	L48 IV, L64 G
L3-1	κ	9	1tet	L90 QNH, L95 P
L3-2	κ	9	2fbj	L94 P, L95 L
L3-3	κ	8	2hfl	L90 Q, L96 P
L3-4	κ	9	7fab	L90 LS, L94 NS, L95 HL
L3-5	κ	11	2fb4	L94 DNG

Table 7.4.

The residues used to define the canonical forms of the LC CDRs.

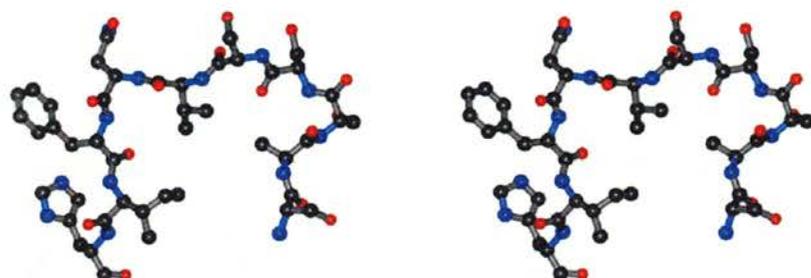
Class	Length	Template	Key Residues
H1-1	10	2fbj	H24 TAVGS, H26 G, H27 FYTG, H29 FLIV, H34 MIVLT, H94 RKTA
H1-1'	10	7fab	H26 G, H27 SD, H29 FI, H34 YW, H94 RN
H1-2	11	1baf	H24 VF, H26 G, H27 GFY, H29 IL, H34 CW, H96 HR
H1-3	12	1ggi	H24 VFG, H26 G, H27 FGD, H29 ILV, H34 WV, H94 HR
H2-1	9	1gig	H55 GD, H71 RKVI
H2-2	10	1bbd	H52A PTA, H55 GS, H71 ALT
H2-3	10	1igc	H52A DP, H54 GNDS, H55 GS, H71 R
H2-4	12	1mcp	H54 KS, H55 Y, H71 R

Table 7.5

The residues used to define the canonical forms of the HC CDRs.

The C α trace of the combining site of 4F5 is shown in Figure 7.7, with the CDR residues highlighted. The contributions of the CDRs (Kabat and contact definition) and the SDRs to the surface of the combining site are shown in Figure 7.8. The distribution of CDRs making up the combining site is standard, with H3 and L3 having a central position, H1 and L1 further out, and H3 and L3 on the periphery. The residues of the Kabat CDRs that have an accessible surface area (ASA) of more than 5 Å² are listed in Table 7.6. Also given is the total ASA of each CDR. H2 contributes the largest surface area of 884 Å² followed by L1 with an ASA of 401 Å².

L1



L2



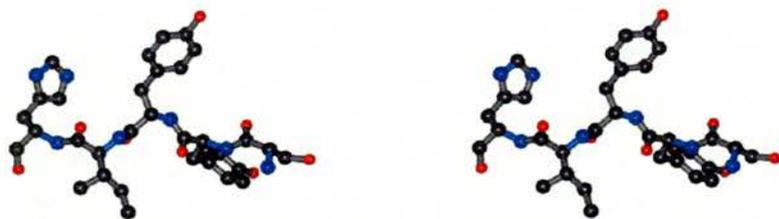
L3



Figure 7.5

Stereo diagrams of the CDR loops of the LC of 4F5.

H1



H2



H3

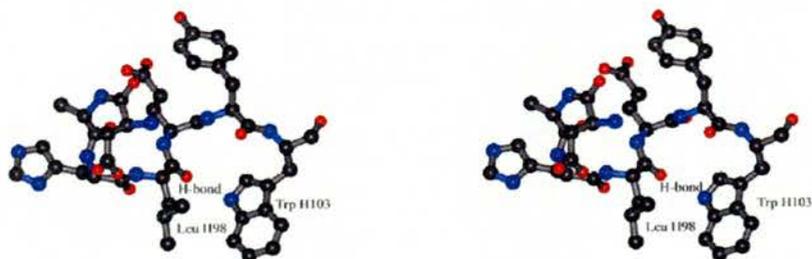


Figure 7.6

Stereo diagrams of the CDR loops of the HC of 4F5.

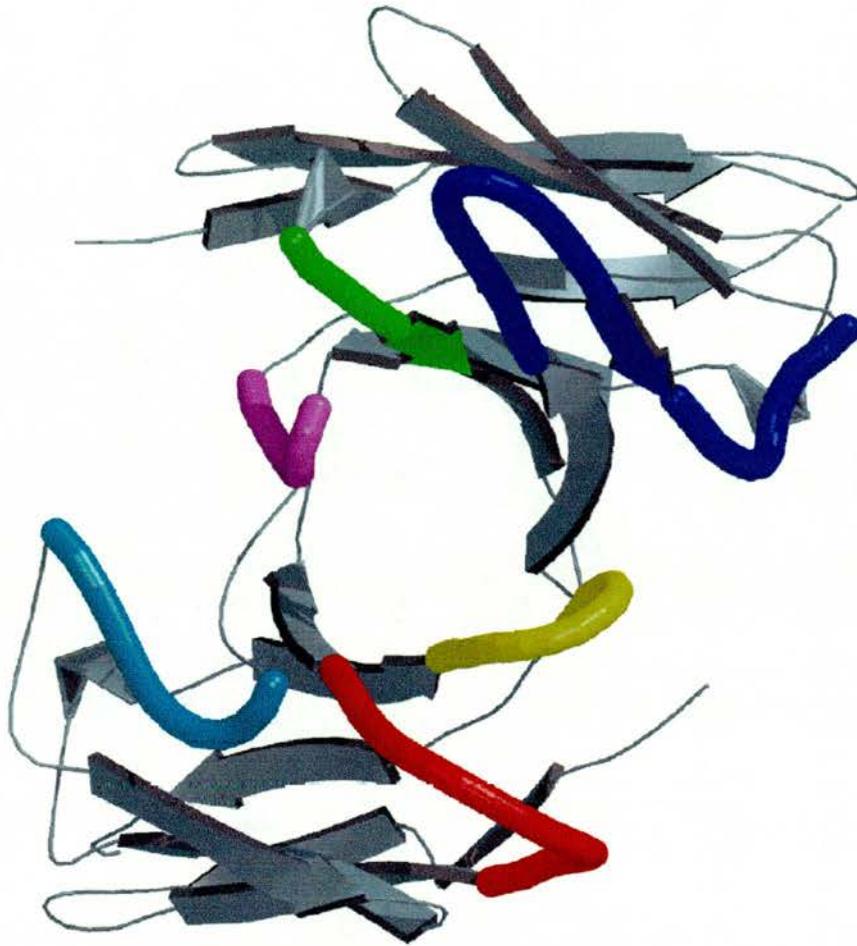


Figure 7.7

The position of the CDR loops using the Kabat definition. CDRs are coloured L1 (red), L2 (light blue), L3 (yellow), H1 (green), H2 (dark blue), H3 (pink).

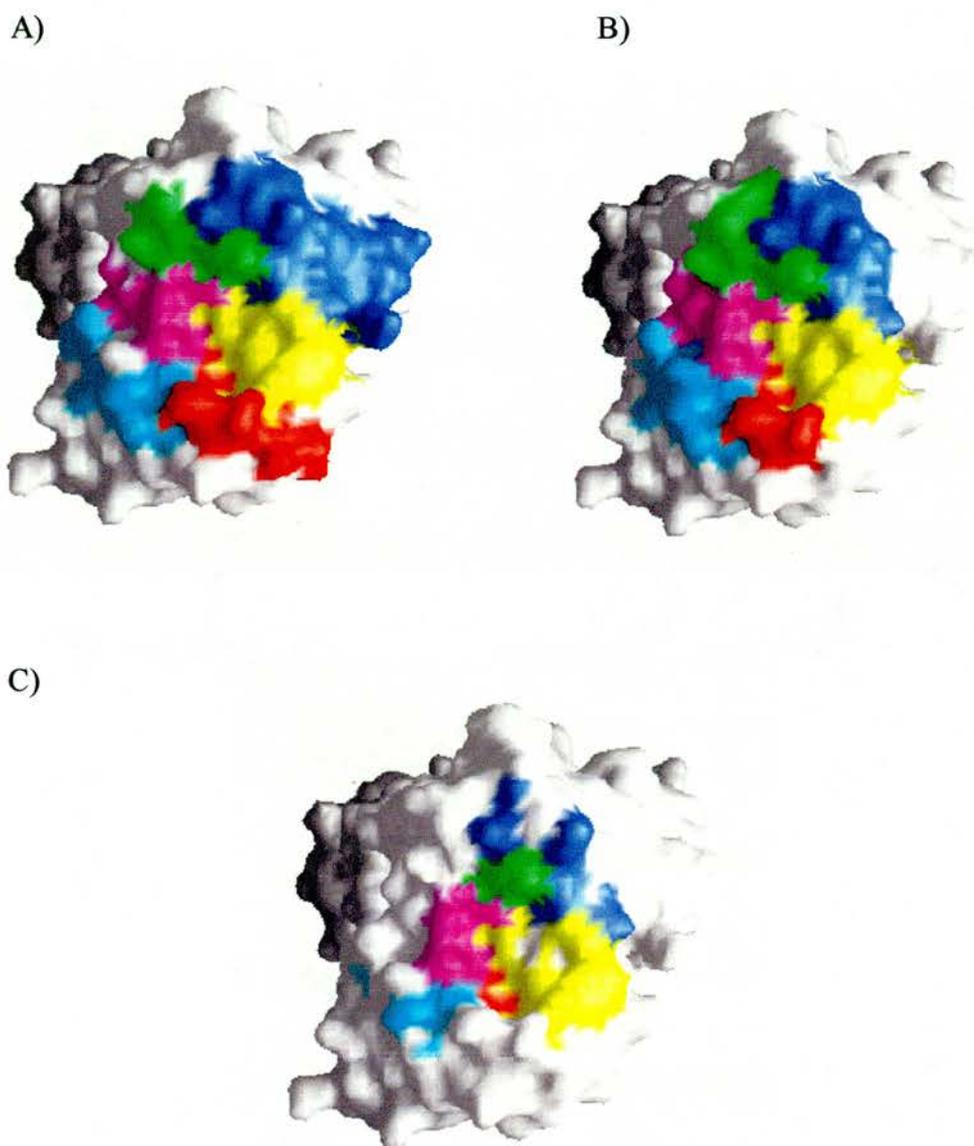


Figure 7.8

Grasp profiles of the combining site of 4F5, coloured by CDR contribution.

A) Kabat definition, B) Contact definition, C) SDRs.

CDRs are coloured L1 (red), L2 (light blue), L3 (yellow), H1 (green), H2 (dark blue), H3 (pink), the pictures are in the same orientation as Figure 7.7.

CDR	Sequence	Total ASA (Å ²)
L1	F	401
L2	LW Y LA	243
L3	R YPY	235
H1	YY	202
H2	W W Y VY KY EKFKD	884
H3	RDA LEY	243

Table 7.6

The accessible surface area of the CDRs of 4F5.

The characteristics of the combining site are shown in Figure 7.9. The surface is made up of a patchwork of acidic and basic regions and is relatively flat. The charged residues that make up the combining site are predominantly from the H2 loop (lysines 58, 62 and 64 and glutamine 61). H3 (aspartate 95) and L3 (arginine 91) each contribute a charged residue, both of which are located in a central hollow. Figure 7.9C highlights the surface curvature, a representation in which the hollow is accentuated. The most striking feature of the combining site is the preponderance of aromatic residues, shown in green in Figure 7.9 B. Aromatic residues, particularly tyrosines have been shown to be common in antibody binding sites [Padlan 1990]. Figure 7.10 shows the tyrosine and phenylalanine residues clustering at the combining site interface, predominantly on the HC.

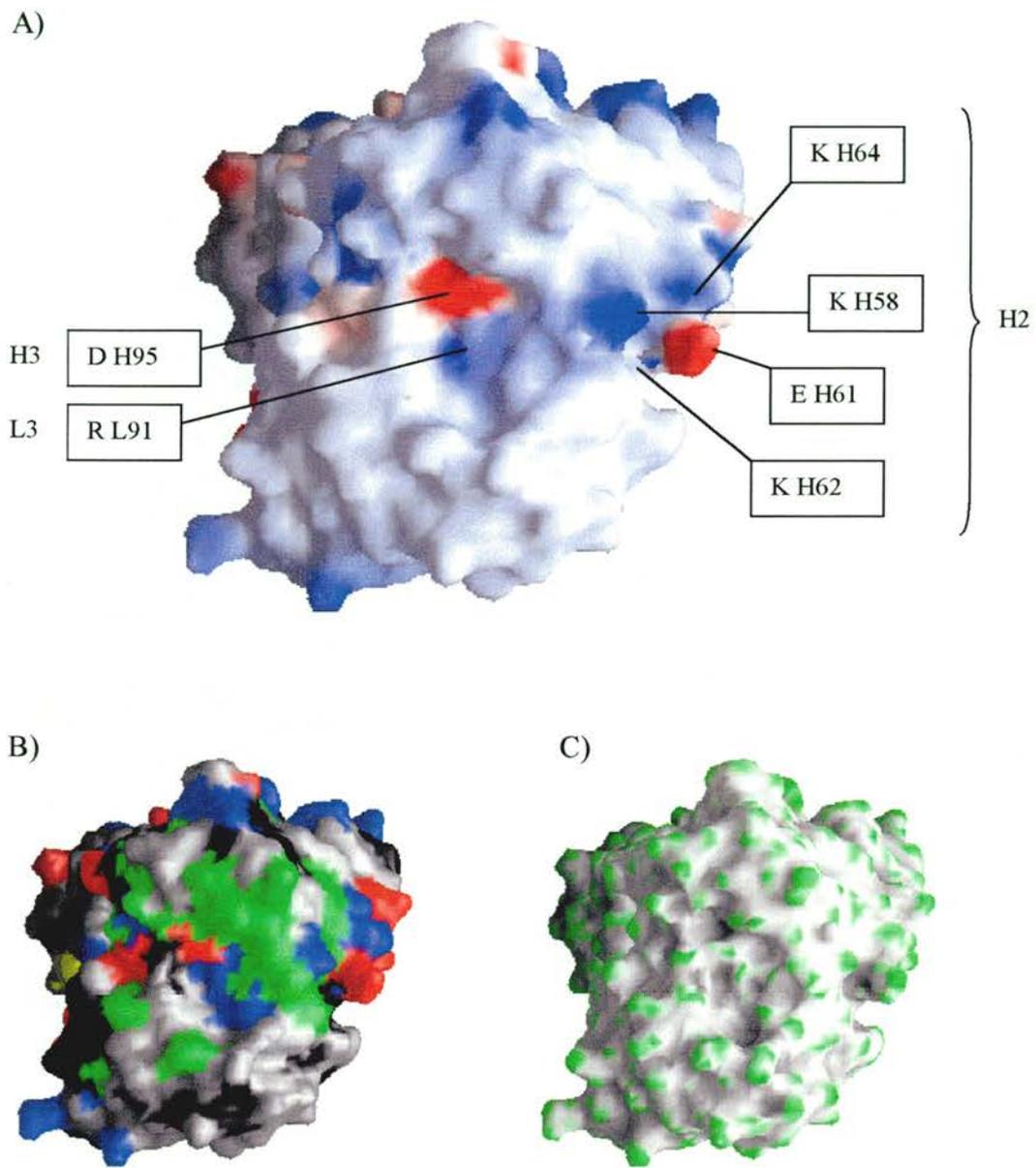


Figure 7.9

The combining site of 4F5.

Grasp profiles coloured by charge (A) and amino acid contribution as defined in Table 7.1, (B) surface curvature (black concave, green convex).



Figure 7.10

The combining site of 4F5 with surface accessible tyrosine and tryptophan residues in the CDRs shown in ball and stick. HC in red, LC in blue.

The CDR residues of one 4F5 molecule in the crystal interact with residues in the C_L domain of a neighbouring molecule (see Figure 6.7). The position of the two molecules in the crystal is shown in Figure 7.11. The interaction involves residues in the L1, H1, H2 and H3 CDRs with helix 2 (L184-190), the N-terminus (L112) and residues L151 and L153 of the C_L domain. The CDR residues are mapped onto the combining site in Figure 7.12A. The interface accessible surface area is 471 Å² for V_H and 457 Å² for C_L.

46% of the V_H atoms and 58% of the C_L atoms in the interface are polar, the remainder are nonpolar. The details of the contacts between the molecules are shown in Table 7.7. There are nine hydrogen bonds, three van der Waals interactions and one hydrophobic interaction. A close up of the interaction of the proteins, showing residues that form hydrogen bonds, is shown in Figure 7.13. There are two bridging water molecules forming hydrogen bonds with both 4F5 molecules, one of which is shown in Figure 7.14.

Sequencing showed that 12 of the amino acids in the 4F5 CDRs are a result of somatic mutation (see Figure 5.4). The position of these residues on the surface of 4F5 is shown in Figure 7.12B. It can be seen that there is a striking similarity between the somatically mutated residues and those involved in crystal contacts. The exception is the pink patch corresponding to residues in H3. However two out of three of the contacting residue in H3 interact via water molecules. It is likely that these residues normally bind to a feature of TPO that is missing in the crystal contact and instead compensate by using bridging waters. Of the seven residues that are within 4 Å of the neighbouring 4F5 molecule in the crystals, four have been somatically mutated (Table 7.7). Somatic mutation is evidence for antigen driven selection. It is therefore likely that the CDRs involved in the crystal contact interaction are involved in contacting TPO.

However, it is unlikely that this interaction is fully representative of that with TPO as the interaction surface area on 4F5 is less than 500 Å², whereas that in antibody-

antigen complexes is 700-900 Å² [Padlan 1994]. In addition, only CDR residues from L1, H1, H2 and H3 are involved, whereas L3 has been shown to play a significant part in interactions with antigen [Wilson & Stanfield 1994].

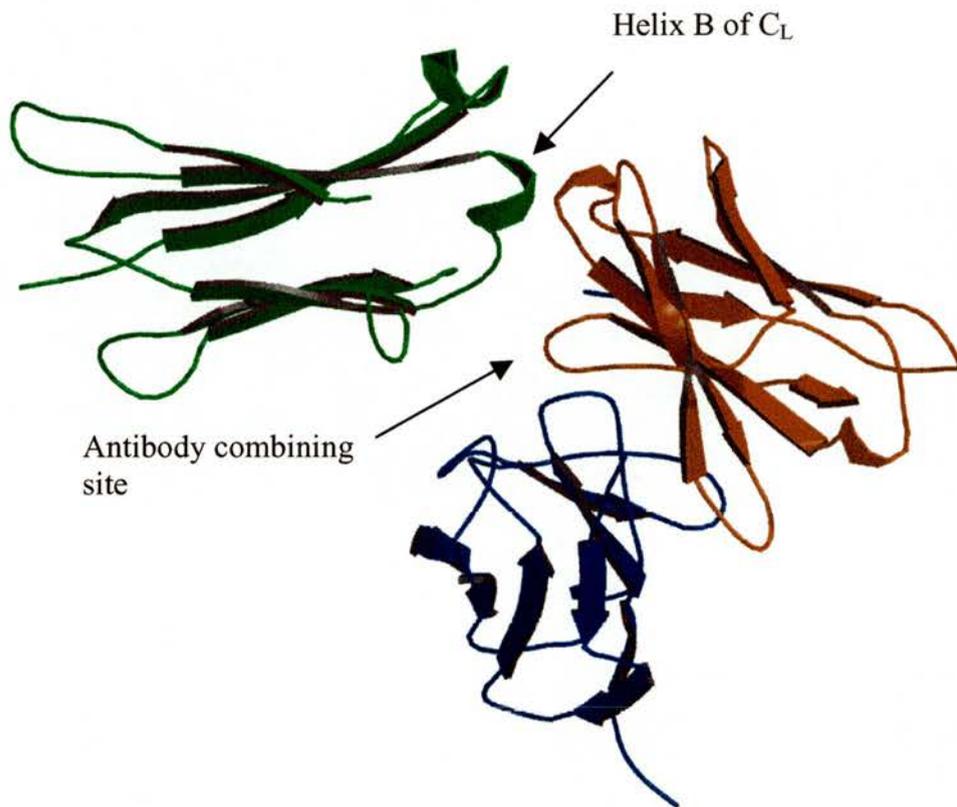
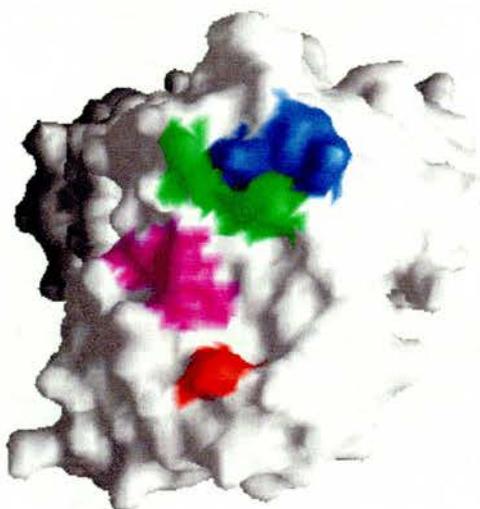


Figure 7.11

The interaction of the CDRs of one molecule of 4F5 (HC red, LC blue) with the C_L domain of a neighbouring molecule in the crystal (green).

A)



B)

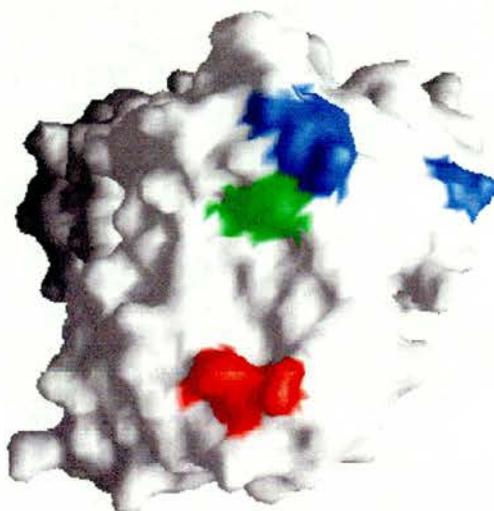


Figure 7.12

The combining site of 4F5, L1 (red), H1 (green), H2 (dark blue), H3 (pink).

A) Residues involved in the crystal contacts.

B) Residues that have undergone somatic mutation.

Loop	CDR residues	C _L residues	Contact	Distance (Å)
L1	* Phe L32 CE1	Ser L153 OG	Van der Waals	3.3
H1	Ser H31 OG, O	Glu L187 OE2, Arg L211 NH2	H bonds	3.2, 2.8
H1	* Tyr H33 OH	Asp L184 O, Arg L188 NE	H bonds	2.6, 3.4
H2	Tyr H52 OH	Glu L187 OE1	H bond	2.5
H2	* Asn H54 ND2	Asp L184 CB	Van der Waals	3.2
II2	* Tyr II56 CE1	Asp L184 OD2	Van der Waals	3.6
H3	Ala H96 CB	Asn L190 CB	Hydrophobic	3.4
H3	His H97 NE2	Asp L151 OD1	H bond via water	2.8, 2.8
H3	Glu 101 OE2	Asn L212 OD1	H bond via water	2.5, 3.6

Table 7.7

Contact residues between the CDRs of 4F5 and the adjacent molecule in the crystal lattice. Starred residues are those that have undergone somatic mutation.

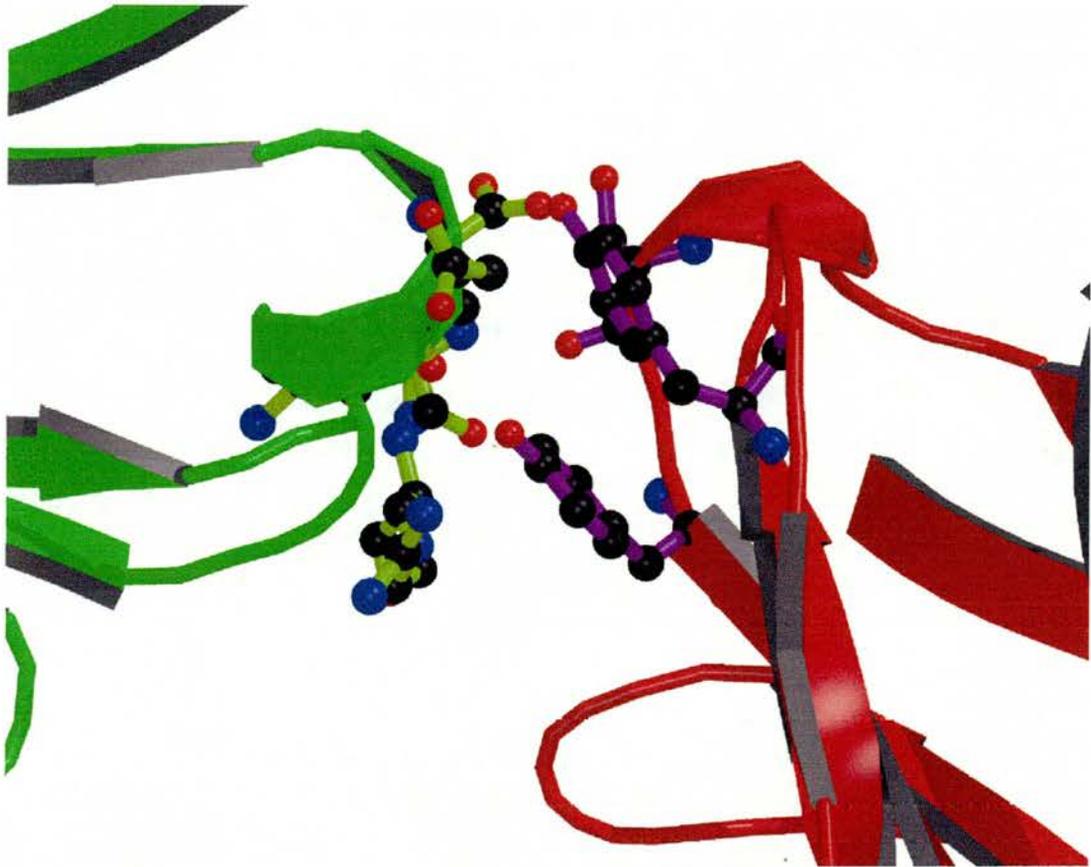


Figure 7.13

The residues involved in the crystal contacts between the 4F5 V_H domain (red) and helix A of the C_L domain (green). Residues involved in hydrogen bonds are shown as ball-and-stick (V_H purple, C_L light green).

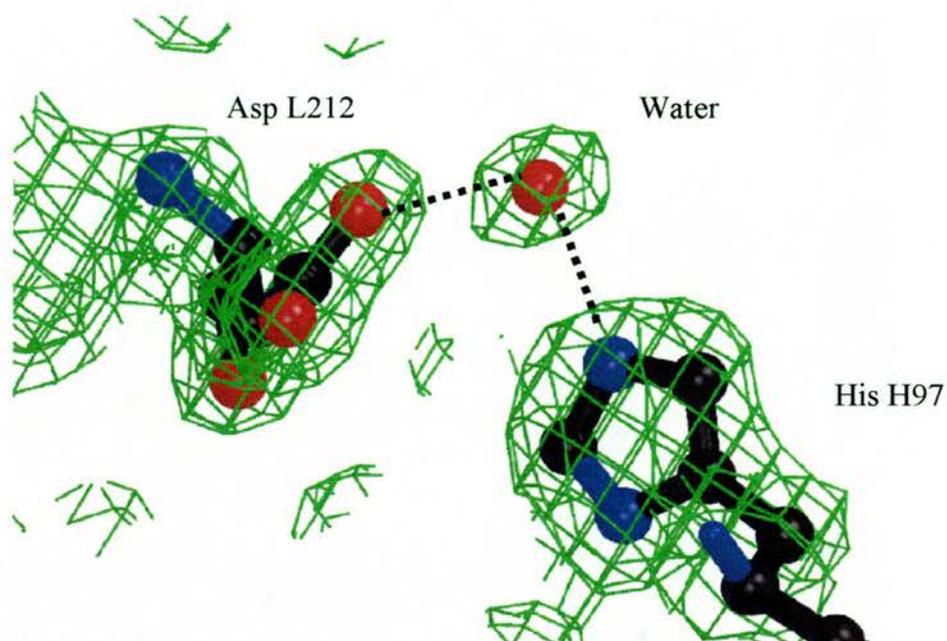


Figure 7.14

An electron density map ($2F_{\text{obs}} - F_{\text{calc}}$) of a bridging water molecule in the interface between adjacent molecules in the 4F5 crystal, contoured to 1 sigma. Hydrogen bonds are shown by broken lines.

7.4 Comparison with TR1.9

TR1.9 is a high-affinity IgG1 κ human autoantibody to TPO. The crystal structure of the Fab fragment has been solved by crystallography to 2.0 Å resolution [Chacko 1996]. TR1.9 has a flat combining site and flexible H2 and H3 loops that may result from the mutation of germline asparagines and the presence of several glycines. A superimposition of 4F5 and TR1.9 is shown in Figure 7.15. The rmsd between the structures is 1.0 Å. The elbow angle of TR1.8 is 138.7°, 2.5° smaller than 4F5. For comparison, TR1.9 has been renumbered according to the Kabat system.

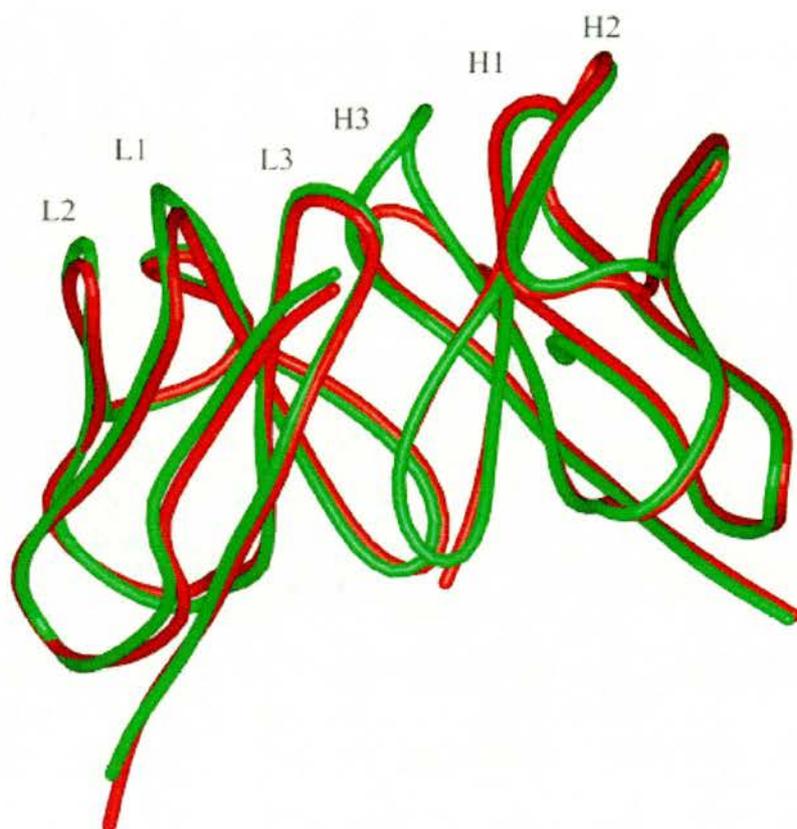


Figure 7.15

Superimposition of the variable domain of 4F5 (red) and TR1.9 (green).

The canonical class of the CDRs of TR1.9 are reported to be 2, 1, 1, 1, 3 compared with 1, 1, 1, 1, 2 for 4F5. It can be seen in Figure 7.15 that the L1 canonical form templates for class 1 and 2 are very similar (rmsd 0.62 Å), with the latter being a single residue longer. L2, L3 and H1 show very little deviation between 4F5 and TR1.9 (rmsd 0.20 Å, 0.27 Å, 0.17 Å respectively), as would be expected by the fact that they are of the same canonical class. H3 is six residues longer in TR1.9 than 4F5 and they both share the same stem conformation.

H2 of TR1.9 has been assigned to canonical class 3, however, the conformation of the loop is very similar to 4F5 which is class 2 (rmsd 0.72 Å). Analysis of the sequence of TR1.9 indicates that H2 of TR1.9 may in fact be class 2. Using the canonical templates, H2 has two outliers if class 3, compared with one outlier if class 2 (arginine H71). Superposition of the example templates for each canonical class showed a rmsd of 0.531 Å for class 2 (1bbd) and 1.125 Å of class 3 (1icg) using residues from H50 to H65. Therefore it appears that the canonical class of H2 in 4F5 and TR1.9 is the same. The canonical classes of 4F5 and TR1.9 are compared in Table 7.8.

CDR	Rmsd Å	Canonical class	
		4F5	TR1.9
L1	0.62	1	2
L2	0.20	1	1
L3	0.27	1	1
H1	0.17	1	1
H2	0.72	2	2 (3)

Table 7.8

A comparison of the canonical class of the CDRs of 4F5 and TR1.9.

The Kabat and contact definitions of the CDR residues, the SDRs and the ASA of the TR1.9 combining site are given in Table 7.9. The Kabat CDRs of TR1.9 are of similar length to 4F5, apart from H3 which is 12 residues long in TR1.9, compared to only six in 4F5. L1 and L2 possess negative charges in TR1.9 but not 4F5, whereas L3 has a positive charge in 4F5 alone. H2 of 4F5 has two negative charged residues in addition to the three positive charges in both 4F5 and TR1.9.

H3 of TR1.9 has three negative and one positive charge, compared to the two negative of 4F5. The contact definition shows similar differences in charge distribution, with 4F5 having a total of two negatively charged residues, compared with the five of TR1.9, and both sharing a total of three positively charged residues.

Grasp profiles of the combining site of TR1.9 are shown in Figure 7.16. The surface is more highly charged than 4F5 and slightly less flat, due to the increased length of H3. Like 4F5, TR1.9 has a high number of aromatic residues in the combining site. As would be expected from the different CDR sequences, the overall charge distribution and topography of the antibody combining sites are very different. It therefore seems unlikely that 4F5 and TR1.9 recognise the same structure on TPO.

7.5 Summary

The structure of 4F5 is that of a standard antibody Fab fragment, and has CDRs of canonical class 1, 1, 1, 1 and 2. The combining site is relatively flat and is characterised by a large number of tyrosine residues.

Crystal packing leads to the interaction of the CDRs of one molecule of 4F5 with the constant domain of a different molecule. Residues from CDRs L1, H1, H2 and H3 are involved in the interaction. Over half of the interacting residues have arisen due to somatic mutation, suggesting the contact may have physiological significance. However, the interaction surface is not extensive enough to be fully representative of the binding to TPO.

CDR		Sequence
L1	Kabat	RA...GIC...ALA
	Contact	...L...ALAWY
	SDR	...A A
	ASA	R...A
L2	Kabat	...DA...LE
	Contact	LLIYDA...LE
	SDR	D...E
	ASA	YD...LE
L3	Kabat	...F...YPL
	Contact	...F...YPL
	SDR	...Y L
	ASA	F...YP
H1	Kabat	...YGL
	Contact	...YGL
	SDR	...YG
	ASA	...Y
H2	Kabat	W...AGT...G...KY...KFRG
	Contact	W...AGT...G...K
	SDR	W...AGT...K...K
	ASA	W......NTKY...KFR
H3	Kabat	DPYGGGK...EFDY
	Contact	ARDPYGGGK...EFDY
	SDR	PYGGGK...E
	ASA	R...PY...K...DY

Table 7.9

An analysis of the residues in the CDRs of TR1.9.

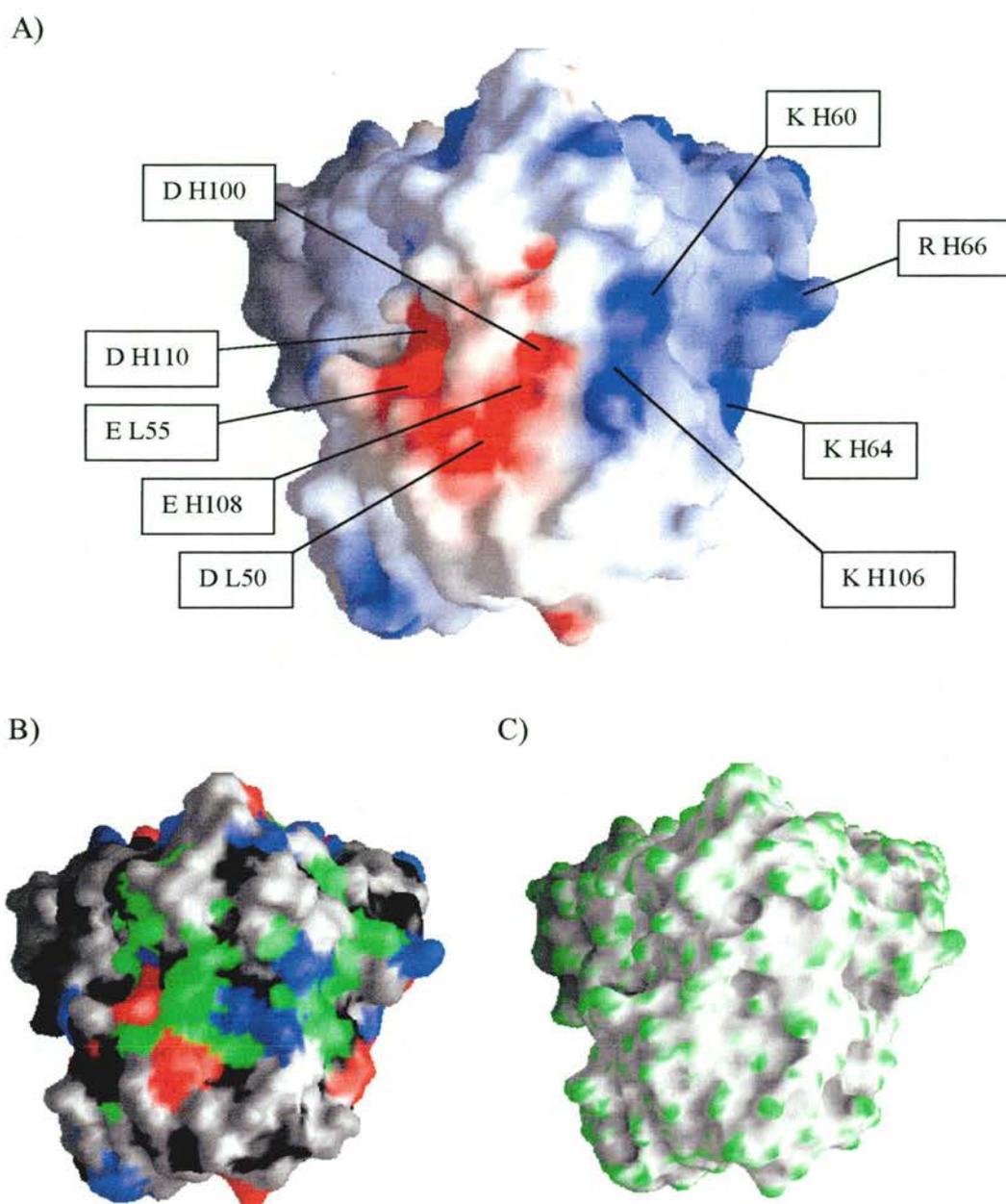


Figure 7.16

The combining site of TR1.9.

Grasp profiles coloured by charge (A) and amino acid contribution as defined in Table 7.1, (B) surface curvature (black concave, green convex).

Comparison of the structure of 4F5 with TR1.9 shows that the variable and constant domains have the same overall structure and a similar elbow angle. The CDR structures of both Fabs also have a strong resemblance, reflected in the similar canonical classifications. However, the 4F5 and TR1.9 CDR sequences are different, leading to distinct surface profiles. The differences indicate that 4F5 and TR1.9 are unlikely to recognise the same structure on TPO, although due to the static nature of crystal structures, the possibility of induced fit has not been taken into account.

Chapter 8

Discussion

8.1 Thyroid Peroxidase Heterogeneity

Heterogeneity has been shown to be present in the TPO preparation used for crystallisation. The heterogeneity is visible as a broad band or doublet on SDS-PAGE gels, and as a number of species of different pI on IEF gels. Heterogeneity in the protein sample is known to contribute to poorly diffracting crystals [Ducruix & Giege 1992]. It is therefore likely that the identified heterogeneity is responsible for the low diffraction limits of the TPO crystals. The determination of the reasons for the heterogeneity, and the elimination of its causes are necessary before crystals can be grown that diffract to high resolution.

Evidence has been provided, at least for this recombinant TPO preparation, that a degree of heterogeneity results from cleavage at N-terminus of the protein. N-terminal sequencing of the individual bands on a non-reduced SDS-PAGE indicates that there are at least two forms of TPO, which differ in the length of the N-terminal domain. This is confirmed by MS analysis which gives a molecular mass estimate that is consistent with degradation of expressed protein.

In addition, it has been shown that the heterogeneity is unlikely to be due to multiple glycoforms as previously thought. IEF gels show that a range of charged species is present, and that these remain following deglycosylation. MS analysis of released glycans confirmed that the carbohydrates on TPO are uncharged and so would not contribute to the multiple bands on IEF gels, even if deglycosylation were incomplete. The lack of charged sugars on TPO is in disagreement with a lectin binding study of the same recombinant preparation, which showed sialic acid

to be present [Grennan Jones 1996]. However, it is unlikely that sialic acid is present on recombinant TPO as insects are not thought to have the ability to synthesize sialylated proteins [Altman 1993].

Despite the addition of protease inhibitors to the cell culture, it appears that at some stage in the expression and purification of TPO there is more than one proteolytic cleavage event that results in a population of TPO molecules. N-terminal sequencing showed that cleavage at residues 76 and 109 had occurred. A trypsin recognition site is found at residue 109, and cleavage here has been experimentally shown by N-terminal sequencing of pig and human TPO [Yokoyama & Taurog 1988; Taurog 1990]. However, the sequence around residue 76 does not correspond to any commonly known protease recognition sites [Wilkins 1997].

N-terminal sequencing of other TPO preparations has rarely been reported. Therefore, it is not clear if similar proteolysis is responsible for the doublet seen on native and other recombinant TPO gels. N-terminal sequencing of recombinant TPO produced in CHO cells showed that the first amino acid was phenylalanine 19, however, this TPO ran as a single band at 106 kDa and not as a doublet [Fan 1996]. Interestingly, the sequencing of trypsin purified native TPO identified a N-terminus at residue 112 (in addition to 109), which does not correspond to a trypsin cleavage site [Taurog 1990] and it was postulated that the N-terminal domain of TPO may be cleaved by an endogenous protease. Such proteolysis occurs in the closely related peroxidase, MPO, in which the N-terminal 141 residues are removed by post-translational processing [Olsen & Little 1983]. The alignment of N-terminal domains of TPO sequences from human, pig, mouse and rat (see Figure 4.15) shows that the first 60 residues are poorly conserved, implying that they may not have a physiological role.

Assuming N-terminal heterogeneity is responsible for the disordered crystals, the N-terminus must be made more homogeneous before well diffracting crystals can be obtained. One approach would be to purify an individual band of the doublet,

however, the results presented here show that each doublet band may be made up of more than one species. Treatment of the expressed protein with a protease to cleave the N-terminal domain in the same position in all the molecules is a potential solution. However, a unique and accessible cleavage site on TPO is necessary and this method is likely to introduce more heterogeneity than it removes. Alternatively, use of additional protease inhibitors during the production of TPO may inhibit the proteolysis and produce TPO molecules with the N-terminus intact.

A new DNA construct of TPO could be made, in which the N-terminus was truncated C-terminal to residue 109. However, this may lead to expression problems, as there is evidence that the N-terminus is necessary for translocation to the cell surface and proper folding [Andersson 1998]. In addition, expression of only half of the N-terminus may prevent the domain from folding correctly and so introduce further heterogeneity. Alternatively, the cleavage sites present in the N-terminus could be removed by mutation, although it is not clear which protease is responsible for cleavage at residue 75. An accessible cleavage site at the beginning of the peroxidase domain could be introduced into TPO, and then the N-terminus post-translationally cleaved, mirroring the naturally occurring situation in MPO. It is unlikely that the N-terminal domain is antigenic as there was no change in autoantibody binding when the N-terminal domain was replaced with that of MPO [Nishikawa 1993; Nishikawa 1994b]. Therefore it is likely that molecules mutated in this domain would have the same autoantibody binding properties as the full-length molecule.

Other than the N-terminal heterogeneity, the presence of multiple glycoforms and carbohydrate or interdomain flexibility are possible reasons for the low-resolution diffraction of the TPO crystals. The carbohydrate residues could be removed by deglycosylation, or by the mutation of sugar binding asparagine residues. Deglycosylation studies of TPO were attempted in this study, but not enough TPO was available to carry out a fully exhaustive study. The mutation of sugar binding residues may be problematic as carbohydrates are thought to be necessary for the

correct trafficking and folding of proteins [Elbein 1991]. Comprehensive MS analysis to identify the occupation and heterogeneity of each N-glycosylation site could be used to target mutations to the most heterogeneous sites.

Interdomain movement could be overcome by expressing individual domains of TPO. The cleavage of the N-terminal domain described above may go some way to reducing flexibility. There is evidence that the CCP/EGF domain pair may contain an antigenic site [Estienne 1998]. Separate construct could be made of the CCP/EGF pair and of the peroxidase domain. Autoantibody binding could then be mapped to the expressed domains which would be expected to be fully folded. If the isolated peroxidase domain did not bind to autoantibodies, the structure of this domain would still shed light on the enzyme active site.

8.2 Thyroid Peroxidase Structure

The 7 Å dataset collected from a recombinant, human TPO crystal was processed and a potential molecular replacement solution found using MPO as a phasing model. Crystal packing led to two different TPO dimers formed by crystallographic two-folds. One set of crystal contacts involves residues 605 to 618. Interestingly, these are the same as those shown by Hobby *et al.* to comprise an immunodominant region on TPO [Hobby 2000]. The other set of crystal contacts are mediated by the loop that is involved in the interchain disulphide bond in MPO [Davey & Fenna 1996].

An alignment of the TPO and MPO sequences shows that the position of the disulphide bonded cysteine is conserved in both proteins. Therefore, assuming one of the dimers is representative of the physiological dimer, it is more likely to be the dimer formed by the cysteine containing loop. In this dimer, the association between monomers is not as close as in MPO, and the monomers are at a different relevant orientation to each other. The cysteines are not close enough to form a

disulphide bond, however the bond may be able to form as TPO has a ten residue insertion in this loop compared with MPO. The solution provides evidence that the dimer interface in TPO uses the same interchain disulphide bridge as in MPO, but that the monomers are at a different orientation to each other. The association of domains from homologous proteins using different surfaces has been previously identified in the C-type lectin domain family [Drickamer 1999].

The electron density map of TPO did not show any additional features beyond the MPO model. However, given the low resolution and the quality of the data this is not entirely surprising. Therefore, the molecular replacement solution must remain tentative. In order to trace the chain, data to higher resolution is required, and it may also be necessary to locate the position and orientation of the N- and C-terminal domains by molecular replacement. The high AMORE correlation coefficient and the generation of a crystallographic symmetry related dimer involving the interchain cysteine is evidence that the solution may be correct.

In the absence of high resolution data, a molecular model of TPO has been built. The peroxidase domain was modelled using MPO, the EGF domain with fibrillin, and the CCP domain with vaccinia CCP. The high degree of sequence identity and the conservation of structurally important residues in the template structures suggest that the predictions made by the model may be treated with some confidence. The identity between EGF and CCP domains of TPO and the homology models are lower than that of the peroxidase domain with MPO (39% and 35% respectively, compared with 46% for MPO), but still high enough to lead to reliable models. The N-terminal domain was threaded onto the structure of sperm lysin. This is the most tentative region of the model as it was built by *de novo* techniques which are inherently less reliable than homology modelling.

The model of the peroxidase domain of TPO shows that all seven key residues in the active site of MPO are within 1 Å of their position in MPO. This suggests that the mechanism of oxidation of iodide by TPO is likely to be similar to that of

chloride by MPO. A surface representation of the model shows that a bridge is formed in TPO, close to the active site, by the peptide that is cleaved in MPO to give a light chain and a heavy chain. If this region of the model is correct, the bridge may be involved in thyroglobulin binding. The peroxidase model also explains the inactivity of TPO-2, due the structurally integral nature of the missing peptide.

The model provides evidence that the dimer interface in TPO is different from that of MPO. Firstly, the loop containing the interchain disulphide is longer in TPO, secondly, the glycosylation sites that binds sugars important in stabilising the MPO interface is absent in TPO, and thirdly residues that form hydrogen bonds across the MPO interface are mutated in TPO. The model is consistent with the possibility, suggested by the low resolution TPO structure, that the two-fold rotational symmetry centered on the interchain disulphide of the MPO dimer is maintained in TPO, but that the monomers are at a different relative orientation.

The different dimer interface in TPO may be due to constraints on the dimer formation in TPO that are not present in MPO. TPO has three addition domains and a membrane anchorage. The position of the N- and C-termini of the domains and the length of the domain linkers will only allow certain peroxidase dimers to form. The presence of transmembrane helices restricts the position of the C-terminus of each monomer to a two dimensional plane. In addition, TPO interactions with a large substrate (thyroglobulin), and access to the active site may restrict positions of the monomers in the dimer.

The identification of a four-helix structure as a model for the N-terminal domain agrees with a previous secondary structure prediction [Banga 1990]. The role of the N-terminal domain of TPO is unclear. It has been proposed that in MPO it is important in the trafficking of the mature protein [Andersson 1998]. However, post-translationally cleavage of the N-terminal domain in TPO has not been reported, suggesting that the domain may play some additional role. A four-helix

bundle could be involved in interacting with thyroglobulin during thyroid hormone synthesis, for example mediating interactions between TPO and thyroglobulin. It is therefore interesting that the scaffold used to thread the N-terminal domain, sperm lysin, is an adhesion protein.

N-terminal sequencing indicates that the recombinant TPO preparation has been cleaved in the centre of the N-terminal domain. The model of the N-terminal domain positions both the N-terminal start sites in loop regions, suggesting that they are accessible to protease cleavage. Therefore, the N-terminus of the recombinant preparation is likely to be extended tail rather than a fully folded domain. Heterogeneity in the TPO preparation due to different lengths of N-terminal sequence would explain the low resolution diffraction obtained from the TPO crystals. The crystals would not be able to pack in a sufficiently regular array to diffract to higher resolution due to the differing lengths, and possible flexibility of the N-termini.

The 'divide and conquer' modelling strategy does not provide any information on the relative orientation of the domains. A model of the potential domain orientation, considering the interchain disulphide bond, the proximity of the N- and C-termini of the peroxidase domain, and the short length of the linkers between domains has been proposed. The model shows that, due to steric considerations, there are certain regions of the structure that would not be accessible to autoantibodies. Inaccessible regions are defined by the interface between domains, between the protein and the membrane, and between dimers. In order to visualise the sizes of the molecules involved, a full length IgG molecule is drawn at the same scale as an MPO dimer in Figure 8.1. The model of the quaternary structure of TPO explains observations that dimeric TPO is more antigenic than monomeric TPO [Baker 1994b] and that autoantibodies do not interact with the TPO dimer interface [Nishikawa 1994b].

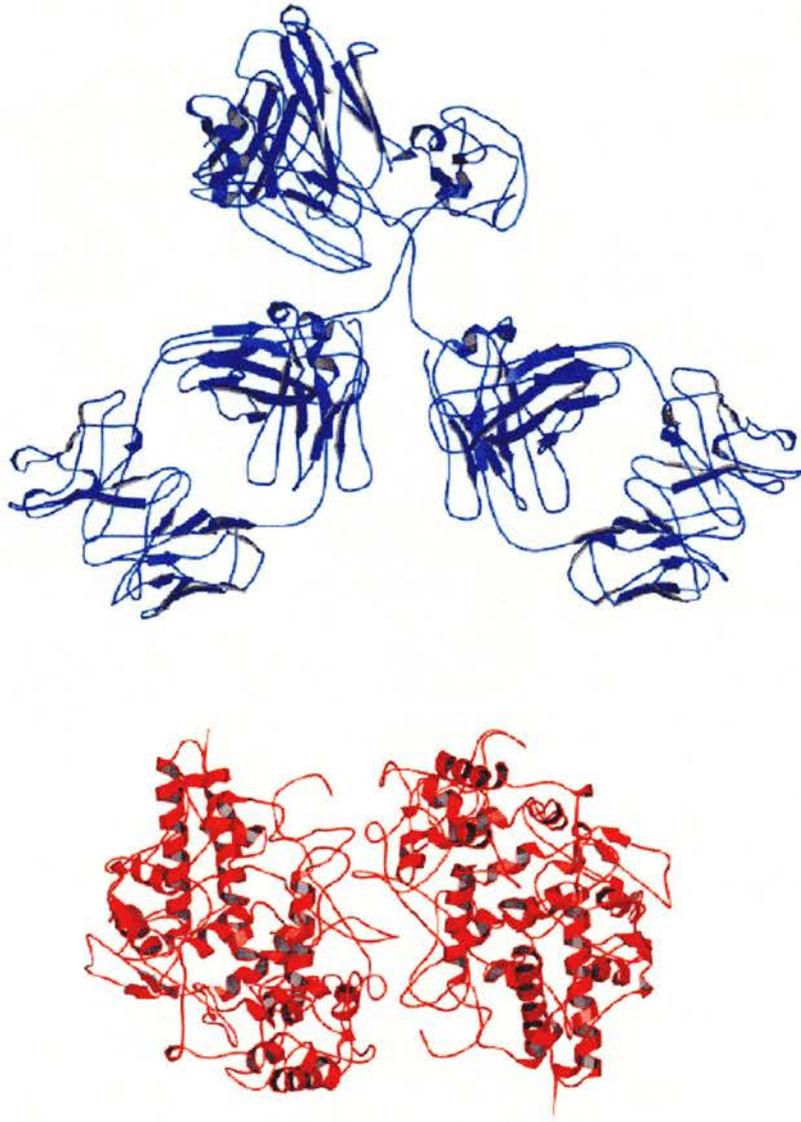


Figure 8.1

The structure of IgG (blue) drawn to the same scale as an MPO dimer (red).

Recently, it has been shown that cellular TPO but not purified TPO and adjuvant induces antibodies in mice that resemble autoantibodies [Jaume 1999b]. This may be a reflection of the monomeric nature of purified TPO compared with dimeric cellular TPO. Purified TPO is used for epitope mapping studies [Rapoport 1995] and to generate the widely used panel of mouse monoclonals [Ruf 1989]. The oligomeric nature of purified TPO is not clear, but the recombinant TPO used in this study behaves as a monomer on SDS-PAGE gels and on a gel filtration column. The disruption of dimer interface may be caused by the lack of membrane anchorage and, in the case of recombinant TPO, altered glycosylation. Membrane anchorage may promote the formation of oligomers by limiting diffusion from three to two dimensions. Recombinant TPO has been shown to have a different carbohydrate content from native protein [Guo 1998b], which may disrupt the dimer interface. Catalytically active recombinant MPO expressed in CHO cells has been shown to be monomeric, whereas the native enzyme is dimeric [Moguilevsky 1991].

In contrast, it is thought that, *in vivo*, autoantibodies bind to dimeric, membrane bound TPO. Lymphocyte infiltration in the thyroid gland is common in AITD patients and the site of synthesis of TPO autoantibodies is believed to be the thyroid [Weetman & McGregor 1994]. It is possible that extrathyroidal sites of autoantibody synthesis, such as the lymph nodes draining the thyroid gland, may contain monomeric TPO. However, TPO-reactive Fab isolated from the thyroid and lymph node of the same patient showed similar genes sequences and recognition of native TPO [Czarnocka 1997]. In addition, the low concentrations of autoantibodies that recognise denatured TPO [Guo 1999] suggests that the selection of the B-cell repertoire occurs at the surface of the thyroid cells. Therefore, it is possible that the A and B immunodominant domains on TPO, identified by competition studies with patient sera, may simply be those regions of the structure that are accessible to autoantibody binding in the membrane bound dimer.

It can be speculated that TPO contains an ancestral adhesion site. It has been proposed that the animal peroxidase family may have evolved from proteins with adhesion functions [Taurog 1999]. The site of adhesion in peroxinectin has been localised to a KGD peptide [Johansson 1995], and a protein sequence alignment shows the homologous sequence in TPO is RGD, a related motif known to bind integrins [Ruoslahti 1987]. The RGD motif is also present in the same position in MPO. However, the RGD motif is not positioned on the surface of the TPO model, as would be expected for an adhesion site. This may indicate that the model is incorrect in this region, or that the RGD motif is no longer a functional adhesion site.

Assuming the RGD motif is on the surface of TPO and has a role in adhesion, it is possible that it may form an inherently 'sticky' patch that is a target for autoantibodies. In support of this hypothesis is the fact that MPO is also an autoantigen, with autoantibodies to MPO found in patients with systemic vasculitis [Kallenberg 1998]. This possibility could be tested by site-directed mutagenesis of the RGD motif.

8.3 4F5 Structure

4F5 is a murine antibody that binds with high affinity to TPO. Competition studies have shown that, despite being a murine antibody, 4F5 competes with autoantibodies from AITD patient sera for binding to TPO. Therefore, the structure of 4F5 provides information that can be directly applied to the interaction of autoantibodies with TPO. Antibody molecules are susceptible to cleavage into fragments which exhibit some of the properties of the intact molecule. The Fab fragment of 4F5 contains the antigen binding site and experiments have shown that it binds to TPO in a similar fashion to the full-length IgG.

The structure of the combining site of 4F5 provides a hint of the nature of the autoantigenic surface of TPO. The centre of the antigen binding site contains a small basic and acidic cavity, the charges contributed by aspartate H95 and arginine L91. Next to the cavity is a protruding negative (glutamate H61) and positive (lysine H58) charge. In common with other antibody structures, there are many tyrosine residues with accessible side chains in the combining site. The complementary surface on TPO may contain a ridged surface with a negative and a positive charge juxtaposed with a hollow containing a positive and negative charge. If the structure of TPO becomes available it may be possible to carry out docking simulations in order to identify such a surface.

It has been shown that some antibodies undergo a conformational change, or induced fit, on binding antigen and that water molecules are involved in the interface [Davies & Cohen 1996]. If this occurs with 4F5, then direct extrapolations from antibody to antigen structure are more difficult. However, there is evidence that induced fit may be less common in antibodies that have undergone extensive maturation [Wedemayer 1997]. Given the presence of somatic mutations and the high affinity of 4F5 it is likely that it may have matured to a structure that binds to TPO using a lock-and-key mechanism.

The interaction in the crystal of the CDR residues of one 4F5 molecule with the constant domain of another may be representative of antigen binding. Head-to-tail packing is a common feature of antibody fragment crystals [Cygler 1987], and intermolecular contacts have been used to propose a model for antigen binding of the Fv fragment MFE-23 to its protein ligand [Boehm 2000]. The 4F5 interaction involves seven CDR residues that are within 4 Å of the C_L domain of a neighbouring molecule, and two residues that bind to bridging water molecules. Four residues that are involved in the interaction have undergone somatic mutation (phenylalanine L32, tyrosines H33 and H56 and asparagine H54), and are therefore likely to be involved in the physiological interaction with TPO. It would be

interesting to perform site-directed mutagenesis of these residues in order to study their role in binding to TPO.

It is unlikely that this interaction is fully representative of that with TPO as the interaction surface area is less than 500 \AA^2 , whereas antibody-protein complexes have been shown to bury $700\text{-}900 \text{ \AA}^2$ [Padlan 1994]. In addition, only CDR residues from L1, H1, H2 and H3 are involved in the interaction, whereas L3 has been shown to play a significant part in interactions with antigen [Wilson & Stanfield 1994].

The region of the C_L domain of 4F5 that interacts with the CDRs is composed predominantly of helix B. Three charged residues are involved in hydrogen bonds, they form the sequence X Asp X X Glu Arg X X (where X is an amino acid). The TPO sequence contains a related motif, X Glu X X Glu Arg X X (residues 558 to 565), which corresponds to helix 11 of MPO. This sequence is on the surface of the peroxidase domain model and the side chains of the two glutamates and the arginine are accessible to the solvent. Extrapolating from the crystal contacts, it is possible 4F5 binds to this region of TPO. This could be studied using site directed mutagenesis of the two glutamates and the arginine on TPO.

The combining site of 4F5 can be compared with that of TR1.9, an autoantibody to TPO of known structure [Chacko 1996]. TR1.9 defines the A domain on TPO [Nishikawa 1994a] and 4F5 has been shown to experimentally to recognise one of two autoantigenic domains on TPO. However, it is not known how the domain bound by 4F5 relates to the A and B domains. 4F5 and TR1.9 have a similar structure in all CDRs apart from H3, but the surface charge profile differs significantly. It is likely, due to differences in charge and topology between the combining site of the two antibodies, that they do not bind to the same region on the surface of TPO. However, the possibility that either or both of the antibodies undergo induced fit in order to bind to a similar structure TPO cannot be completely excluded. It is also possible that the epitopes recognised by 4F5 and TR1.9 are

overlapping. It would be interesting to see how 4F5 and TR1.9 compete with each other for binding to TPO.

The analysis of the antibody combining site of 4F5 has been undertaken using the assumption that the conventional site (i.e. the CDRs) is used for antigen binding. The structure of rheumatoid factor bound to an Ig has shown that other regions are involved in binding to antigens [Corper 1997]. However, rheumatoid factor is a low affinity antibody which may use the conventional site for binding a different antigen, whereas the interaction between TPO and 4F5 is monospecific and of high affinity.

8.4 Summary

This work provides information on the structure of TPO and the structure of an anti-TPO antibody. A potential molecular replacement solution to low resolution TPO data has been found, indicating that a different dimer interface from MPO may be present. Characterisation of TPO has shown that N-terminal heterogeneity may be responsible for the low diffraction of the crystals, and ways to increase the resolution have been proposed. A model of TPO has been built which provides evidence that the dimer interface is different from MPO, and indicates that regions of membrane bound TPO are likely to be inaccessible to autoantibodies. The high resolution structure of an antibody to TPO has been solved and residues in the combining site that are likely to be involved in binding TPO have been identified. It is hoped that this study has provided groundwork that will allow the solution of crystal structures of TPO-autoantibody complexes.

Appendix A

Protein Crystallography Theory¹

Protein Crystallisation

Individual protein molecules interact with X-rays weakly, therefore the diffraction signal must be amplified by a three-dimensional array of molecules. Such an array can be found in a crystal. Protein crystals are held together by weak non-covalent interactions, and contain 30% to 80% solvent. Growth is achieved by steadily increasing the protein concentration using a precipitant to cause removal of water. To obtain ordered crystals, the protein must be maintained in a state of homogeneous content and conformation. Due to the large number of variables in a crystallisation experiments, developing a method is largely empirical.

A saturated solution contains the maximum possible amount of solute under equilibrium conditions. The aim of a crystallisation experiment is to induce the formation of a metastable supersaturated solution which contains more than the equilibrium amount of solute. At low ionic strength protein solubility increases with ionic strength and the protein's net charge, a process called salting-in. At high salt concentrations salting-out occurs, whereby protein solubility decreases logarithmically with increasing ionic strength. Precipitants cause the formation of a precipitant-poor layer near the protein surface, either due to the higher affinity of the protein for water than for the precipitant (for example ammonium sulphate) or solvent exclusion effects (polyethylene glycol). The formation of a precipitant-

¹ Reviewed in Drenth, J. (1994). Principles of Protein X-ray Crystallography, Springer.

Rhodes, G. (1993). Crystallography made Crystal Clear, Academic Press..

depleted layer is energetically unfavourable and promotes molecular associations that decrease the total protein surface area exposed to solvent.

The first step in the growth of a crystal is nucleation. This is the assembly of a thermodynamically stable ordered aggregate, which, on reaching a critical size (10-100 molecules) favours growth over dissolution. Growth is the addition of molecules to the crystal and competes with the formation of other non-crystalline aggregates. Nucleation and growth both require supersaturation, the extent of which changes during the experiment. The process is illustrated on a phase diagram (Figure 1).

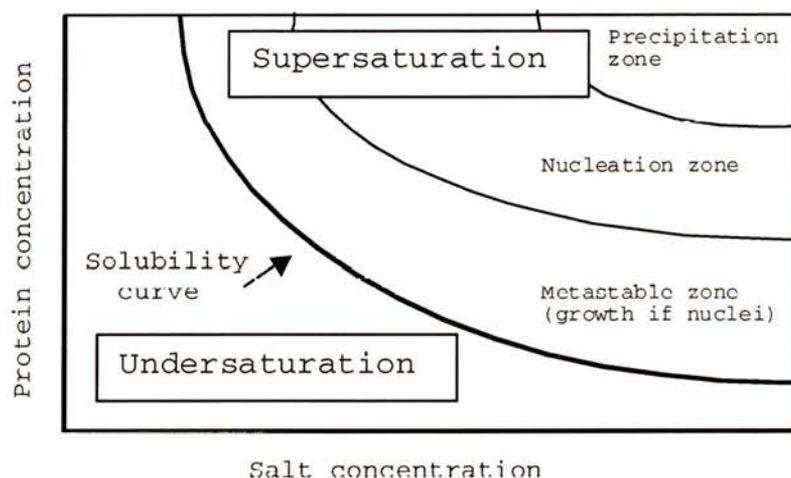


Figure 1

A phase diagram for a protein solution.

In a successful crystallisation experiment the conditions will be gently changed so as to allow the solution to enter the nucleation zone, whilst avoiding precipitation. During nucleation a small number of stable nuclei are formed which leads to removal of molecules from solution. The lower protein concentration promotes return to the metastable zone where the nuclei can grow.

Supersaturation is normally achieved by slow changes in precipitant and protein concentration caused by diffusion. Initial screening experiments sample a wide range of reagents, pH, protein concentration and temperature, often Hampton Research commercial screen are used. Crystals or crystalline precipitates can then be optimised by adjusting several experimental parameters. Vapour diffusion is the most common method for growing protein crystals and the hanging drop setup is shown in Figure 2.

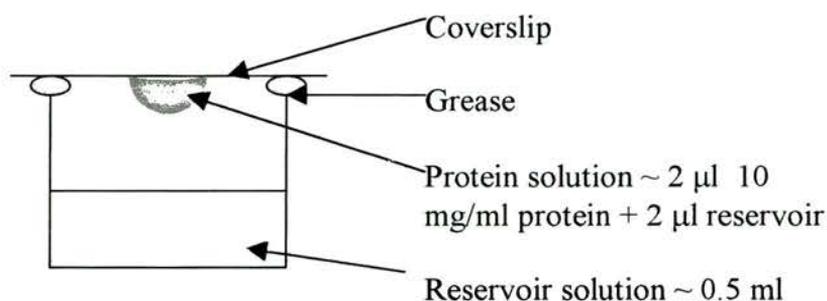


Figure 2

A hanging drop crystallisation experiment.

Protein Crystals

Definitions of the terms used to describe protein crystals are given in Table 1. The positions within a crystal are described by vectors, a , b and c , that define the unit cell (Figure 3). Symmetry considerations leads the possibility of seven different three-dimensional crystal systems, these are reflected in the external symmetry of the crystal. Internally, the protein crystals are restricted to 14 Bravais lattices (P = primitive, I = body-centered, F = face-centered, C = single-face centred). There are

32 crystal classes, or point groups, 11 of which are found in protein crystals. The combination of Bravais lattices and point groups leads to the 65 space groups that are found in protein crystals. These are listed in Table 2.

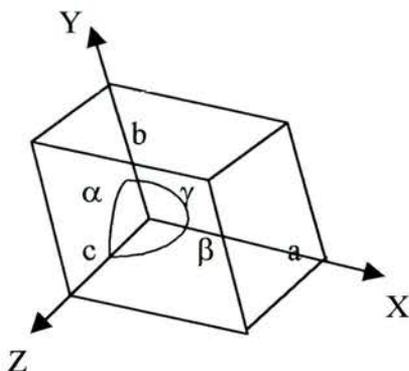


Figure 3
The crystallographic coordinate system.

Crystal system	Cell parameters	Bravais lattices	Point groups	Space groups
Cubic	$a=b=c, \alpha=\beta=\gamma=90^\circ$	P, I, F	23, 432	With/without screws
Tetragonal	$a=b, \alpha=\beta=\gamma=90^\circ$	P, I	4, 422	With/without screws
Orthorhombic	$\alpha=\beta=\gamma=90^\circ$	P, I, C, F	222	With/without screws
Triclinic	No restrictions	P	1	P1 only
Monoclinic	$\alpha=\beta=90^\circ$	P, C	2	With/without screws
Trigonal	$a=b, \alpha=\beta=90^\circ, \gamma=120^\circ$	P	3, 32	With/without screws
Hexagonal	$a=b, \alpha=\beta=90^\circ, \gamma=120^\circ$	P	6, 622	With/without screws

Table 2
The space groups of protein crystals

Crystal	Atoms or molecules arranged in a pattern that repeats periodically in three dimensions.
Unit Cell	The fundamental volume of the crystal.
Asymmetric Unit	Geometric entities or subsets within the unit cell that are equivalent by crystallographic symmetry.
Lattice	A simplified geometric representation of the translational repetition within the crystal, each unit cell is represented by a single point.
Symmetry Operation	An action that converts between indistinguishable states.
Symmetry Element	A geometric entity (point, line or plane) associated with a symmetry operation.
Crystallographic Sym.	Symmetry that is a property of the crystal lattice.
Non-crystallographic Sym.	Symmetry only valid within a single part of the unit cell, a property of the protein.
Rotational Symmetry	A rotation by $360^\circ/n$ about an axis or a line.
Screw Operation	The combination of rotation and translation parallel to the axis of rotation, symbol n_p (rotation by $360^\circ/n$ followed by translation of tp/n , where t is the shortest lattice translation parallel to the screw axis).
Point Group	A symmetry operation that leaves one point in space unmoved: rotation, reflection or rotatory-inversion.
Space Group	A description of the internal structure of the crystal, there are 230 three-dimensional space groups. The symmetry operations of the space group are sufficient to build the unit cell from the asymmetric unit.

Table 1.

Definitions of terms used to describe protein crystals

The X-ray Diffraction Experiment

In order for an object to diffract light, the wavelength of the light must be shorter than the object. X-rays are in the same region of the electromagnetic spectrum as bonded atoms in a protein, which are around 1.5 Å apart. X-rays can be produced in-house by a metal-anode generator, which emits a continuous spectrum of radiation with sharp lines of Cu-K α X-rays ($\lambda = 1.542$ Å). An alternative source is the very intense, polarised, and tunable radiation emitted by bending the path of accelerated electrons in a synchrotron. Optical elements such as monochromators, filters, mirrors and focusing elements define the beam size and energy, and a goniometer head is used to position the crystal. Electronic area detectors with phosphor screens and video or CCD readout capture the position and intensity of the diffracted spots.

X-ray scattering is the redirection of a small fraction of a beam of X-rays by the electrons in a sample. The extent of elastic scattering (scattering without a change in energy) is recorded by the detector. The discrete directions of the diffracted beams are determined by the repetitional symmetry of the crystal lattice, the intensity distribution is determined by the distribution of atoms in the unit cell.

Diffraction can be depicted as occurring in reciprocal space, where the distances are reciprocally related to those in real space. A schematic diagram of the generation of a diffraction pattern is shown in Figure 4. A sphere with a radius equal to the reciprocal wavelength is drawn, called the Ewald sphere. Diffraction only occurs when the Ewald sphere passes through the reciprocal lattice points.

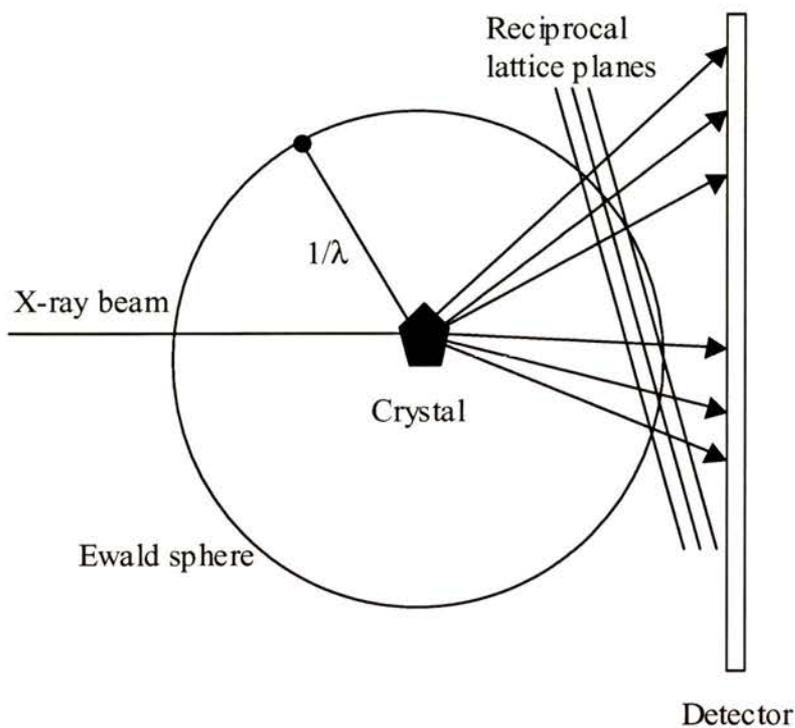


Figure 4

The generation of an X-ray diffraction pattern in reciprocal space.

Two or more interacting waves may show interference. If two waves arrive at a point with the same phase, then the net amplitude is the sum of the individual amplitudes. If the difference in phase is a non-integral number of wavelengths, then the phase will be modulated and the combined amplitude will be smaller than the sum of the two amplitudes. If the pathlength difference is an integral number of wavelengths, as happens in a regular array of scatterers, then the waves completely interfere and there will be no net scattering. The result of interference caused by diffraction of the atoms arranged in a protein crystal is restriction in the directions of scattering, giving rise to an ordered pattern of diffraction spots on the detector.

A wave can be regarded as being reflected against a plane at angle θ (Figure 5). Vector s_0 is the incoming wave of length $1/\lambda$, and s is the reflected wave. S is the difference between s and s_0 and is perpendicular to the reflecting plane. Therefore, using the rules of trigonometry:

$$S = 2(\sin\theta)/\lambda$$

[1]

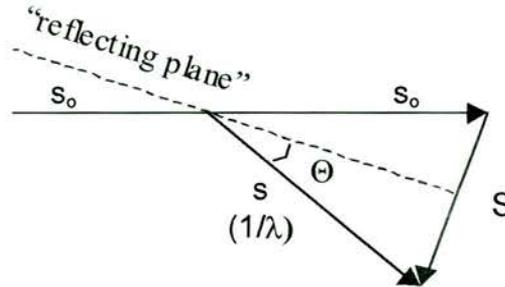


Figure 5

The reflection of a primary wave s_0 against a plane.

A unit cell has a total scattering function $F(S)$ derived from the scattering of all its atoms. A crystal made up of unit cells spaced by the vectors a , b and c will lead to interference with respect to the three spacings. This leads to the Laue equations, which must simultaneously be satisfied:

$$a.S = h \quad b.S = k \quad c.S = l \quad [2]$$

where h , k , and l are independent integers called Miller indices. In practice, it is more convenient to use a coordinate system of reciprocal dimensions, and a reciprocal unit cell of vector a^* , b^* , and c^* .

Bragg's law explains the location of maxima in the diffraction pattern by analogy to the reflection of waves from plane surfaces. If the reflecting plane in Figure 4 is taken to be the lattice plane (hkl) and the distance between planes to be d , then using the Laue condition $d.S = 1$ and substituting into [1]:

$$\lambda = 2d\sin\theta \quad [3]$$

In order to sum individual structure factors they are described by periodic functions:

$$f(x) = F \cos 2\pi (hx + \alpha) \quad [4]$$

$f(x)$ = vertical height of the wave at any horizontal position x , where x is one full wavelength

F = amplitude

h = frequency

α = phase in radians

For mathematical reasons it is more useful to convert equation [4] to the form of a complex number. The phase α is not shown, but is implicit in the combination of cosine and sine functions:

$$f(x) = F [(\cos 2\pi (hx) + i \sin 2\pi (hx))] = F \exp [2\pi i (hx)] \quad [5]$$

The summation of all the atomic structure factors can then be written as a Fourier series in which each term gives the contribution of one atom (j) to the reflection hkl . This leads to the structure factor equation:

$$F_{hkl} = \sum_j F_j \exp [2\pi i (hx_j + ky_j + lz_j)] \quad [6]$$

The structure factors must be multiplied by a temperature dependent factor, B , to account for the vibration of atoms around an equilibrium position:

$$-B_j [(\sin\theta) / \lambda]^2 \quad [7]$$

B is related to the mean square displacement of the atomic vibration (u^2) such that a B-factor of 30 \AA^2 corresponds to a root mean square displacement of 0.62 \AA of the atoms from their equilibrium position:

$$B = 8 \pi^2 u^2 \quad [8]$$

The Fourier transform of the structure factor equation gives the electron density equation:

$$\rho(x, y, z) = 1/V \sum_{hkl} F_{hkl} \exp[-2\pi i (hx + ky + lz)] \quad [9]$$

Using these equations it is possible to get from experimentally determined intensities (which are equal to the amplitude, or F_{hkl} , squared) to the electron density, if the phase information is known.

The Phase Problem

Phase information is not measured in the diffraction experiment, therefore three methods have been developed that used a reference vector with a 'known' phase. These are isomorphous replacement, anomalous diffraction and molecular replacement. Direct methods can be used in some circumstances where very high resolution data have been collected.

In multiple isomorphous replacement, atoms of high atomic number are added to the crystal by soaking pre-grown crystals in a solution containing a heavy-atom reagent. The unit cell must remain isomorphous to the native crystal. Substitutions of highly diffracting atoms generate a measurable change in the intensity pattern, from which the added atoms can be located by Patterson methods. Patterson functions (F_P) are amplitudes (structure factors squared) with phase angles equal to zero, and Patterson maps displays peaks at locations corresponding to vectors

between atoms. F_P is calculated from the native data and F_{PH} from the heavy atom data. A difference Patterson function is used to locate heavy atoms, this function is a map of vectors between the heavy atoms:

$$P(u, v, w) = 1/V \sum_{hkl} (F_{PH} - F_P)^2 \cos [2\pi (hu + kv + lw)] \quad [10]$$

When combined with the measured amplitudes from the native crystal, the reference vector from the heavy atoms provides phasing information for the protein.

Anomalous diffraction signals arise from the variation in atomic scattering as a function of X-ray wavelength that is associated with absorption edges. A tunable X-ray source, such as a synchrotron, is necessary to access the absorption edge. The signal most commonly comes from selenium atoms that replace the sulphur atoms in cysteine residues. Phasing by anomalous diffraction requires the measurement of several difference signals, a technique called multiwavelength anomalous diffraction, or MAD.

The method of molecular replacement uses the phases of a correctly positioned and orientated model structure to provide initial phases. The first step is a cross-rotation function to determine the orientation of the known and unknown molecules with respect to each other. A rotation function is the angular relationship between identical molecules (self-rotation function) or between closely related or identical molecules in two different crystal forms (cross-rotation function). This is followed by translation to superimpose the two molecules. Patterson maps are independent of the position of the structure in the unit cell, so the maps of the measured data and the phasing model will have the highest correlations when the molecules are in the same orientation. The search is monitored by calculating the R-factor between the two sets of structure factors (where F_{obs} are the observed structure factor amplitudes and F_{calc} are the calculated structure factor amplitudes):

$$R = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs} \quad [11]$$

The choice of the known structure and the content of the phasing model are important considerations.

Phases can be improved by map modification techniques that use various assumptions about the electron density. The density in solvent channels should be constant and lower than average protein density, so solvent density can be flattened if a protein-solvent boundary can be calculated. Histogram matching is a related technique where the distribution of densities in the protein is matched to an expected distribution. The electron density of protomers in a multimeric protein are usually equivalent. If the protomers are crystallographically independent then the related densities may be averaged to construct an improved map. Molecular averaging can also be used if the protein crystallizes in more than one crystal form.

Model Building and Refinement

Information about protein structure such as geometry, stereochemistry and fragments of solved structures are used to model the amino acids into the electron density map. The most useful type of map for model building is a $2F_{\text{obs}} - F_{\text{calc}}$ map which reduces the overall model influence and produces a continuous trace. The maps are weighted according to the reliability of the atom positions, for example using sigma-A. Sigma-A is an error measurement that can be approximated by the square root of the correlation coefficient between the observed and calculated normalised structure factors.

Refinement is composed of cycles of reciprocal space calculations, and real space model building. It is an iterative process in which the reliability of the atomic parameters, x , y and z (the position of the atoms) and B (the isotropic temperature factor) are improved, in order to minimise the difference between F_{obs} and F_{calc} . The minimized function, Φ , includes terms for stereochemical restraints. These are bond lengths, bond angles, planarity of aromatic rings, chirality, van der Waals

interactions, and torsion angles. Terms for bond lengths and angles are shown below:

$$\begin{aligned} \Phi = & \sum_{hkl} w_{hkl} (|F_{obs}| - |F_{calc}|)^2_{hkl} \\ & + \sum_i^{bonds} w_i (d_i^{ideal} - d_i^{model})^2 \\ & + \sum_i^{angles} w_i (\phi_i^{ideal} - \phi_i^{model})^2 \end{aligned} \quad [12]$$

w = weight corresponding to the reliability of the reflections

d_i = length on bond i

ϕ_i = bond angle at location j

Simulated annealing is a molecular dynamics technique that removes regions of structure trapped in local minimum. Atoms are assigned velocities in a random direction, from a Boltzman distribution, which states that the number of structures with potential energy ϵ_{pot} is proportional to:

$$\exp [-\epsilon_{pot} /kT] \quad [13]$$

k = Boltzmann's constant

T = temperature K

The increase in energy is sufficient to allow the structure to escape from the local energy minima. The temperature is then gradually dropped in steps to approach the global minimum.

The progression of refinement is monitored by the R-factor, calculated as in equation [9]. The free R-factor uses a test set of 5-10% of the observed reflections so is unbiased by the refinement procedure.

Appendix B

Amino Acid Sequences of TPO Autoantibodies

Anti-TPO autoantibody sequences are from the KabatMan database (last update September 30 1999) [Martin 1996] and the NCBI database (last update 1 July 2000) [NCBI 2000], and collated from papers. The 4F5 sequence is added for comparative purposes. CDRs were defined (Kabat definitions) and the canonical class identified (Chothia templates) automatically by KabatMan. Lengths of CDRs are given with canonical class in brackets, a question mark signifies that the canonical class was not automatically identified. Canonical class assignment could only be performed if the sequence of the entire V domain was available, incomplete sequences are indicated with a star. The colour code used is that used to analyse the structures of 4F5 and TR1.9, and is shown in Table 7.1.

The subset of the sequences which have been mapped to autoantigenic domains of TPO are compared separately. The immunodominant domains are named A and B as identified in [Ruf 1989] or 2G4/not 2G4 as identified in [Hexham 1994]. Minor domains are shown in brackets, and both domains separated by a slash if they are equally important.

Light Chain CDRs

Name	CDR Sequence			CDR length (canonical class)			Reference
	L1	L2	L3	L1	L2	L3	
10I	RASQSISSYLN	AASSLQS	QGSYSTPFF	11(?)	7(1)	9(1)	[Hexham 1994]
1261P5	RASQSISSYLN	AASSLQS	QGSYSTIP	11(?)	7(1)	8(?)	[McIntosh 1997]
126A	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126C	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126D	RASHRISSWLA	AASHLQW	QCANSFPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126F	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126G	RASQGISSYLA	AASTLQS	QCVHSYPTV	11(?)	7(1)	9(1)	[McIntosh 1997]
126H	RASQGISSYLA	AASTLQS	QCLHSYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126I	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TO1	RASQSISSYLN	AASSLQS	QGSYSTPFF	11(?)	7(1)	9(1)	[McIntosh 1997]
126TO2	RASQSISSYLN	AASSLQS	QGSYSTIP	11(?)	7(1)	8(?)	[McIntosh 1997]
126TO6	RVSQSISSYLN	AASSLQS	QGSYSTIP	11(?)	7(1)	8(?)	[McIntosh 1997]
126TO8	RASQGISSYLA	AASTLQS	QCVHSYPTV	11(?)	7(1)	9(1)	[McIntosh 1997]
126TO9	RASQGISSYLA	AASTLEH	QCVNHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TO10	RATQGISSYLA	VASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TO15	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP1	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP2	RASQSISSYLN	AASSLQS	QGSYSTPFF	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP5	RASQSISSYLN	AASSLQG	QGSYSTPWV	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP6	RATQSISSYLN	AASSLHS	QCVYHY*P*	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP7	RATQGISSYLA	AASTLHT	QCVNHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP8	RATQGISSYLA	AASTLHS	QCVYHYPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP9	RASQSISSYLN	AASSLQS	QGSYSTIP	11(?)	7(1)	8(?)	[McIntosh 1997]
126TP13	RASQSISSYLN	AASSLQS	QGSYSTPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
126TP14	RASQSVSSNLA	GASTRAY	QCYHWLPPF	11(?)	7(1)	10(?)	[McIntosh 1997]
126TP15	RASQSISSYLN	AASSLQS	QGSYSTPLT	11(?)	7(1)	9(1)	[McIntosh 1997]
131TP5	RASQSISSYLN	AASSLQS	QGSYSTP*	11(?)	7(1)	9(1)	[McIntosh 1997]
131TP6	RASQSISSYLN	AASSLQS	QGSYSTL*	11(?)	7(1)	9(1)	[McIntosh 1997]
131TP14	RASQSVSSNLA	GASTRAY	QCYHW*P*	11	7	9	[McIntosh 1997]
2G4	RISAAIRAYLT	GSSCRAA	QNYDMPPGIT	12(?)	7(1)	11(?)	[Hexham 1994]
4F5	EAASVVFIL	STSNLAH	LGRSYPYT	10(?)	7(1)	9(1)	Unpublished
ICA	EGSSSSIGNKYVF	DINKRPS	CKAAGNYIV	13(5)	7(?)	10(?)	[Chapel 2000]
ICA5	TGSSSSNIGAGYDVH	GINKRPS	AAWDDLNGPV	15(6)	7(1)	11(?)	[Chapel 2000]
ICB7	EGSSSHTGMYVS	DINKERPS	GTWDSRLAYV	13(?)	7(1)	11(?)	[Chapel 2000]
JA.9	RASQSISSYLN	AASSLQS	QGSYSTPLT	11(?)	7(1)	9(1)	[Chazenbalk 1993]
Lambda	GGNIGSKVH	DDDRPS	QVWDRSSDRYV	11	7	11	[Prummel 1994]
SP1.2	RASENISRYSH	AASTLQS	QGSYSTPFF	11(?)	7(1)	9(1)	[Portolano 1993]
SP1.4	RASQTVGTYLN	AASTLQS	QGSYSTPWV	11(?)	7(1)	9(1)	[Portolano 1993]
SP1.5	RASQNIQKYLH	GTSTLQS	QGSYSTP*	11	7	9	[Portolano 1993]
SP1.12	RASQSISSYLN	AASSLQS		11	7		[Portolano 1993]
SP1.13	RTEQTISSRYLN	AASSLQS		11	7		[Portolano 1993]
SP1.14	RASQNIQKYLH	GTSTLQS		11	7		[Portolano 1993]
SP1.16	RASQDISSRYLN	GASTLEH		11	7		[Portolano 1993]
SP1.17	RASENISRYSH	AASTLQS		11	7		[Portolano 1993]
SP1.18	RARAISTYLN	GTSTLQS		11	7		[Portolano 1993]
SP1.20	RASQSISSYLN	AASSLQS		11	7		[Portolano 1993]
TR1.3	RASQTISSDYLN	AAANLHT	QGSYATPWV	11(?)	7(1)	9(1?)	[Portolano 1993]
TR1.5	RASQSISSYLN	AASTLQS	QGSYSTP*	11(?)	7(1)	9(1?)	[Portolano 1993]
TR1.6	RASQSLINGGYLYL	LGNRAS	MAALPPY	16(?)	7(1)	9(1)	[Portolano 1993]

	D							
TR1.8	RASQSLLSNGYNYL LSSRAA	MALOPPYT	16(?)	7(1)	9(1)	[Portolano 1993]		
	D							
TR1.9	RASQGISALA	DASLE	11(?)	7(1)	9(1?)	[Chazenbalk 1993]		
TR1.10	RASQIKYLA	DASLIG	11(?)	7(1)	8(?)	[Chazenbalk 1993]		
TR1.13	RASQIGRGLA	DASLE	12(?)	7(1)	9(1)	[Chazenbalk 1993]		
TR1.21	RASQISSYLN	AASLES	11(?)	7(1)	9(1)	[Portolano 1993]		
TR1.22	RASQISSYLN	AASLES	11(?)	7(1)	9(1)	[Portolano 1993]		
TR1.23	RASQISSYLN	AASLES	11(?)	7(1)	9(1)	[Portolano 1993]		
TR1.32	RASQISSYLN	AASLES	11(?)	7(1)	9(1?)	[Portolano 1993]		
TR1.37	RASQSLLSNGYNYL LGSTRAA	MARIPYT	16(?)	7(1)	9(1)	[Portolano 1993]		
	D							
TR1.41	GGDHIGK VRL	YDTRPS	11(?)	7(1)	11(?)	[Prummel 1994]		
WR1.7	RASQIIGRYLN	TASLLS	11	7	9	[Chazenbalk 1993]		
WR1.102	TGTSSDVGGYKYVS	EVSRPS	14(?)	7(1)	6(?)	[Portolano 1993]		
WR1.107	SGDALPK YAN	KDERPS	11(?)	7(1)	8(?)	[Portolano 1993]		
WR1.112	SGDALPK YAN	KDERPS	11(?)	7(1)	8(?)	[Portolano 1993]		
WR1.201	RASQVGRILA	GASTRAA	11	7	9	[Jaume 1996]		
WR1.202	RASQVGRILA	GASTRAA	11	7	9	[Jaume 1996]		
WR1.204	RASQIANYLN	AASLES	11	7	9	[Jaume 1996]		
WR1.213	RASQISSYLN	AASLES	11	7	9	[Jaume 1996]		
WR1.214	RASQISSYLN	AASLES	11	7	9	[Jaume 1996]		
WR1.222	RASQSLNWLA	KTSSLES	11	7	9	[Jaume 1996]		
WR1.223	KRQSIILYSSIRK Y	WASTRES	17	7	9	[Jaume 1996]		
	LA							
WR4.2	RASQISSYLN	ATNLLS	11(?)	7(1)	?	[Chazenbalk 1993]		
WR4.3	RASQIIGRYLN	ATNLLS	11(?)	7(1)	?	[Chazenbalk 1993]		
WR4.4	RASQVGRYLN	ATNLLS	11(?)	7(1)	?	[Chazenbalk 1993]		
WR4.5	RASQISSYLN	ATNLLS	11	7	9	[Chazenbalk 1993]		
WR4.7	RASQIIGRYLN	TASLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.8	RASQIIGRYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.9	RASQIIGRYLN	TASLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.10	RASQIIGRYLN	ATNLLS	11(?)	7(1)	11(?)	[Chazenbalk 1993]		
WR4.12	RASQVGRYLN	ATNLLS	11(?)	7(1)	?	[Chazenbalk 1993]		
WR4.21	RASQISSYLN	ATNLLS	11(?)	7(1)	?	[Chazenbalk 1993]		
WR4.22	RASQIIGRYLN	ATNLLS	11	7	9	[Chazenbalk 1993]		
WR4.25	RASQVGRYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.32	RASQIIGRYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.33	RASQMVYLN	ATNLLS	11(?)	7(1)	9(?)	[Chazenbalk 1993]		
WR4.35	RASQISSYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.36	RASQIIGRYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		
WR4.37	RASQVGRYLN	ATNLLS	11(?)	7(1)	9(1)	[Chazenbalk 1993]		

Heavy Chain CDRs

Name	CDR Sequence			CDR length (canonical class)			Reference
	H1	H2	H3	H1	H2	H3	
10I	YAMT	EFANGDFAYYAD	VKG AGRILGVVLWY LYY	5(1)	17(3)	19	[Hexham 1994]
126A	TYMM	SIDGGAFVYYAD	VKG GGWLQLDFFD	5(1)	17(3)	13	[McIntosh 1997]
126B	TFMM	SIGGGAYLYYAD	VKG GGWLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126C	WHSMM	SINSGAFIYYAD	VKG GGWLQLDFFDT	5(1)	17(3)	13	[McIntosh 1997]
126D	TYMM	SIDGGAYVYYAD	VKG GGWLQLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126F	TYMM	SIDGGAYVYYAD	VKG GGWLQLDFFD	5(1)	17(3)	13	[McIntosh 1997]
126G	TYMM	SIDPGSAYVYYAD	VKG GGWLQLDFFD	5(1)	17(3)	13	[McIntosh 1997]
126H	TYMM	SIDGGAYYYAD	VKG GGFLPLDWWL	5(1)	17(3)	13	[McIntosh 1997]
126J	TKMM	SIDPGSAYYYAD	VKG GGWLQLDWFDT	5(1)	17(3)	13	[McIntosh 1997]
126TO1	NYAIH	WINAGDGR KY	EKLGG RDEGFFSGMDV	5(1)	17(3)	13	[McIntosh 1997]
126TO2	NYAIH	WINAGSGR KY	EKLGG RENAFFSGMDV	5(?)	17(3)	13	[McIntosh 1997]
126TO3	NYAIH	WINAGSGR KY	EKLGG RENAFFSGMDV	5(?)	17(3)	13	[McIntosh 1997]
126TO6	NYAIH	WINAGSGR KY	EKLGG RENAFFSGMDV	5(?)	17(3)	12	[McIntosh 1997]
126TO8	TYMM	PIDPGSAYVYYAD	VKG GGWLQLDFFDP	5(1)	17(3)	13	[McIntosh 1997]
126TO9	TFMM	SIGGGAYLYYAD	VKG GGCLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TO10	TFMM	SIGGGAYLYYAD	VKG GGWLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TO15	TFMM	SIGGGAYLYYAD	VKG GGWLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TP1	TFMM	SIGGGAYLYYAD	VKG GGWLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TP5	NYAIH	WINAGSGR KY	EKLGG SEVLFSGMAV	5(1)	17(3)	12	[McIntosh 1997]
126TP6	TYMM	SIDPGSAYVYYAD	VKG GGWLQMDFFDP	5(1)	17(3)	13	[McIntosh 1997]
126TP7	TFMM	SIGGGAYLYYAD	VKG GGWLPMDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TP8	TYMM	SIDPGSAYVYYAD	VKG GGWLQLDFFDT	5(1)	17(3)	13	[McIntosh 1997]
126TP9	NYAIH	WINAGSGR KY	EKLGG RREAFFSGMDV	5(1)	17(3)	12	[McIntosh 1997]
126TP10	TFMM	SIGGGAYLYYAD	VKG GGWLPLDFFDL	5(1)	17(3)	13	[McIntosh 1997]
126TP13	NYAIH	WINAGSGR KY	EKLGG RDEGFFSGMDV	5(1)	17(3)	12	[McIntosh 1997]
126TP14	NYAIH	WINAGSGR KY	EKLGG RENAFFSGIDV	5(?)	17(3)	12	[McIntosh 1997]
126TP15	NYAIH	WINAGSGR KY	EKLGG RENAFFSGIDV	5(?)	17(3)	12	[McIntosh 1997]
131TP14	SYEMH	YIPYGR IYYAD	VRG AFLRFYYYGMDV	5(1)	17(3)	14	[McIntosh 1997]
131TP15	NYAMB	TINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
131IP2	NYAMB	TINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
131TP5	NYAMB	TINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
131TP6	NYAMB	TINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
131TP7	NYAMB	TINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
131TP8	NYVMS	SINNSGGSTYAE	VKG RGPIPYYYYALDV	5(?)	17(3)	13	[McIntosh 1997]
2G4	IKYLS	LLNFDGIPFYAD	VRG DFNELLAS	5(1)	16(1)	9	[Hexham 1994]
4F5	YYIH	WIYPGVVY KY	EKFKD DANLEY	5(1)	17(?)	6	Unpublished
6F	SHDIN	WIDRGTFRYACKFG	GAGAGGWGMDV	5(1)	17(3)	12	[Hexham 1994]
7F	DGDYYW	SIYYGTYYPPLTG	GRAALFGSEYPLD	7(?)	16(1)	15	[Hexham 1994]
ICA	IYNIH	WINAGSGR KY	SCVFD DPNFGDFD	5(1)	17(?)	9	[Chapel 2000]
ICA5	SRPVS	GIIPIFRKNYA	KYGD TRMKITVFASLTDY	5(?)	17(?)	14	[Chapel 2000]
ICB7	IYNIH	WINAGSGR KY	SCVFD DPNFGDFD	5(1)	17(?)	9	[Chapel 2000]
JA.9	GYMH	WISPRGARIFA	KFGG TRAYYGMDV	5(1)	18(?)	10	[Chazenbalk 1993]
Lambda	NYAIH	GIIPIFGANYA	KFGG DRGGPTRRGDAFDV	5	17	14	[Prummel 1994]
SP1.2	GYMH	WISPRGARIFA	KFGG TRAYYGMDV	5(1)	17(?)	10	[Portolano 1993]
SP4.6	DYIVV	WISPKAGRY	CKFGG GVGVG	5	17	6	[Portolano 1993]
TR1.3	LNYMT	VIFDGGIPYAD	VRG TQGRSYYYIDV	5(?)	16(1)	12	[Portolano 1993]
TR1.5	LNYMT	VIFDGGIPYAD	VKG SQGRSYYYIDL	5(?)	16(1)	12	[Portolano 1993]

TR1.6	KFAIH	GFIPMFGTNYA	KFDG	GDRGPVA	FGVFDL	5(1)	17(2)	15	[Portolano 1993]
TR1.8	FAIS	GFIPMFGATYYA	KFDG	ANDRGAYA	FGGFDV	5(1)	17(2)	15	[Portolano 1993]
TR1.9	YGLH	WIDAGTGNKY	KFRG	DPYGGGK	EFDY	5(1)	17(?)	12	[Portolano 1993]
TR1.10	YHLY	WIDPGKGNKY	KFDG	VLGIIAAD		5(1)	17(?)	9	[Portolano 1993]
TR1.21	DYIQ	WIRPARGTIRYA	KFDG	GDTWGFYD		5(1)	17(?)	8	[Portolano 1993]
TR1.22	DYIQ	WIRPNSGATRFA	KFDG	GDTWGLDY		5(1)	17(3)	8	[Portolano 1993]
TR1.23	YIYH	WIDAAAGKRY	PNFA	SGVGSDF	FALGY	5(1)	17(2)	14	[Portolano 1993]
TR1.32	LNYMT	VLYDGTYYAD	VKG	SGGTR	YYYIDL	5(?)	16(1)	12	[Portolano 1993]
TR1.37	RPAIS	GFIPMFGATYYA	KFDG	GDRGALAS	FGGFDV	5(?)	17(2)	15	[Portolano 1993]
TR1.41	NYVIT	GVIPFGTAYAKK	FLG	DRGGPTRR	DALDI	5(?)	17(2)	14	[Prummel 1994]
WR1.7	NYMIH	WIDADGKIRY	KFDG	RGDNI		5(?)	17(?)	6	[Chazenbalk 1993]
WR1.9	NYMIH	WIDAGNKIRY	KFDG	RGDNI		5(?)	17(?)	6	[Chazenbalk 1993]
WR1.102	NYGMS	VISGGTYYAD	VKG			5(1)	17(3)		[Portolano 1993]
WR1.107	DYIH	WIDPNSGGTRFA	KFDG			5(1)	17(?)		[Portolano 1993]
WR1.112	NGYY	IGYINYGSTYYA	SLK			5(?)	18(?)		[Portolano 1993]
WR1.201	NYAMS	GISGGTYYRD	VKG			5	17		[Jaume 1996]
WR1.202	SYMH	IIMPNSGTTYY	KFDG			5	17		[Jaume 1996]
WR1.204	TYIQ	WIDGDTGGKY	KFDG			5	17		[Jaume 1996]
WR1.213	TYAIH	WIDAGTGNKY	KFDG			5	17		[Jaume 1996]
WR1.214	TYVH	WIDGGTGNKY	KFDG			5	17		[Jaume 1996]
WR1.222	NYRIH	WIDVWGNTRY	KL	D		5	17		[Jaume 1996]
WR1.223	RYAIC	GISGGSTYYD	VKG			5	17		[Jaume 1996]
WR4.2	DYIH	WIDPKAGTRY	EKFDG	GVGVT		5(1)	17(?)	6	[Chazenbalk 1993]
WR4.5	DYIH	WIDPNSGGTNYA	KFR			5	17		[Portolano 1993]
WR4.10	DFYIH	WIDPKAGTRFERF	GG	RGLGVT		5(1)	18(?)	4	[Chazenbalk 1993]
WR4.12	DFYIH	WIDPKAGTRFERF	G	GLGVT		5(1)	17(?)	6	[Chazenbalk 1993]
WR4.25	DYYIH	WIDPKAGTRFERF	AG	GLGVT		5(1)	18(?)	6	[Chazenbalk 1993]
WR4.27	DYIH	WIDPKAGTRFERF	G	GLGVT		5(1)	17(?)	6	[Chazenbalk 1993]
WR4.28	DYYIH	WIDPKAGTRFERF	AG	GLGVT		5(1)	18(?)	6	[Chazenbalk 1993]
WR4.31	DYYIH	WIDPKAGTRFERF	G	GVGVT		5(?)	17(?)	6	[Chazenbalk 1993]
WR4.32	DYYIH	WIDPKAGTRY	EKFDG	GVPVT		5(?)	17(?)	6	[Chazenbalk 1993]
WR4.34	DYYIH	WIDPKAGTRY	EKFDG	GVGVT		5	17	6	[Chazenbalk 1993]
WR4.35	DYYIH	WIDPKAGTRY	EKFDG	GVGVT		5(?)	17(?)	6	[Chazenbalk 1993]

Heavy and Light Chain CDRs of TPO Autoantibodies with Mapped Autoantigenic Domains

Name	L1	L2	L3	H1	H2	H3	Immunodominant domain
2G4	RISQAAIRAYLT	GSSCRAA	QH ^Y DM ^F PPGIT	IKYLS	LLHTDGY ^F YAD ^S VRG	DFSSLLA ^H	2G4
6F	RASCRIRISYLN	AASSLQS	QO ^S YS ^T PT	SHDIN	W ^I INRGT ^T RYA ^K Q ^F Q ^G	GAGAGG ^I W ^G MD ^V	2G4
7F	RASQSVSSN ^F LA	GASSRAT	QHYGTPR ^T	DGD ^Y Y	HI ^Y YSG ^T W ^Y Y ^N PS ^L ITG	GRAALF ^G S ^E Y ^P LD ^H	2G4
4F5	SASSSV ^F FIH	STSNLAS	LORSSYP ^T	SY ^Y I ^H	WI ^Y PG ^N V ^Y TK ^S E ^K FK ^D	DAH ^L E ^Y	NOT 2G4
10I	RASQSISSYLN	AASSLQS	QO ^S YS ^T PT	SYAM ^T	S ^F SA ^N G ^D FAY ^Y AD ^V K ^G	AGRILGV ^V W ^L W ^Y SL ^Y Y	NOT 2G4
WR4.5	RASQSISSQYLH	ATSNLLS	QO ^S Y ^I HP [*]	DY ^H I ^H	WIN ^P NEGG ^T W ^Y AQ ^K E ^R	G ^F D ^V	B
SP1.2	RAS ^E NI ^R YSN	AASTLQS	QO ^T YSSP ^F T	G ^H Y ^M H	WIS ^P NRGAT ^R FAQ ^K E ^F Q ^G	TR ^T AY ^Y GM ^D V	B
WR4.2	RASQSIGRYLH	ATSNLVS	QO ^S Y ^T TP [*]	DY ^F I ^H	WIN ^P K ^N AG ^T RY ^S E ^K F ^G	GVGVGT	B
WR4.10	RASQSIGRYLH	ATSNLVS	QO ^S F ^T TP ^Y T	DFY ^I H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^G	RGLGVGT	B
WR4.12	RASQSVGRYLN	ATSNLLS	QO ^S F ^T TP ^Y *	DFY ^I H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^G	GLGVGT	B
WR4.21	RASQSIGRYLH	ATSNLVG	QO ^S Y ^T TP [*]	DY ^Y I ^H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^A G	GLGVGT	B
WR4.25	RASQSVGRYLN	ATSD ^L LS	QO ^S F ^T TP ^Y T	DY ^Y I ^H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^A G	GLGVGT	B
WR4.32	RASQSIGRYLH	ATSNLVG	QO ^S Y ^T TP ^Y T	DY ^Y I ^H	WIN ^P K ^N AG ^T RY ^S E ^K F ^G	GV ^P VGT	B
WR4.35	RASQSISSQYLH	ATSNLLS	QO ^S Y ^I TP ^Y T	DY ^Y I ^H	WIN ^P K ^N AG ^T RY ^S E ^K F ^G	GVGVGT	B
SP4.6				DY ^H VH	WIN ^P K ^N AG ^T RY ^S Q ^K F ^G	GVGVGT	B
WR4.27				DY ^H I ^H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^G	GLGVGT	B
WR4.28				DY ^Y I ^H	WIN ^P K ^N AG ^T RF ^S ER ^F Q ^A G	GLGVGT	B
WR4.31				DY ^Y I ^H	WIN ^P K ^N AG ^T RF ^S E ^K F ^G	GVGVGT	B
WR4.34				DY ^Y I ^H	WIN ^P K ^N AG ^T RY ^S E ^K F ^G	GVGVGT	B
WR4.22	RASQSIGRYLH	ATSNLVG	QO ^S Y ^T TP ^Y T				B
WR4.3	RASQSIGRYLH	ATSNLV	QO ^S Y ^T TP [*]				B

WR4.33	RASQ MV HYLH	ATSNLVS	QQSY FT PYT	SYHLY	W	IN	P	G	K	N	T	K	Y	S	Q	K	F	Q	Q	G	V	L	G	I	A	A	D	H	B(A)										
WR4.4	RASQSVGRYLN	ATSSLLS	QQS F TPY *	QYTIH	W	I	N	A	A	N	G	K	T	R	Y	S	Q	N	F	Q	A	S	G	V	S	S	D	F	F	A	L	G	B(A)						
WR4.8	RASQIGRYLH	ATSNLVG	QQSY FT PYT	NYMIH	W	I	N	A	D	N	G	K	I	R	Y	S	Q	K	F	Q	Q	R	G	D	S	N	I					B(A)							
SP1.4	RASQTVGTYLN	TASPLQS	QQSY ST PWT	NYAIH	W	I	N	N	D	G	R	T	K	F	S	E	N	L	Q														B(A)						
SP1.5	RASQNIKYLN	GTSTLQS	QQSY ST PWT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q													B(A)						
SP1.12	RASQISAYLN	SASSLQS		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q														B(A)					
SP1.13	RTSQIIRYLN	AASSLQT		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q														B(A)					
SP1.14	RASQNIKYLN	GTSTLQS		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q															B(A)				
SP1.16	RASQDIRYLN	GASTLES		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																B(A)			
SP1.17	RASQENIRYLN	AASTLQS		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																B(A)			
SP1.18	RASQRAISTYLN	GTSTLQS		NYAIH	W	I	N	N	D	G	R	T	K	Y	S	E	N	L	Q																	B(A)			
SP1.20	RASQISAYLN	SASSLQS		NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																	B(A)		
TR1.10	RASQITKYLA	DASTLQG	QQAY FT WT	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																B(A)				
TR1.23	RASQISSYLN	AASSLQG	QQSY ST PWT	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																	B(A)			
WR1.7	RASQISRYLN	TASTLLS	QQGY D HPNW	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																	B(A)			
126T01	RASQISSYLN	AASSLQS	QQSY ST PWT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																	B(A)		
126T02	RASQISSYLN	AASSLQS	QQSY ST PWT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																		B(A)	
126T06	RVSQISSYLN	AASSLQS	QQSY ST IT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																		B(A)	
126TP5	RASQISSYLN	AASSLQS	QQSY ST PWT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																		B(A)	
126TP9	RATQIDSYLA	AASTLHS	QQVY H YPLT	NYAIH	W	I	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																		B(A)	
126TP13	RASQISSYLN	AASSLQS	QQSY ST PWT	NYAIH	W	I	N	N	D	G	R	T	K	Y	S	E	N	L	Q																			B(A)	
126TP14	RASQVSSNLA	GASTRAT	QQYNN W LPTT	NYAIH	W	L	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																		B(A)	
126TP15	RASQISSYLN	AASSLQS	QQSY ST PWT	NYAIH	W	L	N	A	G	S	G	R	T	K	Y	S	E	K	L	Q																			B(A)
131TP5	RASQISSYLN	AASSLQS	QQSY ST P*	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																		B(A)		
131TP6	RASQISSYLN	AASSLQS	QQSY ST L*	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																			B(A)	
131TP7	RASQISSYLN	AASSLQS	QQSY ST L*	NYAMS	T	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																			B(A)	
131TP8	RASQISSYLN	AASSLQS	QQSY ST L*	NYVMS	S	I	N	N	S	G	G	S	T	Y	A	E	S	V	K																			B(A)	

131TP15				NYAMS	TINNSGGSTSYAESVKG	RGPIPIYYIALDV	B(A)
126TO3				NYAIH	WINAGSGRKYSEKIQG	RQENAFFSGMDV	B(A)
TR1.3	RASQTI	GDYLN	AASNLHT	QSYATPWT		TQGRSYIYIDV	B/A
TR1.5	RASQINSYLN		AASTLLS	QOSYITP*		SGIRSYIYIDL	B/A
131TP14	RASQSVSSNLA		GASTRAT	QQYNNWP*		AFSLRFSYYGMDV	A(B)
126TO8	RASQGISSYLA		AASTIQS	QQVNSYPVT		GGWLQLDNFFDP	A(B)
126H	RASQGISSYLA		AASTIQS	QQLASYPIT		GGHFLPLDNLNS	A(B)
TR1.6	RSSQELLHGNGYN	LGSNRRAS		MOALQPPYT		GMDRGPVASFVFDL	A
	YLD						
TR1.8	RSSQELLHSHNGYN	LSSRRAS		MOALQPPYT		ANDRGAYASFVFDV	A
	YLD						
TR1.9	RASQGISSALA	DASNLES		QQFNISYPLT		DPYGGGKSEFDY	A
TR1.41	GGDNIGIKSVH	YDTARPS		QVWDRSSNHVY		DRGGPTRRSDALDI	A
126B						GGHFLPLDNLNS	A
WR1.102	TGTS	DVGKYKYV	EVENRPS	SSYTNS		GGHFLPLDNLNS	A
	S						
WR1.223	KSSQSILYSSNKNK	WASTRES		QQYFNSP			A
	NILA						
TR1.13	RASQGISRGLA	DASTLES		QQFNISYPLT			A
WR1.107	SGDALPKQYAH	KDTERPS		QSADSSGT			A

Heavy and Light Chain CDRs (Length and Canonical Class) of TPO
Autoantibodies with Mapped Autoantigenic Domains

Name	CDR Length (Canonical Class)						Immunodominant domain
	L1	L2	L3	H1	H2	H3	
2G4	12 (?)	7 (1)	11(?)	5(1)	16(1)	9	2G4
6F	11(?)	7 (1)	9 (1)	5 (1)	17 (3)	12	2G4
7F	12 (?)	7 (1)	8 (?)	7 (?)	16 (3)	15	2G4
4F5	10 (?)	7(1)	9(1)	5 (1)	17(?)	5	NOT 2G4
10I	11(?)	7 (1)	9 (1)	5 (1)	17 (3)	19	NOT 2G4
WR4.5	11	7	?	5	17		B
WR4.2	11(?)	7 (1)	?	5(1)	17(?)	6	B
WR4.3	11(?)	7 (1)	?				B
WR4.4	11(?)	7 (1)	?				B
WR4.8	11(?)	7 (1)	9(1)				B
WR4.10	11(?)	7 (1)	11(?)	5(1)	18(?)	7	B
WR4.12	11(?)	7 (1)	?	5(1)	17(?)	6	B
WR4.21	11(?)	7 (1)	?	5	18	6	B
WR4.22	11	7	9				B
WR4.25	11(?)	7 (1)	9(1)	5(1)	18(?)	6	B
WR4.27				5(1)	17(?)	6	B
WR4.28				5(1)	18(?)	6	B
WR4.31				5(?)	17(?)	6	B
WR4.32	11(?)	7(1)	9(1)	5(?)	17(?)	6	B
WR4.33	11(?)	7 (1)	9(1)				B
WR4.34				5(?)	17(?)	6	B
WR4.35	11(?)	7(1)	9(1)	5(1)	17(?)	6	B
SP1.2	11(?)	7 (1)	9 (1)	5(1)	17(?)	10	B
SP1.4	11(?)	7 (1)	9 (1)				B
SP1.5	11(?)	7 (1)	9 (1)				B
SP1.12	11	7					B
SP1.13	11	7					B
SP1.14	11	7					B
SP1.16	11	7					B
SP1.17	11	7					B
SP1.18	11	7					A
SP1.20	11	7					A
SP4.6				5	17	6	A
TR1.10	11(?)	7 (1)	8 (?)	5 (1)	17 (?)	9	B(A)
IK1.23	11(?)	7 (1)	9 (1)	5 (1)	17 (2)	14	B(A)
WR1.7	11	7	9	5	17	6	B(A)
126TO1	11(?)	7(1)	9(1)	5(1)	17(3)	13	B(A)
126TO2	11(?)	7(1)	8 (?)	5 (?)	17(3)	12	B(A)
126TO3				5 (?)	17(3)	12	B(A)
126TO6	11(?)	7 (1)	8 (?)	5(?)	17(3)	12	B(A)
126TP5	11(?)	7 (1)	9 (1)	5(1)	17(3)	12	B(A)
126TP9	11(?)	7 (1)	9 (?)	5(1)	17(3)	12	B(A)
126TP13	11(?)	7(1)	9 (1)	5(1)	17 (3)	12	B(A)
126TP14	11 (?)	7(1)	10 (?)	5(?)	17(3)	12	B(A)

126TP15	11(?)	7 (1)	9 (1)	5 (?)	17(3)	12	B(A)
131TP5	11(?)	7(1)	9(1)	5(?)	17(3)	13	B(A)
131TP6	11(?)	7(1)	9(1)	5(?)	17(3)	13	B(A)
131TP7				5 (?)	17(3)	13	B(A)
131TP8				5(?)	17(3)	13	B(A)
131TP15				5 (?)	17(3)	13	B(A)
TR1.3	11(?)	7 (1)	9 (1?)	5(?)	16 (1)	12	B/A
TR1.5	11(?)	7(1)	9(1?)	5(?)	16(1)	12	B/A
126TO8	11(?)	7(1)	9(1)	5(1)	17(3)	13	A(B)
126H	11(?)	7 (1)	9 (1)	5 (1)	17 (3)	13	A(B)
131TP14	11	7	9	5 (1)	17 (3)	14	A(B)
TR1.6	16(?)	7(1)	9(1)	5(1)	17(2)	15	A
TR1.8	16 (?)	7 (1)	9 (1)	5 (1)	17 (2)	15	A
TR1.9	11(?)	7 (1)	9 (?)	5 (1)	17 (?)	12	A
TR1.13	12(?)	7(1)	9(1)				A
TR1.41	11(?)	7(1)	11(?)	5(?)	17(2)	14	A
126B				5(1)	17(3)	13	A
WR1.102	14 (?)	7 (1)	6 (?)	5 (1)	17 (3)	0	A
WR1.107	11(?)	7 (1)	8 (?)	5 (1)	17(?)	0	A
WR1.223	17	7	9	5	17		A

References

ABG (2000). Directory of 3D structures of antibodies. http://www.ibt.unam.mx/vir/structure/stat_aim.html.

Abler, G., Flaswinkel, H., Kim, K., Weiser, P. and Reth, M. (1992). Structure and function of the B-cell antigen receptors. Progress in Immunology VIII. J. Gergely. Berlin, Springer-Verlag.

Abramowicz, M. J., Targovnik, H. M., Verala, V., Cochaux, P., Krawiec, L., Pisarev, M. A., Propeto, F. V. E., Juvenai, G., Chester, H. A. and Vassart, G. (1992a). Identification of a mutation in the coding sequence of the human thyroid peroxidase gene causing congenital goiter. Am. Soc. Clin. Invest. **90**: 1200-1204.

Abramowicz, M. J., Vassart, G. and Christophe, D. (1992b). Functional study of the human thyroid peroxidase gene promoter. Eur. J. Biochem. **203**: 467-473.

Adler, T. R., Beall, G. N., Curd, J. G., Heiner, D. C. and Shabharwal, U. K. (1984). Studies of complement activation and IgG subclass restriction of anti-thyroglobulin. Clin. Exp. Immunol. **56**: 383-389.

Aichinger, G., Fill, H. and Wick, G. (1995). In situ immune complexes, lymphocyte subpopulations, and HLA-DR positive epithelial cells in Hashimoto's thyroiditis. Lab. Invest. **52**: 132.

Al-Lazikani, B., Lesk, A. M. and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J. Mol. Biol. **273**: 927-948.

Altman, R., Kornfeld, T., Dalik, E., Staudacher, E. and Glossl, J. (1993). Processing of asparagine linked oligosaccharides in insect cells. Glycobiology **3**: 619-625.

Amit, A. G., Mariuzza, S. E. V., Phillips, S. E. V. and Poljak, R. J. (1986). Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. Science **233**: 747-753.

Andersson, E., Hellmann, L., Gullberg, U. and Olsson, I. (1998). The role of the propeptide for processing and sorting of human myeloperoxidase. J. Biol. Chem. **273**(8): 4747-4753.

Arcott, P. L., Koenig, R. J., Kaplan, M. M., Glick, G. D. and Baker, J. R., Jr. (1996). Unique autoantibody epitopes in an immunodominant region of thyroid peroxidase. J. Biol. Chem. **271**(9): 4966-73.

Atassi, M. Z. (1984). Antigenic structures of proteins. Their determination has revealed important aspects of immune recognition and generated strategies for synthetic mimicking of protein binding sites. Eur. J. Biochem. **145**: 1-20.

Bach, J. F. (1994). Predictive medicine in autoimmune diseases: from the identification of genetic predisposition and environmental influence to precocious immunotherapy. Clin. Immunol. Immunopathol. **72**: 156-161.

Baker, H. M., Day, C. L., Norris, G. E. and Baker, E. N. (1994a). Enzymatic deglycosylation as a tool for crystallization of mammalian binding proteins. Acta Cryst. **D50**: 380-384.

Baker, J. R., Arcott, P. and Johnson, J. (1994b). An analysis of the structure and antigenicity of different forms of human thyroid peroxidase. Thyroid **4**(2): 173-8.

Balfour, B. M., Donaich, D., Roitt, I. M. and Couchman, K. G. (1961). Fluorescent antibody studies in human thyroiditis: autoantibodies to an antigen of the thyroid colloid distinct from thyroglobulin. Br. J. Exp. Pathol. **42**: 307-316.

Banga, J. P., Mahadevan, D., Barton, G. J., Sutton, B. J., Saldanha, J. W., Odell, E. and McGregor, A. M. (1990). Prediction of domain organisation and secondary structure of thyroid peroxidase, a human autoantigen involved in destructive thyroiditis. FEBS Lett. **266**(1-2): 133-41.

Baniahmad, A., Ha, I., Reinberg, D., Tsai, S., Tsai, M.-J. and O'Malley, B. W. (1993). Interaction of human thyroid hormone receptor β with transcription factor TFIIB may mediate target gene derepression and activation by thyroid hormone. PNAS **90**: 8832.

Barlow, P. N. and Campbell, I. D. (1994). Strategy for studying modular proteins: application to complement modules. Meth. Enzymol. **239**: 464-485.

Baudry, N., Jejeune, P.-J., Delom, F., Vinet, L., Carayon, P. and Mallet, B. (1997). Role of multimerized porcine thyroglobulin in iodide storage. Biochem. Biophys. Res. Comm. **242**: 292-296.

Beever, K., Bradbury, J., Phillips, D., McLachlan, S. M., Pegg, C., Goral, A., Overbeck, W., Feifel, G. and Rees Smith, B. (1989). Highly sensitive assays of autoantibodies to thyroglobulin and to thyroid peroxidase. Clin. Chem. **35**(9): 1949-1954.

Behring, E. A. and Kitasato, S. (1890). Uber das Zustandekommen der Diphtherie-Immunitat und der Tetanus-Immunitat bei Thieren. Dtsch. Med. Wochenschr. **49**: 1113-1114.

Belyavin, G. and Trotter, W. R. (1959). Investigations of thyroid antigens reacting with Hashimoto's sera. Lancet **March 28**(648-652).

Benjamin, D. C. (1984). The antigenic structure of proteins, a reappraisal. Ann. Rev. Immunol. **2**: 67-101.

Benjamin, D. C. (1995). B-cell epitopes: fact and fiction. Inhibitors to Coagulation Factors. L. M. Aledort. New York, Plenum Press.

Benjamin, D. C. and Perdue, S. S. (1996). Site-directed mutagenesis in epitope mapping. Methods **9**: 508-515.

Benner, R., Hijmans, W. and Haaijman, J. J. (1981). The bone marrow: the major source of serum immunoglobulins, but still a neglected site of antibody formation. Clin. Exp. Immunol. **46**: 1-8.

Bentley, G. A. (1996). The crystal structures of complexes formed between lysozyme and antibody fragments. EXS **75**: 301-319.

Bermann, M., Magee, M., Koenig, R. J., Kaplan, M. M., Arscott, P., Maastricht, J., Johnson, J. and Baker, J. R., Jr. (1993). Differential autoantibody responses to thyroid peroxidase in patients with Graves' disease and Hashimoto's thyroiditis. J. Clin. Endocrinol. Metab. **77**(4): 1098-101.

Bhat, T. N., Bentley, G. A., Fishmann, T. O., Boulot, G. and Poljak, R. J. (1990). Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. Nature **347**: 483-385.

Billingham, R., Brent, L. and Medawar, P. B. (1953). Nature **172**: 603.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. and Wiley, D. C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. Nature **329**: 506-512.

Blankenstein, T., Zobelein, G. and Krawinkel, U. (1984). Analysis of immunoglobulin heavy chain V-region genes belonging to the V NP-gene family. Nucleic Acid Res. **12**: 6887-6990.

Boehm, B. O., Kuhl, P., Loliger, C., Ketzler-Sasse, U., Holzberger, G., Seidl, S., Bauerle, R., Schifferdecker, E. and Usadel, K. H. (1993). HLA-DR3 and HLA-DR5 confer risk for autoantibody positivity against the thyroperoxidase (mic-TPO) antigen in healthy blood donors. Clin. Invest. **71**(3): 221-5.

Boehm, M. K., Corper, A. L., Wan, T., Sohi, M. K., Sutton, B. J., Thornton, J. D., Keep, P. A., Chester, K. A., Begent, R. H. J. and Perkins, S. J. (2000). Crystal structure of the anti-(carcinoembryonic antigen) single-chain Fv antibody MFE-23 and a model for antigen binding based on intermolecular contacts. Biochem. J. **346**: 519-528.

Bogner, U., Badenhoop, K., Peters, H., Schmiege, D., Mayr, W. R., Usadel, K. H. and Schleusener, H. (1992). HLA-DR/DQ gene variation in nongoitrous autoimmune thyroiditis at the serological and molecular level. Autoimmunity **14**(2): 155-8.

Bogner, U., Hegedus, L., Hansen, J. M., Finke, R. and Schleusener, H. (1995). Thyroid cytotoxic antibodies in atrophic and goitrous autoimmune thyroiditis. Eur. J. Endocrinol. **132**(1): 69-74.

Bogner, U., Kotulla, P., Peters, H. and Schleusener, H. (1990). Thyroid peroxidase/microsomal antibodies are not identical with thyroid cytotoxic antibodies in autoimmune thyroiditis. Acta Endocrinol. (Copenh) **123**(4): 431-7.

Bogner, U., Schleusener, H. and Wall, J. R. (1984). Antibody-dependent cell mediated cytotoxicity against human thyroid cells in Hashimoto's thyroiditis but not Graves' disease. J. Clin. Endo. Met. **54**: 734-738.

Bork, P., Holm, L. and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J. Mol. Biol. **242**: 309-320.

Boulet, G., Guillon, V., Mariuzza, R. A., Poljak, R. J., Riottot, M.-M., Souchon, H., Spinelli, S. and Tello, D. (1988). Crystallization of antibody fragments and their complexes with antigen. J. Crystal Growth **90**: 213-221.

Braden, B. C., Fields, B. C., Ysern, X., Goldbaum, F. A., Dall'Acqua, W., Schwarz, F. P., Poljak, R. J. and Mariuzza, R. A. (1996). Crystal structure of the complex of the variable domain of antibody D1.3 and turkey egg white lysozyme: a novel conformational change in antibody CDR-L3 selects for antigen. J. Mol. Biol. **257**: 889-94.

Braden, B. C., Goldman, E. R., Mariuzza, R. A. and Poljak (1998). Anatomy of an antibody molecule: structure, kinetics, thermodynamics and mutational studies of the antilysozyme antibody D1.3. Immunol. Rev. **163**: 45-57.

Bradford, M. M. (1976). A rapid sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein dye binding. Anal. Biochem. **72**: 248-252.

Braverman, L. E. and Utiger, R. D., Eds. (1996). Werner and Ingbar's The Thyroid. Philadelphia, New York, Lippincott-Raven.

Brenner, S. A., Badcoe, I., Cohen, M. A. and Gerloff, D. (1994). Bona Fide prediction of aspects of protein conformation. J. Mol. Biol. **235**: 926-958.

Bresler, H. S., Burek, C. L., Hoffman, W. H. and Rose, N. R. (1990). Autoantigenic determinants on human thyroglobulin. Clin. Immunol. Immunopath. **54**: 76-86.

Brunger, A. T. (1992). X-PLOR, Version 3.1. A system for X-ray crystallography and NMR. New Haven, Yale University Press.

Brunger, A. T., Adams, P. T., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kuntzle, R. W., Jiang, J., Kuszewski, J., Nigles, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). Crystallography and NMR System: a new software suite for macromolecular structure determination. Acta Cryst. **D53**: 905-921.

Bundle, D. R., Baumann, H., Brisson, J. R., Gagne, S. M., Zdanov, A. and Cygler, M. (1994). Solution structure of a trisaccharide-antibody complex: comparison of NMR measurements with a crystal structure. Biochem. **33**: 5183-5192.

Burmeister, W. P., Huber, A. H. and Bjorkman, P. J. (1994). Crystal structure of the complex of rat neonatal Fc receptor with Fc. Nature **372**: 379.

Burnet, F. M. (1959). The Clonal Selection Theory. London, Cambridge University Press.

Campbell, I. D. and Bork, P. (1993). Epidermal growth factor-like modules. Curr. Opin. Struct. Biol. **3**: 285-392.

Capeillere-Blandin, C. (1998). Oxidation of guaiacol by myeloperoxidase: a two-electron-oxidized guaiacol transient species as a mediator of NADPH oxidation. Biochem. J. **336**: 395-404.

Cetani, F., Costagliola, S., Tonacchera, M., Panneels, V., Vassart, G. and Ludgate, M. (1995). The thyroperoxidase doublet is not produced by alternative splicing. Mol. Cell. Endocrin. **115**: 125-132.

Chacko, S., Padlan, E. A., Portolano, S., McLachlan, S. M. and Rapoport, B. (1996). Structural studies of human autoantibodies. Crystal structure of a thyroid peroxidase autoantibody Fab. J. Biol. Chem. **271**(21): 12191-8.

Chan, C. C. and Mochizuki, M. (1999). Sympathetic ophthalmia: an autoimmune ocular inflammatory disease. Springer Semin. Immunopathol. **21**: 125-134.

Chan, C. T. J., Byfield, P. G. H., Himsworth, R. L. and Shepherd, P. (1987). Human autoantibodies to thyroglobulin are directed towards a restricted number of human specific epitopes. Clin. Exp. Immunol. **70**: 516-523.

Chan, J. Y., Lerman, M. I., Prabhakar, B. S., Isozaki, O., Santisteban, P., Koppers, R. C., Oates, E. L., Notkins, A. L. and Kohn, L. D. (1989). Cloning and characterisation of a cDNA that encodes a 70 kDa novel human thyroid autoantigen. J. Biol. Chem. **264**: 3651-3654.

Chang, B. and Casali, P. (1994). The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement. Imm. Today **15**(8): 367-373.

Chang, C. C., Huang, C. N. and Chuang, L. M. (1998). Autoantibodies to thyroid peroxidase in patients with type 1 diabetes in Taiwan. Eur. J. Endocrinol. **139**(1): 44-8.

Charlton, B. and Lafferty, K. J. (1995). The Th1/Th2 balance in autoimmunity. Curr. Opin. Immunol. **7**: 793-798.

Chazenbalk, G. D., Costante, G., Portolano, S., McLachlan, S. M. and Rapoport, B. (1993a). The immunodominant region on human thyroid peroxidase recognized by autoantibodies does not contain the monoclonal antibody 47/c21 linear epitope. J. Clin. Endocrinol. Metab. **77**(6): 1715-8.

Chazenbalk, G. D., Portolana, S., Russo, D., Hutchison, J. S., Rapoport, B. and McLachlan, S. (1993b). Human organ-specific autoimmune disease: Molecular cloning and expression of an autoantibody gene repertoire for a major autoantigen reveals an antigenic immunodominant region and restricted immunoglobulin gene usage in the target organ. J. Clin. Invest. **92**: 62-74.

Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanite-phenol-chloroform extraction. Anal. Biochem. **162**(1): 156-159.

Chothia, C. and Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. **196**: 901-917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. and Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions. Nature **342**: 877-883.

Cohen, S. B. and Weetman, A. P. (1988). The effect of iodine depletion and supplementation in the Buffalo strain rat. J. Endocrin. Invest. **11**: 625-627.

Collaborative Computer Project Number 4 (1994). The CCP4 Suite: programs for protein crystallography. Acta Cryst. **D50**: 760-763.

Costante, G., Portolana, S., Nishikawa, T., Juame, J. C., Chazenbalk, G. D., Rapoport, B. and McLachlan, S. M. (1994). Recombinant thyroid peroxidase-

specific autoantibodies II. Role of individual heavy and light chains in determining epitope recognition. Endocrin. **135**(1): 25-30.

Corper, A. L., Sohi, M. K., Bonagura, V. R., Steinitz, M., Jefferis, R., Feinstein, A., Beale, D., Taussig, M. J. and Sutton, B. (1997). Structure of human IgM rheumatoid factor Fab bound to its autoantigen IgG Fc reveals a novel topology of antibody-antigen interaction. Nat. Struct. Biol. **4**(5): 374-381.

Corper, A. L., Stratmann, T., Apostolopoulos, V., Scott, C. A., Garcia, K. C., Kang, A. S., Wilson, I. A. and Teyton, L. (2000). A structural framework for deciphering the link between I-Ag7 and autoimmune disease. Science **288**: 505-511.

Courcelle, E. (2000). ESPript. <http://www-pgm1.ipbs.fr:8080/cgi-bin/ESPript-exe.cgi>.

Coutinho, A. (1989). Beyond clonal selection and network. Immunol. Rev. **110**: 63-87.

Coyle, P. V., Wyatt, D., Connolly, J. H. and O'Brien, C. (1989). Epstein-Barr virus infection and thyroid dysfunction. Lancet **8608**: 899.

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998). Jpred: a consensus secondary structure prediction server. <Http://jura.ebi.ac.uk:8888/>. Bioinformatics **14**: 892-893.

Cygler, M., Boodhoo, A., Lee, J. S. and Anderson, W. F. (1987). Crystallisation and structure determination of an autoimmune anti-poly(dT) immunoglobulin Fab fragment at 3.0 Å resolution. J. Biol. Chem. **262**(2): 643-648.

Czarnocka, B., Janota-Bzowski, M., McIntosh, R. S., Asghar, M. S., Watson, P. F., Kemp, E. H., Carayon, P. and Weetman, A. P. (1997). Immunoglobulin G κ

antithyroid peroxidase antibodies in Hashimoto's thyroiditis: epitope-mapping analysis. J. Clin. Endocrinol. Metab. **82**(8): 2639-2644.

Czarnocka, B., Pastuszko, D., Carayon, P., Ruf, J. and Gardas, A. (1996). Majority of thyroid peroxidase autoantibodies in patients with autoimmune thyroid disease are directed to a single TPO domain. Autoimmunity **23**(3): 145-54.

Czarnocka, B., Ruf, J., Ferrand, M., Carayon, P. and Lissitzky, S. (1985). Purification of the human thyroid peroxidase and its identification as the microsomal antigen involved in autoimmune thyroid diseases. FEBS Lett. **190**(1): 147-52.

Czarnocka, B., Szabolcs, I., Pastuszko, D., Feldkamp, J., Dohan, O., Podoba, J. and Wenzel, B. (1998). In old age the majority of thyroid peroxidase autoantibodies are directed to a single TPO domain irrespective of thyroid function and iodine intake. Clin. Endocrinol. **48**(6): 803-8.

Davey, C. A. and Fenna, R. E. (1996). 2.3 Å resolution X-ray crystal structure of the bisubstrate analogue inhibitor salicylhydroxamic acid bound to human myeloperoxidase. Biochemistry **35**: 10967-10973.

Davie, J. M., Paul, W. E., Mage, R. G. and Goldman, M. B. (1971). Membrane-associated immunoglobulin of rabbit peripheral blood lymphocytes: allelic exclusion at the β locus. J. Immunol. **103**: 163-169.

Davies, D. R. and Cohen, G. H. (1996). Interactions of protein antigens with antibodies. PNAS **93**: 7-12.

Davies, D. R. and Padlan, E. A. (1990). Antibody-antigen complexes. Ann. Rev. Biochem. **59**: 439-473.

de Luis, D. A., Varela, C., de La Calle, H., Canton, R., de Argila, C. M., San Roman, A. L. and Boixeda, D. (1998). Helicobacter pylori infection is markedly increased in patients with autoimmune atrophic thyroiditis. J. Clin. Gastroenterol. **26**(4): 259-63.

Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9 and 2.8 Å resolution. Biochemistry **20**: 2361-2370.

Derwahl, M., Seto, P. and Rapoport, B. (1989). Complete nucleotide sequence of the cDNA for thyroid peroxidase in FRTL5 rat thyroid cells. Nucleic Acids Res. **17**: 8380.

Dighiero, G. P., Lymberi, B., Guilbert, B., Ternynk, T. and Avrameas, S. (1986). Natural autoantibodies constitute a substantial part of normal circulating immunoglobulins. Ann. N. Y. Acad. Sci. **475**: 135-145.

Doble, N. D., Banga, J. P., Pope, R., Lalor, E., Kilduff, P. and McGregor, A. M. (1988). Autoantibodies to the thyroid microsomal/thyroid peroxidase antigen are polyclonal and directed to several distinct antigenic sites. Immunology **64**(1): 23-9.

Donath, J. and Landsteiner, K. (1904). Munch. Med. Wochenschr. **1**: 1590-1595.

Dong, Q., Ludgate, M. and Vassart, G. (1991). Cloning and sequencing of a novel 64-kDa autoantigen recognized by patients with autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **72**(6): 1375-81.

Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D. and Handford, P. A. (1996). Solution structure of a pair of calcium-binding epidermal growth factor-like domains: Implications for the Marfan syndrome and other genetic disorders. Cell **85**: 597-605.

- Drickamer, K. (1999). C-type lectin-like domains. Curr. Opin. Struct. Biol. **9**: 585-590.
- Druet, P. (1992). Diagnosis of autoimmune diseases. J. Imm. Meth. **150**: 177-184.
- Ducruix, A. and Giege, R. (1992). Crystallization of Nucleic Acids and Proteins. Oxford, IRL Press.
- Edman, P. (1970). Sequence determination. Mol. Biol. Biochem. Biophys. **1970(8)**: 211-255.
- Ehrlich, P. and Morgenroth, J. (1901). Uber Hamolysine funfte Mitteilung. Berlin klin Wochenschr **38**: 251-256.
- Ekholm, R. and Bjorkman, U. (1984). Localization of iodine binding in the thyroid gland *in vitro*. Endocrinology **115**: 1558.
- Elbein, A. D. (1991). The role of N-linked oligosaccharides in glycoprotein function. Trends Biotech. **9**: 346-352.
- Elisei, R., Vassart, G. and Ludgate, M. (1991). Demonstration of the existence of the alternatively spliced form of thyroid peroxidase in normal thyroid. J. Clin. Endocrin. Metab. **72(3)**: 700-2.
- Endo, Y., Onogi, S., Umeki, K., Yamamoto, I., Kotani, T., Ohtaki, S. and Fujita, T. (1995). Regional localization of the gene for thyroid peroxidase to human chromosome 2p25 and mouse chromosome 12C. Genomics **25(3)**: 760-1.
- Esnouf, R. (1999). Further additions to Molscript version 1.4, including reading and contouring of electron density maps. Acta Cryst. D **55**: 938-940.

Estienne, V., Blanchet, C., Niccoli-Sire, P., Duthoit, C., Durand-Gorde, J. M., Geourjon, C., Baty, D., Carayon, P. and Ruf, J. (1999a). Molecular model, calcium sensitivity, and disease specificity of a conformational thyroperoxidase B-cell epitope. *J. Biol. Chem.* **274**(50): 35313-7.

Estienne, V., Duthoit, C., Costanzo, V. D., Lejeune, P. J., Rotondi, M., Kornfeld, S., Finke, R., Lazarus, J. H., Feldt-Rasmussen, U., Franke, W. G., Smyth, P., D'Herbomez, M., Conte-Devolx, B., Persani, L., Carella, C., Jourdain, J. R., Izembart, M., Toubert, M. E., Pinchera, A., Weetman, A., et al. (1999b). Multicenter study on TGPO autoantibody prevalence in various thyroid and non-thyroid diseases; relationships with thyroglobulin and thyroperoxidase autoantibody parameters. *Eur. J. Endocrinol.* **141**(6): 563-9.

Estienne, V., Duthoit, C., Vinet, L., Durand-Gorde, J. M., Carayon, P. and Ruf, J. (1998). A conformational B-cell epitope on the C-terminal end of the extracellular part of human thyroid peroxidase. *J. Biol. Chem.* **273**(14): 8056-62.

Ewins, D. L., Rossor, M. N., Butler, J., Roques, P. K., Mullan, M. J. and McGregor, A. M. (1991). Association between autoimmune thyroid disease and familial Alzheimer's disease. *Clin. Endocrinol.* **35**(1): 93-6.

Falcone, M., Lee, J., Patstone, G., Yeung, B. and Sarvetnick, N. (1998). B lymphocytes are crucial antigen-presenting cells in the pathogenic autoimmune response to GAD65 antigen in nonobese diabetic mice. *J. Immunol.* **161**: 1163-1168.

Fan, J.-L., Patibandla, S. A., Kimura, S., Rao, T. N., Desai, R. K., Seetharamaiah, G. S., Kurosky, A. and Prabhakar, B. S. (1996). Purification and characterization of a recombinant human thyroid peroxidase expressed in insect cells. *J. Autoimmun.* **9**: 529-536.

Farid, N. R., Sampson, L., Moens, H. and Barnard, J. M. (1981). The association of goitrous autoimmune thyroiditis with HLA-DR5. Tissue Antigens **17**(3): 265-8.

Fayadat, L., Niccoli-Sire, P., Lanet, J. and Franc, J. L. (1998). Human thyroperoxidase is largely retained and rapidly degraded in the endoplasmic reticulum. Its N-glycans are required for folding and intracellular trafficking. Endocrinol. **139**(10): 4277-85.

Fayadat, L., Niccoli-Sire, P., Lanet, J. and Franc, J. L. (1999). Role of heme in intracellular trafficking of thyroperoxidase and involvement of H₂O₂ generated at the apical surface of thyroid cells in autocatalytic covalent heme binding. J. Biol. Chem. **274**(15): 10533-8.

Fenna, R., Zeng, J. and Davey, C. (1994). Structure of the green heme in myeloperoxidase. Arch. Biochem. Biophys. **316**(1): 653-656.

Fiedler, T. J. (2000). X-ray crystal structure and characterisation of halide-binding sites of human myeloperoxidase at 1.8 Å resolution. J. Biol. Chem. **275**(16): 11964-11971.

Finke, R., Seto, P. and Rapoport, B. (1990). Evidence for the highly conformational nature of the epitope(s) on human thyroid peroxidase that are recognized by sera from patients with Hashimoto's thyroiditis. J. Clin. Endocrinol. Metab. **71**(1): 53-9.

Finke, R., Seto, P., Ruf, J., Carayon, P. and Rapoport, B. (1991). Determination at the molecular level of a B-cell epitope on thyroid peroxidase likely to be associated with autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **73**(4): 919-21.

Finzel, B. C., Poulos, T. L. and Kraut, J. (1984). Crystal structure of yeast cytochrome c peroxidase refined at 1.7 Å resolution. J. Biol. Chem. **259**: 13027-13036.

Foti, A. and Rapoport, B. (1990). Carbohydrate moieties in recombinant human thyroid peroxidase: role in recognition by antithyroid peroxidase antibodies in Hashimoto's thyroiditis. Endocrinol. **126**(6): 2983-2987.

Foti, D., Kaufman, K. D., Chazenbalk, G. D. and Rapoport, B. (1990). Generation of a biologically active secreted form of human thyroid peroxidase by site-directed mutagenesis. Mol. Endocrinol.: 786-791.

Frohman, M., Francfort, J. W. and Cowing, C. (1991). T-dependent destruction of thyroid isografts exposed to IFN- γ . J. Immunol. **146**(2227-2234).

Furmaniak, J., Bradbury, J. and Rees Smith, B. (1987). Antibodies to membrane antigens in autoimmune thyroid disease. Acta. Endocrinol. **116**(1): 13-20.

Furmaniak, J., Sanders, J. and Rees Smith, B. (1999). Autoantigens in the autoimmune endocrinopathologies. Contemporary Endocrinology: Autoimmune Endocrinopathies. R. Volpe. Towata, NJ, Humana Press Inc: 183-216.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. and Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature **384**: 134-141.

Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L. and Wilson, I. A. (1996). An $\alpha\beta$ T-cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. Science **274**: 209-219.

Garcia, S., Vassko, V., Henry, J. F. and De Micco, C. (1998). Comparison of thyroid peroxidase expression with cellular proliferation in thyroid follicular tumors. Thyroid **8**(9): 745-9.

Gardas, A., Domek, H. and Czarnocka, B. (1990). The effect of dithiotreitol on thyroid peroxidase and microsomal antigen epitopes recognized by auto and monoclonal antibodies. Autoimmunity **7**(2-3): 149-56.

Gardas, A., Sohi, M. K., Sutton, B. J., McGregor, A. M. and Banga, J. P. (1997). Purification and crystallisation of the autoantigen thyroid peroxidase from human Graves' thyroid tissue. Biochem. Biophys. Res. Commun. **234**(2): 366-70.

Gardas, A., Sutton, B. J., Piotrowska, U., Pasięka, Z., Barnett, P. S., Huang, G., McGregor, A. M. and Banga, J. P. (1999). Distinct immunological and biochemical properties of thyroid peroxidase purified from human thyroid glands and recombinant protein produced in insect cells. Biochim. Biophys. Acta. **1433**(1-2): 229-39.

Garman, S. C., Wurzburg, B. A., Tarchevskaya, S. S., Kinet, J.-P. and Jardetzky, T. D. (2000). Structure of the Fc fragment of human IgE bound to its high-affinity receptor FcεRIα. Nature **406**: 259-266.

Germain, R. N. (1994). MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocytes activation. Cell **76**: 287-299.

Gernstein, H. C. (1990). How common is post-partum thyroiditis? Arch. Inter. Med. **150**: 1397-1400.

Giraud, A., Franc, J. L., Long, Y. and Ruf, J. (1992). Effects of deglycosylation of human thyroperoxidase on its enzymatic activity and immunoreactivity. J. Endocrinol. **132**(2): 317-23.

Golub, E. S. (1987). Somatic mutation: diversity and regulation of the immune response. Cell **48**: 723-724.

Gotch, F., Gallimore, A. and McMichael, A. (1996). Cytotoxic T cells - protection from disease progression - protection from infection. Immunol. Lett. **51**: 125-128.

Gowans, J. L., McGregor, D. D., Cowen, D. M. and Ford, C. E. (1962). Initiation of immune responses by small lymphocytes. Nature **196**: 651-655.

Gray, D. (1992). Immunological memory. Seminars in Immunology **4**: 1.

Grennan Jones, F., Wolstenholme, A., Fowler, S., Smith, S., Ziemnicka, K., Bradbury, J., Furmaniak, J. and Rees Smith, B. (1996). High-level expression of recombinant immunoreactive thyroid peroxidase in the High Five insect cell line. J. Mol. Endocrinol. **17**(2): 165-74.

Grennan Jones, F., Ziemnicka, K., Sanders, J., Wolstenholme, A., Fiera, R., Furmaniak, J. and Rees Smith, B. (1999). Analysis of autoantibody epitopes on human thyroid peroxidase. Autoimmunity **30**(3): 157-69.

Guo, J., McIntosh, R. S., Czarnocka, B., Weetman, A. P., Rapoport, B. and McLachlan, S. M. (1998a). Relationship between autoantibody epitopic recognition and immunoglobulin gene usage. Clin. Exp. Immunol. **111**: 408-414.

Guo, J., McLachlan, S. M., Hutchison, S. and Rapoport, B. (1998b). The greater glycan content of recombinant human thyroid peroxidase of mammalian than of insect cell origin facilitates purification to homogeneity of enzymatically protein remaining soluble at high concentration. Endocrinol. **139**(3): 999-1005.

Guo, J., Quaratino, S., Jaume, J. C., Costance, G., Londei, M., McLachlan, S. M. and Rapoport, B. (1996). Autoantibody-mediated capture and presentation of autoantigen to T cells via the Fc ϵ receptor by a recombinant human autoantibody Fab converted to IgE. J. Immun. Meth. **195**: 81-92.

Guo, J., Rapoport, B. and McLachlan, S. M. (1997). Thyroid peroxidase autoantibodies of IgE class in thyroid autoimmunity. Clin. Immunol. Immunopathol. **82**(2): 157-62.

Guo, J., Wang, Y., Jaume, J. C., Rapoport, B. and McLachlan, S. M. (1999). Rarity of autoantibodies to a major autoantigen, thyroid peroxidase, that interact with denatured antigen or with epitopes outside the immunodominant region. Clin. Exp. Immunol. **117**(1): 19-29.

Guo, J., Wang, Y., Rapoport, B. and McLachlan, S. M. (2000). Evidence for antigen presentation to sensitized T cells by thyroid peroxidase (TPO)-specific B cells in mice injected with fibroblasts co-expressing TPO and MHC class II. Clin. Exp. Immunol. **119**(1): 38-46.

Gut, P., Grennan Jones, F., Sullivan, A., Ziemnicka, K., Smith, S., Jaskolski, D., Furmaniak, J. and Rees Smith, B. (2000). Recombinant human thyroid peroxidase produced in insect cells has similar properties to native human thyroid peroxidase. Thyroid **10**(7): 543-550.

Hamada, N., Jaeduck, N., Portmann, L., Ito, K. and DeGroot, L. J. (1987). Antibodies against denatured and reduced thyroid microsomal antigen in autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **64**(2): 230-8.

Hames, B. D. and Rickwood, D. (1990). Gel Electrophoresis of Proteins. Oxford, IRL Press.

Hanafusa, T., Pujol-Borrell, R., Chiovato, L., Russell, R. C. G., Doniach, D. and Bottazzo, G. F. (1983). Aberrant expression of HLA-DR antigen on thyrocytes in Graves' disease: relevance for autoimmunity. Lancet **8359**: 1111-1115.

Harris, L. J., Larson, S. B., Hasel, K. W., Day, J., Greenwood, A. and McPherson, A. (1992). The three dimensional structure of an intact monoclonal antibody for canine lymphoma. Nature **360**: 369-372.

Harris, L. J., Larson, S. B., Hasel, K. W. and McPherson, A. (1997). Refined structure of an intact Ig2 α monoclonal antibody. Biochem. **36**: 1581-1597.

Harris, L. J., Skaletsky, E. and McPherson, A. (1998). Crystallographic structure of an intact IgG1 monoclonal antibody. J. Mol. Biol. **275**: 861-872.

Hata, J., Yamashita, S., Yagihashi, S., Kato, H., Kabeno, S., Hirai, K., Kimura, S., Umeki, K., Kotani, T. and Ohtaki, S. (1989). Stable high level expression of human thyroid peroxidase in cultured chinese hamster ovary cells. Biochem. Biophys. Res. Comm. **164**(3): 1268-1273.

Haubruck, H., Mauch, L., Cook, N. J., Steffens, U., Hunt, N., Berthold, H., Niemann, H., Wirbelauer, C. and Northemann, W. (1993). Expression of recombinant human thyroid peroxidase by the baculovirus system and its use in ELISA screening for diagnosis of autoimmune thyroid disease. Autoimmunity **15**(4): 275-84.

Henchozlecoanet, S., Jeannin, P., Aubry, J. P., Graber, P., Bradshaw, C. G., Pochon, S. and Bonnefoy, J. Y. (1996). The Epstein Barr virus-binding site on CD21 is involved in CD23 binding and interleukin-4-induced IgE and IgG4 production by human B-cells. Immunology **88**: 35-39.

Hendrick, S. M., Cohen, D. I., Neilsen, E. A. and Davies, M. M. (1984). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature **308**: 149-153.

Hexham, J. M., Patridge, L. J., Furmaniak, J., Petersen, V. B., Collis, J. C., Pegg, C., Rees Smith, B. and Burton, D. R. (1994). Cloning and characterisation of TPO autoantibodies using combinatorial phage display libraries. Autoimmunity **17**: 167-179.

Hexham, M., Pegg, C., Burton, D., Peterson, V. B., Horimoto, M., Furmaniak, J. and Rees Smith, B. (1992). Variable region sequence of a human monoclonal thyroid peroxidase autoantibody. Autoimmunity **14**: 169-172.

Hillenkamp, F., Karas, M., Beavis, R. C. and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal. Chem. **63**: 1194A-1203A.

Hirayu, H., Seto, P., Magnusson, R. P., Filetti, S. and Rapoport, B. (1987). Molecular cloning and partial characterization of a new autoimmune thyroid disease-related antigen. J. Clin. Endocrinol. Metab. **64**(3): 578-84.

Hobby, L., Gardas, A., Radomski, R., McGregor, A. M., Banga, J. P. and Sutton, B. J. (2000). Identification of an immunodominant region recognised by human autoantibodies in a three-dimensional model of thyroid peroxidase. Endocrinol. **141**(6): 2018-2026.

Hodgkin, P. D., Yamashita, L. C., Seymour, B., Coffman, R. L. and Kehry, M. R. (1991). Membranes from both Th1 and Th2 T cell clones stimulate B cell proliferation and prepare B cells lymphokine-induced differentiation to secrete Ig. J. Immunol. **147**: 3696-3702.

Horimoto, M., Petersen, V. S., Pegg, C. A., Fukuma, N., Wakabayashi, N., Kiso, Y., Furmaniak, J. and Rees Smith, B. (1992). Production and characterisation of a human monoclonal thyroid peroxidase autoantibody. Autoimmunity **14**(1): 1-7.

Hosoya, T. and Morrison, M. (1965). The peroxidase and other hemoproteins of thyroid microsomes. Biochem. Biophys. Res. Commun. **20**: 27-32.

Huang, D.-F., Olee, T., Masuho, Y., Matsumoto, Y., Carson, D. A. and Chen, P. P. (1992). Sequence analysis of three immunoglobulin G anti-virus antibodies reveal their utilization of autoantibody-related immunoglobulin Vh genes but not VI genes. J. Clin. Invest. **90**: 2197-2208.

Hunt, D. F., Michel, H., Dickenson, T. A., Shabanowitz, J., Cox, A. L., Sakaguchi, K., Appella, E., Grey, H. M. and Sette, A. (1992). Peptides presented to the immune system by the murine class II MHC molecule I-Ad. Science **256**: 1817-1920.

Igblast (2000). BLAST for immunoglobulins. <http://www.ncbi.nlm.nih.gov/igblast>.

Jaume, J. C., Burek, C. L., Hoffman, W. H., Rose, N. R., McLachlan, S. M. and Rapoport, B. (1996a). Thyroid peroxidase autoantibody epitopic 'fingerprints' in juvenile Hashimoto's thyroiditis: evidence for conservation over time and in families. Clin. Exp. Immunol. **104**(1): 115-23.

Jaume, J. C., Costante, G., Nishikawa, T., Phillips, D. I. W., Rapoport, B. and McLachlan, S. M. (1995a). Thyroid peroxidase autoantibody fingerprints in hypothyroid and euthyroid individuals. I. Cross-sectional study in elderly women. J. Clin. Endocrin. Metab. **80**: 994-9.

Jaume, J. C., Guo, J., Pauls, D. L., Zakarija, M., McKenzie, J. M., Egeland, J. A., Burek, C. L., Rose, N. R., Hoffman, W. H., Rapoport, B. and McLachlan, S. M.

(1999a). Evidence for genetic transmission of thyroid peroxidase autoantibody epitopic "fingerprints". J. Clin. Endocrinol. Metab. **84**(4): 1424-1431.

Jaume, J. C., Guo, J., Wang, Y., Rapoport, B. and McLachlan, S. M. (1999b). Cellular thyroid peroxidase (TPO), unlike purified TPO and adjuvant, induces antibodies in mice that resemble autoantibodies in human autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **84**(5): 1651-7.

Jaume, J. C., McLachlan, S. M., Burek, C. L., H., H. W., Rose, N. and Rapoport, B. (1994). Epitopic 'fingerprints' of thyroid peroxidase autoantibodies in Hashimoto's thyroiditis: evidence for conservation over time and in families. Thyroid **4**: 5.

Jaume, J. C., Parkes, A. B., Lazarus, J. H., Hall, R., Costante, G., McLachlan, S. M. and Rapoport, B. (1995b). Thyroid peroxidase autoantibody fingerprints. II. A longitudinal study in postpartum thyroiditis. J. Clin. Endocrinol. Metab. **80**(3): 1000-5.

Jaume, J. C., Portolano, S., Rapoport, B. and McLachlan, S. M. (1996b). Influence of the light chain repertoire on immunoglobulin genes encoding thyroid autoantibody Fab from combinatorial libraries. Autoimmunity **24**: 11-23.

Jenner, E. (1798). An inquiry into the causes and effects of the variola vaccinia, a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of Cowpox. London, Sompson Lew.

Jerne, N. K. (1960). Ann. Rev. Microbiol. **14**: 301-307.

Johansson, M. W., Lind, M. J., Holmblad, T., Thornqvist, P. and Soderhall, K. (1995). Peroxinectin, a novel cell adhesion protein from crayfish blood. Biochem. Biophys. Res. Comm. **216**(3): 1079-1087.

Johansson, M. W., Patarroyo, M., Oberg, F., Siegbahn, A. and Nilsson, K. (1997). Myeloperoxidase mediates cell adhesion via the $\alpha\text{M}\beta\text{2}$ integrin (Mac-1, CD11b/CD18). J. Cell. Science **110**: 1133-1139.

Jones, S. and Thornton, J. M. (1996). Principles of protein-protein interactions derived from structural studies. PNAS **93**: 13-20.

Jones, T. A., Zou, J. Y., Cowan, S. and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Cryst. **A47**(110-119).

Jukes, T. H. and King, J. L. (1979). Evolutionary nucleotide replacements in DNA. Nature **281**: 605-606.

Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**: 2577-2637.

Kallenberg, C. G. (1998). Autoantibodies to myeloperoxidase: clinical and pathophysiological significance. J. Mol. Med. **76**(10): 682-687.

Karjalainen, J., Knip, M., Mustonen, A., Ilonen, J. and Akerblom, H. K. (1986). Relationship between insulin-antibody and complement-fixing islet cell antibody at clinical diagnosis of IDDM. Diabetes **35**: 620.

Katjita, Y., Morgan, D., Parces, A. B. and Rees Smith, B. (1985). Labelling and immuno-precipitation of thyroid microsomal antigen. FEBS Lett. **187**: 334.

Kato, H. and Enjyoji, K. (1991). Amino acid sequence and location of the disulphide bonds in bovine beta 2 glycoprotein I: the presence of five Sushi domains. Biochemistry **30**: 11687-11694.

Kaufman, K. D., Filetti, S., Seto, P. and Rapoport, B. (1990). Recombinant human thyroid peroxidase generated in eukaryotic cells: a source of specific antigen for the immunological assay of antimicrosomal antibodies in the sera of patients with autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **70**(3): 724-8.

Kaufman, K. D., Foti, D., Seto, P. and Rapoport, B. (1991). Overexpression of an immunologically-intact, secreted form of human thyroid peroxidase in eukaryotic cells. Mol. Cell. Endocrinol. **78**: 107-114.

Kaufmann, K. D., Rapoport, B., Seto, P., Chazenbalk, G. D. and Magnusson, R. P. (1989). Generation of recombinant, enzymatically active human thyroid peroxidase and its recognition by antibodies in the sera of patients with Hashimoto's thyroiditis. J. Clin. Invest.: 394-403.

Kehry, M. R. and Hodgkin, P. D. (1993). Helper T cells: delivery of cell contact and lymphokine-dependent signals to B cells. Semin. Immunol. **6**: 393-400.

Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. <http://www.bmm.icnet.uk/servers/3dpssm>. J. Mol. Biol. **299**: 501-522.

Kendler, D. L., Brennan, V., Davies, T. F. and Magnusson, R. P. (1993). Expression of human thyroid peroxidase in insect cells using recombinant baculovirus. Mol. Cell. Endocrinol. **93**: 199-206.

Kettleborough, C. A., Saldanha, J., Ansell, K. H. and Bendig, M. M. (1993). Optimization of primers for cloning libraries of mouse immunoglobulin genes using the polymerase chain reaction. Eur. J. Immunol. **23**: 206-211.

Khoury, E. L., Hammond, L., Bottazzo, G. F. and Doniach, D. (1981). Presence of organ-specific 'microsomal' autoantigen on the surface of human thyroid cells in culture: its involvement in complement-mediated cytotoxicity. Clin. Exp. Immunol. **45**: 315-328.

Kikkawa, F., Gonzalez, F. J. and Kimura, S. (1990). Characterization of a thyroid-specific enhancer located 5.5 kilobase pairs upstream of the human thyroid peroxidase gene. Mol. Cell Biol. **10**: 6216-6224.

Kimura, S., Hong, Y. S., Kotani, T., Ohtaki, S. and Kikkawa, F. (1989). Structure of the human thyroid peroxidase gene: comparison and relationship to the human myeloperoxidase gene. Biochemistry **28**(10): 4481-9.

Kimura, S. and Ikeda-Saito, M. (1988). Human myeloperoxidase and thyroid peroxidase, two enzymes with separate and distinct physiological functions, are evolutionarily related members of the same gene family. Proteins **3**(2): 113-20.

Kimura, S., Kotani, T., McBride, O. W., Umeki, K., Hirai, K., Nakayama, T. and Ohtaki, S. (1987). Human thyroid peroxidase: complete cDNA and protein sequence, chromosome mapping, and identification of two alternately spliced mRNAs. PNAS **84**(16): 5555-9.

Kiso, Y., Furmaniak, J., Morteo, C. and Smith, B. R. (1992). Analysis of carbohydrate residues on human thyroid peroxidase (TPO) and thyroglobulin (Tg) and effects of deglycosylation, reduction and unfolding on autoantibody binding. Autoimmunity **12**(4): 259-69.

Klein, R. and Zachau, H. G. (1995). Expression and hypermutation of human immunoglobulin κ genes. Ann. N. Y. Acad. Sci **764**: 74-83.

Klickstein, L. B., Barlow, T. J., Miletic, V., Rabson, L. D., Smith, J. A. and Fearon, D. T. (1988). Identification of distinct C3b and C4b recognition sites in the C3b/C4b receptor. J. Exp. Med. **168**: 1699-1717.

Knott, V., Downing, A. K., Cardy, C. M. and Handford, P. (1996). Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1. J. Mol. Biol. **255**: 22-27.

Kohno, Y., Naito, N., Hiyama, Y., Shimojo, N., Suzuki, N., Tarutani, O., Niimi, H., Nakajima, H. and Hosoya, T. (1988). Thyroglobulin and thyroid peroxidase share common epitopes recognized by autoantibodies in patients with chronic autoimmune thyroiditis. J. Clin. Endocrinol. Metab. **67**(5): 899-907.

Kohno, Y., Yamaguchi, F., Saito, K., Niimi, H., Nishikawa, T. and Hosoya, T. (1991). Anti-thyroid peroxidase antibodies in sera from healthy subjects and from patients with chronic thyroiditis: differences in the ability to inhibit thyroid peroxidase activities. Clin Exp Immunol **85**(3): 459-63.

Kootstra, P. R., Wever, R. and de Vijlder, J. J. M. (1993). Thyroid peroxidase: kinetics, pH optima and substrate dependency. Acta Endocrinol. **129**: 328-331.

Kotani, T., Umeki, K., Matsunga, S., Kato, E. and Ohtaki, S. (1986). Detection of autoantibodies to the thyroid peroxidase in autoimmune thyroid diseases by micro-ELISA and immunoblotting. J. Clin. Endo. Met. **62**: 928.

Kotani, T., Umeki, K., Yamamoto, I., Takeuchi, M., Takechi, S., Nakayama, T. and Ohtaki, S. (1993). Nucleotide sequence of the cDNA encoding mouse thyroid peroxidase. Gene **123**(2): 289-90.

Kovari, L. C., Momany, C. and Rossman, M. G. (1995). The use of antibody fragments for crystallization and structure determinations. Structure **3**: 1291-1293.

Kraulis, P. J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. J. App. Cryst. **24**: 946-950.

Kulberg, A. I. (1972). Effector functions of immunoglobulins, molecular mechanism. Allerg. Immunol. **18**(4): 217-240.

Kurosawa, Y. and Tonegawa, S. (1982). Organization of human immunoglobulin heavy chain diversity DNA segments. J. Exp. Med. **115**: 201-218.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685.

Lascombe, M. B., Alzari, P. M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A. and Poljak, R. J. (1989). Three-dimensional structure of Fab R19.9, a monoclonal murine antibody specific for the p-azobenzene-arsenate group. PNAS **2**(607-611).

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. **26**: 283-291.

Laver, W. G., Air, G. M., Webster, R., G. and Smith-Gill, S. J. (1990). Epitopes on protein antigens: misconceptions and realities. Cell **61**: 553-556.

Lee, B. and Richards, F. M. (1971). The interpretation protein structures: estimation of static accessibility. J. Mol. Biol. **55**: 379-400.

Lee, K. M., Chung, E., Griffen, M., Khattri, R., Hong, D. K., Zhang, W., Straus, D., Samelson, L. E., Thompon, C. B. and Bluestone, J. A. (1998). Molecular basis of T cell activation by CTLA-4. Science **282**: 2263-2266.

Lewis, S. M. (1994). The mechanism of V(D)J joining: lessons from molecular immunological, and comparative analyses. Adv. Immunol. **56**: 27-150.

Libert, F., Ludgate, M., Dinsart, C. and Vassart, G. (1991). Thyroperoxidase, but not the thyrotropin receptor, contains sequential epitopes recognized by autoantibodies in recombinant peptides expressed in the pUEX vector. J. Clin. Endocrinol. Metab. **73**(4): 857-60.

Libert, F., Ruel, J., Ludgate, M., Swillens, S., Alexander, N., Vassart, G. and Dinsart, C. (1987a). Complete nucleotide sequence of the human thyroperoxidase-microsomal antigen cDNA. Nucleic Acids Res. **15**(16): 6735.

Libert, F., Ruel, J., Ludgate, M., Swillens, S., Alexander, N., Vassart, G. and Dinsart, C. (1987b). Thyroperoxidase, an auto-antigen with a mosaic structure made of nuclear and mitochondrial gene modules. EMBO J. **6**(13): 4193-6.

Logtenberg, T. (1990). Properties of polyreactive natural antibodies to self and foreign antigens. J. Clin. Immun. **10**: 137-140.

MacCallum, R. M., Martin, A. C. R. and Thornton, J. M. (1996). Antibody-antigen interactions: contact analysis and binding site topography. J. Mol. Biol. **262**: 732-745.

Mackay, I. R. and Gershwin, M. E. (1989). Molecular basis of mitochondrial autoreactivity in primary biliary cirrhosis. Immun. Today **10**(9): 315-318.

Maenaka, K. and Jones, E. Y. (1999). MHC superfamily structure and the immune system. Curr. Opin. Struct. Biol. **6**: 745-753.

Magnusson, R. P., Chazenbalk, G. D., Gestautas, J., Seto, P., Filetti, S., DeGroot, L. J. and Rapoport, B. (1987). Molecular cloning of the complementary deoxyribonucleic acid for human thyroid peroxidase. Mol. Endocrinol. **1**(11): 856-61.

Magnusson, R. P., Gestautas, J., Seto, P., Taurog, A. and Rapoport, B. (1986). Isolation and characterisation of a cDNA clone for porcine thyroid peroxidase. FEBS Lett. **208**: 391-396.

Malby, R. L., Tulip, W. R., Harley, V. R., McKimm-Breschkin, J. L., Laver, W. G., Webster, R. G. and Colman, P. M. (1994). The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. Structure **2**(8): 733-46.

Malthiery, Y. and Lissitzky, S. (1987). Primary structure of human thyroglobulin: location and surroundings of two thyroxine-forming sites. Eur. J. Biochem. **165**: 491.

Mamula, M. J. (1993). The inability to process a self-peptide allows autoreactive T cells to escape tolerance. J. Exp. Med. **177**(2): 567-71.

Marion, T. N., Tillman, D. M., Jou, N. T. and Hill, R. J. (1992). Selection of immunoglobulin variable regions in autoimmunity to DNA. Immunol. Rev. **128**: 123-149.

Mariotti, S., Caturegli, P., Piccolo, P., Barbesino, G. and Pinchera, A. (1990). Antithyroid peroxidase autoantibodies in thyroid diseases. J. Clin. Endocrinol. Metab. **71**(3): 661-9.

Marquart, M., Deisenhofer, J., Huber, R. and Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. **141**: 369-391.

Martin, A. (1998). Antibodies - structure and sequence. <http://www.biochem.ucl.ac.uk/~martin/abs/chothia.html>.

Martin, A., Barbesino, G. and Davies, T. F. (1999). T-cell receptors and autoimmune thyroid disease-signposts for T-cell- antigen driven diseases. Int. Rev. Immunol. **18**(1-2): 111-40.

Martin, A. C. R. (1996). Accessing the kabat antibody sequence database by computer. Proteins **25**: 130-133.

Matechak, E. O., Killeen, N., Hedrick, S. M. and Fowlkes, B. J. (1996). MHC class II-specific T cells can develop in the CD8 lineage when CD4 is absent. Immunity **4**: 337-3347.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. **33**(2): 491-497.

Max, E. E., Maizel, J. V. and Leder, P. (1981). The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse kappa immunoglobulin J and C region genes. J. Biol. Chem. **256**: 511-5120.

McIntosh, R., Watson, P. and Weetman, A. (1998). Somatic hypermutation in autoimmune thyroid disease. Immunol. Rev. **162**: 219-31.

McIntosh, R. S., Aghar, M. S., Kemp, E. H., Watson, P. F., Gardas, A., Banga, J. P. and Weetman, A. P. (1997). Analysis of immunoglobulin G κ antithyroid peroxidase

antibodies from different tissues in Hashimoto's thyroiditis. J.Clin. Endocrinol. Metab. **82**(11): 2818-3825.

McIntosh, R. S. and Weetman, A. P. (1997). Molecular analysis of the antibody response to thyroglobulin and thyroid peroxidase. Thyroid **7**(3): 471-87.

McLachlan, S. M., Atherton, M. C., Nakajima, Y., Napier, J., Jordan, R. K., Clark, F. and Rees Smith, B. (1990). Thyroid peroxidase and the induction of autoimmune thyroid disease. Clin. Exp. Immunol. **79**(2): 182-8.

McLachlan, S. M. and Rapoport, B. (1995). Genetic and epitopic analysis of thyroid peroxidase (TPO) autoantibodies: markers of the human thyroid autoimmune response. Clin. Exp. Immunol. **101**(2): 200-6.

Metcalf, R. A., McIntosh, R. S., Morgan, B. P., Levin, J. L. and Weetman, A. P. (1996). The effect of soluble complement receptor 1 (sCR1) and human thyroid antibodies on the course of experimental autoimmune thyroiditis in rats. Autoimmunity **23**: 1-8.

Metchnikoff, E. (1893). Lectures on the Comparative Pathology of Inflammation. London, Kegan, Paul, Trench, Trubner.

Moguilevsky, N., Garcia-Quintana, I., Jaquet, A., Tournay, C., Fabray, L., Pierard, L. and Bollen, A. (1991). Structural and biological properties of human recombinant myeloperoxidase produced by Chinese hamster ovary cell lines. Eur. J. Biochem. **197**: 605-614.

Monks, C. R., Freiburg, B. A., Kupfer, H., Sciaky, N. and Kupfer, A. (1998). Three-dimensional segregation of supramolecular activation clusters in T cells. Nature **395**: 82-86.

Morea, V., Leplae, R. and Tramontano, A. (1998a). Protein structure prediction and design. Biotech. Ann. Rev. **4**: 177-214.

Morea, V., Tramontano, A., Rustici, M., Chothia, C. and Lesk, A. M. (1998b). Conformations of the third hypervariable region in the VH domain of immunoglobulins. J. Mol. Biol. **275**: 269-294.

Morishita, K., Tsuchiya, M., Asano, S., Kaziro, Y. and Nagata, S. (1987). Chromosomal gene structure of human myeloperoxidase and regulation of its expression by granulocyte colony-stimulating factor. J. Biol. Chem. **262**: 15208-15213.

Mosman, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. and Coffman, R. L. (1986). Two types of murine helper T cell clones. I. Definition according to profiles of lymphokine activities and secreted proteins. J. Immunol. **136**: 2348-2357.

Moss, P. A. H., Rosenberg, W. M. C. and Bell, J. I. (1992). The human T-cell receptor in health and disease. Ann. Rev. Immunol. **10**: 71.

Nagayama, Y. (1995). Continuous versus discontinuous B-cell epitopes on thyroid-specific autoantigens - thyrotropin receptor and thyroid peroxidase. Eur. J. Endocrinol.(132): 9-11.

Nakagwa, H., Kotani, T., Ohtaki, S., Nakamura, M. and Yamazaki, I. (1985). Purification of thyroid peroxidase by monoclonal antibody-assisted immunoaffinity chromatography. Biochem. Biophys. Res. Comm. **127**(1): 8-14.

Nakashima, T. and Taurog, A. (1978). Improved assay procedures for thyroid peroxidase: application to normal and adenomatous human thyroid tissue. Clin. Chem. Acta. **83**: 129-140.

Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. Acta Cryst. **50**: 157-163.

NCBI (2000). National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov>.

Nelson, R. E., Fessler, L. I., Takagi, Y., Blumberg, B., Keene, D. R., Olson, P. F., Parker, C. G. and Fessler, J. H. (1994). Peroxidasin: a novel enzyme-matrix protein of *Drosophila* development. EMBO J. **13**: 3438-3447.

Niccoli, P., Fayadat, L., Panneels, V., Lanet, J. and Franc, J.-L. (1997). Human thyroperoxidase in its alternatively spliced form (TPO2) is enzymatically inactive and exhibits changes in intracellular processing and trafficking. J. Biol. Chem. **272**(47): 29487-29492.

Nicholls, A., Sharp, K. A. and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11**: 281-296.

Nishikawa, T., Costante, G., Prummel, M. F., McLachlan, S. M. and Rapoport, B. (1994a). Recombinant thyroid peroxidase autoantibodies can be used for epitopic fingerprinting of thyroid peroxidase autoantibodies in the sera of individual patient. J. Clin. Endocrinol. Metab. **78**(4): 944-949.

Nishikawa, T., Jaume, J. C., McLachlan, S. M. and Rapoport, B. (1995). Human monoclonal autoantibodies against the immunodominant region on thyroid peroxidase: lack of cross-reactivity with related peroxidases or thyroglobulin and inability to inhibit thyroid peroxidase enzymatic activity. J. Clin. Endocrinol. Metab. **80**(4): 1461-6.

Nishikawa, T., Nagayama, Y., Seto, P. and Rapoport, B. (1993). Human thyroid peroxidase-myeloperoxidase chimeric molecules: tools for the study of antigen recognition by thyroid peroxidase autoantibodies. Endocrinology **133**(6): 2496-501.

Nishikawa, T., Rapoport, B. and McLachlan, S. M. (1994b). Exclusion of two major areas on thyroid peroxidase from the immunodominant region containing the conformational epitopes recognized by human autoantibodies. J. Clin. Endocrinol. Metab. **79**(6): 1648-54.

Nishikawa, T., Rapoport, B. and McLachlan, S. M. (1996). The quest for the autoantibody immunodominant region on thyroid peroxidase: guided mutagenesis based on a hypothetical three- dimensional model. Endocrinology **137**(3): 1000-6.

Nossal, G. J. (1994). Negative selection of lymphocytes. Cell **76**: 229-239.

Ogasawara, M., Di Lauro, R. and Satoh, N. (1999). Ascidian homologs of mammalian thyroid peroxidase genes are expressed in the thyroid-equivalent region of the endostyle. J. Exp. Zool. **285**(2): 158-69.

Ohtaki, S., Kotani, T. and Nakamura, Y. (1986a). Characterization and isolation of thyroid microsomal antigen. J. Clin. Invest. **79**: 819.

Ohtaki, S., Kotani, T. and Nakamura, Y. (1986b). Characterization of human thyroid peroxidase purified by monoclonal antibody-assisted chromatography. J. Clin. Endocrinol. Metab. **63**(3): 570-576.

Ohtaki, S., Nakagawa, H., Kimura, S. and Yamazaki, I. (1981). Analysis of catalytic intermediates of hog thyroid peroxidase during its iodination reaction. J. Biol. Chem. **256**(805-810).

Okamoto, Y., Hamada, N., Saito, H., Ohno, M., Noh, J., Ito, K. and Morii, H. (1989). Thyroid peroxidase activity-inhibiting immunoglobulins in patients with autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **68**(4): 730-4.

Ollier, W. and Symmons, D. P. M. (1992). Autoimmunity, Bios Scientific Publishers.

Olsen, R. L. and Little, C. (1983). Purification and some properties of myeloperoxidase and eosinophil peroxidase from human blood. Biochem. J. **209**: 781-787.

Otwinowski, Z. and Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. C. W. Carter and R. M. Sweet. London, Academic Press. **276**.

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**: 112-124.

Padlan, E. A. (1994). Anatomy of the antibody molecule. Mol. Immunol. **31**(3): 169-217.

Padlan, E. A., Abergel, C. and Tipper, J. P. (1995). Identification of specificity-determining residues in antibodies. FASEB J. **9**: 133-139.

Paolieri, F., Salmaso, C., Battifora, M., Montagna, P., Pesce, G., Bagnasco, M., Richiusa, P., Galluzzo, A. and Giordano, C. (1999). Possible pathogenetic relevance of interleukin-1 beta in "destructive" organ-specific autoimmune disease (Hashimoto's thyroiditis). Ann. N. Y. Acad. Sci. **876**: 221-8.

Parker, D. C. (1993). T cell-dependent B cell activation. Ann. Rev. Immunol. **11**: 331.

Parry-Smith, D. J., Payne, A. W., Mitchie, A. D. and Attwood, T. K. (1998). CINEMA - a novel colour interactive editor for multiple alignments. Gene **221**(1): 57-63.

Paschke, R., Vogg, M., Kristoferitisch, R., Aktuna, D., Wawschinek, O., Eber, O. and Usadel, K. H. (1995). Methimazole has no dose-related effect on the intensity of the intrathyroidal autoimmune process in relapsing Graves' disease. J. Clin. Endocrinol. Metab. **80**(8): 2470-2474.

Phillips, D., Prentice, L., Upadhyaya, M., Lunt, P., Chamberlain, S., Roberts, D. F., McLachlan, S. and Smith, B. R. (1991). Autosomal dominant inheritance of autoantibodies to thyroid peroxidase and thyroglobulin--studies in families not selected for autoimmune thyroid disease. J. Clin. Endocrinol. Metab. **72**(5): 973-5.

Phillips, D. I., Lazarus, J. H. and Butland, B. K. (1990). The influence of pregnancy and reproductive span on the occurrence of autoimmune thyroiditis. Clin. Endocrinol. **32**(3): 301-6.

Phillips, D. I. W., Cooper, C., Fall, C., Prentice, L., Osmond, C., Barker, D. J. P. and Rees Smith, B. (1993a). Fetal growth and autoimmune thyroid disease. Q. J. Med. **86**: 247-253.

Phillips, D. I. W., Shields, D. C., Dugoujon, J. M., Prentice, L., McGuffin, P. and Rees Smith, B. (1993b). Complex segregation analysis of thyroid autoantibodies: are they inherited as an autosomal dominant trait? Human Hered. **43**: 141-146.

Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. and Saul, F. (1973). Three dimensional structure of the Fab' structure of human immunoglobulins. Proc. Natl. Acad. Sci. **70**: 3305-3310.

Pommier, J., Deme, D., Fimiani, E., Gavaret, J. M. and Nunez, J. (1976). Mechanism of thyroid peroxidase catalyzed thyroid hormone synthesis. Thyroid research. J. Robbins. Amsterdam, Excerpta Medica.

Portmann, L., Hamada, N., Heinrich, G. and DeGroot, L. J. (1985). Anti-thyroid peroxidase antibody in patients with autoimmune thyroid disease: possible identity with anti-microsomal antibody. J. Clin. Endocrinol. Metab. **61**(5): 1001-3.

Portolano, S., Chazenbalk, G. D., Hutchinson, J. S., McLachlan, S. M. and Rapoport, B. (1993a). Lack of promiscuity in autoantigen-specific H and L chain combinations as revealed by human H and L chain 'roulette'. J. Immunol. **150**: 880-887.

Portolano, S., McLachlan, S. M. and Rapoport, B. (1993b). High affinity, thyroid specific human autoantibodies displayed on the surface of filamentous phage use V genes similar to other autoantibodies. J. Immunol. **151**: 2839-2851.

Portolano, S., Prummel, M. F., Rapoport, B. and McLachlan, S. M. (1995). Molecular cloning and characterization of human thyroid peroxidase autoantibodies of lambda light chain type. Mol. Immunol. **32**(14-15): 1157-69.

Poulos, T. L. (1993). Peroxidases. Curr. Opin. Biotech. **4**: 484-489.

Premawardhana, L., Kiso, Y., Phillips, D. I., Morteo, C., Furnamiak, J. and Rees Smith, B. (1992). Is TPO detectable in the circulation? Thyroid **3**: 225-228.

Prentice, L., Kiso, Y., Furmaniak, J., Horimoto, M., Petersen, V., Grennan Jones, F. and Rees Smith, B. (1995). Monoclonal thyroglobulin antibodies: variable region analysis and epitope recognition. J. Clin. Endocrinol. Metab. **80**: 977-986.

Prentice, L., Phillips, D. I. W., Premawardhana, L. D. K. E. and Rees Smith, S. (1993). Genetic linkage analysis of thyroid autoantibodies. Autoimmunity **15**: 225-229.

Prentice, L. M., Phillips, D. I., Sarsero, D., Beever, K., McLachlan, S. M. and Smith, B. R. (1990). Geographical distribution of subclinical autoimmune thyroid disease in Britain: a study using highly sensitive direct assays for autoantibodies to thyroglobulin and thyroid peroxidase. Acta Endocrinol. **123**(5): 493-8.

Rabinowe, S. L., George, K. L., Loughlin, R., Soeldner, J. S. and Eisenbarth, G. S. (1986). Congenital rubella. Monoclonal antibody-defined T cell abnormalities in young adults. Am. J. Med. **81**: 779-782.

Radosavljevic, V. R., Jankovic, S. M. and Marinkovic, J. M. (1996). Stressful life events in the pathogenesis of Graves' disease. Eur. J. Endocrin. **134**: 699-701.

Ralff, C., Apriletti, J. W., Wagner, R. L., Feng, W., Kushner, P. J., Nilsson, S., Scanlan, T. S., West, B. L., Fletterick, R. J. and Baxter, J. D. (1998). X-ray crystallographic and functional studies of thyroid hormone receptor. J. Steroid. Biochem. Mol. Biol. **65**: 133-141.

Rapoport, B., Chazenbalk, G. D., Jaume, J. C. and McLachlan, S. M. (1998). The thyrotropin (TSH) receptor: interaction with TSH and autoantibodies. Endocrin. Rev. **19**(6): 673-716.

Rapoport, B., Portolano, S. and McLachlan, S. M. (1995). Combinatorial libraries: new insights into human organ-specific autoantibodies. Immunol. Today **16**(1): 43-49.

Rawitch, A. B., Pollock, G., Yang, S. X. and Taurog, A. (1992). Thyroid peroxidase glycosylation: the location and nature of the N- linked oligosaccharide units in porcine thyroid peroxidase. Arch. Biochem. Biophys. **297**(2): 321-7.

Rees Smith, B. and Hall, R. (1974). Thyroid stimulating immunoglobulins in Graves' disease. Lancet **2**: 427-431.

Rees Smith, B., McLachlan, S. M. and Furmaniak, J. (1988). Autoantibodies to the thyrotropin receptor. Endocr. Rev. **9**: 106-121.

Reinhertz, E. L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y., Hussey, R. E., Smolyar, A., Hare, B., Zhang, R., Joachimiak, A., Chang, H. C., Wagner, G. and Wang, J. (1999). The crystal structure of a T cell receptor in complex with peptide and MHC class II. Science **286**: 1913-1912.

Rini, J. M., Schulze-Gahmen, U. and Wilson, I. A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science **255**(959-965).

Roberts, J., Jenkins, C., Wilson, R., Pearson, C., Franklin, I. A. and Maclean, M. A. (1996). Recurrent miscarriage is associated with increased numbers of CD5/20 positive lymphocytes and an increased incidence of thyroid antibodies. Eur. J. Endocrinol. **134**: 84-86.

Roitt, I. (1994). Essential Immunology. Oxford, UK, Blackwell Science.

Roitt, I., Ling, N. R., Doniach, D. and Couchman, K. G. (1964). The cytoplasmic autoantigen of the human thyroid. I. Immunological and biochemical characteristics. Immunology **7**: 375.

Roitt, I. M., Doniach, D., Campbell, P. N. and Hudson, R. V. (1956). Autoantibodies in Hashimoto's disease. Lancet **2**: 820-821.

Rose, N. R. and Witebsky, E. (1956). Studies on organ specificity: V changes in the thyroid glands of rabbits following acute immunization with rabbit thyroid extracts. J. Immunol. **76**: 417-428.

Rossmann, M. G. and Blow, D. M. (1962). Acta Cryst. **15**: 24-31.

Rudd, P. M., Wormald, M. R., Stanfield, R. L., Huang, M., Mattsson, N., Speir, J. A., DiGennaro, J. A., Fetrow, J. S., Dwek, R. A. and Wilson, I. A. (1999). Roles for glycosylation of cell surface receptors involved in cellular immune recognition. J. Mol. Biol. **293**: 351-366.

Ruf, J., Feldt-Rasmussen, U., Hegedus, L., Ferrand, M. and Carayon, P. (1994). Bispecific thyroglobulin and thyroperoxidase autoantibodies in patients with various thyroid and autoimmune diseases. J. Clin. Endocrinol. Metab. **79**(5): 1404-9.

Ruf, J., Toubert, M. E., Czarnocka, B., Durand-Gorde, J. M., Ferrand, M. and Carayon, P. (1989). Relationship between immunological structure and biochemical properties of human thyroid peroxidase. Endocrinology **125**(3): 1211-8.

Ruoslahti, E. (1987). New perspectives in cell adhesion: RGD and integrins. Science **238**: 491-497.

Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature **286**: 676-683.

Salvi, M., Fukazawa, H., Bernard, N., Hiromatsu, Y., How, J. and Wall, J. R. (1988). Role of autoantibodies in the pathogenesis and association of endocrine autoimmune disorders. Endocrine Rev. **9**(4): 450-466.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci **74**: 5463-5467.

Sanz, I., Dang, H., Takei, M., Talal, N. and Capra, J. D. (1989). VH Sequence of a human anti-Sm autoantibody. J. Immunol. **142**(3): 883-887.

Sarvetnick, N., Shizuru, J., Liggitt, D., Martin, L., McIntyre, B., Gregory, A., Parslow, T. and Stewart, T. (1990). Loss of pancreatic islet tolerance induced by beta-cell expression of interferon-gamma. Nature **346**: 844-847.

Scatchard, G. (1949). The attraction of proteins for small molecules and ions. Ann. N. Y. Acad. Sci **51**: 660-672.

Schardt, C. W., McLaughlan, S. M., Matheson, J. and Rees Smith, B. (1982). An enzyme linked immunoassay for thyroid microsomal antibodies. J. Immunol. Met. **55**: 155-68.

Schwartz, M. and Cohen, I. R. (2000). Autoimmunity can benefit self-maintenance. Immunol. Today **21**(6): 265-268.

Segrest, J. P., Jackson, R. L., Andrew, E. P. and Marchesi, V. T. (1971). Human erythrocyte membrane glycoprotein: a re-evaluation of the molecular weight as

determined by SDS polyacrylamide gel electrophoresis. Biochem. Biophys. Res. Comm. **44**(2): 390-395.

Seto, P., Hirayu, H., Magnusson, R. P., Gestautas, J., Portmann, L., DeGroot, L. J. and Rapoport, B. (1987). Isolation of a complementary DNA clone for thyroid microsomal antigen. Homology with the gene for thyroid peroxidase. J. Clin. Invest. **80**(4): 1205-8.

Seto, P., Nagayama, Y., Foti, D., McLachlan, S. M. and Rapoport, B. (1993). Autoantibodies in the sera of patients with autoimmune thyroid disease recognize a secreted form of human thyroid peroxidase generated in a baculovirus system. Mol. Cell. Endocrinol. **94**(1): R5-8.

Simitsek, P. D., Campbell, D. G., Lanzavecchia, A., Fairweather, N. and Watts, C. (1995). Modulation of antigen processing by bound antibodies can boost or suppress class II major histocompatibility complex presentation of different T cell epitopes. J. Exp. Med. **181**(6): 1957-1963.

Sinha, A. A. (1990). Autoimmune diseases: the failure of self tolerance. Science **248**: 1380-1388.

Smyth, P. P. A., Shering, S. G., Kilbane, M. T., Murray, M. J., McDermott, R. W. M., Smith, D. F. and O'Higgins, N. J. (1998). Serum thyroid peroxidase autoantibodies, thyroid volume, and outcome in breast carcinoma. J. Clin. Endocrinol. Metab. **83**(8): 2711-2716.

Sondermann, P., Huber, R., Osthuizen, V. and Jacob, U. (2000). The 3.2 Å crystal structure of the human IgG1 Fc fragment-FcεRIII complex. Nature **406**: 267-273.

Song, Y., Li, Y. and Maclaren, N. K. (1996). The nature of autoantigens targeted in autoimmune endocrine diseases. Immunol. Today **17**(5): 232-238.

Spitzweg, C. H., A. E. (1997). Update on the thyroid sodium iodide symporter: a novel thyroid antigen emerging on the horizon. Eur. J. Endocrinol. **137**: 22-23.

Stenszky, V., Kozma, L., Balazs, C., Rochlitz, S., Bear, J. C. and Farid, N. R. (1985). The genetics of Graves' disease: HLA and disease susceptibility. J. Clin. Endo. Met. **61**: 735-740.

Sutton, B. J., Corper, A. L., Sohi, M. K., Jefferis, R., Beale, D. and Taussig, M. J. (1998). The structure of a human rheumatoid factor bound to IgG Fc. Glycoimmunology 2. Axford. New York, Plenum Press.

Sutton, B. J. and Phillips, D. C. (1983). The three-dimensional structure of the carbohydrate within the Fc fragment of the immunoglobulin G. Biochem. Soc. Trans. **11**: 130-132.

SWISS-PROT (2000). Annotated protein sequence database. <http://www.ebi.ac.uk/swissprot/>.

Taurog, A. (1999). Molecular evolution of thyroid peroxidase. Biochimie **81**(5): 557-62.

Taurog, A., Dorris, M. L. and Doerge, D. R. (1996). Mechanism of simultaneous iodination and coupling catalyzed by thyroid peroxidase. Arch. Biochem. Biophys. **330**(1): 24-32.

Taurog, A., Dorris, M. L., Yokoyama, N. and Slaughter, C. (1990). Purification and characterisation of a large, tryptic fragment of human thyroid peroxidase with high catalytic activity. Arch. Biochem. Biophys. **278**(2): 333-341.

Taurog, A., Lothrop, M. L. and Esterbrook, R. W. (1970). Improvements in the isolation procedure for thyroid peroxidase: Nature of the heme prosthetic group. Arch. Biochem. Biophys. **139**: 221-229.

Taurog, A. and Wall, M. (1998). Proximal and distal histidines in thyroid peroxidase: relation to the alternatively spliced form, TPO-2. Thyroid **8**(2): 185-91.

Terness, P. and Opelz, G. (1998). Natural anti-immunoglobulin autoantibodies: irrelevant by products of immunoregulatory molecules? Int. Arch. Allergy Immunol. **115**: 270-277.

Theofilopoulos, A. N. (1995a). The basis of autoimmunity: Part I. Mechanisms of aberrant self-recognition. Immunol. Today **16**(2): 90-98.

Theofilopoulos, A. N. (1995b). The basis of autoimmunity: Part II. Genetic predisposition. Immunol. Today **16**(3): 150-158.

Thiebe, R. (1999). The variable genes and gene families of the mouse immunoglobulin κ locus. Euro. J. Immunol. **29**: 2072-2081.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acid Res. **22**: 4673-4680.

Todd, J. A., Bell, J. I. and McDevitt, H. O. (1987). HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature **329**: 599-604.

Tomarev, S. I., Zinovieva, R. D., Weis, V. M., Chepelinsky, A. B., Piatigorsky, J. and McFall-Ngai, M. J. (1993). Abundant mRNAs in the squid light organ encode proteins with a high similarity to mammalian peroxidases. Gene **132**: 219-226.

Tomer, Y. and Davies, T. F. (1993). Infection, thyroid disease, and autoimmunity. Endocrine Rev. **14**(1): 107-120.

Tonegawa, S., Hozumi, N., Matthysens, G. and Schuller, R. (1977). Somatic changes in the content and context of immunoglobulin genes. Cold Spring Harb. Symp. Quant. Biol. **41**: 877-889.

Tormo, J., Stadler, E., Skern, T., Auer, H., Kanzler, O., Betzel, C., Blaas, D. and Fita, I. (1992). Three-dimensional structure of the Fab fragment of a neutralizing antibody to human rhinovirus serotype 2. Prot. Sci. **9**: 1154-1161.

Townsend, A. and Bodmer, H. (1989). Antigen recognition by class I restricted T cells. Ann. Rev. Immunol. **7**: 601-624.

Trotter, W. R., Belyavin, G. and Wadhams, A. (1957). Precipitating and complement fixing antibodies in Hashimoto's disease. Proc. R. Soc. Med. **50**: 961.

Turcotte, M. (2000). Helical wheel JAVA applet. <http://www.sote.uottowa.ca/~turcotte/resources/HelicalWheel>.

Tzartos, S. J., Barkas, T., Cung, M. T., Mamalaki, A., Marraud, M., Orlewski, P., Papanastasiou, D., Sakarellos, C., Sakarellos-Daitsiotis, M., Tsantili, P. and Tsikaris, V. (1998). Anatomy of the antigenic structure of a large membrane autoantigen, the muscle-type nicotinic acetylcholine receptor. Immunol. Rev. **163**: 89-120.

Volpe, R. (1988). The immunoregulatory disturbance in autoimmune thyroid disease. Autoimmunity **2**(1): 55-72.

Volpe, R. (1993). Suppressor T lymphocyte dysfunction is important in the pathogenesis of autoimmune thyroid disease: a perspective. Thyroid **3**(4): 345-52.

von Boehmer, H. (1994). Positive selection of lymphocytes. Cell **76**: 219-228.

Wallis, M. (1973). The pKa values of the α -amino groups of peptides derived from the N-terminus of bovine growth hormone. Biochem. Biophys. Acta. **310**: 388-397.

Wang, D., Liao, J., Mitra, D., Akolkar, P. N., Gruezo, F. and Kabat, E. A. (1991). The repertoire of antibodies to a single antigenic determinant. Mol. Immunol. **28**(12): 1387-1397.

Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G. and Stevens, R. C. (1997). Structural insights into the evolution of an antibody combining site. Science **276**: 1665-1669.

Wedlock, N., Furnmaniak, J., Fowler, S., Kiso, Y., Bednarek, J., Baumann-Antczak, A., Morteo, C., Sudbery, P., Hinchcliff, A. and Rees Smith, B. (1993). Expression of human thyroid peroxidase in the yeasts *Saccharomyces cerevisiae* and *Hansenula polymorpha*. J. Mol. Endocrinol. **10**: 325-336.

Weetman, A. P. (1991). Autoimmune Endocrine Disease. Cambridge, Cambridge University Press.

Weetman, A. P. (1992). Autoimmune thyroiditis: predisposition and pathogenesis. Clin. Endocrinol. **36**(4): 307-23.

Weetman, A. P. (1994). The potential immunological role of the thyroid cell in autoimmune thyroid disease. Thyroid **4**(4): 493-9.

Weetman, A. P., Black, C. M., Cohen, S. B., Tomlinson, R., Banga, J. P. and Reimer, C. B. (1989a). Affinity purification of IgG subclasses and the distribution of thyroid autoantibody reactivity in Hashimoto's thyroiditis. Scand. J. Immunol. **30**: 73-82.

Weetman, A. P., Cohen, S. B., Makgoba, M. W. and Borysiewicz, L. K. (1989b). Expression of an intercellular adhesion molecule, ICAM-1, by human thyroid cells. J. Endocrin. **122**: 185-191.

Weetman, A. P. and McGregor, A. M. (1984). Autoimmune thyroid disease: developments in our understanding. Endocr. Rev. **5**(2): 309-55.

Weetman, A. P. and McGregor, A. M. (1994). Autoimmune thyroid disease: further developments in our understanding. Endocr. Rev. **15**(6): 788-830.

Weissman, A. M., Ross, P., Luong, E. T., Garcia-Morales, P., Jelachich, M. L., Biddison, W. E., Klausner, R. D. and Samelson, L. E. (1988). Tyrosine phosphorylation of the human T cell antigen receptor zeta chain: activation via CD3 but not CD2. J. Immunol. **141**(10): 3532-3536.

Wen, L., Roberts, S. J., Viney, J. L., Wong, F. S., Mallick, C., Findly, R. C., Peng, Q., Craft, J. E., Owen, M. J. and Hayday, A. C. (1994). Immunoglobulin synthesis and generalized autoimmunity in mice congenitally deficient in alpha beta (+). Nature **369**: 654-658.

Wiles, A. P., Shaw, G., Bright, J., Perczel, A., Campbell, I. D. and Barlow, P. N. (1997). NMR studies of a viral protein that mimics the regulators of complement activation. J. Mol. Biol. **272**: 253-265.

Wilkins, M. R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Hochstrasser, D. F. and Appel, R. D. (1997). Detailed peptide characterisation using PEPTIDEMASS - a World-Wide-Web accessible tool. Electrophoresis **18**: 403-408.

Wilson, I. A. and Stanfield, R. L. (1994). Antibody-antigen interactions: new structures and new conformational changes. Curr. Opin. Struc. Biol. **4**(6): 857-867.

Wilson, R., Aincough, R., Anderson, K. and Baynes, C. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature **368**: 32-38.

Wu, T. T., Johnson, G. and Kabat, E. A. (1993). Length distribution of CDRH3 in antibodies. Proteins **16**(1): 1-7.

Wu, T. T. and Kabat, E. A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med **132**: 211-250.

Wu, Z., Podack, E. R., McKenzie, J. M., Olsen, K. J. and Zakarija, M. (1994). Perforin expression by thyroid-infiltrating T cells in autoimmune thyroid disease. Clin. Exp. Immunol. **98**: 470.

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

Xiao, S., Dorris, M. L., Rawitch, A. B. and Taurog, A. (1996). Selectivity in tyrosyl iodination sites in human thyroglobulin. Arch. Biochem. Biophys. **334**: 284-294.

Yeh, J. I. and Hol, W. G. J. (1998). A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. Acta. Cryst. **D54**: 479-480.

Yewdell, J. and Bennick, J. (1999). Immunodominance in major histocompatibility complex class I-restricted lymphocyte responses. Ann. Rev. Immunol. **17**: 51-88.

Yokoyama, N. and Taurog, A. (1988). Porcine thyroid peroxidase: relationship between the native enzyme and an active, highly purified tryptic fragment. Mol. Endocrinol.: 838-844.

Zeng, J. and Fenna, R. E. (1992). X-ray crystal structure of canine myeloperoxidase at 3 Å resolution. J. Mol. Biol. **226**: 185-207.

Zinkernagel, R. M. and Doherty, P. C. (1974). Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytic choriomeningitis. Nature **251**: 547-548.

Zuniga-Pflucker, J. C., Di, J. and Lenardo, M. J. (1995). Requirement for TNF- α and IL-1 α in fetal thymocyte commitment and differentiation. Science **268**: 1906-1909.