



Modification of Human Thyroid Peroxidase for Structural Studies

Daniel Stephen Thomas

A thesis submitted for the degree of doctor of philosophy to
Cardiff University

March 2019

STATEMENT 1 This thesis is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Signed _____ Date _____

STATEMENT 2 This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed _____ Date _____

STATEMENT 3 I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed _____ Date _____

DECLARATION This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed _____ Date _____

WORD COUNT _____ 51,084 _____

ACKNOWLEDGMENTS

This work is the result of a collaboration between RSR Ltd, Cardiff and Cardiff University School of Chemistry. I am grateful to RSR Ltd for funding and allowing me the opportunity to carry out this work. I would like to thank Dr Jadwiga Furmaniak of RSR Ltd and Prof Rudolf Allemann for their supervision over the course of this work.

I would like to thank all of my colleagues at FIRS laboratories, RSR Ltd for their help and support during this project. In particular, I would like to thank Dr Stuart Baker, Mr Andrew Sullivan and Dr Michael Powell for their help and guidance with the practical aspects of this work.

I am also grateful to all of my colleagues at Cardiff University. I would like to thank Dr Pierre Rizkallah who provided invaluable assistance with crystallography and X-ray diffraction experiments, Mr Thomas Williams who carried out the mass spectrometry of samples for the epitope mapping studies, Dr Alexander Nodling who assisted with some chromatography and sample preparation and Dr Robert Mart who offered useful advice throughout the project.

N-terminal sequencing was carried out by Alta Bioscience Ltd, Redditch. Some crystallography and X-ray diffraction experiments were carried out by Charles River Laboratories, Saffron Walden which are described in the relevant sections of the main text.

Finally, I would like to thank Hannah and my parents for supporting me throughout this work.

Abstract

Thyroid peroxidase (TPO) is an enzyme found on the cell surface of thyrocyte cells which is involved in the production of thyroid hormones. TPO is one of three major autoantigens in autoimmune thyroid disease (AITD). Despite investigations by several groups, the three-dimensional structure of the TPO protein is unknown. Solving the structure of TPO would further the understanding of AITD and produce useful data for improving the diagnosis and treatment of the disease.

The focus of this work was to use information from the structures of related peroxidase enzymes to design modified TPO proteins which would be suitable for X-ray crystallography. Five modified proteins were designed and expressed in an insect cell line. Of these five proteins, four were successfully purified and concentrated to the high levels required for crystallographic studies. These proteins were characterised to show that they retained key features of the native protein such as autoantibody binding and enzymatic activity.

The four proteins were tested in sparse matrix crystal screens. Two of the proteins produced large, single crystals suitable for X-ray diffraction experiments. The diffraction data obtained from these crystals had a maximum resolution of 5.2 Å, an improvement on previous work but insufficient to produce a reliable structure. Further experiments were carried out but the diffraction resolution could not be improved. These involved varying the crystallization conditions, modifying the protein by deglycosylation and using ligands to stabilise the protein.

Additional work focused on furthering the understanding of autoantibody binding to TPO. Using an epitope excision technique not previously applied to TPO, it was demonstrated that an anti-TPO monoclonal antibody bound to a region close to or covering one of two immunodominant regions of the protein. These results are in good agreement with data from the current literature.

Contents

Chapter 1- Introduction	1
<u>The immune system</u>	<u>1</u>
<u>Autoimmune disease</u>	<u>8</u>
<u>Autoimmune thyroid disease</u>	<u>12</u>
<u>Thyroid autoantigens</u>	<u>15</u>
<u>Prevalence of autoimmune thyroid disease</u>	<u>16</u>
<u>Causes and risk factors of autoimmune thyroid disease</u>	<u>18</u>
<u>Treatment of autoimmune thyroid disease</u>	<u>23</u>
<u>The thyroid</u>	<u>24</u>
<u>Thyroid peroxidase function</u>	<u>25</u>
<u>Thyroid peroxidase gene structure</u>	<u>30</u>
<u>Related peroxidase enzymes</u>	<u>31</u>
<u>Myeloperoxidase</u>	<u>33</u>
<u>Lactoperoxidase</u>	<u>34</u>
<u>Eosinophil peroxidase</u>	<u>36</u>
<u>Thyroid peroxidase protein structure</u>	<u>36</u>
<u>N-terminal domain</u>	<u>37</u>
<u>Peroxidase domain</u>	<u>37</u>
<u>Complement control protein like domain</u>	<u>39</u>
<u>Epidermal growth factor like domain</u>	<u>41</u>
<u>Transmembrane domain</u>	<u>43</u>
<u>Glycosylation</u>	<u>43</u>
<u>Thyroid peroxidase immunodominant regions</u>	<u>44</u>
<u>Aims of study</u>	<u>47</u>
Chapter 2- Production and Characterization of Modified Thyroid Peroxidase Proteins	49
<u>Introduction</u>	<u>49</u>
<u>Result and discussion</u>	<u>50</u>
<u>Design of modified thyroid peroxidase proteins and DNA cloning</u>	<u>50</u>
<u>Further modification of NΔTPO gene</u>	<u>53</u>
<u>Expression of modified thyroid peroxidase proteins</u>	<u>56</u>
<u>Capture of protein from culture media</u>	<u>59</u>
<u>Purification of modified thyroid peroxidase proteins</u>	<u>60</u>
<u>Final polishing of modified thyroid peroxidase proteins</u>	<u>64</u>
<u>Characterisation of modified thyroid peroxidase proteins</u>	<u>67</u>
<u>Summary</u>	<u>82</u>
<u>Key observations</u>	<u>83</u>
<u>Problems</u>	<u>83</u>
<u>Primers</u>	<u>84</u>
Chapter 3- Crystallography of Modified Thyroid Peroxidase Proteins	86
<u>Introduction</u>	<u>86</u>
<u>Results and Discussion</u>	<u>87</u>
<u>Crystal screening of modified thyroid peroxidase proteins</u>	<u>87</u>

<u>Further crystallography experiments</u>	105
<u>Summary</u>	109
<u>Key observations</u>	109
<u>Problems</u>	110
<u>Chapter 4- Deglycosylation of Modified Thyroid Peroxidase Protein for Crystallography</u>	111
<u>Introduction</u>	111
<u>Results and Discussion</u>	114
Protein deglycosylation method development	114
Preparative scale protein deglycosylation	116
Crystallization of deglycosylated protein	123
<u>Summary</u>	127
<u>Key observations</u>	127
<u>Problems</u>	127
<u>Chapter 5- Co-crystallization of Modified Thyroid Peroxidase Proteins with Ligands</u>	128
<u>Introduction</u>	128
<u>Inhibitor screening</u>	128
Protein:F(ab) complex screening	128
<u>Results and Discussion</u>	131
Initial inhibitor screening	131
Selection of further inhibitors	132
Further inhibitor screening	133
4F5 F(ab) production	135
N Δ TPO:4F5 F(ab) complex production	139
N Δ TPO:4F5 F(ab) complex crystallography	143
<u>Summary</u>	145
<u>Inhibitor screening</u>	145
F(ab) complex screening	145
<u>Key observations</u>	146
<u>Problems</u>	146
<u>Final Summary- Thyroid Peroxidase Crystallography</u>	147
<u>Chapter 6- Thyroid Peroxidase Epitope Mapping Studies</u>	149
<u>Introduction</u>	149
<u>Results and Discussion</u>	152
Antibody selection	152
4F5 epitope excision method development	155
Epitope excision results and discussion	159
<u>Conclusions</u>	165
<u>Summary</u>	167
<u>Key observations</u>	167
<u>Problems</u>	167
<u>Chapter 7- Future Work and Outlook</u>	168

<u>Future crystallography work: short term</u>	168
<u>Longer term studies: crystallography and beyond</u>	170
<u>Outlook</u>	173
<hr/>	
Chapter 8- Materials and Methods	176
<u>DNA cloning</u>	176
<u>Recombinant baculovirus production</u>	179
<u>Protein expression</u>	182
<u>Protein purification</u>	183
<u>Protein analysis</u>	185
<u>Crystallization screening</u>	188
<u>X-ray diffraction experiments</u>	189
<u>Deglycosylation</u>	190
<u>F(ab) complex production</u>	191
<u>Antibody selection</u>	193
<u>Antibody coupling to Sepharose</u>	193
<u>Epitope excision</u>	194
<u>Liquid chromatography-mass spectrometry analysis</u>	194
<hr/>	
References	196

List of figures

- Figure 1-** Diagram of an IgG molecule. **P5**
- Figure 2-** Structure of methimazole and propylthiouracil. **P23**
- Figure 3-** Structure of 3,3',5-triiodothyronine and thyroxine. **P25**
- Figure 4-** Structure of haem in the 3+ resting state and the 4+ oxoferryl form. **P26**
- Figure 5-** Oxidation of the TPO active site by hydrogen peroxide. **P27**
- Figure 6-** Structure of haem in compound II. **P28**
- Figure 7-** Iodotyrosine coupling reaction to form T3/T4. **P30**
- Figure 8-** Alignment of human TPO, human MPO and bovine LPO protein sequences. **P32**
- Figure 9-** MPO structure. **P34**
- Figure 10-** LPO structure. **P35**
- Figure 11-** Human TPO domain organisation. **P37**
- Figure 12-** Active site haem of LPO and MPO. **P39**
- Figure 13-** NMR structure of 16th CCP module of human factor H. **P40**
- Figure 14-** Alignment of TPO CCP domain, factor H 16th CCP domain and VCP 3rd CCP domain protein sequences. **P40**
- Figure 15-** Alignment of TPO EGF domain, fibrillin 33rd EGF domain and factor IX first EGF domain protein sequences. **P41**
- Figure 16-** NMR structure of 33rd EGF domain of human fibrillin and crystal structure of first EGF domain of human factor IX. **P42**
- Figure 17-** Amino acids mapped to TPO immunodominant regions. **P46**
- Figure 18-** Alignment of TPO cDNA with primer 1. **P51**
- Figure 19-** Alignment of TPO cDNA with primer 5. **P52**
- Figure 20-** Alignment of NΔTPO cDNA with primer 7. **P53**
- Figure 21-** Alignment of NΔTPO cDNA with primer 9. **P54**
- Figure 22-** Alignment of TPO cDNA with primer 11. **P55**
- Figure 23-** Mechanism of reaction catalysed by TPO in guaiacol oxidation assay. **P58**
- Figure 24-** Western blot of intact protein from harvest supernatant. **P58**
- Figure 25-** IMAC elution profile for NΔTPO(1). **P61**
- Figure 26-** Stained 8% acrylamide reduced SDS-PAGE of pooled post IMAC NΔTPO (1) elution peaks. **P63**
- Figure 27-** Preparative SEC elution profiles for NΔTPOS and NΔTPO-10. **P65**
- Figure 28-** Stained 8% acrylamide reduced SDS-PAGE of NΔTPOS SEC fraction samples. **P66**
- Figure 29-** Preparative SEC elution profile for NΔTPO-2 and analytical SEC profile for NΔTPO-2 unused fraction pool. **P67**
- Figure 30-** Analytical SEC elution profiles of final pools. **P69**
- Figure 31-** Stained 8% acrylamide reduced SDS-PAGE of final pools. **P70**

- Figure 32-** Stained IEF gel of final pools. **P73**
- Figure 33-** Pooled patient sera antibody binding for individual modified TPO proteins. **P75**
- Figure 34-** 2G4 antibody binding for individual modified TPO proteins. **P76**
- Figure 35-** Pooled patient sera relative antibody binding for modified TPO proteins. **P77**
- Figure 36-** 2G4 relative antibody binding for modified TPO proteins. **P77**
- Figure 37-** Side view diagram of hanging drop and sitting drop crystal screening formats. **P89**
- Figure 38-** Micrographs of hits from manual NΔTPO crystal screening. **P90**
- Figure 39-** Micrographs of Spherulites or small crystals of NΔTPO-2. **P91**
- Figure 40-** Micrographs of NΔTPO crystals grown at Charles River Laboratories. **P94**
- Figure 41-** Symmetry elements in trigonal space groups. **P96**
- Figure 42-** Micrographs of optimised crystals grown at Charles River Laboratories. **P98**
- Figure 43-** Micrographs of NΔTPO crystal grown from experiments set up by robot. **P100**
- Figure 44-** Structure of ethylene glycol and polyethylene glycol. **P100**
- Figure 45-** NΔTPO crystal mounted on I04 beamline. **P101**
- Figure 46-** Example image of X-ray diffraction data collected from NΔTPO crystal. **P101**
- Figure 47-** Symmetry elements in the P6₃22 space group. **P102**
- Figure 48-** Micrographs of NΔTPO crystals grown with and without 5% sucrose. **P106**
- Figure 49-** Diagrams of common insect N-linked glycans. **P111**
- Figure 50-** Lectin blot of NΔTPO trial deglycosylation samples. **P115**
- Figure 51-** SEC elution profile for deglycosylated NΔTPO. **P118**
- Figure 52-** Analytical SEC elution profiles of deglycosylated NΔTPO SEC fractions A11 and E1. **P118**
- Figure 53-** Stained 8% acrylamide reduced SDS-PAGE of deglycosylated NΔTPO. **P119**
- Figure 54-** Analytical SEC elution profile of deglycosylated NΔTPO pool. **P120**
- Figure 55-** Pooled patient sera and 2G4 antibody binding for deglycosylated NΔTPO. **P121**
- Figure 56-** Pooled patient sera relative antibody binding for deglycosylated NΔTPO. **P122**
- Figure 57-** 2G4 relative antibody binding for deglycosylated NΔTPO. **P122**
- Figure 58-** Micrographs of deglycosylated NΔTPO crystals grown at Charles River Laboratories in sparse matrix screens. **P125**
- Figure 59-** Micrographs of deglycosylated NΔTPO crystals grown at Charles River Laboratories in optimisation screens. **P125**
- Figure 60-** Diagram of IgG, F(ab) and F(ab')2 molecules. **P130**
- Figure 61-** Micrographs of NΔTPO crystals grown at Cardiff University. **P132**
- Figure 62-** Inhibitors used for crystal screening. **P133**
- Figure 63-** Micrographs of NΔTPO crystals grown with *p*-toluidine. **P134**
- Figure 64-** Stained 12% acrylamide reduced SDS-PAGE of trial papain digestions. **P136**

Figure 65- Stained 12% acrylamide reduced SDS-PAGE of 4F5 F(ab) (1) production samples. **P137**

Figure 66- Direct binding assay results for 4F5 F(ab) (2) and 4F5 IgG and inhibition assay results for 4F5 F(ab)(1) and 4F5 F(ab) (2). **P138**

Figure 67- Overlaid analytical SEC elution profiles of unpurified NΔTPO:4F5 F(ab) complex and 4F5 F(ab) (1). **P139**

Figure 68- Analytical SEC elution profile of 4F5 F(ab) post preparative SEC. **P140**

Figure 69- Analytical SEC elution profiles of NΔTPO:4F5 F(ab) (2) complex ratio trials. **P141**

Figure 70– Preparative SEC elution profile for NΔTPO:4F5 F(ab) (1). **P141**

Figure 71- Stained 12% acrylamide reduced SDS-PAGE gels of NΔTPO:4F5 F(ab) (1) and (2). **P142**

Figure 72- Analytical SEC elution profiles of NΔTPO:4F5 F(ab) complex final pools. **P143**

Figure 73– Micrograph of NΔTPO:4F5 F(ab)(1) crystals. **P144**

Figure 74- Trypsin reaction mechanism. **P151**

Figure 75- SEC elution profiles for 100:1 protein:trypsin digested 4F5 and 2G4 IgG. **P152**

Figure 76- Stained 8% acrylamide reduced SDS-PAGE of trypsin digested 4F5 and 2G4. **P153**

Figure 77- Analytical SEC elution profiles of rhTPO digested with trypsin at a 100:1 w/w ratio (blue) and 5:1 w/w (red). **P155**

Figure 78- Overview of epitope excision procedure. **P156**

Figure 79- Mechanism of IgG coupling to CNBr Sepharose. **P157**

Figure 80- Epitope excision elution LCMS TIC chromatogram. **P160**

Figure 81- Deconvoluted spectra of peak eluted at 20.38 min in Fig. 77. **P161**

Figure 82- MS/MS spectra of EP1 peptide fragmented by CID. **P162**

Figure 83- EP1 peptide structure in single letter amino acid code. **P163**

Figure 84- Human TPO amino acids 109-738. **P163**

Figure 85- Map of pUC18 plasmid. **P176**

Figure 86- Map of pFastBac1 plasmid. **P181**

Figure 87- Diagram of bacmid DNA production process. **P181**

List of tables

Table 1- Amino acid sequences of modified TPO proteins. **P55**

Table 2- Pooled protein harvests prior to purification. **P57**

Table 3- DEAE purification yields. **P60**

Table 4- IMAC purification yields. **P62**

Table 5- Post IMAC pool enzymatic activities and peak ratios. **P63**

Table 6- Post SEC fractions summary. **P65**

Table 7- Final pool weights and volumes. **P68**

Table 8- Purity of final pools. **P68**

Table 9- Molecular weight of modified TPO proteins. **P71**

Table 10- Isoelectric points of modified TPO proteins. **P73**

Table 11- Enzymatic activity of final pools. **P78**

Table 12- Final pool absorbance ratios. **P78**

Table 13- Modified TPO protein melting temperatures. **P81**

Table 14- Summary of results of modified TPO proteins in sparse matrix crystal screens with manual drop setting. **P90**

Table 15- Summary of results of screening modified TPO proteins in sparse matrix crystal screens carried out by Charles River Laboratories with a crystallography robot. **P93**

Table 16- Summary of modified TPO optimisation screens set up with a crystallography robot by Charles River Laboratories. **P97**

Table 17- Results output from “Matthews_coef” component of the CCP4 software suite. **P104**

Table 18- Trial NΔTPO deglycosylation reaction conditions. **P115**

Table 19- Fractions selected from SEC purification of deglycosylated NΔTPO for analytical SEC. **P119**

Table 20- Summary of deglycosylated NΔTPO crystal screens set up with a crystallography robot by Charles River Laboratories. **P123**

Table 21- NΔTPO:4F5 F(ab) complex final pool weights and volumes. **P141**

List of abbreviations

Aire-	Autoimmune regulator protein	IMAC-	Immobilised metal affinity chromatography
AITD-	Autoimmune thyroid disease	LC-MS-	Liquid chromatography-mass spectrometry
APC-	Antigen presenting cell	LPO-	Lactoperoxidase
BCR-	B cell receptor	MHC-	Major histocompatibility complex
BSA-	Bovine serum albumin	MMI-	Methimazole
cAMP-	Cyclic adenosine monophosphate	MOI-	multiplicity of infection
CCP-	Complement control protein	MPO-	Myeloperoxidase
CCP4-Collaborative Computational Project Number 4		MS-	Mass spectrometry
CDR-	Complementarity determining region	MS/MS-	Tandem mass spectrometry
CHO-	Chinese hamster ovary	NHANES III-	National health and nutrition examination survey III
CID-	Collision induced dissociation	NK-	Natural killer (cell)
CNBr-	Cyanogen bromide	NMR-	Nuclear magnetic resonance spectroscopy
CPM-	Counts per minute	PDB-	Protein data bank
Cryo-EM-	Cryogenic electron microscopy	PEG-	Polyethylene glycol
CV-	Column volume	PI-	Isoelectric point
DEAE-	Diethylaminoethyl	PPM-	Parts per million
dNTP-	Deoxyribonucleotide triphosphate	PPT-	Postpartum thyroiditis
EGF-	Epidermal growth factor	PTU-	Propylthiouracil
ELISA-	Enzyme linked immunosorbent assay	qPCR-	Quantitative polymerase chain reaction
EPO-	Eosinophil peroxidase	RE-	Restriction endonuclease
F(ab)-	Antibody binding fragment	rhTPO-	Recombinant human thyroid peroxidase
F(c)-	Crystallisable fragment	RIA-	Radioimmunosorbent assay
GD-	Graves' disease	RP-	Reverse phase chromatography
GO-	Graves' ophthalmopathy	SDS-PAGE-	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
GU-	Guaiacol unit	SEC-	Size exclusion chromatography
HLA-	Human leukocyte antigen complex	T1DM-	Type 1 diabetes melitus
HT-	Hashimoto's thyroiditis	T3-	Triiodothyronine
IDR-	Immunodominant region	T4-	Thyroxine
IEF-	Isoelectric focusing		
Ig-	Immunoglobulin		
IL-	Interleukin		

TCR- T cell receptor

Tg- Thyroglobulin

TGF- β - Transforming growth factor- β

Tm- Melting point

TPCK- Tosyl phenylalanyl chloromethyl ketone

TPO- Thyroid Peroxidase

Treg- Regulatory T (cell)

TSH- Thyroid stimulating hormone

TSHR- Thyroid stimulating hormone receptor

UPLC- Ultra performance liquid chromatography

VCP- Vaccinia virus complement control protein

XFEL- X-ray free electron laser

Chapter 1- Introduction

The immune system

The human immune system is responsible for preventing pathogens including parasites, viruses and microorganisms such as bacteria and fungi from causing harm to the body's cells. To cause harm, these organisms must first bypass the body's intrinsic defences. These are mechanical, chemical and microbiological in nature. Skin acts as a mechanical barrier to infection as microorganisms cannot pass between the epithelial cells which make up the tissue unless the epithelium is damaged in some way¹. An example of a chemical defence against pathogens is the low pH of the stomach. This acidic environment kills or prevents growth of many microorganisms present in food. The native microbial flora found on the skin, in the lungs or living in the gut of any healthy individual also plays a role in protecting the body from infection. These bacteria are non-pathogenic and are well adapted to growth on or in their human host. As such they deter pathogenic organisms by competing with them for nutrients and by producing antimicrobial substances of their own².

There are 2 types of immunity, innate and adaptive. Innate immunity relies on phagocytic cells known as macrophages which destroy pathogens by engulfing them³. Macrophages are constantly in circulation in the body and recognise many types of molecules commonly found on the surface of pathogens¹. This type of response does not require previous exposure to a pathogen. The innate response aims to control an infection whilst an adaptive response, which takes 4 to 7 days in the first instance of exposure to a pathogen develops.

When a macrophage is activated, it produces cytokines and chemokines which increase the permeability of blood vessels and attract neutrophils to the site of the infection. This increased permeability causes inflammation at the site of the infection. This increases the flow of lymph into the lymphoid tissue and carries dendritic cells to the local lymph nodes⁴. Dendritic cells are a type of macrophage that absorb extracellular molecules for presentation to T cells as

well as engulfing and digesting pathogens. When an immature dendritic cell recognises a pathogen it becomes activated and matures. Mature dendritic cells are antigen presenting cells (APC) which possess the cell surface molecules required to stimulate proliferation of T cells⁵. An antigen is a molecule which triggers an immune response. In this case the antigen presented are small peptides produced by the digestion of proteins found in the extracellular fluid. Recognition of an antigen by B or T cells initiates a response from the adaptive immune system. Mature dendritic cells secrete cytokines and migrate to the lymph nodes to present antigens to T lymphocytes.

Working alongside macrophages is a system of serum proteins known as complement. The complement system is also part of the innate response. When a pathogen is recognised by complement, a cascade is triggered which causes the proteolytic break down of molecules on the cell surface of the pathogen⁶. This process, known as opsonisation, increases the likelihood of a pathogen being recognised by a macrophage. The complement system can also kill cells directly by forming anaphylatoxins or membrane attack complexes which form large pores in cell membranes^{7,8}.

Adaptive immunity occurs as a response to a particular pathogen and involves lymphocytes which target specific antigens. Lymphocytes produce receptors or antibodies which bind to specific sites on an antigen known as an epitope. The adaptive immune system does not respond to a pathogen until after infection, however once exposed, reinfection with the same pathogen will result in a more rapid response⁹.

Lymphocytes are produced in the bone marrow from common lymphoid progenitor cells. These cells, as with the precursors of other cell types found in blood, mature from pluripotent haematopoietic stem cells¹⁰. These stem cells have the potential to become several different types of mature cells. Common lymphoid progenitor cells either mature in the bone marrow and become B cells or migrate to the thymus to mature into T cells.

T cells have cell surface receptors (TCR) which recognise antigens. A large repertoire of different receptors is produced by gene recombination¹¹. A limited number of gene segments recombine randomly until a functional receptor

capable of recognising major histocompatibility (MHC) molecules is produced. Once this occurs the recombination stops, fixing the specificity of that lymphocyte. Each T cell has multiple copies of the same TCR on its surface to recognise MHC bound antigen.

Once a T cell has undergone recombination and has produced a functional receptor it enters circulation. Mature T cells recirculate between the blood stream and lymphatic tissue⁴. In the lymph nodes T cells are presented with antigens by APC. Until the T cell receptor is stimulated it is said to be naïve. If the receptor is simultaneously stimulated by a foreign antigen attached to a MHC molecule and other co-stimulatory molecules found on the surface of APC, the T cell becomes activated and begins to proliferate¹². Its progeny will produce the same cell surface receptors and will thus be activated by the same epitope on the antigen which initiated the proliferation.

Stimulation of the TCR alone is not enough to trigger clonal expansion and differentiation of T cells, co-stimulation by another surface receptor is also required. CD28 is one of the more highly studied of these receptors. During activation CD80 or CD86 molecules on the surface of the APC bind to CD28 on the T cell¹³. This activates the T cell which will then produce cytotoxic T lymphocyte associated factor 4 (CTLA-4). CTLA-4 also binds CD80 and CD86, but produces an inhibitory signal when it does so¹⁴. CTLA-4 has a greater affinity for CD80 and CD86 than CD28, so the inhibitory signal reduces the sensitivity of progeny cells. As these new cells are stimulated to a lesser degree by the APC, interleukin-2 (IL-2) production is limited. IL-2 is an autocrine factor which controls T-cell differentiation and in some cases expansion by affecting the expression of receptors for other cytokines¹⁵. This negative feedback loop regulates T cell growth by preventing excessive proliferation.

Cytotoxic T cells destroy cells infected with intracellular pathogens such as parasites, viruses and some bacteria¹⁶. Antigens from the pathogen are displayed on the infected cells surface by MHC. Interaction with the T cell activates caspase cascades in the infected cell which activate enzymes that break down DNA. This kills the host cell and the pathogen within.

B cells produce proteins known as immunoglobulins (Ig) or antibodies. As with T cells, gene rearrangement allows each cell to produce a unique antibody from a fixed number of gene segments. There are 5 isotypes of immunoglobulins. IgM is the first type of Ig produced by a developing B cell and by a mature cell producing a primary response to a previously unencountered antigen. It is mainly found as a monomer or pentamer although other polymeric forms have been described¹⁷. IgD is produced after IgM by alternative splicing¹⁸. The other 3 isotypes are produced by isotype switching, a process which occurs in response to antigen stimulation or cytokine signalling¹⁹. IgA is dimeric and mostly found in mucosal areas such as the gut or respiratory tract although it is present in small amounts in sera²⁰. It is also found in secreted fluids such as tears or saliva. IgD, E and G are monomers. IgE is mainly involved in immunity against parasites but also plays a role in type 1 hypersensitivity reactions²¹. It is the least abundant Ig in circulation, but produces a potent response as mast cells bind to it with a high affinity. IgG is found in serum and accounts for around 75% of the total Ig produced (Fig. 1). There are 4 subclasses of IgG- IgG1, 2, 3 and 4²². IgG1 are the most abundant of the IgG subclasses. IgG1 are usually directed to soluble and membrane bound protein antigens²³. IgG2 are more commonly directed against polysaccharides such as those found in the capsules surrounding bacterial cells. While IgG 1, 2 and 4 share a quite similar structure IgG3 type antibodies have an extended hinge region relative to the other classes and extra glycosylation sites²⁴. This subclass also has the shortest half-life in circulation. This may be an adaptation to prevent an excessive immune response as IgG3 produces a strong inflammatory response. IgG4 antibodies are formed against antigens to which there is a repeated or long-term exposure such as allergens²⁵. IgG4 is the only IgG subclass which does not fix complement.

Ig molecules bind antigens through complementarity determining regions (CDRs) found at the of the F(ab) regions²⁶. Each F(ab) region has six CDRs in total, three each on both the heavy and light chain. Each CDR is a loop of variable length, particularly in the case of the heavy chain CDR3. The CDRs play a large, although not total, role in determining what antigen the antibody recognises²⁷. These are therefore the sites with the highest rates of variation.

The genes coding for the stretches of the protein, known as the framework region are relatively conserved in comparison.

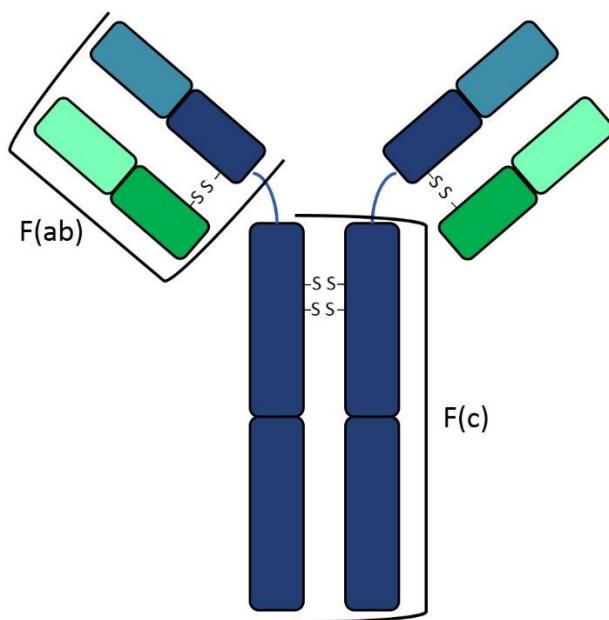


Figure 1- Diagram of an IgG molecule. Heavy chain- blue. Light chain- green. Hinge region of heavy chain- blue line. -S S - disulphide bond. F(ab)- antibody binding fragment (two per molecule). Lighter colours indicate the regions involved in antigen binding. F(c)- crystallisable fragment. This region mediates interactions with effector cells via Fc receptors and complement proteins.

IgD and monomeric IgM molecules are found bound to the surface of the mature B cell although both are also secreted. When stimulated they act as receptors and initiate the production and secretion of more Ig with the same specificity²⁸. The act of B cell receptors (BCR) binding to the antigen may neutralise it if it is a virus or bacterial toxin. Other pathogens such as bacteria or parasites may not be directly neutralised by Ig binding. Natural killer (NK) cells recognise the constant region of Ig molecules which are not involved in antigen binding (Fc). NK cells do not have antigen specific receptors and are part of the innate immune response. Any molecule or cell with Ig bound to it will be recognised by the Fc receptors on NK cells and destroyed by a process known as antibody dependent cell mediated cytotoxicity²⁹.

Few antigens activate a B cell proliferative response directly, therefore another subset of T cells known as helper T cells are required to produce a B cell response. Naïve helper T cells stimulated by an APC differentiate into either

Th1 or Th2 effector cells³⁰. Th1 cells secrete interferon-γ and tumour necrosis factor-α whereas Th2 cells secrete several types of interleukins.

When an antigen is recognised by the BCR it is internalised to an endosomal compartment. Here it is broken down into short peptides which are bound to MHC class II proteins³¹. The MHC bound peptides are then displayed on the cell surface where they are recognised by helper T cells which are activated by the same antigen.

The probability of an activated B cell meeting a T cell with the same specificity in circulation is very low. To increase the likelihood of this event happening activated B cells become trapped in the T cell zone of the spleen while naïve B cells pass through this zone quickly³². The T cell zone of the spleen is where T cells are trapped and differentiate to helper T cells. Once B cells receive the required co-stimulation in the T cell zone they leave the spleen, differentiate and proliferate.

After contact with Th cells bearing the same antigen receptor, B cells migrate to the border between the T cell zone and red pulp of the spleen and establish a primary focus of proliferation and differentiate into memory or plasma cells³³. Some activated B cells migrate to the primary lymphoid follicles where a germinal centre is formed³⁴. Here B cells go through a process of somatic hypermutation. During this process point mutations are introduced to the variable (V), diversity (D) and joining (J) genes of Ig molecules at each cell division³⁵. These genes code for the section of an Ig which binds to the antigen. While there are a limited number of these genes available, by generating different combinations of V(D)J genes (V and J only for light chains) a larger, more diverse repertoire of Ig molecules is achieved³⁶. This changes the structure of the complementarity-determining regions which make up the antibody binding site. These mutations occur at a much higher rate than during normal cell division increasing the chances of producing a higher affinity clone. Mutagenesis is driven by activation-induced cytidine deaminase (AICD). AICD deaminates cytosine to uracil in single stranded DNA, causing mismatches in the sequence³⁷. Depending on how the repair occurs this mismatch can result in a mutation. B cell clones with higher affinity receptors compete more

successfully for the limited amount of antigen available³⁸. These cells receive survival signals from helper T cells in the germinal centre where lower affinity clones or those which no longer produce a functioning receptor die by apoptosis and are cleared by macrophages.

Plasma cells produce and secrete large amounts of Ig, but no longer express the MHC molecules required to present antigen to T cells³⁹. A plasma cell cannot differentiate further or undergo isotype switching to change the type of Ig it produces. Memory B cells are long lived cells which express the high affinity receptors selected for in germinal centres⁴⁰. These cells therefore produce a faster and more potent response if the body is reinfected with a previously encountered antigen.

The immune system has a mechanism to prevent it targeting the body's own cells known as tolerance. In 1953 Billingham *et al* showed that exposing foetal mice or chickens to foreign cells early in development induces tolerance of these cells in adult life⁴¹. Skin grafts from the donor animal to the host which had previously received cells from the donor were successful where they would otherwise be rejected and attacked by the host's immune system. Such hosts retained the ability to mount an immunological response to grafts from other donors not previously encountered.

When the immune system is functioning correctly, lymphocytes which react to epitopes on self antigens are prevented from maturing. This negative selection may occur by destroying the cell by apoptosis or by rendering the cell anergic so it cannot respond to the antigen.

For developing T cells this selection occurs in the thymic cortex. Positive selection removes the majority of the developing T cell population. The developing cells encounter MHC molecules on the surface of stromal cells. If a cell's TCR does not recognise a MHC molecule in the thymus it undergoes programmed cell death⁴². This interaction triggers the down regulation of either CD4 or CD8 receptors, depending on whether the cell recognises class I or class II MHC molecules⁴³. This commits the cell to becoming a helper or regulatory T cell (CD4+ recognising MHC class II) or cytotoxic T cell (CD8+ recognising MHC class I) once mature. The cell population then undergoes

another positive selection step for the expression of one of these co-receptors required to mount a response to an antigen.

Expression of self peptides in the thymus is controlled by the regulatory protein autoimmune regulator (Aire). This gene was discovered in a study of genetic mutations linked to autoimmune polyglandular syndrome type 1⁴⁴. When functioning correctly Aire promotes ectopic expression of peripheral tissue antigens spread over several locations on the nuclear DNA⁴⁵. These self molecules are then presented to immature T cells as part of the selection process.

Once T cells have been positively selected, they are presented with a variety of self peptides bound to MHC molecules in the medulla⁴⁶. Cells which respond to these molecules undergo clonal deletion or become regulatory T cells (Treg cells), which play a role in peripheral tolerance and modulating the levels of immune responses⁴⁷. However under some conditions negative selection can still occur when the T cell moves to the periphery. Anergy may be triggered in the periphery if an antigen is presented without the required co-stimulation of the CD28 receptor⁴⁸.

B cells also undergo selection to remove self-reactive cells during maturation. Once a B cell expresses IgM on its surface, its development can be arrested by the binding of a self peptide before the cell matures and enters the periphery⁴⁹. Under normal circumstances this would be a self peptide encountered as a soluble protein or membrane bound on the surface of stromal cells in the bone marrow⁵⁰. Developmental arrest prevents B cells from expressing the receptors and adhesion molecules required to mature and enter circulation. Prolonged exposure to a recognised antigen causing arrest leads to cell death.

Autoimmune disease

The process of immune tolerance is not always complete and autoimmune diseases can develop as a result. Autoimmune diseases occur when the immune system responds to the body's own naturally occurring molecules. Molecules which trigger this response are known as autoantigens. The effects

and symptoms caused by these inappropriate immune responses vary and depend on the function of the particular autoantigen and the cells which carry it. Antibodies which react with self molecules are referred to as autoantibodies. Evidence for the existence of autoantibodies was first discovered in 1953 when sera from patients with acquired haemolytic anaemia was shown to agglutinate erythrocytes from normal individuals as well as several other mammalian species⁵¹. Prior to this study it was thought it was not possible for the body to produce antibodies against its own molecules.

It is now known that a wide range of self reactive antibodies are present in healthy individuals⁵². This phenomenon has also been demonstrated in rats and pigs and could be common amongst mammals. These low affinity, polyreactive antibodies are known as natural antibodies⁵³. These are constantly in circulation in the body regardless of any exposure to an antigen and are expressed by a subset of B cells known as B-1 cells. Natural antibodies are mostly of the IgM isotype although IgG and IgA are also present⁵⁴.

Autoantibodies differ from natural antibodies in that they recognise a single, specific epitope with high affinity. Cells that produce autoantibodies responsible for a number of diseases have been shown to undergo isotype switching and affinity maturation⁵⁵. These processes produce higher affinity antibodies often, but not always, of the IgG or IgA types commonly associated with autoimmune disease. An example of affinity maturation of an autoantibody is given by comparing the crystal structures of a high and low affinity rheumatoid factor IgM which binds to IgG crystallisable fragment (F(c))⁵⁶. Analysis of the structures showed, in this case, mostly residues in the heavy chain antibody binding region were involved in binding the antigen to the low affinity antibody⁵⁷. Only two light chain residues made any contact with the antigen leaving two of the complementarity determining region loops of this chain completely uninvolved in binding. The higher affinity antibody bound in a more conventional way with several mutated residues across each chain forming contacts to the antigen either by hydrogen bonds or van der Waals interactions, or by inducing conformational changes, bringing other residues into contact with the antigen⁵⁶.

Somatic hypermutation of B cells has been demonstrated by producing the unmutated form of an insulin autoantibody isolated from a patient with type 1 diabetes mellitus (T1DM)⁵⁸. The unmutated antibody was produced from germline V genes and bound insulin with a low affinity in comparison to the mutated or partially mutated sequences.

Autoantibodies may be markers of a disease or the causative agent depending on the disease and its exact mechanisms. Autoantibodies may activate cytotoxic cells such as NK cells causing antibody-dependent cell-mediated cytotoxicity⁵⁹. Both of these processes cause cell death by lysis. B cells may also play a role in autoimmunity by acting as APC⁶⁰.

T lymphocytes and cytokines are also involved in the pathogenesis of autoimmune disease. A 2013 study examined levels of Treg cells in autoimmune thyroid disease (AITD) patients relative to healthy control individuals⁶¹. No significant difference was found in the levels of Treg cells or the ability of Th cells to proliferate. However the ability of the Treg cells from AITD patients to suppress proliferation of Th cells was greatly reduced compared to controls. This lack of suppression may be one factor which allows an autoimmune response to develop.

The role of cytokines in autoimmunity is complex and not fully understood. Disruption of the production of a single cytokine or its associated receptor can cause wide ranging immune dysfunction. Two cytokines of particular importance in suppressing autoimmunity are transforming growth factor-β (TGF-β) and IL-10. TGF-β inhibits naïve T cells from differentiating into Th type cells and, in conjunction with TCR stimulation, induces the expression of FOXP3 to produce Treg cells from circulating CD4+ T cells⁶². Reduced levels of TGF-β have been observed in patients with Hashimoto's thyroiditis and systemic lupus erythematosus, indicating that a reduced immunosuppressive capacity may be a factor in the development of autoimmune disease^{63,64}.

Due to its anti-inflammatory effects much effort has been put into studies on using IL-10 as a treatment for various autoimmune diseases⁶⁵. IL-10 is a regulatory cytokine with both anti-inflammatory and immunostimulatory activity produced by almost all leukocytes⁶⁶. It inhibits the activation of dendritic cells

and macrophages by inhibiting MHC class II expression, reducing the production of several pro-inflammatory cytokines and the co-stimulatory potential of these cells⁶⁷. IL-10 also affects T cells directly by inhibiting CD28 co-stimulation preventing proliferation and cytokine production⁶⁸. Conversely IL-10 has a stimulatory effect on B cells increasing proliferation and antibody production⁶⁹. Disappointingly, studies with this cytokine have proven largely unsuccessful as the complex role of IL-10 in the immune system means simply administrating the cytokine to a patient is generally ineffective.

Autoimmune diseases can be systemic, affecting the body as a whole or localised affecting only a single organ or particular type of cell. In a non-organ specific, systemic autoimmune disease such as SLE several organs are targeted by the immune system. Symptoms of SLE are wide ranging given its systemic nature but include photosensitive skin rashes, inflammation and abnormalities in the blood vessels, most noticeable in the kidneys⁷⁰. Descriptions of the facial rash common in lupus can be found dating from ancient Greece. The term “lupus” was first used in the 1200’s due to the supposed resemblance of the rash to a wolf bite⁷¹. The involvement of autoantibodies in the disease was not demonstrated until 1958 when Friou *et al* used fluorescent anti-human antibodies to visualise IgG from patient sera binding to cell nuclei⁷².

SLE occurs when the immune system becomes sensitised to various components of the cell nucleus such as double stranded DNA and small nuclear ribonucleoprotein⁷³. This is thought to be caused by nucleosomes entering circulation due to inefficient clearance by phagocytes following apoptosis⁷⁴. This could be due to genetic mutations causing impaired function of these cells or from complement deficiencies. Once triggered this could be exasperated by the disease process due to antibodies blocking phagocyte function or depleting complement proteins. If not cleared from circulation these molecules can stimulate receptors on APC which normally react to viral components, leading to the further expansion of autoreactive B and T cells⁷⁵. Autoreactive Th cells stimulate B cells to produce high affinity IgG type antibodies to components of the nucleosome⁷⁰. A reduction in the number or function of Treg cells could

allow these autoreactive Th cells to proliferate, although this may not occur in all patients⁷⁶.

Autoimmune thyroid disease

Autoimmunity to the thyroid can be divided into three main types, hypothyroidism, hyperthyroidism and post-partum thyroiditis. Unlike the systemic disease SLE, autoimmune hypothyroidism is organ specific. The disease known as Hashimoto's thyroiditis (HT) is named after Dr. Hakaru Hashimoto who in 1912, described 4 cases of patients in Japan having undergone partial thyroidectomy as treatment for goitre⁷⁷. Histopathological studies of thyroid tissue samples from these patients showed signs of lymphocytic infiltration. This was distinct from the more common colloid goitre associated with low iodine intake. Modern studies indicate that thyroid atrophy is more frequent in HT although some patients do present with goitre. Other symptoms of hypothyroidism are wide ranging but can include fatigue, impaired memory, depression, weight gain and bradycardia⁷⁸. Patients with overt hypothyroidism have high levels of circulating thyroid stimulating hormone (TSH) but low concentrations of the thyroid hormones 3,3',5-triiodothyronine (T3) and thyroxine (T4).

Examination of the thyroid of HT patients typically reveals evidence of cellular infiltration into the gland⁷⁹. The infiltrating cells are primarily lymphocytes but plasma cells and macrophages are also present. The lymphocytes may organise and form germinal centres inside the thyroid. In some cases this may be accompanied by fibrosis but this is generally rare, representing less than ten percent of cases.

Another, more recently classified form of autoimmune hypothyroidism is IgG4 thyroiditis⁸⁰. In IgG4 thyroiditis the infiltrating cells contain a higher proportion of IgG4 expressing plasma cells. It is possible that IgG4 thyroiditis is part of the more systemic IgG4 related disease which affects the endocrine system more widely⁸¹. This disease is still poorly understood as until quite recently it was thought to be several separate diseases which occurred in isolation.

By comparison autoimmune hyperthyroidism can involve other organs. This condition is known as Graves' disease (GD) after the physician Robert James Graves⁸². Graves described it in 1835 along with at least two other physicians, Caleb Hillier Parry and Karl Adolph von Basedow, also publishing around the same time. GD is the most common cause of hyperthyroidism in areas with sufficient iodine intake and results in the over production of thyroid hormones⁸³. This leads to symptoms such as weight loss despite increased appetite, tremors, irritability, goitre and heat intolerance. Patients with GD have low serum TSH concentrations and high levels of T3 and T4⁸⁴.

In 1958 it was demonstrated that the sera of some patients with hyperthyroidism would have a stimulatory effect on thyroid hormone release in guinea pigs⁸⁵. Initially thought to be an abnormal form of TSH, this substance was shown to stimulate the thyroid but the mechanism by which it did was thought to be different from that of TSH. In 1964 some evidence indicated this substance, which had become known as "long acting thyroid stimulator", was an Ig molecule⁸⁶. In 1974 Smith & Hall showed that autoantibodies to the TSHR have the ability to mimic TSH and stimulate the TSHR⁸⁷.

GD patients may also present with exophthalmos, known as Graves' ophthalmopathy (GO). This condition causes a protrusion of the eyes due to swelling of the extraocular tissue caused by increased adipogenesis or enlargement of the extraocular muscles⁸⁸. Th cells infiltrate the tissue and produce cytokines such as IL-2, interferon γ and tumor necrosis factor. Orbital fibroblasts respond to these cytokines by secreting increased amounts of hyaluronan. This is a hydrophilic glycosaminoglycan found in the matrix between the ocular muscle fibres. Excess production of this molecule draws more fluid into the tissue causing swelling. GO usually occurs within 18 months of the onset of hyperthyroidism in around half of GD patients. Between 3 and 5% of patients suffer a more severe form of the condition which can cause permanent sight impairment⁸⁹. The condition is believed to be mediated by TSHR antibodies acting on TSHR expressed in the retro orbital fibroblasts⁸⁸.

GO proved difficult to study as a reproducible mouse model was not developed until 2013⁹⁰. This model was produced by immunising mice with a plasmid

containing the gene for human TSHR leucine rich repeat domain and produces similar pathologies to those observed in humans. Mice immunized following this protocol may develop stimulating or blocking TSHR antibodies and be hyper or hypothyroid as a result⁹¹. The reason for these different phenotypes is currently unknown.

GO is often mild and self limiting so treatment with topical lubricants and the use of sunglasses can be sufficient to ameliorate the symptoms. In more severe cases steroids or surgical intervention are required to control the condition⁹².

Pregnancy is a major risk factor for thyroid disease. The demand on the thyroid to produce hormones can cause the gland to increase in volume by as much as 40%⁹³. This increase in demand, immune suppression during pregnancy, and the return normal activity in the postpartum period may trigger thyroid disease⁹⁴. The risk of HT and GD developing both increase with pregnancy and the third common form of AITD, postpartum thyroiditis (PPT) occurs in approximately 8% of women⁹⁵. PPT was first noted in the literature in 1948 by Dr H. E. W. Roberton in New Zealand⁹⁶. He found patients suffering from fatigue, irritability and coldness after giving birth improved dramatically when treated with thyroid extracts and noted the similarity of the condition with hypothyroidism.

PPT can occur up to a year after giving birth⁹⁷. Patients alternate between bouts of hyperthyroidism and hypothyroidism with symptoms alternating between those associated with each condition. Hyperthyroid periods are often asymptomatic but hormone replacement may be required during hypothyroid phases depending on the severity of the disease. TSH levels are usually low in PPT patients and T3 and T4 levels may be elevated. The condition is transient in most cases and is usually resolved within a year of developing, however permanent hypothyroidism develops in around 23% of patients⁹⁸.

Thyroid autoimmunity has been linked to adverse pregnancy outcomes in a number of studies⁹⁹. This includes an increased risk of spontaneous and recurrent miscarriage, and preterm delivery. The reasons for this are not completely clear. It is possible that minor hypothyroidism increases risk as treatment with levothyroxine appears to have a beneficial effect¹⁰⁰. There is also

evidence that the offspring of TPO antibody positive mothers are more likely to suffer from impaired neurological development during childhood¹⁰¹. The IQ scores of the children of TPO antibody positive mothers are lower on average than those of antibody negative mothers up to the age of seven¹⁰². Maternal TPO antibody positivity is also associated with an increased risk of sensorineural hearing loss.

Thyroid autoantigens

In AITD autoantibodies are produced which recognise three main autoantigens, thyroid peroxidase (TPO), TSHR and thyroglobulin (Tg)^{103–105}. Antibodies to TPO have proven diagnostic and predictive value for HT and PPT^{79,97}. A large number of TPO autoantibodies have been isolated and studied¹⁰⁶. The majority of these have been produced from combinatorial libraries although at least one monoclonal antibody has been produced from hybridoma cells generated from a patient lymphocyte^{107–109}. TPO antibodies from patient sera are predominantly IgG1 and IgG4 and have been shown to activate complement *in vitro* to induce complement-mediated cytotoxicity^{110,111}.

Antibodies to TSHR found in AITD patients are interesting due to the effect they have on TSH receptor function. TSHR is a G protein coupled receptor stimulated by TSH. It is membrane bound and found on the basal surface of thyrocytes. Stimulation of the receptor triggers production of cyclic adenosine monophosphate (cAMP) which regulates several metabolic processes within the cell and leads to production of thyroid hormones¹¹².

Autoantibodies to the TSHR can be of two types, stimulating or blocking^{113,114}. Stimulating and blocking antibodies disrupt normal thyroid function when they bind to the receptor¹¹⁵. Stimulating antibodies mimic the action of TSH causing the receptor to be activated when bound and consist exclusively of the IgG1 subtype¹¹⁶. Blocking antibodies inhibit the binding of TSH preventing the receptor from functioning¹¹⁷. Both stimulating and blocking antibodies have been demonstrated to bind to the leucine rich repeat domain of the TSHR^{118,119}. Neutral antibodies to the TSHR have also been described and are believed to

bind to the receptor but do not affect the binding of TSH or stimulate cAMP production¹²⁰.

Patients with AITD may simultaneously produce blocking and stimulating autoantibodies and as a result serum concentrations of TSHR autoantibodies may not correlate with disease severity¹²¹. It is not unknown for patients to fluctuate between hyper and hypothyroidism due to changes in the relative concentrations of blocking and stimulating antibodies.

Tg is a very large protein, approximately 330 kDa, produced in the thyroid¹²². It plays a role in iodine storage as well as being the precursor to T3 and T4. Found in the follicular lumen, it makes up a large proportion of the total thyroid protein and is present in several multimeric forms which may be soluble or insoluble. Tg autoantibodies are mostly IgG4 but lower levels of IgG1, 2 and 3 are usually detectable^{110,123}. IgG4 antibodies do not fix complement, but some weak complement activity can be detected from the other IgG classes of Tg autoantibodies. Tg autoantibodies are found in the sera of patients with HT and GD but are generally the least useful clinically for predicting disease progression as they do not fix complement to damage cells directly or correlate with lymphocytic infiltration as TPO antibodies do¹²⁴.

It is possible that Tg has a greater role as a T cell autoantigen. A 2008 paper by Muixí *et al* purified MHC class II molecules from GD patient thyroid tissue¹²⁵. Several different peptides bound to the MHC molecules were shown to originate from Tg, indicating it could play more of a role in the T cell autoimmune response.

Prevalence of autoimmune thyroid disease

A 2008 review of the literature by McGrogan *et al* highlighted the difficulty of studying AITD on a population-wide basis¹²⁶. Most of the data collected on the prevalence of thyroid disease specifically caused by thyroid autoimmunity is from Caucasian populations based in the UK, USA and Scandinavia. Patient selection criteria can also affect results as AITD is often subclinical. Patients with detectable thyroid antibodies may be euthyroid and inclusion of these

cases could skew results as could variations in assay methodologies. Many studies examine the prevalence of thyroid dysfunction in relation to other factors such as pregnancy or iodine intake or study only a limited population which is not generally representative. This limits the amount of data which can be included in any meta-analysis of the literature.

One of the largest scale studies on the prevalence of thyroid disease is the “Whickham survey”¹²⁷. This work was carried out on a sample representative of the United Kingdom general population between 1972 and 1974 and involved an initial survey and a 20 year follow up¹²⁸. The follow up study found a 3.9% prevalence of hyperthyroidism in women. Hyperthyroidism is rare in men, with only a single recurrent case over the study period. The prevalence of hypothyroidism in women was 9.3% and 1.3% in men. While the Whickham survey is one of the most frequently cited sources for thyroid disease prevalence figures, the results of a single study of one population may not be representative. It must also be taken into account that the results of the 20 year follow up were gathered in 1993, making the data over 20 years old at the time of writing. Given the changes in lifestyle and diet of the population over this period the data may no longer give an accurate representation of the prevalence of thyroid disease, even in the UK.

Data was gathered as part of an even larger study in the USA carried out between 1988 and 1994¹²⁹. The “national health and nutrition examination survey III (NHANES III)” is an on-going, large scale survey designed to produce a range of data representative of the general population of America at the time. Its purpose is to allow for the analysis of risk factors and trends over time for a wide variety of diseases. A 2002 study using this data found an overall prevalence of 1.3% for hyperthyroidism and 4.6% for hypothyroidism. While this study looked at differences between age and ethnic groupings, prevalence was not reported for males and females separately. This work suffers the same limitations as the Whickham survey in that the data is already quite old by the time it is collated and analysed. However the large scale of these works means the conclusions drawn from them are more likely to be accurate and representative.

Causes and risk factors of autoimmune thyroid disease

There is unlikely to be a single underlying cause for complex conditions like HT and GD. However, a number of contributing factors have been identified. Despite the difficulties in measuring the prevalence of thyroid disease one clear trend is that both GD and HT affect women more commonly than men. This is a general trend for most autoimmune diseases and may be due to immunological differences between the sexes¹³⁰. In humans and animals females typically have stronger immune responses and higher levels of some serum immunoglobulins^{131,132}. These differences appear after sexual maturity is reached and are linked to the action of sex hormones. This has been demonstrated by performing gonadectomies on mice which then have a reduced disparity in immune responses between the genders¹³³. This is possibly caused by oestrogen, which can stimulate B cell immune responses and therefore reactivity to self antigens¹³⁴.

Skewed X chromosome inactivation is another possible explanation for the higher occurrence of AITD in females. Skewed X inactivation has been found to occur more commonly in females with AITD than in the general population¹³⁵. Under normal circumstances one X chromosome is randomly inactivated and the overall proportion of each copy is roughly equal. In skewed X inactivation the same copy is inactivated in 80% or more of the cells which may lead to reduced expression of some self-antigens in the thymus and loss of tolerance.

Reports of familial clustering of thyroid disease have been made as far back as the 1940's although these may be explained by shared environmental and lifestyle factors as well as a genetic contribution¹³⁶. Twin studies are more useful for determining the influence of genetics on disease progression as monozygotic twins are almost genetically identical. High rates of disease concordance between monozygotic twins indicates a strong genetic influence. The largest of these studies into GD was carried out in Denmark and found significantly higher concordance in monozygotic than dizygotic twins which have greater genetic diversity^{137,138}. Other large studies carried out in America confirmed this trend for GD and HT^{139,140}. In all these studies concordance was

not complete indicating environmental factors still play a role and while important, genetics are not wholly responsible for the disease developing.

Some of the first studies into the contribution of genetics to AITD were genome wide linkage analyses. These studies determine which genetic markers, such as short tandem repeats, are positioned close to the disease causing genes on the chromosomes¹⁴¹. The closer a marker is to the disease causing gene the more likely they are to be co-inherited. This identifies loci of interest for closer study. In most cases this type of study into the genetic influences of AITD have failed to give reproducible data¹⁴². More reliable results have been obtained from studies into the role of individual genes. These studies focus on either thyroid specific or immune regulatory genes based on the hypothesis that these are likely to have some involvement, given what is known about the nature of the disease. These studies use a statistical approach to determine which genetic markers appear more or less frequently in AITD patients compared to healthy controls. Surprisingly given its role as a major autoantigen, no link has been established between the TPO gene and AITD^{143,144}.

The TSHR gene has been shown to be linked with GD and AITD in general in several studies¹⁴⁵. Further work identified five GD associated polymorphisms in intron 1 of the gene. As this region is not protein coding two mechanisms have been suggested for its link to GD. The first proposed mechanism is that disease associated mutations may affect mRNA splicing producing more autoantigenic variants of TSHR¹⁴⁶. Measurements of full-length and truncated mRNA in thyroid tissue by Brand *et al* showed higher levels of truncated transcripts in individuals carrying the disease associated alleles. While this alone is not conclusive, these truncated transcripts could be translated to produce free TSHR extracellular domain which contains the autoantibody epitopes. Alternatively, autoimmunity to the TSHR could arise from decreased thymic expression of particular alleles¹⁴⁷. Colobran *et al* demonstrated differences in mRNA levels of different TSHR alleles in the thymus¹⁴⁷. It was also shown that in heterozygous individuals the disease associated allele was transcribed at lower levels relative to an allele linked to a lower risk of GD. Decreased mRNA transcription will lead to lower protein expression in the thymus and could result in self-reactive T cells escaping negative selection. These two theories are not

mutually exclusive and both effects may occur simultaneously to cause or exacerbate the disease.

The Tg gene is found in one of the few chromosomal loci to show linkage to AITD in multiple studies^{148,149}. Given its known role as an autoantigen, the gene was subjected to closer study. Single nucleotide polymorphisms causing amino acid substitutions in the protein sequence were found which were linked to an increased susceptibility to AITD¹⁵⁰. The authors suggest that some of these amino acid substitutions could cause structural changes which produce a more antigenic form of the protein, although this has yet to be proven.

Mutations in genes coding for some MHC class II proteins have been linked to AITD. Modelling the structural consequences of these mutations have partially explained the mechanisms behind this increased susceptibility^{151,152}. When specifically discussing humans, the MHC is also referred to the human leukocyte antigen complex (HLA). Mutations to a specific MHC class II gene, HLA-DR, can alter the shape and charge of the peptide binding groove in the expressed protein and have been linked to the incidence of both GD and HT. This alters how antigens bind to the molecule and therefore how they are presented to T cells. An example of this is the mutation of residue 74 in the HLA-DR3 β chain to arginine which has been shown to increase the risk of GD¹⁵¹. This residue is more commonly a glutamine or alanine, both of which have smaller, uncharged side chains. Modelling the expressed protein shows this mutation alters the surface charge of the peptide binding groove and thus which peptides bind to it. Hodge *et al* proposed an interaction between this HLA-DR mutation and a polymorphism conferring increased susceptibility in the Tg gene but their results narrowly failed to reach the required statistical significance to confirm a link¹⁵³.

There is a similar correlation between HT and other HLA-DR mutations. In this case the association seems more complex as mutations at several different positions are involved¹⁵². The same mechanism, alteration of the binding characteristics of the peptide binding groove, has been demonstrating by modelling but the full significance of these mutations is not yet fully understood in either disease.

Studies of polymorphisms in the immunoregulatory genes CTLA-4 and protein tyrosine phosphatase-22 have shown clear evidence of particular alleles increasing the risk of AITD and other autoimmune diseases^{154,155}. Both of these proteins down-regulate T cell activation so possession of a less active allele could contribute to the disease process. Similar studies suggest other genes such as CD40 and FOXP3 may also play a role in AITD but are less clearly linked¹⁵⁶. CD40 is expressed on B cells and APCs and is a co-stimulatory signalling receptor involved in antigen presentation and interactions between B and T cells¹⁵⁷. FOXP3 is a DNA binding transcription factor which regulates expression of several genes of the immune system¹⁵⁸. Again, inactive or less active alleles of these genes could cause immune dysfunction leading to autoimmunity.

Environmental factors can also modify the risk of developing AITD. One of these factors is diet. The micronutrients iodine and selenium play an important role in thyroid function^{159,160}. Deficiencies in these elements can be detrimental to general thyroid health and exacerbate conditions such as AITD. TPO oxidizes iodide to a reactive iodine species which iodinates tyrosine residues on Tg to produce thyroid hormones (detailed in “Thyroid peroxidase function”)¹⁶¹. Iodine deficiency can lead to hypothyroidism, goitre, reduced fertility and growth or developmental abnormalities in children¹⁶². Efforts have been made to reduce this problem worldwide through the use of iodized salt with some success¹⁶³. While this is generally beneficial in iodine deficient areas some studies have shown that iodine supplementation may increase the risk of developing thyroid autoantibodies^{164,165}. The increase in prevalence of thyroid conditions following interventions such as iodine supplementation appears to be mostly transient if iodine intake remains consistent. It seems sudden, dramatic changes in iodine intake have a greater effect on the likelihood of thyroid problems developing than consistently high or low levels¹⁶⁶. Increased iodination of Tg has been demonstrated to alter recognition by human autoantibodies and increase its potential for triggering autoimmunity in animal models which is one possible cause for this effect^{167,168}. Alternatively high iodine intake has been demonstrated to damage the thyroid in rats, possibly due to the production of high levels of free radicals¹⁶⁹.

Selenium is present as selenocysteine at the active site of several enzymes involved in regulating redox states and in catalysing the activation or inactivation of thyroid hormones¹⁷⁰. Without a sufficient supply of selenium to produce these enzymes the thyroid may be at risk of damage by free radicals. This damage could potentially trigger autoimmunity to the thyroid or exacerbate an existing condition. GD patients have been shown to have lower serum levels of selenium compared to healthy controls and while this could also be the case for HT, so far it has not been proven^{171,172}. Selenium supplementation is commonly prescribed to AITD patients, although a recent review of the literature concluded the current evidence does not support its use routinely except when a patient is shown to have a deficiency¹⁷³.

A 2002 meta-analysis of the literature at the time showed a clear link between smoking and an increased risk of GD and GO in particular¹⁷⁴. The same meta-analysis found no link between smoking and HT, however more recent studies disagree¹⁷⁵. The NHANES III study found lower levels of hypothyroidism among the smokers as did a large population study from Norway^{176,177}. The NHANES III study also found a lower occurrence of thyroid autoantibodies amongst smokers compared to non-smokers. Although this was not examined in the Norwegian study, it has been corroborated by work carried out in Denmark and the Netherlands^{178,179}.

Current estimates put the number of chemicals present in tobacco smoke at around 5000¹⁸⁰. This makes it very difficult to study the effect of smoking on AITD and as a result little is known about the mechanisms by which it influences the disease process. Cigarette smoke has been shown to increase adipogenesis and glycosaminoglycan expression by orbital fibroblasts obtained from GO patients *in vitro*¹⁸¹. This only occurs in the presence of IL-1 which suggests smoking may enhance the symptoms of rather than initiate the disease. Further work is required to fully understand the clinical relevance of these findings.

A general hypothesis used to explain the triggering of autoimmunity is molecular mimicry. This theory suggests that a bacterial or viral infection could initiate an autoimmune response by producing an antigen of similar structure to a self

molecule. There is growing evidence for this phenomenon generated from basic local alignment search tool searches in a range of human autoimmune diseases¹⁸². Studies on proteins from bacteria possibly associated with AITD have found many examples of sections homologous to thyroid autoantigens¹⁸³.

Treatment of autoimmune thyroid disease

Current treatments for GD and HT involve modifying thyroid function to alleviate symptoms¹⁸⁴. For GD this is achieved using drugs such as methimazole (MMI) or propylthiouracil (PTU) (Fig. 2) which decrease thyroid hormone synthesis by inhibiting TPO¹⁸⁵. In most cases MMI is the preferred choice of drug as most patients experience fewer side effects than with PTU¹⁸⁶.

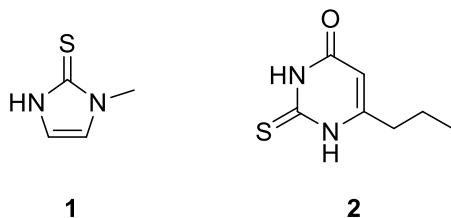


Figure 2- Structure of 1-methyl-2-mercaptopimidazole (methimazole, MMI, 1) and 6-propyl-2-thiouracil (PTU, 2).

If sufficient control of hormone levels cannot be obtained with drugs, alternative treatments are a full or partial thyroidectomy or radioactive iodine ablation of the thyroid¹⁸⁷. Both treatments permanently decrease the output of thyroid hormones. Thyroidectomy is a minimally invasive surgery also used in some cases of thyroid cancer¹⁸⁸. The gland is commonly accessed by a small horizontal incision at the base of the neck. Alternatively a transoral approach can be taken where access is gained endoscopically through the mouth¹⁸⁹. This method is relatively new and its use is not currently widespread but has several advantages including not leaving a visible scar. Thyroidectomy has the benefits of being a definitive treatment which also remedies goitre almost immediately¹⁹⁰. However this must be balanced with the costs and risks associated with surgery and the then lifelong need for treatment to replace thyroid hormones.

The alternative method of permanently stopping the production of the thyroid hormones is to destroy the thyroid with radioactive iodine¹⁹¹. Iodine naturally accumulates in the thyroid where it is essential for the production of thyroid hormones (detailed in “Thyroid peroxidase function”). On administering the radioactive isotope iodine-131 to the patient the dose of radiation is concentrated in the thyroid, limiting any adverse effects to the surrounding tissue. While this treatment can take six months or longer the majority of patients are rendered euthyroid with one treatment¹⁹². This treatment is not suitable for pregnant or breast feeding women and patients need to avoid prolonged close contact with other people, particularly children, for a few days following treatment. As with a surgical thyroidectomy this procedure frequently results in lifelong hypothyroidism requiring appropriate treatment¹⁹⁰.

The treatment for HT involves administering synthetic or, historically, porcine derived T4 to bring the circulating thyroid hormone levels up to the normal range. This does not completely mimic physiological conditions and in some cases T4 levels need to be increased up to or beyond the upper end of the normal range to normalise T3 and TSH levels¹⁹³. However, there is currently no evidence that a combination T3 and T4 replacement regime has any advantages over T4 only. Current treatments only deal with the symptoms caused by AITD, there are no proven curative treatments for AITD at present.

The thyroid

The thyroid is an endocrine gland found at the base of the neck in humans, curving around the trachea¹⁹⁴. It consists of 2 lateral lobes joined by a band of tissue known as the isthmus. The gland contains many follicles which are surrounded by thyrocyte cells.

The function of the thyroid is to produce the hormone T3 and its prohormone T4 (Fig. 3) when stimulated by TSH. T4 is converted to T3 by deiodinase enzymes in muscle, kidney and liver tissue¹⁹⁵. T3 binds to the thyroid hormone nuclear receptor which then binds to enhancers in the promoter regions of target genes¹⁹⁶. This produces a number of physiological effects including increasing

heart rate, gut motility, muscle contraction and relaxation and bone resorption. Thyroid hormones are also crucial during foetal development for regulating the development of the heart, brain and lungs as well as playing a role in controlling metabolism and the endocrine system¹⁹⁷.

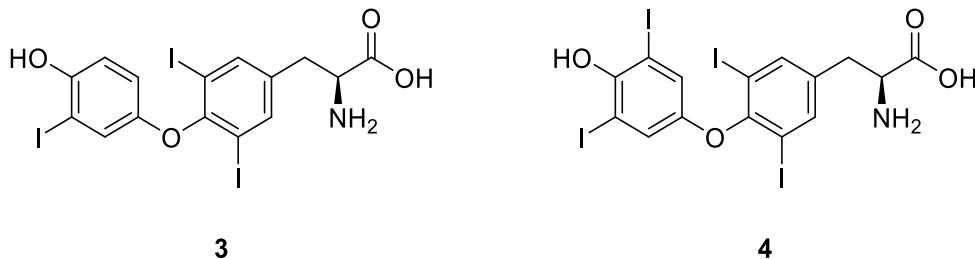


Figure 3- Structure of 3,3',5-triiodothyronine (T3, 3) and thyroxine (T4, 4).

Thyroid hormone synthesis takes place in the follicular lumen of the thyroid. Most of the thyroid hormone output is in the form of T4¹⁹⁸. T4 is less physiologically active than T3 but has a longer half-life in circulation¹⁹⁹. Both hormones are the product of two TPO mediated reactions on Tg (detailed in “Thyroid peroxidase function”).

Once T3 and T4 are synthesised they are released from Tg and enter circulation. Tg is internalized in vesicles that enter the thyrocytes either by pinocytosis or receptor mediated endocytosis²⁰⁰. Once internalized Tg containing vesicles fuse with lysosomes which contain proteolytic enzymes. These enzymes digest Tg, releasing the hormonogenic residues²⁰¹.

Thyroid peroxidase function

TPO is a membrane bound enzyme found on the apical membrane of thyrocytes. It is orientated so its active site extends into the follicular lumen²⁰². Its function is to catalyse both the iodination of tyrosine and the coupling of iodotyrosines to produce T3 and T4²⁰³. Many aspects of the catalytic cycle of TPO have not been studied directly, but can be inferred from work with other enzymes as key residues and structural elements of the active site appear to be conserved amongst other members of the peroxidase family²⁰⁴.

TPO contains a covalently bound haem prosthetic group which is vital for correct folding, intracellular trafficking and enzymatic activity²⁰⁵. The first step in the catalytic cycle is the oxidation of the iron atom at the centre of the haem group. The resting state of the enzyme is the 3+ ferric form (Fig. 4)¹⁶¹. Hydrogen peroxide binds to the haem iron and undergoes heterolytic cleavage to produce a 4+ oxoferryl group and a single water molecule in a two electron oxidation with the second electron is donated by the protoporphyrin ring which forms a π cation radical (Fig. 5).

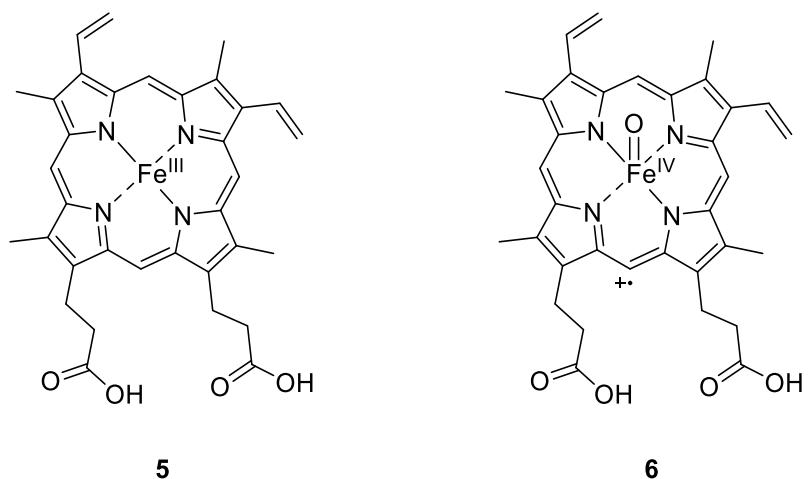


Figure 4- Structure of haem in the 3+ resting state (**5**) and the 4+ oxoferryl form following oxidation by hydrogen peroxide (**6**).

The side chains of Histidine 239 and Arginine 396 play an important role in this reaction (Fig. 5)²⁰⁶. The histidine residue serves as an acid-base catalyst transferring a proton from the iron bound oxygen atom of H₂O₂ to the second oxygen atom, which then leaves as water²⁰⁷. The arginine residue polarises the bond between the two oxygen atoms which promotes the heterolytic cleavage of that bond²⁰⁸.

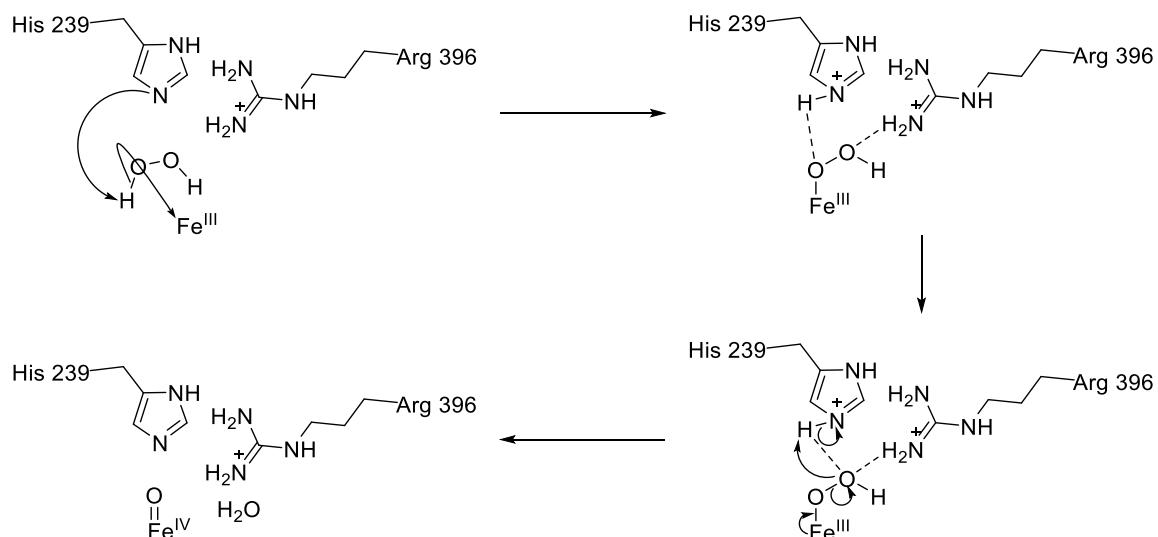


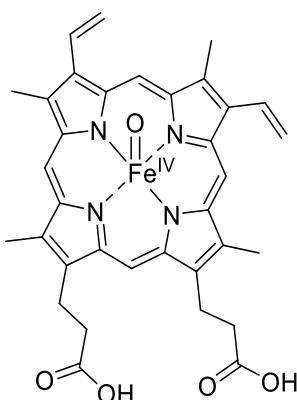
Figure 5- Oxidation of the TPO active site by hydrogen peroxide²⁰⁸.

When describing peroxidase activity the two electron oxidised form of the enzyme is referred to as compound I (active site haem as Fig. 4(6)). If a suitable substrate is not available the π cation radical form of compound I will isomerise to a more stable protein radical²⁰⁹. An amino acid, likely a tyrosine or tryptophan residue close to the active site, accepts an electron from the protoporphyrin ring as has been shown to occur in related peroxidase enzymes myeloperoxidase (MPO) and lactoperoxidase (LPO)^{210,211}. This process has been shown to occur in TPO, but the acceptor residue has not been identified.

Both the iodination and coupling reactions are thought to be performed by the π cation radical form of compound I¹⁶¹. The coupling reaction can be inhibited by high concentrations of iodide, indicating the two reactions are competing for the same form of compound I. Given the constant supply of iodide and other substrates in the thyroid, it has also been suggested that the protein cation radical does not form under physiological conditions^{161,212}.

Compound I can return to the resting state enzyme by at least two series of reactions. A single, two electron oxidation of iodide (I^-) to hypoiodite (IO^-) returns the enzyme directly to the resting state²¹³. If a single electron oxidation of a substrate occurs, the protoporphyrin ring is reduced to form compound II (Fig 7)²¹⁴. Compound II still contains one oxidizing equivalent in the form of the 4+ oxoferryl group. A second, single electron redox reaction returns compound

II to the resting state of the enzyme. These reactions can produce molecular iodine (I_2) by oxidizing two iodide ions. There are several other potential iodine species which could be produced by iodide oxidation and subsequent reactions of the products which would be capable of iodinating Tg. However iodine and hypoidous acid (IOH) have been shown to be the most abundant products produced by TPO-mediated iodide oxidation²¹⁵.



7

Figure 6- Structure of active site haem in compound II (7)²¹⁶. Compound II is formed following a single electron oxidation of a substrate by compound I. Note the protoporphyrin ring has gained an electron when compared to compound I (Fig. 4(6)).

Whilst not proven conclusively, iodination of a substrate by TPO is very likely to be a non-specific reaction. No crystal structure exists for TPO but high resolution structures of MPO and LPO have been obtained^{217,218}. These enzymes both share a sequence identity of around 48% with the peroxidase domain of TPO (detailed in “Related peroxidase enzymes”). In both of these structures the active site is accessed through a channel approximately 15 Å deep and 10-15 Å at the widest point. This would make it difficult for a tyrosine residue in Tg to access the active site to react specifically with an enzyme bound iodine intermediate.

Enzymatic iodination of tyrosine can occur at multiple consensus sequences the simplest of which requires only a glutamic acid or aspartic acid residue preceding the tyrosine²¹⁹. As such there is potential for many proteins other than thyroglobulin to be iodinated and act as a substrate for hormone

generation. This was demonstrated in early work by Coval & Taurog which showed a peroxidase enzyme isolated from porcine thyroid could catalyse the iodination of both Tg, bovine serum albumin (BSA) and free tyrosine residues²⁰³. Iodinated proteins other than Tg and lipids can be detected in thyroid follicles further indicating this reaction is non-specific^{220,221}. The structure of the thyroid follicular lumen confines the iodination reaction, preventing inappropriate iodination of proteins. It has been hypothesised that iodination of other thyroid follicle proteins may provide an iodine storage capacity which is independent of the turnover of Tg, although this has not been proven.

The structure of Tg allows the next step of hormone synthesis, the coupling reaction, to take place²²². During this step each of the two iodotyrosines must be in the correct orientation for the coupling reaction to take place. Tg can be iodinated at many different sites but despite containing 67 tyrosine residues per monomer only 4 are directly involved in producing hormones²¹⁹. The iodinated but non-hormonogenic sites of Tg are thought to act as an iodine store. Tg is present at very high concentrations in thyroid follicles, mostly in an insoluble, multimeric form which is solubilised when the supply of dietary iodine is insufficient.²²³.

The iodinated side chains of the two tyrosine residues are joined by the coupling reaction resulting in a peptide bound T3 or T4 residue and a dehydroalanine residue^{224,225}. Free diiodothyronine has been shown to increase the amount of hormone generated by this reaction *in vitro*^{226,216}. It is likely that the free diiodothyronine residues, which would also be found in thyroid follicles, have a role in transferring electrons to the Tg bound iodotyrosines. As discussed above it would be very difficult for Tg bound residues to access the active site and react directly with an enzyme bound intermediate²²⁷.

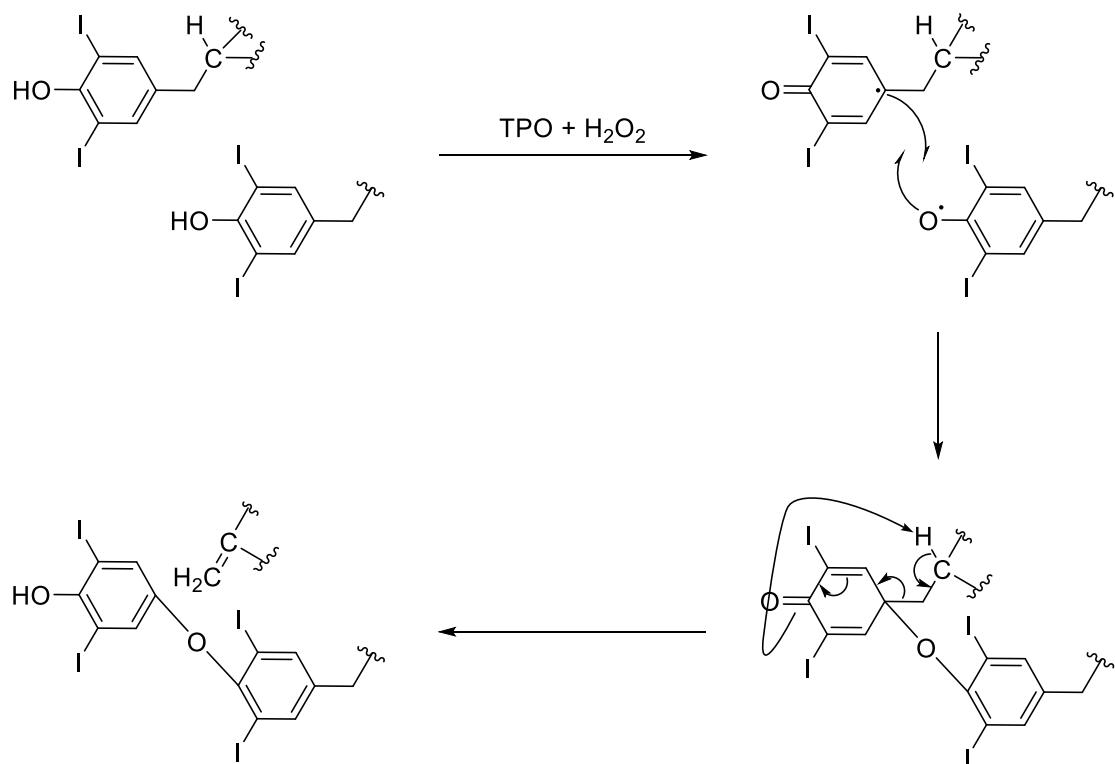


Figure 7- Iodotyrosine coupling reaction to form T3/T4 as proposed by Taurog *et al*²⁵.

Thyroid peroxidase gene structure

The gene coding for TPO was first isolated from GD patient thyroid tissue and sequenced in 1987²²⁸. It is located on chromosome 2 in humans and consists of 17 exons and 16 introns²²⁹. Two cDNA sequences were isolated and designated hTPO-1 and hTPO-2. Analysis of the sequences showed hTPO-2 contained a deletion of 171 nucleotides corresponding to exon 10. This was attributed to alternative mRNA splicing after transcription of a single gene. Later work showed that both hTPO-1 and hTPO-2 mRNA was also present in normal thyroid tissue with hTPO-1 being the predominant form²³⁰. Expression of both mRNAs in Chinese hamster ovary (CHO) cells showed hTPO-2 coded for an inactive form of the protein which was not trafficked to the cell surface²³¹. Comparison of translated protein sequence of hTPO-1 and hTPO-2 to the MPO protein sequence suggests that the missing section of hTPO-2 may alter the structure of the active site rendering this form inactive²³².

A number of other TPO isoforms have subsequently been identified, generated through alternative splicing of mRNA. These include hTPO-3 missing residues encoded by exon 16, hTPO-4 missing exon 14 and hTPO-5 missing exon 8^{233,234}. RNA species with multiple missing exons have also been detected²³⁴. These are all expected to be inactive as one of the missing exons is exon 10 in each case, the same exon missing that results in the inactive hTPO-2 isoform.

Genetically it appears that TPO evolved separately from MPO, LPO and eosinophil peroxidase (EPO). The genes for these other peroxidases all appear closely clustered on chromosome 17. The intron-exon structure of all three are very similar, but differ from that of TPO. This suggests that these enzymes share a common ancestor and arose from gene duplication²³⁵.

Related peroxidase enzymes

TPO is a member of the peroxidase-cyclooxygenase superfamily of haem peroxidases²³⁶. Historically this family was first named the “mammalian”, then “animal” peroxidase family as more homologous enzymes were identified²³⁷. All members of this family contain a covalently bound haem group and most are multi-domain proteins. In other haem peroxidases the haem group is bound to the protein through non-covalent interactions²³⁶. At the time of writing over 15,000 protein sequences were assigned to this family in the InterPro database²³⁸.

The most closely related proteins to TPO are MPO, LPO and EPO. These enzymes are involved in the immune response, producing oxidizing agents to damage pathogens^{239–241}. They are found in neutrophils, milk and saliva and eosinophils respectively. Each of these proteins is cationic. Unlike the membrane bound TPO, MPO, LPO and EPO are all soluble proteins lacking the complement control protein like (CCP), epidermal growth factor like (EGF) or membrane spanning domains found in TPO²⁰⁴. All four enzymes oxidize a range of substrates at varying rates. These include halides, pseudohalides such as thiocyanate and aromatic molecules such as tyrosine or guaiacol²⁰⁴. The structures of MPO and LPO have been solved by X-ray crystallography^{218,242}.

TPO	1	-----MRAALVLSVTIVMACTEAFFPFISRGK-ELWCK
MPO	1	MGVPFFSSLRCMVLDLGPWCAGGLTAE MKILLALAGLLAILAT ---POP SEGAAPAVICE
LPO	1	-----MRVLLHLPALLASLIL---LOA---AASTRRAQT
TPO	34	P ESRVSSVIEESKR LVDTAMYATMQRNLLKK---RGII SPAQLLSFSKLPEPTSGVIARA
MPO	57	V ITSLVLSSMEEAKQ LVDKAYKFR-REST IKQR LRSGSASPME LLSYFKQPV ATRTAVRA
LPO	29	TRTSAIS SDTVSQAKVQVNKAFLDS -RTRLKTAMSSE FTSRQLSEYLKHAK RTRTAIRN
TPO	91	AEIMETS IQAMKRKVNLK TQQSQHPTDAI SED-LLS IIIANMSGGLPYMLPPKCPNTCLAN
MPO	116	ADYLHVADILERK KURSLWRRPFNVTDVLTBA-QLN VLSKSSGCA YQDVGWTICPE---QD
LPO	88	GQVWEES SKRLRQKASL -----TNVTD ESLDLTSLSLEV GCCAPAPVVRCDP---CS
TPO	150	KYRPIT GCACNNRD PRWGASNTALARWLP PVYEDGFSQPRGWNP GFLYNGFPLPPVREVT
MPO	172	KYRTIT GMCNNRRS PTL GASNRAFVRWLPAEYEDGFSLPY GWTPGVKRNGFP VALARAVS
LPO	137	Y RTITG C CNRRK P ALGAANRALARWLPAEYEDG SLPFGWT PGKTRNGFP PLAREVS
TPO	210	R HVIQVSNEVVTD DRYSDL IMA WGQY IDHDIAFT TPQ STSKAEGGGAD C QMTCE NQNPC
MPO	232	N EVVR F PTQLTPDQE RS LMFMQW GQL LDHLDFTP FAARAS EVIGVNC E TSCVQOP PC
LPO	197	N KIVGY LN EGVLDQNRSLL FMQ WQ QIV DHLDFA P DTELGSS E YSKAQCD EYCI QGDNC
TPO	270	FPIQL PEE-ARP A GTA CLP FYRS SAAC CTG D QGAL F GNL STAN RQMN ELTS SF LDAST
MPO	292	F PLK I PPND P RIK N QAC I PFFR SC ACP GSN-----IT RNQ INAL TSF V DASM
LPO	257	FPIMF PPNDPK A G T QCKC M PFFR AGF V CPT PPYK -----SL ARE QINAL TSF LDAS F
TPO	329	V YGS SP AER QLRN WTS AE GLL RV HARI RD S GRAY LPFVPPR A AA ACA PEPGIPGET RG
MPO	342	V YGS SE P LA RN LR NMS SQL GLL AVN QR F QDN GR AL LF FDNLH --DDEC---LLT NR SAR IP
LPO	309	V Y SSE PSL AS RLRN SS PLGL MA VNOE V DHGLP Y LPY DS SKK -E SPC ---EE INT TAR VP
TPO	389	C FLAGD G RASE V PSL T AHL TL L REHN R LA A ALK A LN A H W S ADA V YY Q EAR K V G GAL H QII
MPO	398	C FLAGD I TR S EMPE LT SMHT LL L REHN R LA T EL K SL N P RD G E R LY Q EAR K V G GAM V QII
LPO	365	C FLAGD R RASE H HIL LA T SH T FL R EHN R LA R E L K R LN Q W D G E R LY Q EAR K V G GAF V QII
TPO	449	T L R DY I P R I L G P E A F Q Q Y V G P Y E Y G D S T A N H T V S N V F T A FR F G H A T I P L V R R L D A S E
MPO	458	T YR D Y L P L V L G P T AM R K Y L P T Y R S Y N D V P R I A N V F T -NA F RY G H T L I Q P F M F R L D N Y
LPO	425	T F R D Y Y L P I I L G D H -H MQ K W I P P Y Q G Y S E S V D P R I S N V F T -F A F R F G H L E V P S S M F R L D E N Y
TPO	509	Q EH P D L P C I W L H Q A F F S P W T L L R G G G I D P L I R G L A R P A K L O V Q D Q L M N E E I T R E L F V L S
MPO	517	Q P M E P N P R V I PL S R V F A S W R V V L E G G I D P E I I R G L M A P K A L N R Q N Q I A V D E I T R R E L F E Q V
LPO	483	Q P W G P E P P I I L F F F T W R M V K D G G I D P L I R G L A K S K L M K Q N K M T G E L R N K L F Q P T
TPO	569	N S S-T L D L A S I N L Q R G D H G L P G Y N E W R F C G G L P R E T P A D L S I A J A S R S M A D K I L D L Y K
MPO	577	M R I I G L D I A P L A N N M Q R S R D H G H L P G Y N A W R R F G P Q P E T V G Q L G T V R L N K L A R K L M E Q Y G
LPO	543	H R I H G F D L A A I I N T Q C R D H G Q P G Y N S W R A F C D L S Q P O T L E E L N T V L K S K M L A R K L L G Y G
TPO	628	H P D N I D W I G G L A E N F P R A T G P L F A C L I G K Q M A L R D G D W F W E N H V F T DA Q R R E L
MPO	636	T P D N I D I W M G G V SE P I K R G R V G P L L A C I I G T Q F Q R L D G D R F W W E N E G V F S M Q Q R Q A A
LPO	603	T P D N I D I W I G A I E P L V R G V P L L A C I I G K Q F Q Q I R D G D F W W E N P G V F T N E Q K D S I Q
TPO	688	K H S L S R V I C D N T G I T E V P M D -A F Q V G K F P E D F E C D S I T G M N I E A W R E
MPO	696	Q I SL P R I I C D N T G I T V S K N I E M S N Y P R D F V N C S T I P A L N I S W R E
LPO	663	K M S F S R L V C D N T R I T K V P R D -E F W A N S Y P P D F V D C S A I D K L D I E S W A S

Figure 8- Alignment of human TPO, MPO and LPO protein sequences, ■ identical residues □ similar residues.

Human TPO shares 43.7% amino acid sequence identity with human MPO and 41.7% with human LPO if the entire protein sequence is considered. The most homologous regions of these proteins are the central residues, particularly around the active site. The sequence identity of the TPO peroxidase domain, comprising of amino acids 142-734 to the MPO and LPO sequences are 48.4% and 47.4% respectively.

Myeloperoxidase

Canine MPO was the first to be successfully crystallized and its structure was solved in 1992. The human protein sequence was fitted to the electron density data to produce a structure with a resolution of 3.0 Å. While this gave a good overview of the protein tertiary structure, this data lacked electron density in some regions and the resolution was too low to fully characterise the active site haem. Later work produced higher quality diffraction data with a 1.8 Å structure of human MPO being produced in 2000²¹⁷.

The mature MPO protein is a homodimer of 146 kDa formed from two copies of a single, 745 amino acid translation product²⁴². Each half of the protein undergoes post-translational cleavage to remove a small, six amino acid peptide²⁴³. This produces a large and small polypeptide linked by covalent bonds to the haem prosthetic group. The protein has six intermolecular disulfide bonds with the two halves of the dimer joined by a single disulfide bond between cysteine 319²⁴². The protein secondary structure consists predominantly of α-helices. There are 19 in total with five surrounding the active site, four from the large polypeptide and one from the small. MPO has five potential sites for N-linked glycosylation all of which are glycosylated with complex or high-mannose type glycans under normal conditions²⁴⁴. The MPO crystal structures suggested that only some of these actually have sugars attached, but this was disproved by later studies utilizing mass spectrometry. It is hypothesised that glycosylation of MPO affects the overall enzyme structure as deglycosylation of MPO has been shown to decrease enzymatic activity²⁴⁵. There is also a calcium binding site with pentagonal bipyramidal geometry²¹⁷. This is situated in a loop between residues 168-174 with asparagine 96 also forming a bond to the calcium atom. This site is structurally important and

conserved amongst all the related peroxidases including TPO²⁰⁴. Mutating asparagine 96 to alanine in MPO or making the equivalent mutation in LPO has been shown to almost completely inactivate the enzyme²⁴⁶.

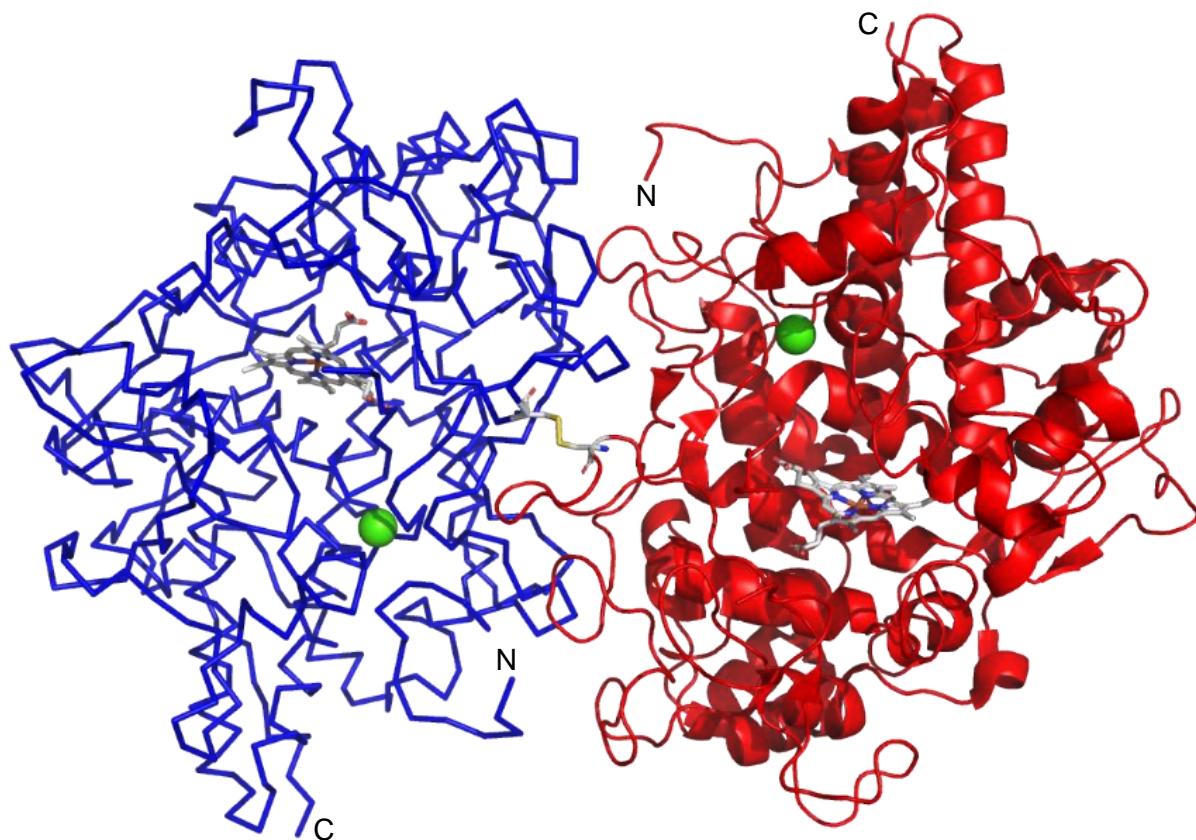


Figure 9- MPO structure (PDB 1CXP). Identical subunits shown as ribbon (left, blue) and cartoon (right, red) depictions. Haem group, calcium and cystine 319 coloured by element (Carbon-grey, nitrogen-blue, oxygen-red, sulphur-yellow, iron-orange, calcium-green).

Lactoperoxidase

LPO is present in secretions such as saliva, tears and milk so a common method of obtaining the protein is to purify it from unpasteurized milk. For this reason most studies on LPO use the enzyme from bovine or caprine sources. Due to its antimicrobial action and the ease of obtaining it LPO, is used commercially in both the food and medical industries²⁴⁷. A 2.4 Å resolution crystal structure for caprine LPO was first produced in 2008²¹⁸. There are now over 60 lactoperoxidase structures deposited in the protein data bank at the time of writing. These include enzymes from *Bos taurus* (cattle), *Capra hircus* (goat) and *Bubalus bubalis* (water buffalo) but not human LPO. These structures

have been produced of the protein alone and in complex with a variety of inhibitors and substrates.

LPO consists of a single polypeptide chain of 712 amino acids^{235,248}. The human enzyme has a molecular weight of 80 kDa but the bovine form is slightly smaller at 78.5 kDa. Overall the structure of LPO is similar to the MPO monomer, particularly the active site as evidenced by crystal structures and the multiple shared enzyme substrates²⁴⁷. The LPO secondary structure contains 20 α-helices with a short α-helix between amino acids 124 and 132 having no equivalent in MPO²¹⁸. This sequence aligns with the six amino acids cleaved from MPO post translation. No cleavage of LPO occurs explaining this small structural difference. LPO has seven disulfide bonds, however in LPO the equivalent residue to cysteine 319 in MPO is a glycine, preventing dimerization at this site²⁴⁷. The only unpaired cysteine residue in LPO is at position 441 but this is buried in the protein structure and unavailable for disulfide bonding. All five potential N-glycosylation sites are glycosylated with complex, hybrid or high-mannose type glycans²⁴⁹.

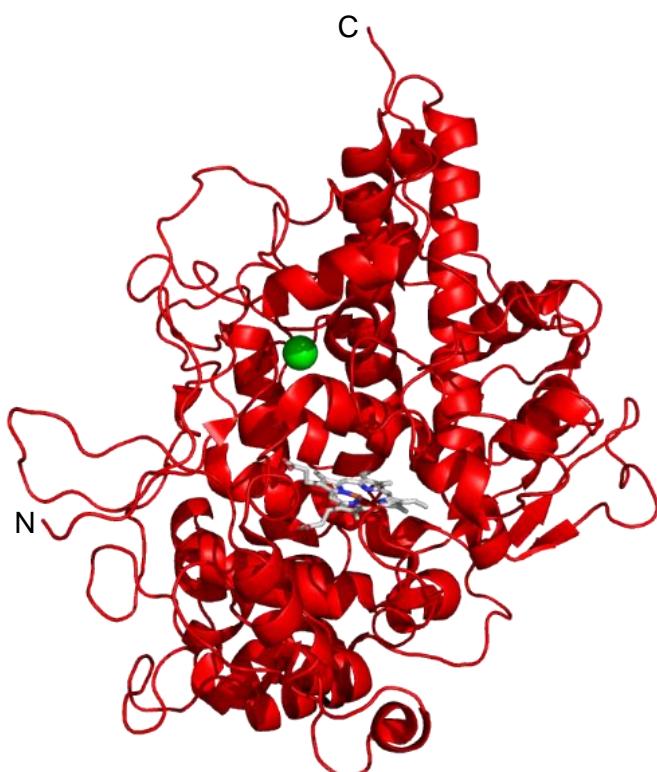


Figure 10- Caprine LPO structure (PDB 5FF1) in cartoon depiction. Haem group and calcium coloured by element (Carbon-grey, nitrogen-blue, oxygen-red, sulphur-yellow, iron-orange, calcium-green).

Eosinophil peroxidase

There is currently no crystal structure for EPO. The protein is monomeric and consists of two subunits produced from a single chain precursor in a similar manner to MPO²⁵⁰. EPO is slightly smaller than MPO, consisting of 715 amino acids with a molecular weight of approximately 70 kDa²⁵¹. Biochemical investigations into the structure of EPO have determined that the pattern of disulfide bonding is similar to that found in MPO²⁵². It is glycosylated at two sites from a potential four although the nature of these glycans has not been investigated. The haem group has been shown to be bound to aspartic acid 93 and glutamic acid 241 through ester bonds²⁵³. This agrees with predictions from sequence alignments and indicates the active site homology observed in MPO and LPO is also likely to be conserved in EPO.

Thyroid peroxidase protein structure

The TPO protein consists of 933 amino acids when expressed and can be divided into six domains²²⁸. Although TPO has previously been crystallized, there is currently no high resolution crystal structure for the protein^{254,255}. Crystals produced have either been unsuitable for X-ray diffraction studies or have not diffracted to sufficiently high resolution for the structure to be solved. Much of the current knowledge of the structure of TPO has been gained from sequence alignments and homology models based on related peroxidase enzymes^{256,257}. Models of the native TPO protein, such as those produced by Hendry and Hobby *et al*, utilise separate template structures for each domain as there is no single protein known which shares a high level of homology across the entirety of TPO^{256,258}. The positioning of the individual domains of the protein relative to each other is difficult to determine precisely by modelling in this way. More recent work by Le *et al* has considered information on enzyme function and autoantibody binding sites to generate improved models²⁵⁷.

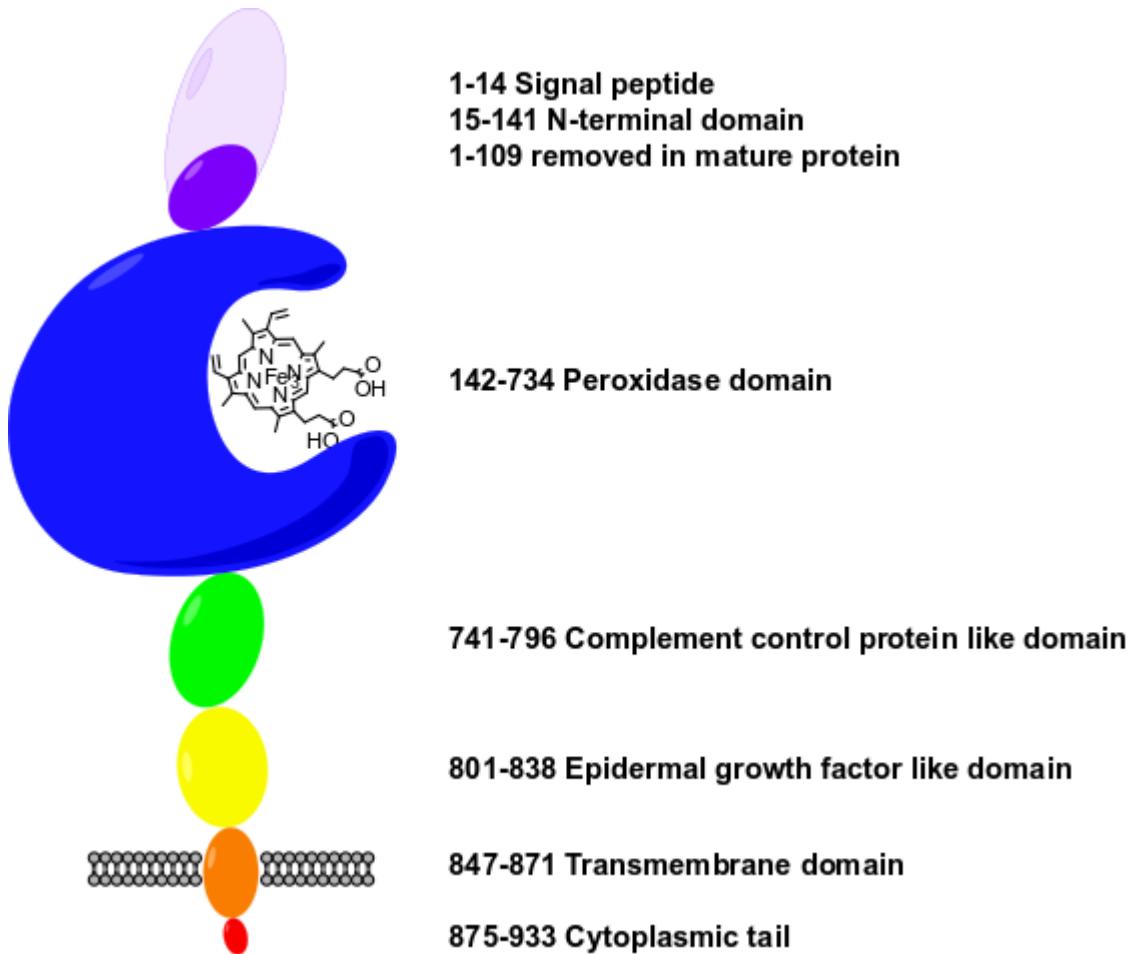


Figure 9- Human TPO domain organisation.

N-terminal domain

The TPO N-terminal domain shares little homology with other peroxidases. This domain consists of 127 amino acids when the protein is first expressed but is trimmed to 32 amino acids in the mature protein²⁵⁹. This is thought to occur in the endoplasmic reticulum and as a result the N-terminal amino acid of the mature protein is threonine 109. A 14 amino acid signal peptide required for trafficking and targeting for secretion is also removed during this modification²⁶⁰. The propeptide spanning amino acids 15-108 does not play an obvious role in expression, as enzymatically active and correctly folded TPO is still trafficked to the cell surface if this section of the gene is removed²⁶¹. Levels of expression, glycosylation, peroxidase activity and reactivity with patient autoantibodies were all comparable to the wild type protein when TPO was expressed in CHO cells.

Peroxidase domain

The largest domain of TPO is the peroxidase domain consisting of 592 amino acids. This region contains the enzyme active site and covalently bound haem. In this region there are two extensions to the TPO protein sequence which are not present in MPO or LPO between amino acids 303-310 and 376-378. The location of these extensions would suggest an elongation of a surface exposed loop which is unlikely to cause a major change in the overall protein structure²⁵⁶.

The haem group found at the TPO active site is thought to take the form of a modified haem b as found in LPO²⁰⁴. In LPO and EPO the haem is bound to the peptide chain through ester bonds to an aspartic acid residue and a glutamic acid residue^{218,253}. From sequence alignments the haem group in TPO is thought to be bound to the enzyme by ester bonds with aspartic acid 237 and glutamic acid 399²⁰⁴. The haem group of MPO forms two equivalent ester bonds to the peptide and a sulfonium ion linkage with a methionine residue²¹⁷. This bond is not formed in TPO, LPO or EPO as there is no equivalent methionine residue in the protein sequence²⁰⁴.

The presence of haem in proteins produces a peak in the Soret region of the UV-Vis absorbance spectra²⁶². In TPO, LPO and EPO the maxima of this peak is around 412 nm, in MPO the soret peak is shifted to 430 nm²⁰⁴. This is due to the third covalent bond distorting the pyrrole from planar²⁶³. Disrupting the sulfonium ion bond by mutating methionine 243 to the equivalent residue found in TPO, LPO or EPO (valine, glutamine or threonine respectively) shifts the soret band closer to the value observed in the other peroxidases²⁶⁴.

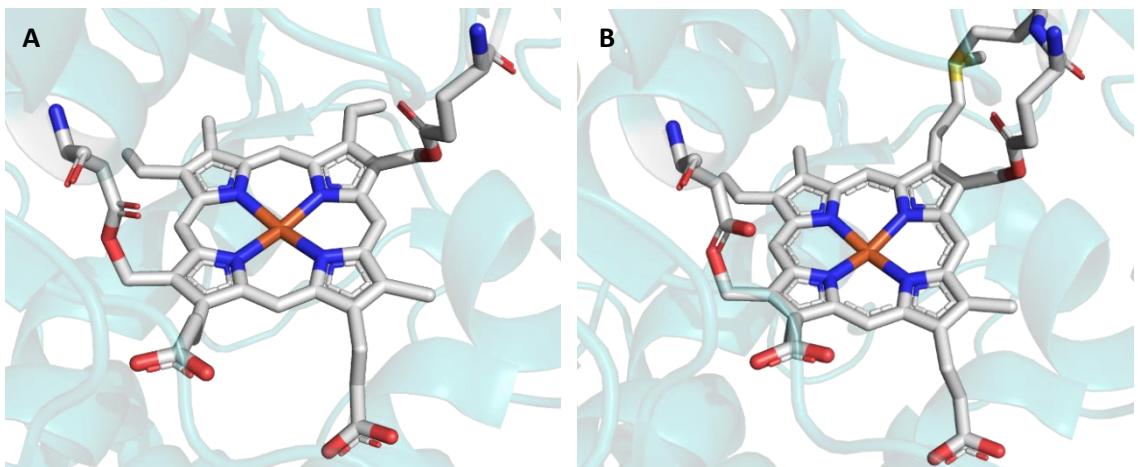


Figure 12- Active site haem of LPO (PDB: 5FF1, **A**) and MPO (PDB: 1CXP, **B**). Protein-teal, haem and connecting residues coloured by element (Carbon-grey, nitrogen-blue, oxygen-red, iron-orange).

Complement control protein like domain

The extra-cellular portion of TPO contains a CCP domain positioned at the C-terminal end of the peroxidase domain. CCP domains are also known as short consensus repeats or sushi domains. Klickstein *et al* used DNA probes to demonstrate that regions of the human complement protein CR1 are made up of repeating sequences of this domain²⁶⁵. Further work by Reid & Day used a computer program to identify other proteins, including TPO, which contained this sequence and showed it is not exclusive to proteins involved in the complement system²⁶⁶.

CCP domains are usually around 60 amino acids in length²⁶⁷. They consist of several β-sheets running roughly parallel to the long axis of the molecule with several conserved hydrophobic residues in the centre. One of the characteristic features of these domains is two disulfide bonds joining four cystine residues in a 1-3, 2-4 pattern. CCP-like domains are often found as a series of repeating domains, although there is only a single CCP domain in TPO. Two proteins are known to consist entirely of repeating CCP domains; the complement protein factor H and vaccinia virus complement control protein (VCP). Factor H is one of the many proteins involved in regulating the complement cascade. Its role is to prevent complement proteins from damaging normal host cells²⁶⁸. VCP also

inhibits complement activation. In this case, virus survival is improved as it prevents infected host cells, which are producing more virus, from being attacked by the complement system²⁶⁹. Structures for individual and pairs of CCP domains from both of these proteins have been produced from nuclear magnetic resonance (NMR) data^{270,271}. These show that despite the relatively low sequence similarity between each domain (between 14% and 28%) the overall structure of each domain overlaps quite closely²⁷¹.

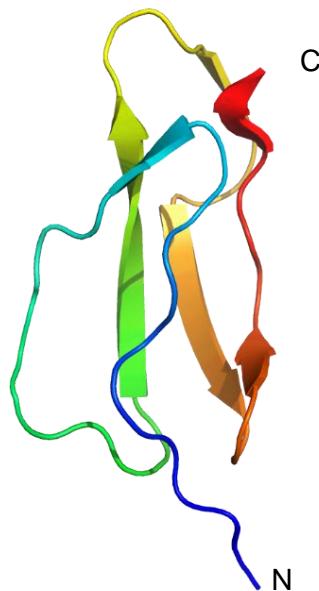


Figure 13- NMR structure of 16th CCP module of human factor H (PDB: 1HCC) coloured blue to red from N to C termini.

The CCP domain found in TPO consists of 56 amino acids. It shares 26.0% amino acid sequence identity with the 16th CCP domain of factor H and 34.6% with the 3rd CCP domain from VCP.

TPO	741	---	KCGFPESVENGDFVHCEES	--GRRVLEVYSCRHGYELQGREQITCTQEGWDFQPFPLCKD	796
FH	909		EGLPCKSPPEISHGVVAHMSDSYQYGEETVTYKCFEGEGIDGPAIAKCLGEKWSHP	-PSCI-	966
VCP	146	--	VKCQSPPSISNGRHNGYEDFYTDGSVVTYSNNSGYSLIGNSGVILCSGGEWSDP	-PTQOI	203

Figure 14- Alignment of TPO CCP domain, factor H 16th CCP domain and VCP 3rd CCP domain protein sequences. █ identical residues █ similar residues.

Epidermal growth factor like domain

The last extra-cellular domain of TPO is an EGF domain which is located on the cell surface. EGF-like domains share structural homology with epidermal growth factor and are found in a large variety of animal proteins. These include extracellular matrix proteins, blood clotting factors and other growth factors²⁷². The amino acid sequence of murine EGF was deduced by Edman degradation in 1970²⁷³. Once the EGF cDNA sequence was obtained database searches showed several proteins contained sections with a similar sequence²⁷⁴. More recently, structures of EGF and EGF-like domains have been produced by NMR and X-ray crystallography^{275,276}.

EGF domains appear most commonly in the extracellular region of membrane proteins or in secreted proteins²⁷². The domain is typically around 30-40 amino acids long and contains two β-sheets. There are three disulfide bonds between cysteine residues 1-3, 2-4 and 5-6. Some EGF domains also contain a calcium binding site made up of five amino acids. The calcium binding motif takes the form aspartic acid/asparagine-x-aspartic acid/asparagine-glutamine/glutamic acid-X-aspartic acid/asparagine-X-tyrosine/phenylalanine where x is a single amino acid and X is a variable number of amino acids²⁷⁷.

The tertiary structure and calcium binding sites of EGF domains have been shown to be functionally important, as mutations disrupting these features are linked to disease. For example, mutations in the blood coagulation factor, factor IX cause haemophilia B²⁷⁸. Loss of function in this protein disrupts the coagulation cascade which produces blood clotting in response to injury. Mutations to another EGF domain containing protein, fibrillin-1, are linked to Marfan syndrome²⁷⁹. Marfan syndrome causes weakening of the connective tissue throughout the whole body, the most serious complication of this being damage to the heart valves²⁸⁰.

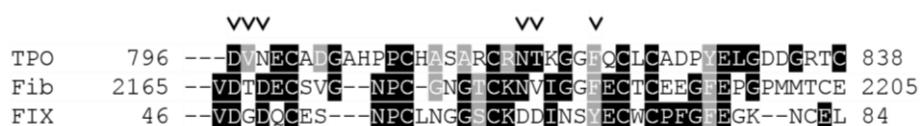


Figure 14- Alignment of TPO EGF domain, fibrillin 33rd EGF domain and factor IX first EGF domain protein sequences. █ identical residues █ similar residues v calcium binding motif residues.

In human TPO the EGF like domain consists of 43 amino acids. It shares 41.4% amino acid sequence identity with the 33rd EGF domain of fibrillin and 26.3% with the first EGF domain from factor IX. Based on sequence alignments and the structures of these proteins, the calcium binding motif is expected to consist of aspartic acid 796, asparagine 798, asparagine 816 and phenylalanine 821 in the TPO EGF domain, with additional contacts to the calcium atom made through the main chain carbonyl oxygens of valine 797 and threonine 817.

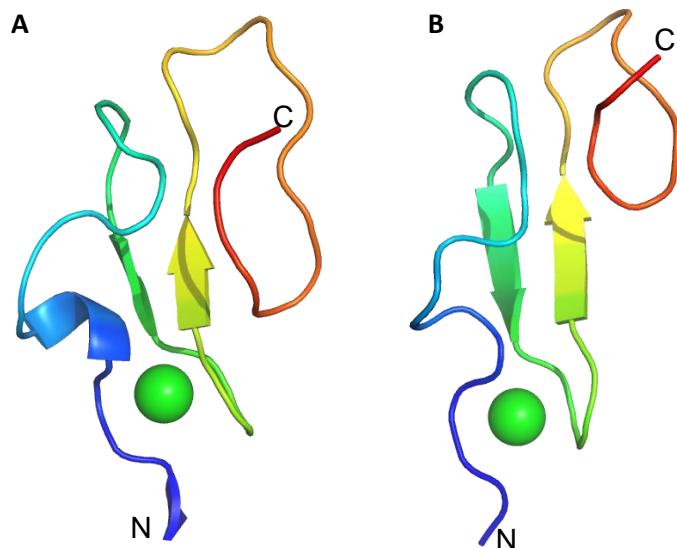


Figure 16- NMR structure of 33rd EGF domain of human fibrillin (PDB: 1EMN, **A**), crystal structure of first EGF domain of human factor IX (PDB: 1EDM, **B**). Coloured blue to red from N to C termini, calcium shown as green sphere.

Unlike the peroxidase domain, the CCP and EGF domains do not have a clearly defined function in TPO. The TPO isoform hTPO-4 which lacks the EGF domain due to alternative mRNA splicing has been shown to be trafficked to the cell surface where it is enzymatically active and recognised by monoclonal TPO antibodies²³⁴. A plausible model of this isoform has been produced suggesting it can adopt a similar confirmation to the full-length protein²⁵⁷. One theory is that these domains act as physical spacers to shift the peroxidase domain away from the cellular membrane as they play a similar role in other proteins^{266,272}. This has yet to be confirmed but is possibly weakened by the evidence from the study of hTPO-4.

Transmembrane domain

The transmembrane domain of TPO consists of 24 amino acids which are predicted to form an α -helix²⁵⁷. This helix contains a glycine-x-x-x-glycine motif between residues 860 and 864, a pattern shown to be common in associating transmembrane α -helices²⁸¹. The fourth residue after glycine 864 is threonine (868) which increases the likelihood of dimerization. The sequence glycine-x-x-x-glycine-x-x-x-threonine is one of the more commonly found extensions of the glycine-x-x-x-glycine motif present in associated helices. While TPO appears to form dimers when in its native, membrane-bound state, the extra-cellular portion of TPO expressed recombinantly has been shown to be a monomer^{282–284}. This supports the theory that formation of a TPO homodimer occurs due interactions between the transmembrane domains of two separately expressed TPO molecules.

At the C terminus there is a 58 amino acid cytoplasmic tail extending into the thyrocyte cytoplasm. This section of the protein is hydrophilic and is predicted to have a random coil secondary structure²⁸⁵.

Glycosylation

The TPO protein sequence contains four N-linked glycosylation sites at asparagine 129, 307, 342 and 569²²⁸. TPO purified from human thyroid glands contains a mixture of complex and high-mannose type glycans²⁸⁶. In contrast, porcine TPO appears to only have high-mannose type glycans attached²⁸⁷. TPO purified from tissue or obtained from recombinant sources commonly produces a double band when visualised by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)^{288,289}. The cause of this has not been proven conclusively, but co-purification of different glycoforms has been suggested as a possible explanation²⁸⁶. Studies have shown that N-linked glycosylation of TPO is essential for successful trafficking to the cell surface in CHO cells, but that deglycosylation does not affect recognition by autoantibodies or enzymatic activity^{286,290}. This differs from studies on MPO which found deglycosylation reduced enzymatic activity²⁴⁵. However, this has

not been discrepancy may be due to the different methodologies employed by each study.

Thyroid peroxidase immunodominant regions

There are two main regions of TPO which contain B-cell epitopes. These have been mapped by several groups using varying methods. Around 80% of AITD patient sera antibodies recognise two overlapping sections on the extra-cellular domains of the protein¹⁰⁷. These regions are known as immunodominant regions (IDR) A and B. The nomenclature used for these two regions is the source of some confusion in the literature²⁹¹. While the regions involved are generally agreed upon the labelling of the two IDRs are often transposed. Throughout this work the nomenclature of Ruf *et al* will be used²⁹².

When initially defining the IDRs human TPO was purified from thyroid glands removed from GD patients. This material was used to immunize mice and generate a panel of anti-TPO monoclonal antibodies²⁹². These monoclonal antibodies were used in an enzyme-linked immunosorbent assay (ELISA) to demonstrate that autoantibodies from GD and HT patients recognise two immunodominant regions of TPO. Other early studies showed the conformations of residues forming the TPO epitopes are important, as patient sera autoantibodies show reduced reactivity with reduced and/or denatured TPO^{293,294}. The epitopes are known to be discontinuous, although monoclonal antibodies to some linear sections of the protein have been shown to inhibit patient sera binding²⁹⁵. Complete inhibition is not observed with these antibodies even at high concentrations, indicating the epitope may be only partially blocked. Despite the sequence similarities, TPO autoantibodies do not react with MPO or LPO and the binding of antibodies to the protein does not inhibit enzymatic activity²⁹⁶.

Other methods used to roughly map the TPO IDRs have included producing short fragments of TPO, or TPO-MPO chimeric proteins to determine the regions involved in autoantibody binding²⁹⁷⁻²⁹⁹. The data produced by these studies helped narrow down the IDRs particularly as most studies pointed

towards the same general areas of the protein being involved despite the varying methods used.

Bresson *et al* used short synthesised peptides as a substrate in a spot blot format to compare the binding of various polyclonal and monoclonal antibodies with patient autoantibodies³⁰⁰. By synthesising eight amino acid long sections of a region suspected to be part of the IDR, strong evidence was produced towards amino acids 597-604 and 611-618 being key sequences for antibody binding to IDR-B. This improved on earlier work in which amino acids 590-622 were identified³⁰¹.

Phage display is another technique which has proven quite successful when studying the TPO epitopes. Its application has allowed several short sequences of the IDRs to be identified³⁰². Sequences are selected from a library of short peptides by binding to a molecule of interest, in this case a TPO monoclonal antibody³⁰³. The peptides making up the library are attached to the surface of a bacteriophage containing the DNA encoding the peptide which can be amplified in *Escherichia coli*. The process is repeated several times to generate a population of phages with high binding affinity to the selecting agent. These can then be analysed to determine the peptide sequences in the selected population. By comparing the sequences present in a library selected using a TPO monoclonal antibody to the TPO protein sequence several short stretches of IDR-A were identified³⁰². These include amino acids 353-363, 377-386, 713-720 and 766-775.

Later studies into the TPO IDRs used site-directed mutagenesis to determine the role of individual amino acids in autoantibody binding. The most common approach has been to substitute alanine for the target residue, although in some cases the mutation was to the equivalent residue in MPO³⁰⁴. The TPO mutants are assayed while displayed on the cell surface of CHO cells. Antibody binding relative to unmodified TPO can then be assessed by flow cytometry to determine the effect, if any, a mutation has had^{304,305}. Alternatively the mutated proteins can be extracted and analysed by methods such as Western blotting, ELISA or radioimmunosorbent assay (RIA)^{306,307}. Residues identified as having a role in autoantibody binding from mutagenesis studies include amino acids

225, 353, 358, 359, 361, 646 and 707 in IDR-A and amino acids 620, 627, and 630 in IDR-B^{304,305,307}. Tyrosine 772 has been found to influence autoantibody binding by two separate groups but its role is not completely clear^{306,307}. Work by Gora *et al* in 2004 found that mutating this residue to alanine affected the binding of antibodies to both IDRs³⁰⁷. The authors suggest this residue may have a currently unknown role in the tertiary structure of the protein. The mutation may disrupt the protein structure in this particular location, reducing autoantibody binding rather than by the loss of a direct contact with the antibody when that residue is absent.

Guo *et al* used a unique method to study the epitope of a particular TPO monoclonal antibody, TR1.9³⁰⁸. Their approach was to use F(ab) from the TPO monoclonal antibody to protect the epitope during biotinylation. Trypsin is inhibited by lysine biotinylation and by comparing the digestion patterns of TPO biotinylated with and without F(ab) present and non-biotinylated TPO, they were able to isolate one peptide containing a lysine residue which was protected from biotinylation by the F(ab). This peptide was then identified by N-terminal sequencing as comprising amino acids 704-733, with the lysine at position 713 being protected from biotinylation by the F(ab).

```

109 TQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNNRDHPRWGA
169 SNTALARWLPPVYEDGFSQPRGWNPFLYNGFPLPPVREVTRHVIQVSNEVVTDDDRYSD
229 LLMAWGQYIDHDIAFTPQSTSAAFGGGADCQMTCENQNPCFPQLPEEARPAAGTACLP
289 FYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSSPALERQLRNWTSAEG
349 LLRVHARLRDGSGRAYLPFVPPRAPSACAPEPGIPGETRGPCFLAGDGRATEVPSLTALHT
409 LWLREHNRLAAALKALNAHWSADAVYQEARKVVGALHQIITLRDYIPRILGPEAFQQYVG
469 PYEGYDSTANPTVSNVFSTAARFREGHATIHPLVRRLDASFQEHPDPLGLWLHQAFFSPWT
529 LLRGGGLDPLIRGLLARPAKLQVQDQLMNEELTERLFVLSNSSTLDLASINLQRGRDHGL
589 PGYNEWREFCGLPRLETPADLSTAISRSVADKILDLYKHPDNIDVWLGGLAENFLPRAR
649 TGPLFACLIGKQMKALRDGFWWENSHVFTDAQRRELEKHSLSRVICDNTGLTRVPMDA
709 FQVGKFPEDFESCDSIPGMNLEAWRETFPQDDKGFPESVENGFVCEESGRRLIVYSC
769 RHGYELQGREQLTCTQEGWDFQPPLC

```

Figure 17- Amino acids mapped to TPO immunodominant regions. Sequence- human TPO amino acids 109-796 (N-terminal, peroxidase and CCP domains).

■ Immunodominant region A ■ Immunodominant region B.

In summary, the IDR-A involves residues from both the peroxidase and CCP domains. Residues 225, 353-363, 377-386, 646, 707 and 713-720 of the peroxidase domain and amino acids 776-775 in the CCP domain contribute to IDR-A. By comparison, the residues making up IDR-B are relatively clustered,

consisting of amino acids 597-604, 611-618, 620, 627 and 630 of the peroxidase domain. Despite intensive study, the exact nature of autoantibody binding to TPO is still not fully characterised. Having multiple domains involved which appear to be joined by flexible hinge regions complicates the crystallization of an immunologically active form of TPO^{302,306}.

Aims of study

The overall aim of this work is to learn more about the structure of the human TPO protein to gain a better understanding of its role in AITD. An accurate structure of the TPO protein has yet to be produced so this work initially focused on obtaining one. X-ray crystallography was deemed the most suitable technique for this. Both native and recombinant TPO have previously been shown to produce poorly diffracting crystals so the initial aims were;

1. Design and produce modified forms of TPO which retain key features of the native protein.
2. Crystalize the modified TPO proteins.
3. Solve the structure of these proteins by X-ray diffraction.

The modifications made to the TPO protein were designed based on existing knowledge of the protein structure with the aim of removing or changing features which may impair crystallization. Four modified TPO proteins were successfully expressed and purified. Two of these proteins were crystallized but did not diffract to a high enough resolution to solve the structure. Several methods of improving the diffraction resolution of the crystals produced were tested but no improvement could be made.

As a structure of the TPO protein could not be produced during the course of this work, efforts were made to address other gaps in the current knowledge of TPO. The nature of autoantibody binding to TPO is of particular interest and clinical relevance. To help further understand this, another aim was added;

4. Identify the epitopes of the monoclonal TPO antibodies 4F5 and 2G4 by epitope excision.

Epitope excision had not previously been applied to the study of TPO antibodies so this was an opportunity to assess the ability of this technique to characterise TPO antibodies and produce unique data regarding the nature of the TPO IDRs. 4F5 and 2G4 were chosen for this study as they were readily available, have been previously characterised and, in the case of 4F5, the structure has been solved by X-ray crystallography. Despite this the exact IDR to which each antibody binds was yet to be determined and therefore of interest.

Chapter 2- Production and Characterization of Modified Thyroid Peroxidase Proteins

Introduction

As a major autoantigen, a structure TPO has been long sought after. However, despite the protein being first crystallized in 1997, the structure is still yet to be solved²⁵⁴. Human thyroid tissue is difficult to obtain in large quantities and generally does not yield large amounts of the protein. The group which crystallized the native enzyme obtained only 20 mg of pure TPO from 2.5 kg of thyroid tissue. For this reason, recombinant forms of the enzyme have been produced using several different eukaryotic expression systems^{284,309,310}. The full length wild type protein has been produced recombinantly, but for many studies only the extra-cellular domains (amino acids 1-838) are required^{283,284}. By truncating the protein after the EGF domain it can be produced in a soluble form which may be advantageous, depending on the nature of the study. The extra-cellular region of TPO, expressed in insect cells, was crystallized shortly after the native human protein²⁵⁵. Unlike the native protein which produced only small, two-dimensional crystals, crystals of insect expressed recombinant TPO were the first TPO crystals suitable for X-ray diffraction studies. These larger, three-dimensional crystals diffracted up to 6 Å which was not sufficient to solve the structure, but did allow some features to be identified.

Several features of the protein which could be responsible for the difficulty in growing well-ordered crystals of TPO. The first steps of this work focused on producing a modified protein more amenable to crystallization. These protein constructs contained modifications designed based on information from various MPO and LPO crystal structures. When modifying the protein, care was taken not to disrupt any key features. Much of the research into TPO relates to its role in autoimmunity and therefore any modified TPO proteins produced needed to retain the autoantibody binding characteristics of the native enzyme. If the autoantibody epitopes were disrupted it is likely that the overall protein fold would have been largely altered. This would severely limit the usefulness of any

structure produced. While not as essential as the retention of autoantibody binding, an enzymatically active protein is also desirable. This would provide further insight into the mechanism of enzymatic activity of TPO and would also indicate that the protein was similar in structure to the native enzyme.

Results and Discussion

Design of modified thyroid peroxidase proteins and DNA cloning

Key points

- DNA coding for TPO aa 1-14,109-796+6x his (NΔTPO) was produced in the pUC18 vector.
- The NΔTPO gene was cloned into the pFastBac vector for further modification and recombinant baculovirus production.

Initially a protein was designed based on the TPO extra-cellular domains (amino acids 1-838) only. A method for expressing the extra cellular domains of TPO in High Five (*Trichoplusia ni*) insect cells and purifying the truncated protein is well established²⁸⁴. Referred to as recombinant human TPO (rhTPO, amino acids 1-838) this protein is soluble and can be purified and concentrated to the high levels required for crystallography. It has been crystallized previously although the crystals diffracted poorly²⁵⁵. Using rhTPO as a starting point, two modifications were introduced to produce a modified protein, designated “NΔTPO”. These were a deletion in the N-terminal domain of the protein and a truncation at the C-terminus, after the CCP domain.

The N-terminal domain of TPO is largely removed in the mature protein²⁵⁹. Evidence has been produced that removal of much of this cleaved section from the DNA sequence has little effect on the structure or trafficking of the expressed protein²⁶¹. N-terminal sequencing of TPO often produces mixed results suggesting that cleavage is not uniform and there are other potential cleavage sites besides the predominant site after amino acid 108²⁵⁸. A mixture of protein isoforms with variation in the N-terminal residue could be difficult to separate resulting in heterogeneity in the final purified preparation. This could inhibit crystal formation or lower the diffraction resolution if the protein does

crystallise³¹¹. X-ray diffraction data is essentially an average of all the molecules making up a crystal, thus heterogeneity in the protein forming the crystal will lower the overall resolution of the data produced. To reduce the potential for heterogeneity at this site, a section of the N-terminal domain was deleted from the DNA sequence so amino acid 109 followed directly after the signal peptide. The signal peptide, consisting of amino acids 1-14, is required for intracellular trafficking, so the DNA coding for amino acids 15-108 was removed to allow the retention of this sequence²⁵⁹. While a case could be made for removing the entire N-terminal domain this was not attempted as it contains two cysteine residues which are expected to be disulphide bonded to other residues elsewhere in the protein^{256,257}. Removal of these two residues could have a large effect on the overall protein fold so was not attempted.

A gene encoding for rhTPO in a pUC18 vector was modified by site directed mutagenesis to introduce a unique *Sall* recognition sequence. The recognition sequence was introduced after the codon for lysine 108 using primers 1 and 2 (sequences of all mutagenesis primers are detailed in the primers section). This new plasmid was designated pUC18/TPO Sal.

TPO	300	310	320	330	340
	5' GCGATGAAAAGAAAAGTCAACCTG	AAAAC	ACTCAACAATCACAGCATCC	3'	

Primer 1 5'	GCGATGAAAAGAAAAGTCAACCTG	TCGACTCAACAATCACAGCATCC	3'		

Figure 18- Alignment of TPO cDNA with primer 1, : indicates identical bases. █ *Sall* recognition sequence, ▼ designates cleavage site. Base pairs numbered above.

JM110 *Escherichia coli* were transformed with the pUC18/TPO Sal plasmid, cultured and plasmid DNA purified. The recovered pUC18/TPO Sal DNA was cut with *EcoRI* and *Sall*, separated by gel electrophoresis and the >5kbp band purified from the gel. A linker sequence was produced by annealing two linker oligonucleotides ("primers"). Linker oligonucleotides were designed so that when annealed, each end had the correct unpaired single stranded end for ligation into an *EcoRI/Sall* cut vector without the need for prior restriction endonuclease (RE) digestion. The linker was ligated into the cut vector using T4 DNA ligase to produce a plasmid containing the sequence for TPO amino acids 1-14, 109-838, i.e. the TPO signal peptide followed immediately by the first amino acid of the mature protein. The *Sall* recognition site was then removed by

another round of site-directed mutagenesis using primers 3 and 4. This plasmid was designated pUC18/Nmod TPO.

The C-terminus was then truncated by site-directed mutagenesis using primers 5 and 6. This incorporated a six histidine tag, stop codon and a *Hind*III recognition sequence after the codons for amino acid 796, the last residue of the TPO CCP domain. This resulted in the removal of the EGF domain when the gene was cloned into the pFastBac (Invitrogen, USA) vector for recombinant baculovirus production. This plasmid was designated pUC18/NΔTPO.

	2370	2380	2390	2400	2410	2420	
TPO	5'	CAGCCTCCCTCTGCAAAGATGTGAACGAGTGTGCAGACGGTG	CCCACCCCCCTGC	3'			
		:::::::	:: :	::	:	:	
Primer 5	5'	CAGCCTCCCTCTGCAAAGATCATCACCATCACCATCACTAGA	VAGCTTG	GGCACTGGC	3'		

Figure 19- Alignment of TPO cDNA with primer 5, : indicates identical bases. □ Insertion, ▼ designates *Hind*III cleavage site. Base pairs numbered above.

The EGF domain has been shown to not be involved with autoantibody recognition or enzymatic activity, the two main features of interest in TPO³¹². When expressed in a soluble form there is potential for the individual domains of the protein to move independently, adopting multiple different confirmations particularly during crystal growth. By removing this domain, it was hoped that this inherent flexibility would be reduced, easing crystallization while still retaining the structure of the main areas of interest. The TPO IDRs are known to extend into the CCP domain so this section could not be removed without altering the antibody binding characteristics of the protein³¹³.

The gene for NΔTPO was amplified by PCR, separated and purified before digestion and ligation into the pFastBac vector (Invitrogen, UK) between the EcoRI and *Hind*III RE recognition sequences. Further modifications to produce other proteins were carried out starting from pFastBac/NΔTPO.

Further modification of NΔTPO gene

Key points

- Four further modifications were made to the NΔTPO gene.
- Recombinant baculoviruses containing each of the five modified TPO genes were produced for protein expression.

Three features of NΔTPO were identified for further modification based on the existing literature and examination of the MPO and LPO crystal structures; an unpaired cysteine residue predicted to be solvent exposed and two flexible loops on the protein surface.

The thiol group on the side chain of cysteine is highly reactive, commonly forming disulfide bonds, coordinating ligands or acting as a nucleophile during catalysis³¹⁴. Cysteine 146 in TPO is predicted to be unpaired and solvent exposed based on sequence alignments and homology modelling. This reactive residue could be involved in a number of unwanted interactions which could lead to protein aggregation, variable disulphide bond arrangement or other forms of heterogeneity amongst molecules, all of which are undesirable during crystallization. To investigate the effects of modifying this residue, primers were designed to mutate cysteine 146 to serine.

Introducing this mutation requires only a single base change; a guanine base was mutated to cytosine to change the TGC codon for cysteine to TCC coding for serine using primers 7 and 8. This was designated NΔTPOS. Serine is structurally similar to cysteine however, the replacement of the sulphur atom with a less nucleophilic oxygen renders it, generally, less reactive.

NΔTPO	150 160 170
	5' TGCCCAAACACTTGCTGGCGAACAAATACAGG 3'
	::::::: :::::::::::::::::::::
Primer 7 5'	TGCCCAAACACTTCCTGGCGAACAAATACAGG 3'
	::::::: :::::::::::::::::::::

Figure 20- Alignment of NΔTPO cDNA with primer 7,: indicates identical bases. █ Mutated codon- TGC codes for cysteine where TCC codes for serine. Base pairs numbered above.

When comparing the protein sequences of the TPO peroxidase domain with that of MPO and LPO there are two notable extensions (Fig. 8). The largest lies between amino acid 292-317 in the TPO protein sequence. This aligns with a

large flexible loop on the surface of MPO and LPO. As this loop is even larger in TPO it was targeted for modification by partial deletion. Flexible sections of a protein can be problematic during crystallography as they may inhibit the formation of crystal contacts. If they adopt multiple confirmations or are free to move inside a crystal then they can lower the overall diffraction resolution. Two deletions to this loop were tested. The first, NΔTPO-10, was a 10 amino acid deletion removing amino acids 299-308. This was an attempt to produce a similar structure to that found in MPO, as the loop in TPO is 10 amino acids longer than the MPO equivalent. The second, NΔTPO-14, was a longer but overlapping deletion of 14 amino acids removing amino acids 298-311. Here a homology model of the TPO peroxidase domain, produced by RSR Ltd (unpublished), was examined to estimate how much of the loop could be removed without altering the overall protein structure. By measuring distances between amino acids on the model it was determined that glycine 297 and asparagine 312 were in a reasonable position to form a peptide bond if the amino acids between these residues were removed. Both of these mutations also removed the glycosylation site at asparagine 307, leaving only three potential sites for N-linked glycosylation.

To remove the section of DNA coding for this loop, a *Kpn*I site was introduced by mutagenesis with primers 9 and 10. This involved making a silent mutation to the codon for valine 406 in the NΔTPO DNA sequence by site-directed mutagenesis.

NΔTPO	5'	910 920 930
		GCCACCGAGG GTCC CCTCCCTGACG 3'
		: : : : : : : : : : : : : : : : : : : :
		Primer 9 5' GCCACCGAGG GTAC ▼ CCTCCCTGACG 3'

Figure 21- Alignment of NΔTPO cDNA with primer 9, : indicates identical bases. █ Mutated codon- GTC to GTA, silent mutation as GTG codes for valine,▼ designates *Kpn*I cleavage site. Base pairs numbered above.

Synthetic genes were obtained from Life Technologies, U.K, coding for TPO amino acids 286-408. Gene 1 had the sequence coding for amino acids 299-308 removed to produce NΔTPO-10, Gene 2 had the sequence coding for amino acids 298-311 removed to produce NΔTPO-14. These were cut from the supplied plasmid and ligated into the NΔTPO sequence in the pFastBac vector between the introduced *Kpn*I site and a naturally occurring *Fse*I site in the TPO

sequence. This removed either 10 or 14 amino acids from the translated sequence. As the mutation to produce the *Kpn*I site was silent, it was not necessary to remove it once the desired sequences had been produced.

The final modification made was a deletion to a loop between amino acids 248-257. This loop aligns with a problematic section of the MPO structure. The first diffraction data produced from canine MPO did not show any electron density in the region containing the equivalent loop²⁴². While this was later resolved in a higher resolution structure of human MPO it does suggest this loop may be flexible or prone to adopting different confirmations²¹⁷. Furthermore this loop is extended in TPO relative to MPO increasing the likelihood of it having a negative effect during crystallography.

Using the same methodology as per NΔTPO-14 it was determined that removing more than two amino acids was likely to put strain on the confirmation of the protein, possibly altering the folding. Removing alanine 251 and alanine 252 in this loop was achieved by site directed mutagenesis of the NΔTPO sequence with primers 11 and 12 to produce the gene for NΔTPO-2.

460	470	480		
NΔTPO	5 GCACCAGCAAA	GCTGCC	TTCGGGGGAGGGC	3
	:::::::	:::::	:::::::	
Primer 11	5 GCACCAGCAAA-----	TTCGGGGAGGGC	3	

Figure 22- Alignment of TPO cDNA with primer 11, : indicates identical bases, - indicates the absence of a base in the primer, included to maintain clarity of alignment. █ Removed section coding for aa 251 & 252.

Once the correct DNA sequence for each protein was cloned into the pFastBac vector and confirmed by sequencing, bacmid DNA and recombinant baculoviruses were generated using the Bac-to-Bac baculovirus expression system (Invitrogen, USA) using the manufacturers' protocols.

Table 1- Amino acid sequences of modified TPO proteins

Designation	amino acids
NΔTPO	1-14, 109-796+6 x his
NΔTPOS	As NΔTPO, cys 146→ser
NΔTPO-10	As NΔTPO, aa 299-308 removed
NΔTPO-14	As NΔTPO, aa 298-311 removed
NΔTPO-2	As NΔTPO, aa 250-251 removed

Amino acid numbering refers to the protein sequence of full-length human TPO.

Expression of modified thyroid peroxidase proteins

Key points

- Modified TPO proteins were expressed in High Five insect cells.
- All proteins expressed at good levels, showed some enzymatic activity and appeared stable.

All modified TPO proteins used in this work were expressed recombinantly in High Five cells. This cell line was selected as it has previously been shown to produce the extra-cellular domains of TPO in an immunologically and enzymatically active form with good yields²⁸⁴. Expression and purification protocols were based on those previously established for rhTPO and produced adequate expression levels with little to no modifications required. Average expression levels determined by RIA and total expression by weight and volume are shown in Table 2. To provide ample material for characterisation and crystallography experiments it was estimated around 240 mg of each protein would need to be expressed assuming final recovery would be around 20-25%. In some cases, multiple expression and purification runs were carried out for the same modified TPO protein. In these instances, the circumstances are explained below.

Table 2- Pooled protein harvests prior to purification

Modified TPO protein (run)	Volume of culture supernatant (L)	Weight of TPO (mg)	Average TPO concentration ($\mu\text{g/mL}$)	Average TPO enzymatic activity (GU/mg)
NΔTPO (1)	14	159.0	11.3	71
NΔTPO (2)	12	225.8	18.8	43
NΔTPO (3)	22	319.1	14.5	60
NΔTPOS	16	255.7	16.0	83
NΔTPO-10	16	279.3	17.5	41
NΔTPO-14 (1)	14	374.8	26.8	34
NΔTPO-14 (2)	14	234.5	16.8	24
NΔTPO-2	8	293.1	36.6	31

Each row details a pool of culture supernatant which formed the starting material for an anion exchange purification run. In cases where multiple expression and purification runs were carried out, each separate purification is numbered in brackets to differentiate.

The RIA used to measure protein expression in culture supernatants utilised a polyclonal rabbit anti-TPO sera. This allowed the quantification of the protein of interest in a complex mixture. At later stages of the purification process, when the proteins were mostly purified, optical absorbance was used to quantify protein concentration (Chapter 6).

Enzymatic activity of expression supernatants was measured by a guaiacol oxidation assay³¹⁵. This assay was used throughout the production and analysis process to measure the ability of the enzymes to catalyse the oxidation of guaiacol in the presence of hydrogen peroxide. Oxidized guaiacol dimerizes to form 3,3'-dimethoxy-4,4'-biphenoquinone (Fig. 23)³¹⁶.

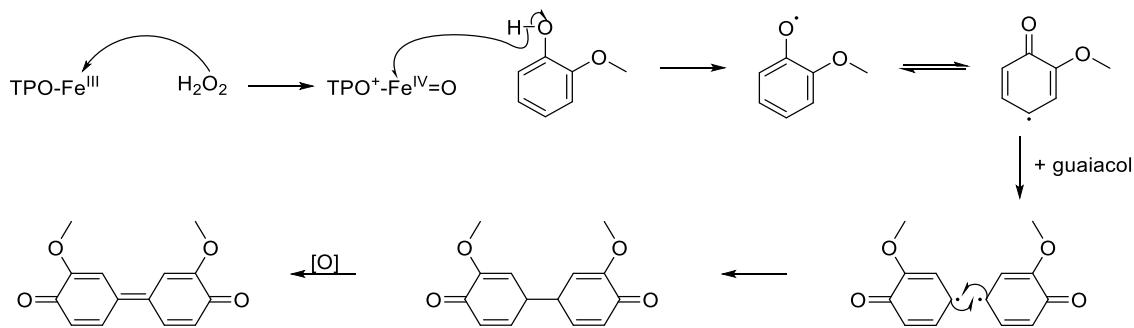


Figure 23- Mechanism of reaction catalysed by TPO in guaiacol oxidation assay³¹⁶. TPO oxidises guaiacol which then dimerizes to produce 3,3'-dimethoxy-4,4'-biphenoquinone which absorbs at 470 nm.

Samples of protein harvested from expression media were also subjected to SDS-PAGE, Western blotted and probed with another, anti-TPO polyclonal rabbit serum. As with the RIA, the use of a polyclonal TPO antibody sera allowed for the analysis of the protein of interest in a complex mixture prior to purification. If the TPO protein in a supernatant has been degraded, then bands with a much lower than expected molecular weight are observed by Western blotting. This technique was used to exclude supernatants containing degraded protein.

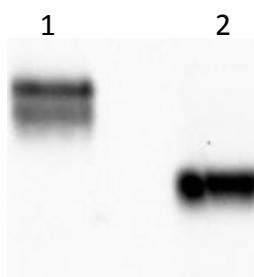


Figure 24- Western blot of intact protein from harvest supernatant. Lane 1- rhTPO, lane 2- NΔTPO-14 (1). This analysis was carried out to check for protein degradation prior to purification. Native and recombinant TPO commonly appears as a doublet when visualised by SDS-PAGE where the NΔTPO proteins resolve as a single band.

Expression of the modified TPO proteins was within expected ranges for recombinant soluble proteins expressed in High Five cells and was consistent with those observed previously for rhTPO^{284,317}. It is possible that protein expression could have been further optimised but this was not attempted as the expression volumes required to reach or exceed the 240 mg target weight were well within the capacity of a single anion exchange column.

Initial expression of NΔTPO (NΔTPO (1), Table 2) gave the lowest expression levels (average concentration in harvested supernatant 11.3µg/mL), but this

pool encompassed several early expression optimisation experiments. Once infection conditions and harvest times were established the second pool (Δ TPO (2)) and all other Δ TPO proteins had overall higher expression levels (average concentrations observed by RIA in harvested supernatant were between 14.5 and 36.6 μ g/mL).

Capture of protein from culture media

Key point

- Modified TPO proteins were captured efficiently using fluidized bed, anion exchange chromatography.

The modified TPO proteins were captured from culture supernatant by anion exchange (diethylaminoethyl, DEAE) using a fluidized bed system. The fluidized bed allows large particles to pass through removing the need for filtration of the culture supernatant provided the majority of cell debris has been removed by centrifugation. The protocol used was identical to that established previously for rhTPO by RSR Ltd (Chapter 6) and gave good recovery yields (Table 3)²⁸⁴. Protein was eluted from the column using a single step increase in sodium chloride concentration. Enzymatically active fractions, as determined by guaiacol oxidation assays, were pooled immediately and analysed following each column run using the same techniques as the post-harvest material. In some cases (Δ TPO-10 and -2) the weight of protein recovered was calculated to increase above the amount loaded onto the column. This may be due to the removal of other proteins and molecules which interacted non-specifically with the expressed modified TPO protein or other assay components when measuring the concentration by RIA. This may have produced a false low value when assaying culture supernatant samples.

Table 3- DEAE purification yields

Modified TPO protein (run)	Elution volume (L)	Weight of TPO (mg)	TPO concentration ($\mu\text{g/mL}$)	Recovery (%)	TPO enzymatic activity (GU/mg)
NΔTPO (1)	3.22	151.1	46.9	95.0	273
NΔTPO (2)	3.21	173.2	53.9	77.0	335
NΔTPO (3)	3.24	293.9	90.7	92.1	247
NΔTPOS	3.46	232.1	67.1	90.8	238
NΔTPO-10	3.14	303.5	96.7	108.7	131
NΔTPO-14 (1)	3.24	337.7	104.2	90.1	167
NΔTPO-14 (2)	3.06	221.9	72.5	94.6	240
NΔTPO-2	3.08	318.3	103.3	108.6	135

Recovery % calculated from column/weight of TPO loaded to column (Table 2).

The only run in which the recovery of modified TPO protein dropped below 90% by RIA was NΔTPO (2). This was attributed to a column leak during the purification run resulting in the loss of an unknown volume of material. As this material was to be combined with the NΔTPO (1) batch later in the purification process and a reasonable weight of protein was recovered, this was deemed acceptable.

Purification of modified thyroid peroxidase proteins

Key points

- Modified TPO proteins were further purified by metal affinity chromatography.
- Each protein eluted in two peaks, analysis showed that material in each peak had different levels of purity and enzymatic activity.

The only major deviation from previously established protocols was in the second of the three purification steps. Antibody affinity chromatography has previously been used successfully for the purification of TPO^{288,318}. While this has been shown to be a very effective method, there are some disadvantages. A long contact time is required for protein binding, making this method relatively time consuming compared to the other steps. Also, producing antibody affinity matrix requires large amounts of IgG which also has to be expressed and purified before use.

A six histidine tag was present at the C-terminus of all proteins produced to facilitate purification by nickel immobilised metal affinity chromatography (IMAC). Purification by this method was quicker and less costly than antibody affinity chromatography. A further advantage of this method was that by eluting the bound protein with an imidazole gradient, two separate peaks of material could be separated (Fig. 25). Both elution peaks were shown to contain the protein of interest by RIA. However, when examined further the peak that eluted at a lower imidazole concentration consisted of mostly enzymatically inactive material. This material was also of slightly lower purity than the second peak based on stained SDS-PAGE. This material was judged unsuitable for further purification as an enzymatically active protein was desired. A mixture of active and inactive material in the final pool could also reduce the chances of producing well-diffracting crystals. In all cases the majority of the protein bound to the column was of the more active type, eluting in the second peak (Table 5). Post IMAC yields given in Table 4 are for the second elution peak only. The DEAE elution pool NΔTPO (3) was split into two equal pools of ~120 mg to maximise recovery, and are designated NΔTPO (3a) and NΔTPO (3b) in Tables 4 and 5.

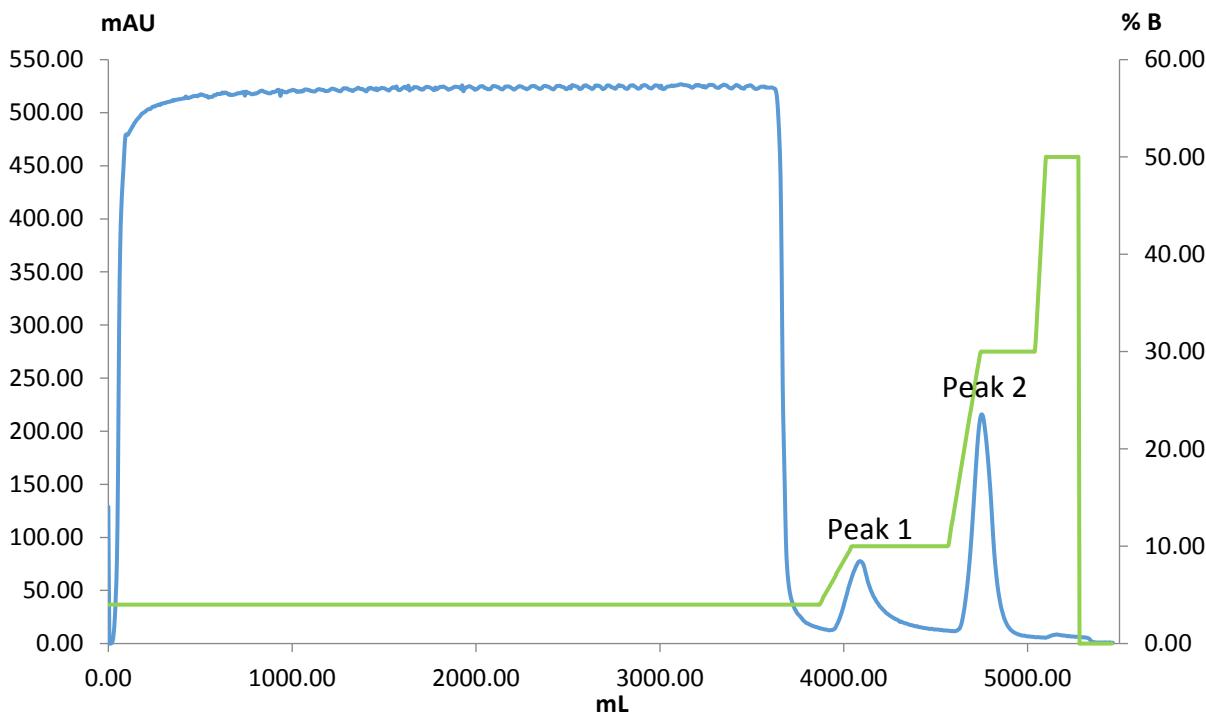


Figure 25- IMAC elution profile for NΔTPO(1). Blue- absorbance at 280 nm, green- % buffer B concentration (buffer A + 500 mM imidazole). NΔTPO proteins elute in two peaks by this method with the first peak containing material of low enzymatic activity and purity. Only protein from the second peak was used for further purification and study.

Table 4- IMAC purification yields

Modified TPO protein (run)	Elution volume(mL)	Weight of TPO (mg)	TPO concentration ($\mu\text{g/mL}$)	Recovery (%)
NΔTPO (1)	260	84.32	324.3	55.8
NΔTPO (2)	256	90.09	351.91	52.0
NΔTPO (3a)	320	98.88	309.00	67.3
NΔTPO (3b)	310	100.32	323.61	68.2
NΔTPOS	390	165.36	424.00	71.2
NΔTPO-10	210	118.67	565.09	39.1
NΔTPO-14 (1)	212	64.07	302.21	19.0
NΔTPO-14 (2)	262	120.58	460.23	54.3
NΔTPO-2	266	183.00	687.97	57.5

Figures for 2nd elution peaks only. Recovery % calculated from the weight of TPO calculated for each post DEAE purification pool (Table 3 and main text).

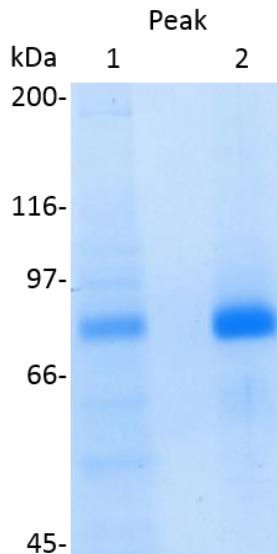


Figure 26- Stained 8% acrylamide reduced SDS-PAGE of pooled post IMAC NΔTPO (1) elution peaks. Material from the first peak was of slightly lower purity and not used for further purification.

Table 5- Post IMAC pool enzymatic activities and peak ratios

Modified TPO protein (run)	Absorbance peak area ratio	peak 1 GU/mg	peak 2 GU/mg
NΔTPO (1)	2.2	36	444
NΔTPO (2)	2.7	42	380
NΔTPO (3a)	2.0	24	320
NΔTPO (3b)	2.2	40	322
NΔTPOS	2.8	44	315
NΔTPO-10	2.6	47	197
NΔTPO-14 (1)	1.1	51	236
NΔTPO-14 (2)	2.6	42	391
NΔTPO-2	4.3	41	198

Peak ratio calculated as absorbance (at 280 nm) peak 2 area/ Absorbance (at 280 nm) peak 1 area.

The recoveries from IMAC purifications were lower than the initial DEAE capture, reflecting the greater selectivity of this step. The analysis of the first peak material from each run shows this step separated some lower activity or inactive protein from the active enzyme (Table 5). The RIA used up to this point to quantify protein yields cannot differentiate active and inactive protein, therefore the yields reported do not take into account the removal of inactive protein.

The yield from the NΔTPO-14 (1) IMAC run was significantly lower than any other run. The low yield was due to an equipment malfunction causing a premature advance of the elution gradient. This caused the appearance of a small absorbance peak between the two normally observed peaks. The time taken for the absorbance to drop between peaks was also extended in this run due to the advancement of the gradient. The use of an incorrect elution gradient is reflected in the peak area ratio for this run (Table 5), the area of the second elution peak was much lower relative to the first peak in this run. Analysis of the material obtained from this run showed no detectable difference in the quality of the protein obtained from the second elution peak. A second expression run was carried out with this protein as the low yield did not appear to stem from any inherent defect cause by this particular modification and gratifyingly, NΔTPO-10 (2) gave a yield and peak area ratio in line with the other modified TPO proteins.

Final polishing of modified thyroid peroxidase proteins

Key points

- Modified TPO proteins were purified by size exclusion chromatography to ensure the final preparations used for crystallography were as homogeneous as possible.
- Individual fractions were analysed separately prior to pooling to ensure only the best material was selected for further study.
- The NΔTPOS protein (cys146→ser) was excluded at this point due to excessive aggregation and apparent instability.

While the TPO proteins produced appeared >90% pure by SDS-PAGE following IMAC, aggregated protein was still present in the pools which would be detrimental to crystal growth. This was removed from the post-IMAC pools by a final polishing step by size exclusion chromatography (SEC). Samples were taken from fractions collected from preparative SEC columns and analysed individually by analytical SEC, SDS-PAGE and guaiacol oxidation assay. Analytical SEC was used to quantify how much, if any, aggregated material was present in each fraction. SDS-PAGE gels were stained to visually check purity and ensure there was no protein degradation. The guaiacol oxidation assay was

used to ensure all material included in final pools had an acceptable level of enzymatic activity.

Table 6- Post SEC fractions summary

Modified TPO protein (run)	No. of fractions	TPO Concentration range ($\mu\text{g/mL}$)	TPO Enzymatic activity range (GU/mg)	Minimum purity (%)
NΔTPO (1+2)	22	171-745	359-498	97.6
NΔTPO (3)	18	59-695	153-716	97.3
NΔTPOS	7	352-394	168-209	97.2
NΔTPO-10	28	124-882	107-283	98.4
NΔTPO-14 (1)	19	69-738	169-619	98.8
NΔTPO-14 (2)	20			
NΔTPO-2	29	80-927	121-347	98.0

Fractions selected for further use from size exclusion chromatography columns only. Minimum purity is the % of monomeric protein present in the fraction determined from analytical SEC profile. Value reported is for the lowest purity fraction included in the final pool.

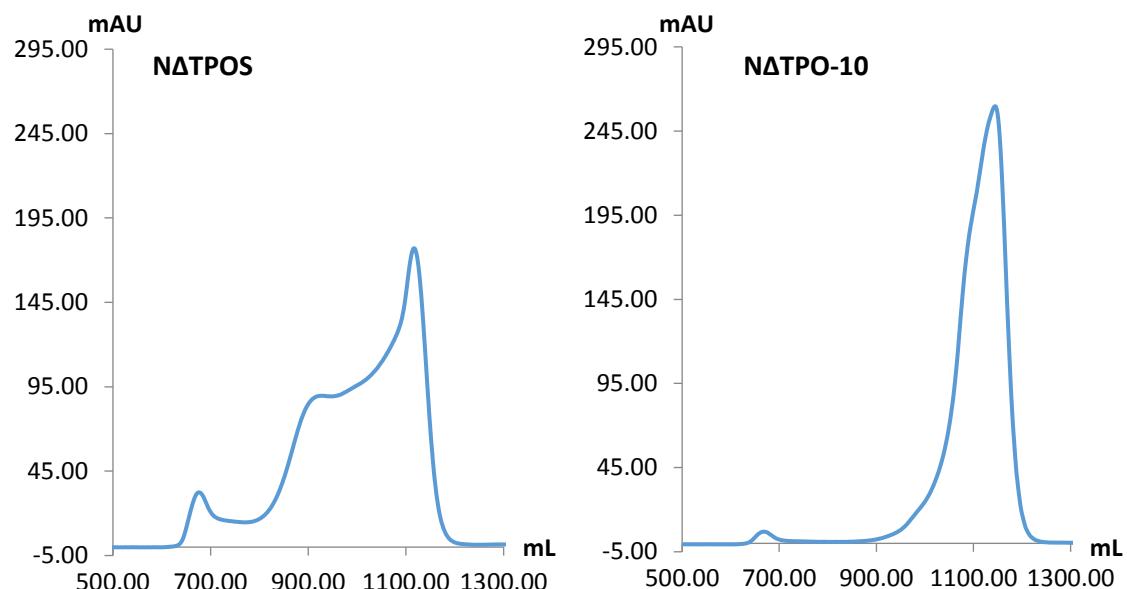


Figure 27- Preparative SEC elution profiles (absorbance at 280 nm) for NΔTPOS and NΔTPO-10. The profile of NΔTPO-10 is representative of the other modified proteins. NΔTPOS contained an unusually high proportion of high molecular weight aggregates than the other four proteins.

Prior to SEC purification the NΔTPO (1) and (2) post-IMAC pools were combined to produce a single, larger final pool of this protein. The NΔTPO (3) SEC run was of a smaller scale and used 160 mL of NΔTPO (3a) IMAC elution pool as the starting material.

NΔTPOS showed the highest recovery following IMAC purification but the SEC elution profile of this material shows much of the protein was in an aggregated

form (Fig. 27). The SEC elution profile of NΔTPO-10 is more typical, as none of the other proteins suffered such extensive levels of aggregation. Analysis of the NΔTPOS SEC elution fractions by stained SDS-PAGE also showed many of the fractions contained lower molecular weight indicating degraded protein was present (Fig. 28). The cause of protein degradation is unknown. Material loaded and eluted from this column was handled in the same manner as all other pools, none of which showed such substantial protein degradation. It is possible this protein was inherently unstable, but the breakdown products could only be detected when the protein was pure. The size of the intact (86 kDa) and degraded (60 kDa) protein is close enough to make separation by SEC difficult so co-purification is a possibility. However, the presence of <45 kDa fragments in some fractions indicates some protein degradation probably occurred following purification.

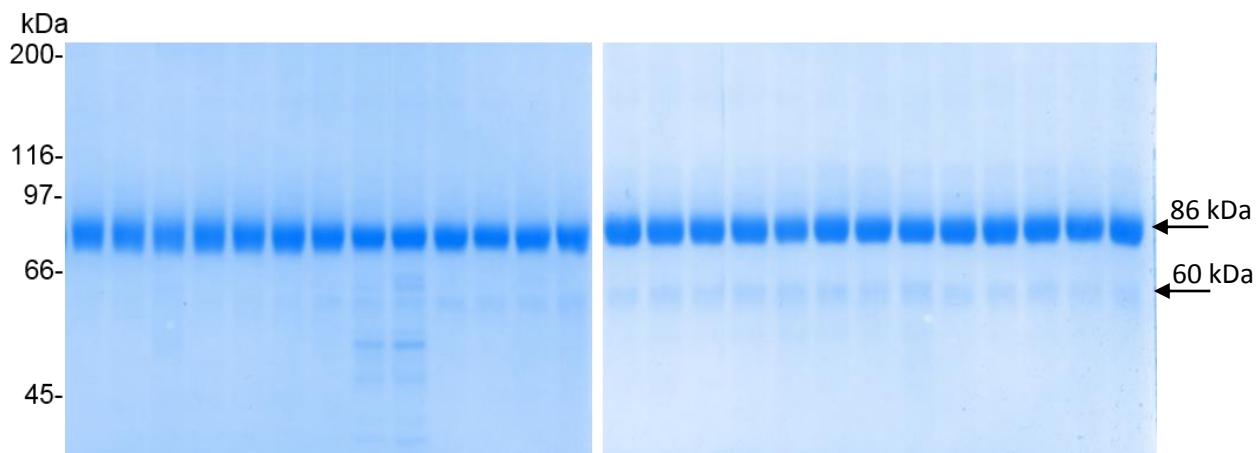


Figure 28- Stained 8% acrylamide reduced SDS-PAGE of NΔTPOS SEC fraction samples (2 gels). The visible bands at 60 kDa and below demonstrate that the majority of samples appear to be of poor quality when assessed by this method.

As only seven NΔTPOS post-SEC fractions were deemed suitable for further use a final pool of this protein was not prepared. Further expression was not attempted due to the questionable stability of this protein.

Post-IMAC NΔTPO-14 (1) and (2) were SEC purified separately to avoid having to over-concentrate the protein to overcome the larger starting volume when loading the column. Fractions from each run were combined when producing a final pool for this protein.

A pool of NΔTPO-2 fractions, not selected for further use, was prepared to investigate the composition of the unused fractions. The fractions selected for

this pool (Fractions A1 to C2 inclusive, Fig. 29A) were not suitable for crystallography due to a high level of aggregated material present. Analytical SEC analysis of this pool showed that a large amount of the material present in these fractions was monomeric. This appears to be easily separated from the analytical SEC profile but would prove difficult on a larger scale as the preparative SEC columns have a lower separation resolution than the smaller, analytical columns. Refolding this material was not attempted due to the large size of the protein. With the exception of NΔTPOS, sufficient protein was recovered from preparative SEC purification to produce a suitable final pool so recovering material from unused fractions was not investigated further.

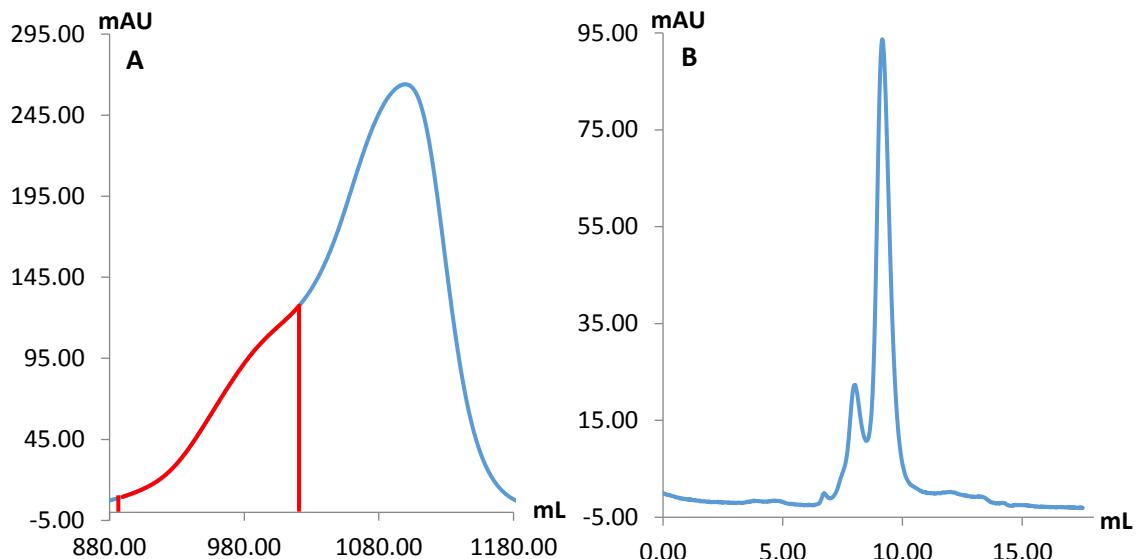


Figure 29- **A)** Preparative SEC elution profile (absorbance at 280 nm) for NΔTPO-2. Fractions selected for unused fraction pool highlighted (red). **B)** Analytical SEC profile (absorbance at 280 nm) for NΔTPO-2 unused fraction pool. This shows that the protein in these fractions was predominantly of the correct molecular weight but could not be recovered due to limitations of the resolution of the column used. While disappointing this loss of material was not significant enough to affect the project.

Final pools were prepared for each protein, with the exception of NΔTPOS, from selected SEC fractions and were concentrated to ≥ 10 mg/mL for crystallography. Each pool was aliquoted to prevent freeze-thaw cycles and stored at -80°C . Details of these pools are shown in Table 7. This material was used for characterising the modified TPO proteins and crystal screening.

Table 7- Final pool weights and volumes

Modified TPO protein (run)	Volume(mL)	Weight(mg)	Concentration (mg/mL)
NΔTPO (1)	4.8	53.52	11.15
NΔTPO (2)	1.1	11.31	10.28
NΔTPO-10	5.8	75.86	13.08
NΔTPO-14	7.8	83.85	10.75
NΔTPO-2	5.3	63.18	11.92

Characterisation of modified thyroid peroxidase proteins

Key points

- The four fully-purified modified TPO proteins were assessed by a number of methods to ensure they were homogeneous and still shared key features with native TPO.
- All four proteins were deemed suitable for use in crystallography experiments.

While some analysis was carried out for quality control purposes during expression and purification more extensive analysis was carried out on the final pools to characterise each modified form of the TPO protein further.

Proteins used for crystallography need to be of high purity to produce crystals which diffract well³¹¹. This was assessed in the final protein pools by three methods, analytical SEC, SDS-PAGE and isoelectric focusing (IEF). Analytical SEC allowed the amount of higher molecular weight protein aggregates in the final pools to be measured. A value for % purity for each pool was produced based on the percentage of the main peak as the total peak area (Table 8).

Table 8- Purity of final pools

Protein (run)	Purity by SEC (%)
NΔTPO (1)	98.3
NΔTPO (2)	96.8
NΔTPO-10	98.1
NΔTPO-14	98.0
NΔTPO-2	97.3

Purity determined from absorbance (280 nm) peak area on analytical SEC profiles.

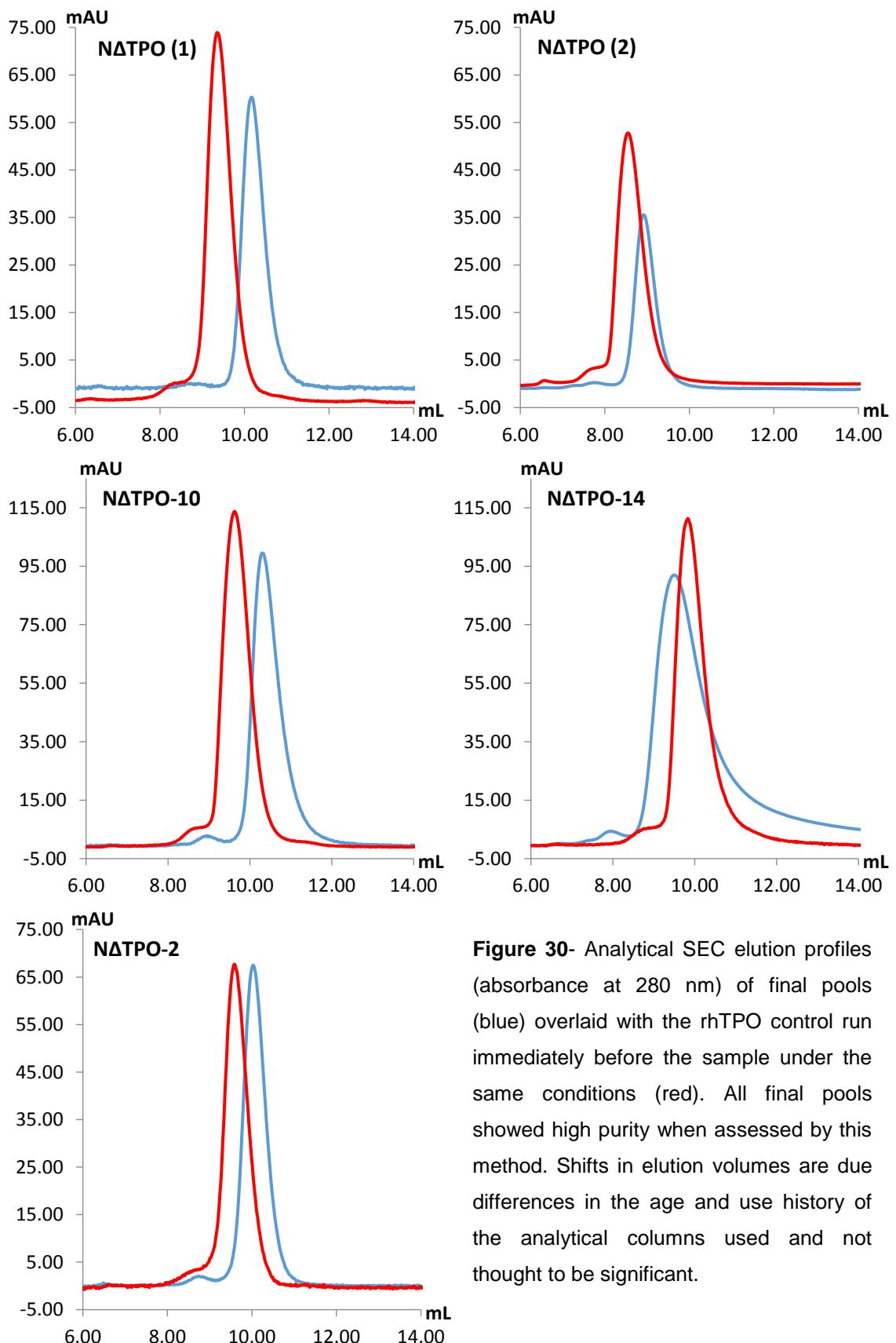


Figure 30- Analytical SEC elution profiles (absorbance at 280 nm) of final pools (blue) overlaid with the rhTPO control run immediately before the sample under the same conditions (red). All final pools showed high purity when assessed by this method. Shifts in elution volumes are due differences in the age and use history of the analytical columns used and not thought to be significant.

Levels of aggregated protein were low in all the final pools and there was no protein degradation detectable by this method. The retention time of the proteins varied more than would be expected for proteins with such similar mass (Fig. 30). Each analytical SEC run was carried out separately and while the same make and model column was used in each case, the exact lot number and previous use history of the columns did vary. The position of the absorbance at 280 nm peaks for each protein relative to the positive control, a preparation of rhTPO which was identical for each run, is very similar. This suggests the change in retention time is caused by the slight variances in chromatography conditions for each run rather than actual unexpected differences in mass between the proteins.

Stained SDS-PAGE gels also gave no evidence of the presence of any impurities or degraded protein in any of the final pools produced, in agreement with the SEC results. One interesting finding was that the NΔTPO proteins appeared as a single band, not the characteristic doublet observed with native or recombinant TPO. This single band was also observed by western blotting during purification (Fig. 24). Identifying and separating the two TPO isoforms visible by SDS-PAGE has previously proven difficult and the reason for their appearance has not been proven conclusively³¹⁹. Expression of alternatively spliced mRNA species or variations in post-translational modifications have been suggested as reasons for this. Having only a single band visible by SDS-PAGE is a promising sign which, indicates these protein samples are more homogeneous than native or recombinant TPO.

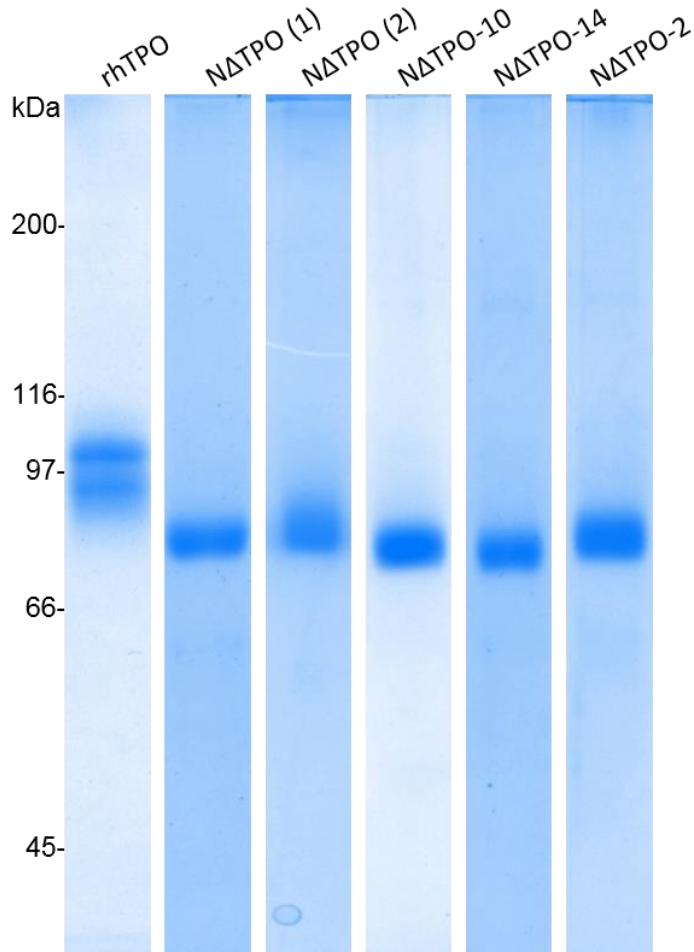


Figure 31- Stained 8% acrylamide reduced SDS-PAGE of final pools (results combined from separate gels to demonstrate purity, see Table 9 for molecular weights calculated from individual gels).

Table 9- Molecular weight of modified TPO proteins

Modified TPO protein (run)	MW by SDS-PAGE kDa	Predicted MW kDa
NΔTPO (1)	86.0	77.4
NΔTPO (2)	86.2	77.4
NΔTPO-10	85.6	76.4
NΔTPO-14	84.0	76.0
NΔTPO-2	85.0	77.2

Calculated from reduced SDS-PAGE.

Molecular weights of each protein, determined by SDS-PAGE, are shown in Table 9. The molecular weights were also predicted from the amino acid sequence using Protparam³²⁰. The observed molecular weight in each case is between 7.8-9.2 kDa (9.2-10.7%) higher than predicted from the protein sequence. The predicted mass of insect expressed rhTPO is also inconsistent

with that observed by SDS-PAGE (12.9-16.9 kDa, 13.7-17.2% difference for the lower and upper doublet bands respectively). This reduction in mobility is likely caused by the presence of N-linked glycosylation on the proteins. Partial deglycosylation of N Δ TPO with endoglycosidase H and F3 (Chapter 3, “Deglycosylation experiments”) caused a drop in the observed molecular weight by SDS-PAGE of 4.8 kDa to 81.2 kDa. This is a 4.7% difference which is reasonable given the limitations of SDS-PAGE for calculating molecular weight³²¹. Glycoproteins are known to bind less SDS than non-glycosylated protein and this can cause altered mobility during electrophoresis experiments³²².

The isoelectric point (pl) of each modified TPO protein was determined by isoelectric focusing (IEF) (Fig. 32 and Table 10). The observed pl for all TPO proteins was similar or the same as that determined for rhTPO (5.4). This result was as expected, as the modifications made to the protein should not have caused major changes in the tertiary structure and therefore the surface exposed residues should be largely identical between each protein. The slight increase in the pl of N Δ TPO-10 and -14 may be caused by the loss of an aspartic acid residue (amino acid 300) from the truncated loop which was predicted to be surface exposed. Each protein produced a slightly smeared band when visualising the IEF gels which may indicate some minor heterogeneity in the pools. There are a number of potential explanations for this such as differences in protein glycosylation and differential deamidation of the protein³²³. Unfortunately it was not possible to compare the deglycosylated N Δ TPO to the glycosylated form due to the limited availability of the deglycosylated material (Chapter 4). While this result was disappointing the other analysis carried out of the final pools suggested an otherwise good level of purity. For this reason none of the modified TPO proteins were excluded from crystal screening on the basis of the IEF results.

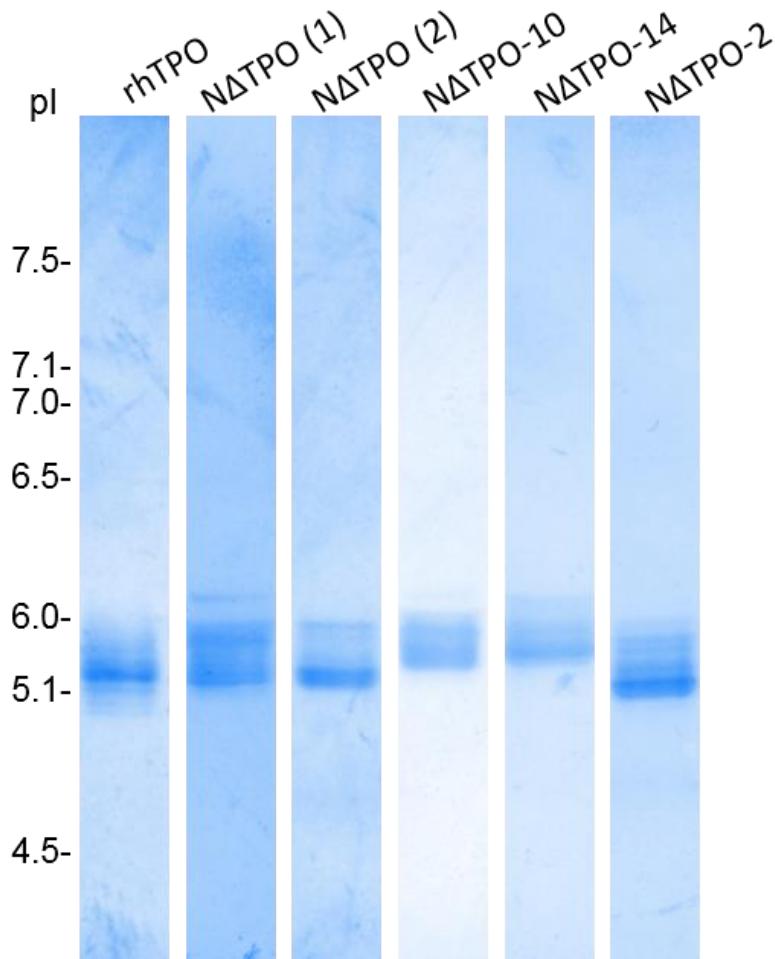


Figure 32- Stained IEF gel of final pools (results combined from separate gels to demonstrate purity, see Table 10 for calculated pl values from individual gels).

Table 10- Isoelectric points of modified TPO proteins

Modified TPO protein (run)	Isoelectric point
NΔTPO (1)	5.4
NΔTPO (2)	5.4
NΔTPO-10	5.5
NΔTPO-14	5.5
NΔTPO-2	5.4

Isoelectric points determined by isoelectric focusing.

Antibody binding of final pool protein was assessed using a pool of AITD patient sera and 2G4, a human monoclonal anti-TPO antibody. 2G4 is a human IgG1 that originates from thyroid infiltrating lymphocytes isolated from a HT patient¹⁰⁹.

It has previously been shown to be highly specific to human TPO and its binding to TPO is conformation dependent.

To compare the antibody binding of each protein, binding curves over a range of protein concentrations were prepared from RIA data and compared to the curve produced by rhTPO. Both the patient sera and 2G4 assays were inhibition assays. In this type of assay the test protein (each of the modified TPO proteins) inhibits the binding of a fixed amount of either patient antibodies or 2G4 IgG to radiolabelled TPO. This produces a measurable decrease in signal (ionizing events or “counts” per minute, CPM) proportional to the concentration of the unlabelled test protein. Results in Figs. 33-36 are displayed as percentage of CPM in test sample compared to the total CPM added (% bound), by test protein concentration. As the concentration of test protein increases, the CPM decreases as there is less antibody available to bind the labelled TPO.

Binding to pooled patient sera antibodies is shown in Figs. 33 and 35 and to 2G4 IgG in Figs. 34 and 36. Each modified TPO protein was analysed in a separate assay but utilising the same rhTPO standards as controls. The individual binding curves from each assay generally show a similar trend for the rhTPO control and NΔTPO protein. The actual % binding values vary between proteins but this is due to inter-assay variation as it was not possible to test each protein in a single assay. When plotting the relative % binding values with the rhTPO values averaged across all assays (to account for inter-assay variation), most points for each protein fall within or close to one standard deviation of the control. This data shows that each of the modified TPO proteins produced fold in a similar manner to the full-length extra-cellular domain of TPO and therefore any structural data generated from these proteins is likely to be a good representation of the *in vivo* occurrence.

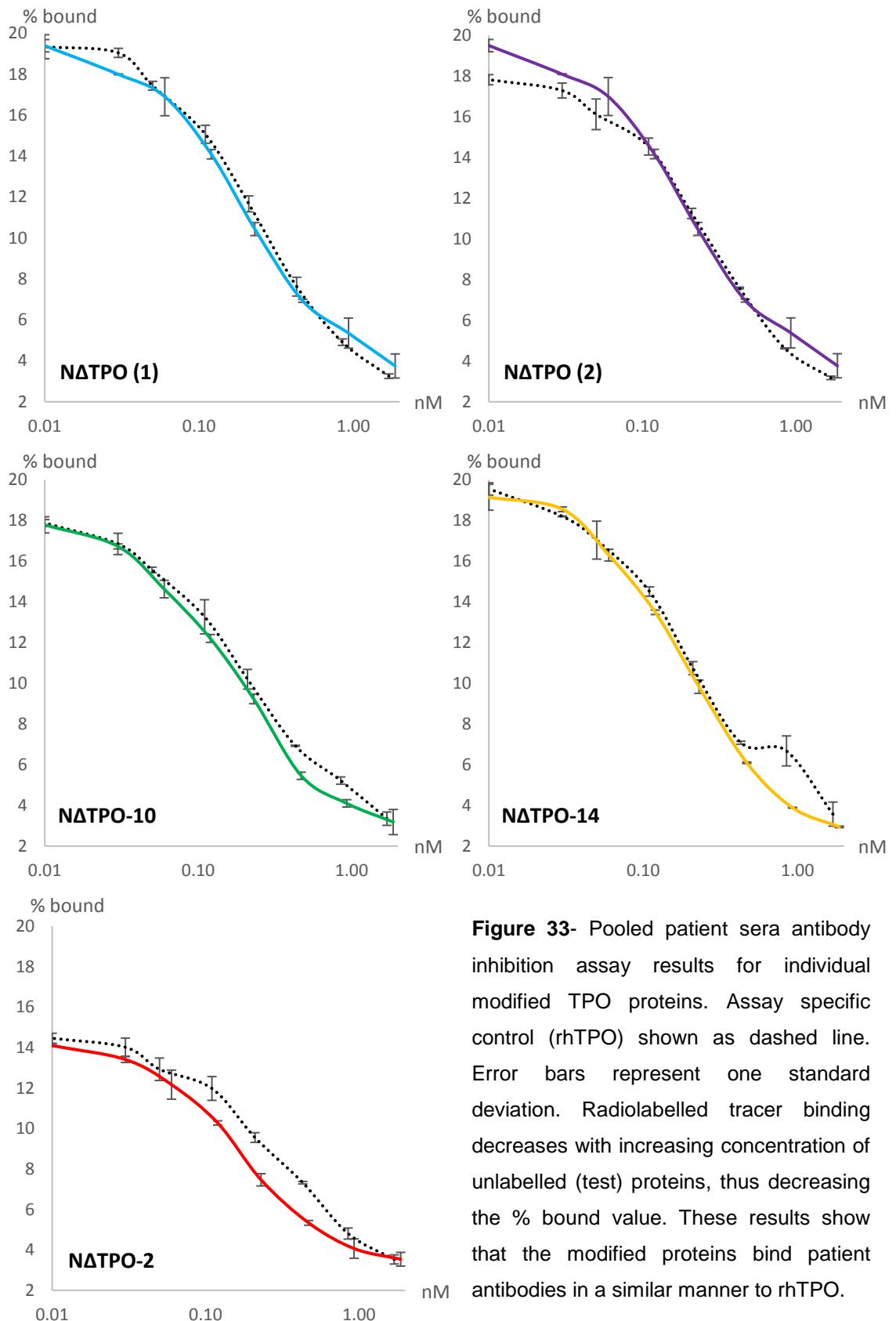


Figure 33- Pooled patient sera antibody inhibition assay results for individual modified TPO proteins. Assay specific control (rhTPO) shown as dashed line. Error bars represent one standard deviation. Radiolabelled tracer binding decreases with increasing concentration of unlabelled (test) proteins, thus decreasing the % bound value. These results show that the modified proteins bind patient antibodies in a similar manner to rhTPO.

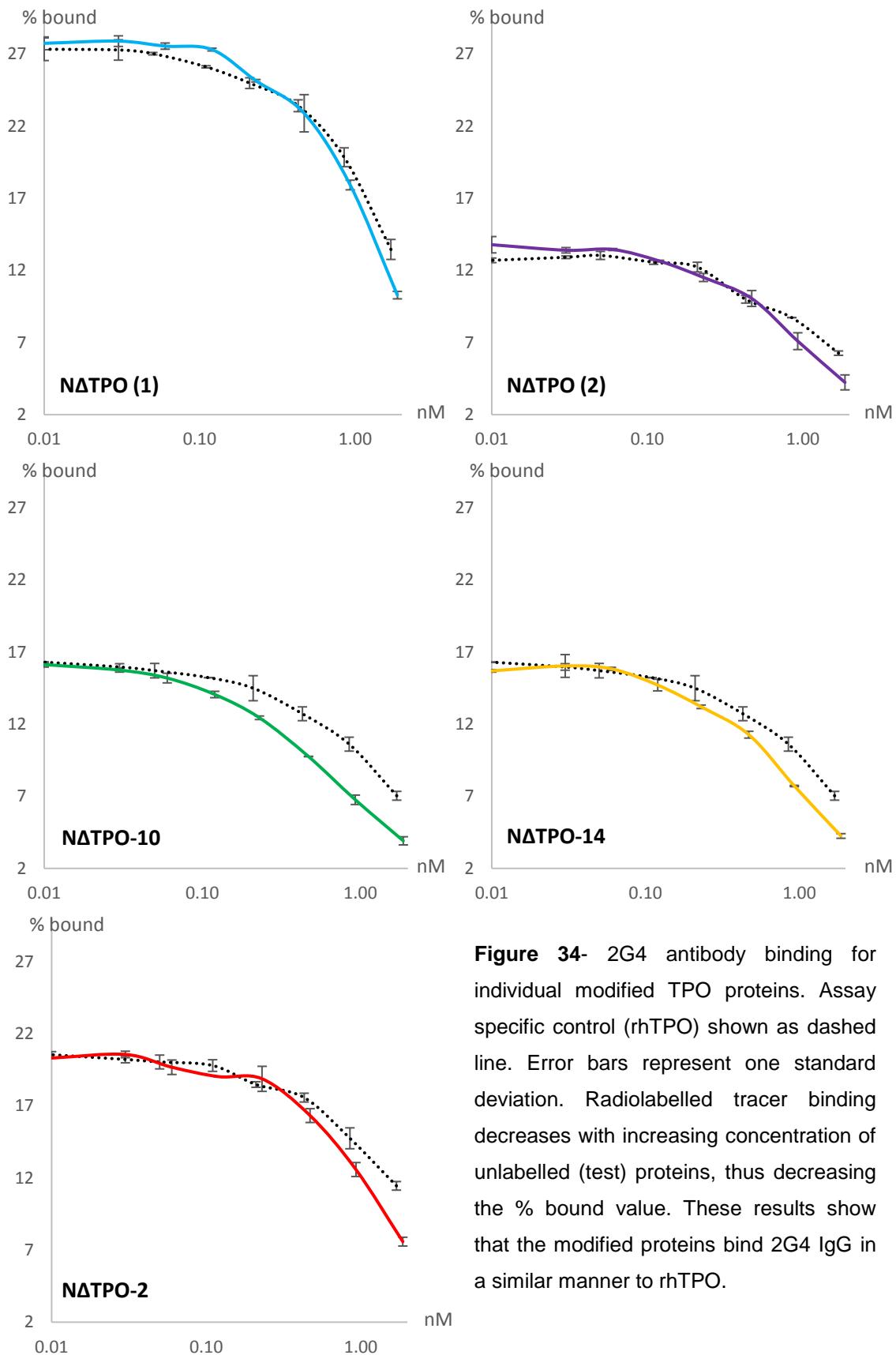


Figure 34- 2G4 antibody binding for individual modified TPO proteins. Assay specific control (rhTPO) shown as dashed line. Error bars represent one standard deviation. Radiolabelled tracer binding decreases with increasing concentration of unlabelled (test) proteins, thus decreasing the % bound value. These results show that the modified proteins bind 2G4 IgG in a similar manner to rhTPO.

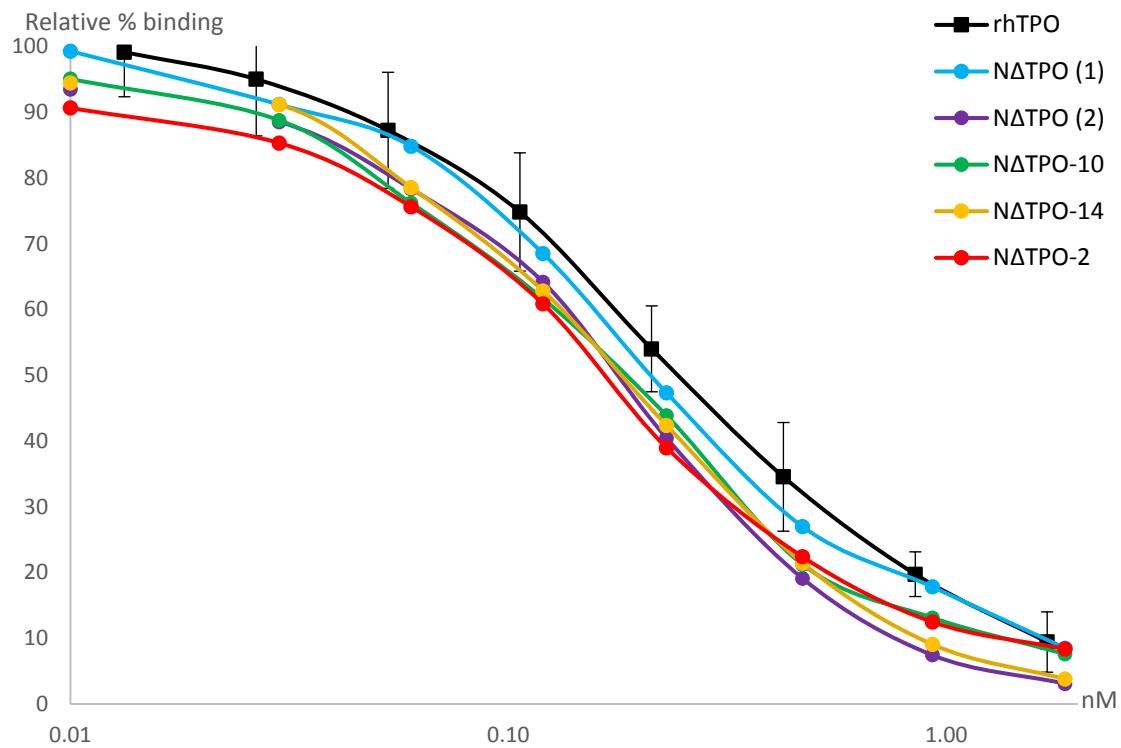


Figure 35- Pooled patient sera relative antibody binding for modified TPO proteins. Control (rhTPO) averaged across all assays. Error bars represent one standard deviation of control.

Collated from data in Fig. 33 for comparison.

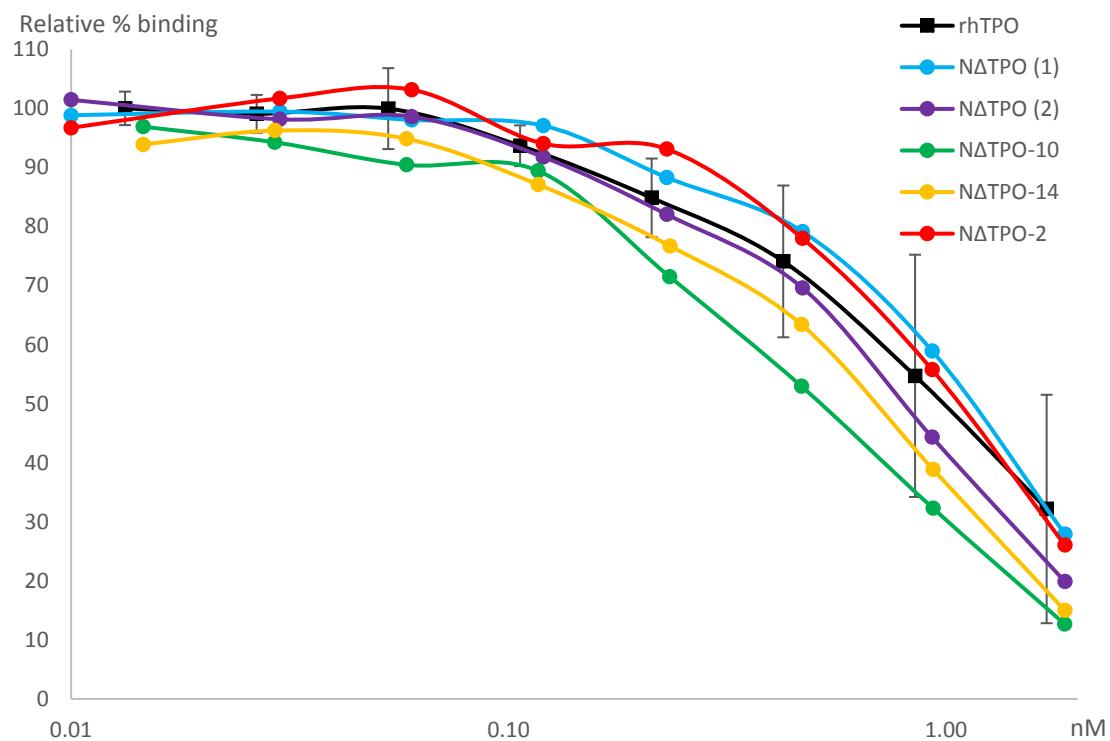


Figure 36- 2G4 relative antibody binding for modified TPO proteins. Control (rhTPO) averaged across all assays. Error bars represent one standard deviation deviation of control. Collated from data in Fig. 34 for comparison.

Enzymatic activities for the final pools determined by this assay are shown in Table 11. NΔTPO was the most active with each of the two pools of this protein produced giving a similar result despite being produced separately.

Table 11- Enzymatic activity of final pools

Modified TPO protein (run)	Enzymatic activity (GU/mg)
NΔTPO (1)	430
NΔTPO (2)	436
NΔTPO-10	367
NΔTPO-14	373
NΔTPO-2	393

Determined by guaiacol oxidation assay.

While each protein produced was active and able to oxidize guaiacol, the levels of enzymatic activity were reduced compared to reported values for rhTPO and native TPO purified from human tissue obtained by the same method³²⁴. The reduced enzymatic activity of the proteins produced for this work may be the result of poor haem incorporation. For haem peroxidases this is quantified by calculating the ratio between the absorbance of the haem group at 412 nm and the absorbance at 280 nm used to calculate protein concentration. Ratios for 412/280 nm absorbance of the protein final pools produced are shown in Table 12.

Table 12- Final pool absorbance ratios

Modified TPO protein (run)	412/280 ratio
NΔTPO (1)	0.252
NΔTPO (2)	0.246
NΔTPO-10	0.143
NΔTPO-14	0.221
NΔTPO-2	0.126

412/280 ratio calculated as absorbance (412 nm) divided by absorbance (280 nm).

Gut *et al* reported a 412/280 ratio of 0.5 for rhTPO produced in insect cells³²⁴. This preparation showed a much higher level of enzymatic activity, 1200 GU/mg, than the proteins produced here. Values reported for native TPO range

between 0.38 and 0.54 when purified from a porcine source and 0.27 to 0.4 for the human enzyme^{324–326}. The enzymatic activity levels reported for these preparations once purified are 900 GU/mg or higher.

The reason for this reduction in haem incorporation and resulting drop in enzymatic function is unknown. Godlewska *et al* previously examined the effect of removing the DNA sequence coding for amino acids 21-108 on the mature protein and found no effect on the enzymatic activity²⁶¹. However, this work was based on the membrane bound protein expressed in CHO cells, so may not be directly comparable.

Haem incorporation into TPO is thought to be an autocatalytic process as observed with LPO³²⁷. Hydrogen peroxide is a required co-factor for this process and for catalytic activity of the mature protein. Insufficient levels of hydrogen peroxide in insect cells have been suggested to cause reduced activity of peroxidases expressed using this system, although rhTPO has been previously expressed with haem incorporation and enzymatic activity levels similar to the native enzyme^{284,327}. There are also reports of inactive TPO expressed in mammalian cells gaining enzymatic function after incubation with hemin³²⁸. Samples of NATPO were incubated with hydrogen peroxide, hemin and both hydrogen peroxide and hemin together to try to improve enzymatic activity and haem incorporation levels, but no improvement was detected.

The stability of the proteins produced in this work was investigated as this can affect crystallization³²⁹. Unstable proteins may degrade in the crystallography drop, resulting in heterogeneity and reducing the likelihood of obtaining good quality crystals. This is important as crystallization often occurs on the time scale of days and weeks or longer. Crystals can be grown at lower temperatures to prevent protein degradation but generally the rate of crystal growth is also reduced.

A fluorescence based thermal shift assay was used to compare thermal stability between proteins to determine what, if any, effect the modifications had on stability. This involved mixing each protein with a dye then subjecting the sample to an increasing temperature gradient whilst monitoring the fluorescence in a qPCR instrument. For this work a commercially available kit

was used (Protein thermal shift dye kit, Applied Biosystems, USA) so the exact dye chemistry is unknown. Generally, the dyes used for thermal shift assays are quenched in a polar solvent³³⁰. As proteins unfold during the temperature gradient, the dye molecules enter the inner structure of the protein which is usually hydrophobic and buried from the outer aqueous environment. The dye is no longer quenched due to the less polar environment and produces a measurable fluorescence increase which correlates to the folding state of the protein molecules in the sample. The T_m value, defined as the peak first derivative of the raw fluorescence signal is used for comparison between samples³³¹. This represents the point at which half the protein molecules in a sample would be in the unfolded state.

The values for T_m obtained for each protein are shown in Table 13. A change in T_m of 2 °C or greater is considered significant³³². For comparison, rhTPO which gives a T_m around 52 °C, was run in each assay. Standard deviations were calculated for four replicates within the same assay (Table 13). Inter-assay standard deviation for the rhTPO control over the five assays was 0.17 °C indicating a good level of reproducibility. The values obtained for NΔTPO and NΔTPO-2 were very similar indicating there were no significant changes in stability from removing the EGF domain, modifying the N-terminus or the deletion of amino acids 251 and 252. The reduced T_m values indicate that NΔTPO-10 and -14 are slightly less stable than rhTPO. While gross misfolding of these two modified TPO proteins seems unlikely as they are recognised by TPO antibodies, it is possible the loop truncations in these proteins could weaken the overall tertiary structure making it more prone to unfolding. While significant, the reduction in stability was small so NΔTPO-10 and -14 were not excluded from crystal screening.

Table 13- Modified TPO protein melting temperatures

Modified TPO protein (run)	Apparent melting point (°C)	Standard deviation (°C)
NΔTPO (1)	52.0	0.21
NΔTPO (2)	51.2	0.15
NΔTPO-10	49.6	0.09
NΔTPO-14	49.9	0.13
NΔTPO-2	52.4	0.13

T_M values determined by thermal shift assay.

During the analysis of the four proteins purified to the final pool stage no data was produced which would exclude any of the proteins from crystal screening. While this analysis was not exhaustive it was shown that the proteins appeared to share the key features of native TPO. Analysis by techniques such as mass spectrometry (MS) and circular dichroism may have produced more interesting data on the nature of these proteins. However this would not necessarily indicate the likelihood of successful crystallization and at some point crystal screening would have to be carried out as the final test of this. For these reasons no further analysis was carried out at this point with the view that if a structure of any of the proteins was obtained, any further analysis required could be carried out at retrospectively.

Summary

A modified form of human TPO, NΔTPO, was designed with the aim of producing a protein that would form crystals that diffract to a high resolution while retaining key features of native TPO. Four further modifications to NΔTPO were also designed to target features of the protein which may have been detrimental to diffraction resolution. All proteins were designed using knowledge gained from previous work with TPO and data from sequence alignments and crystal structures of related peroxidase enzymes.

DNA coding for each of the five modified TPO proteins were produced and used to generate recombinant baculoviruses to express the proteins in an insect cell line. All five modified TPO proteins were successfully expressed, producing good yields.

Four of the five modified TPO proteins were purified using a series of ion exchange, metal affinity and size exclusion chromatography. Final pools of the four proteins were produced and concentrated to 10 mg/mL or higher. These pools were shown to be highly pure and homogeneous.

One protein, NΔTPOS was shown to be both unstable and prone to aggregation during the later stages of purification and therefore was not examined further. Ample amounts of the four other proteins were produced for characterisation and crystallography experiments.

Analysis of the four proteins showed that each bound TPO antibodies from AITD patient sera. This was a key requirement as the overall goal of this work is to learn more about the involvement of TPO in AITD on the molecular scale. The four proteins were also shown to be enzymatically active. The effect of the modifications made to each of the four purified proteins was quantified using a thermal shift assay. Modifications had little or no, negative effect on the stability of the protein compared to the native TPO extra-cellular domains. Analysis carried out on the four purified proteins indicated they should be suitable for crystallography experiments.

Key observations

- The TPO protein can be expressed in a soluble form with aa 15-108 and 797 onwards removed in an insect cell expression system (N Δ TPO).
- This truncated form of TPO binds autoantibodies and shows some enzymatic activity. Stability as determined by thermal shift assay is not affected relative to the insect expressed TPO extra-cellular domains.
- Further truncation of aa 250-251 or 298-311 is possible without effecting antibody binding in the resulting protein. Enzymatic activity is also still present although slightly reduced.

Problems

- Mutation of cysteine 146 to serine destabilised the protein.
- Enzymatic activity was reduced in all modified proteins produced compared to the unmodified insect expressed TPO extra-cellular domains. The reasons for this are unclear.
- IEF suggests there may be some slight heterogeneity amongst the final pools of modified protein.

Primers

Primers 1&2

5' GCG ATG AAA AGA AAA GTC AAC CTG TCG ACT CAA CAA TCA CAG
CAT CC 3'

5' GGA TGC TGT GAT TGT TGA GTC GAC AGG TTG ACT TTT CTT TTC
ATC GC 3'

Linker oligonucleotides

5'ATT TCA TGA GAG CGC TCG CTG TGC TGT CTG TCA CGC TGG TTA
TGG CCG 3'

5'TCG ACG GCC ATA ACC AGC GTG ACA GAC AGC ACA GCG AGC GCT
CTC ATG 3'

Primers 3&4

5'CGC TGG TTA TGG CCA CTC AAC AAT CAC AGC ATC C 3'

5'GGA TGC TGT GAT TGT TGA GTG GCC ATA ACC AGC G 3'

Primers 5 & 6

5'CAG CCT CCC CTC TGC AAA GAT CAT CAC CAT CAC CAT CAC TAG
AAG CTT GGC ACT GGC 3'

5'GCC AGT GCC AAG CTT CTA GTG ATG GTG ATG GTG ATG ATC TTT
GCA GAG GGG AGG CTG 3'

Primers 7 & 8

5'TGC CCA AAC ACT TCC CTG GCG AAC AAA TAC AGG 3'

5'CCT GTA TTT GTT CGC CAG GGA AGT GTT TGG GCA 3'

Primer 9 & 10

5'GCC ACC GAG GTA CCC TCC CTG ACG 3'

5'CGT CAG GGA GGG TAC CTC GGT GGC 3'

Primer 11 & 12

5'GCA CCA GCA AAT TCG GGG GAG GGG C 3'

5'GCC CCT CCC CCG AAT TTG CTG GTG C 3'

Virus titration primers

5'CCA GTG GGT CAA AGG CAA A3'

5'GTG GTG CGC AAA GTG ATT GT 3'

Chapter 3- Crystallography of Modified Thyroid Peroxidase Proteins

Introduction

After designing, producing and analysing four modified proteins based on TPO, these proteins were next tested to determine if they could be used to obtain a 3D structure. To date, efforts to produce a structure for TPO have focused on single crystal X-ray diffraction. This is the most widely used technique to obtain structures for proteins of this size (105 kDa for the full-length, wild type protein).

Even if only the extra-cellular domains of TPO are considered, at 94 kDa this is well above the size of the proteins which are routinely solved by nuclear magnetic resonance spectroscopy³³³. While some structures for some larger proteins have been produced, the majority of NMR structures in the PDB are of proteins of 20 kDa or less. Many of the larger protein structures solved by NMR are of heteromeric proteins, in which individual subunits have been solved separately. Producing isotopically labelled TPO for NMR studies could also prove difficult. Expression of full length or the extra-cellular domains of TPO in yeast gives very low yields³⁰⁹. Large quantities of immunologically active TPO have only been successfully produced in insect or higher eukaryotic cells^{283,284}. Isotopic labelling with ¹³C, ¹⁵N and, to some extent, ²H of proteins produced in insect cells is possible, however it can be difficult and prohibitively expensive³³⁴.

An alternative protein imaging technique which is currently emerging is cryogenic electron microscopy (Cryo-EM). Cryo-EM as a technique for imaging protein molecules has been in development since the 1970's, with atomic resolution structures first produced in 1990^{335,336}. As for protein crystallography, samples need to be highly pure and homogeneous but crystals are not required for cryo-EM³³⁷. Cryo-EM images are produced by directing a beam of electrons through a protein sample in vitrified ice. Data sets, typically containing hundreds of thousands of images are classified into different views corresponding to 2D projections and a 3D model structure that can give rise these projections is generated. The resolution of cryo-EM structures is constantly improving, at the

time of writing 23% of the structures deposited in the electron microscopy data bank (<https://www.ebi.ac.uk/pdbe/emdb/>) had an effective resolution equal to or better than 5 Å. However, most cryo-EM structures are of large proteins or complexes, with only 2% of the structures deposited being of molecules ≤100 kDa. Wild-type TPO would be around the lower size limit of what it is possible to image by cryo-EM at the moment. Atomic resolution structures of proteins at the lower end of the molecular weight limits are rare, so improving on previous results generated by X-ray diffraction studies would be difficult. Protein heterogeneity, suspected to be contributing to the low diffraction resolution of TPO crystals, would also negatively affect the resolution of any cryo-EM data obtained³³⁷. For these reasons, X-ray crystallography remained the method of choice when studying the modified TPO proteins produced in this work.

Results and Discussion

Crystal screening of modified thyroid peroxidase proteins

Key points

- Sparse matrix crystal screening was carried out for the four modified TPO proteins.
- NΔTPO and NΔTPO-2 produced promising results.
- NΔTPO-2 crystals diffracted to 8 Å.
- NΔTPO crystals diffracted to 5.2 Å.

Crystallization conditions for each of the modified TPO proteins produced in Chapter 2 were screened using a sparse matrix approach. Sparse matrix crystal screening involves testing a wide range of salts, buffers and additives to find conditions that produce crystals. Large, single crystals suitable for diffraction studies are unlikely to be generated in these screens for most proteins. Sparse matrix screens are more likely to identify conditions which produce promising signs, such as polycrystalline precipitates. These conditions then act as a starting point for further screening. Successive rounds of screening around an increasingly narrow set of conditions allow crystal growth to be improved to the point where single crystals suitable for diffraction experiments are grown.

Three commercial sparse matrix screens were selected for the initial screening of conditions for crystallization. These were structure screens 1 & 2 (considered a single screen), JCSG+ and PACT Premier all produced by Molecular Dimensions (UK). Structure screen 1 & 2 was one of the earliest sparse matrix screens to be designed³³⁸. The conditions included were chosen based on successful crystallography experiments from published literature at the time. They were then modified to avoid repeating similar conditions in the same screen and to cover as wide a range of conditions as possible. This gave crystallographers a standard set of conditions with which to test new proteins when beginning new experiments.

JCSG+ and PACT Premier were produced more recently and were designed to be used in conjunction. The conditions in the JCSG+ screen were selected by examining which precipitants from earlier screens produced the most crystal hits when tested against a large number of proteins from a single organism³³⁹. The PACT screen was designed more systematically, and contains polyethylene glycol (PEG) in every condition. A wide range of PEG sizes and concentrations are frequently used in crystal screening. PEG is thought to promote crystallization by excluding water molecules from the protein surface³⁴⁰. Data seems to indicate, in most cases, the presence of PEG has a beneficial effect on crystallography outcomes. This screen is essentially made up of several sub-screens designed to examine the effects of pH and cation or anion presence separately. The use of these three screens together did result in some duplicated or very similar conditions being tested, but this was not so extensive as to be considered a waste of resources.

While conditions for the crystallization of insect derived rhTPO have been determined previously, the crystals produced diffracted poorly²⁵⁵. The modifications made to the proteins produced for this work meant these conditions may not have been suitable. A small screen around the conditions which produced rhTPO crystals was tested with each modified TPO protein to confirm this. This consisted of twelve conditions ranging between 100 mM Tris-HCl pH 6.5-7.5 containing 12.5-17.5% PEG 4K.

Crystal screens were set up in either a sitting or hanging drop format. In each case vapour diffusion between a reservoir of precipitant and a drop of protein mixed with the same precipitant causes the drop to lose volume, concentrating the protein and, if successful, inducing crystal nucleation. Some screening was carried out by Charles River Laboratories (UK) as a drop-setting crystallography robot was not available during some stages of the project.

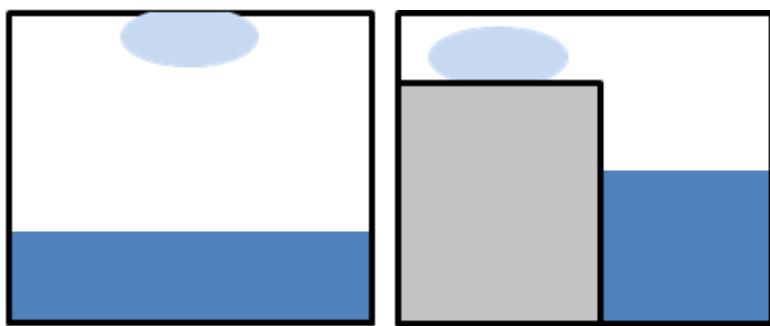


Figure 37– Side view diagram of hanging drop (left) and sitting drop (right) crystal screening formats showing protein/precipitant mixed drop (light blue), precipitant reservoir (dark blue), section through plastic plate (grey). In each format water vapour diffuses from the drop to the reservoir causing the protein to concentrate and precipitate.

Four modified TPO proteins (Δ TPO (1), Δ TPO-10, Δ TPO-14 and Δ TPO-2, Chapter 2) were tested in each of the three sparse matrix screens and the rhTPO based screen detailed above. Proteins were screened at the concentration of the final pool (i.e. ≥ 10 mg/mL, Chapter 2, “Protein production”). Initial screens were set up by hand with a total drop volume of 2 μL consisting of protein and precipitant in a 1:1 ratio, suspended over a 100 μL reservoir.

Screens set up by hand were overall less successful for all the modified TPO proteins than those set up by robot (discussed later) with Δ TPO-10 and Δ TPO-14 producing no hits in any of the three screens. However, Δ TPO did produce some potential crystal hits, as shown below (Fig. 38). The full growth conditions are listed in Table 14, but all contained PEG suggesting the presence of PEG could be important for the crystal growth.

Table 14– Summary of results of modified TPO proteins in sparse matrix crystal screens with manual drop setting.

Protein	Screen	Outcome
NΔTPO 11 mg/mL	Structure screen 1&2	Hit in B3: 200 mM magnesium acetate, 100 mM sodium cacodylate (pH 6.5), 20% w/v PEG 8000 (Fig. 38A)
NΔTPO 11 mg/mL	JCSG+	Hit in A12: 200 mM potassium nitrate, 20% w/v PEG 3350 (Fig. 38B)
NΔTPO 11 mg/mL	PACT Premier	Hit in D11: 200 mM calcium chloride, 100 mM Tris-HCl (pH 8.0), 20% w/v PEG 6000 (Fig. 38C) Hit in E4: 200 mM potassium thiocyanate, 20% w/v PEG 3350 (Fig. 38D)
NΔTPO-2 12 mg/mL	PACT Premier	Hit in D11: 200 mM calcium chloride, 100 mM Tris-HCl (pH 8.0), 20% w/v PEG 6000 (Fig. 39A) Hit in E4: 200 mM potassium thiocyanate, 20% w/v PEG 3350 (Fig. 39B)

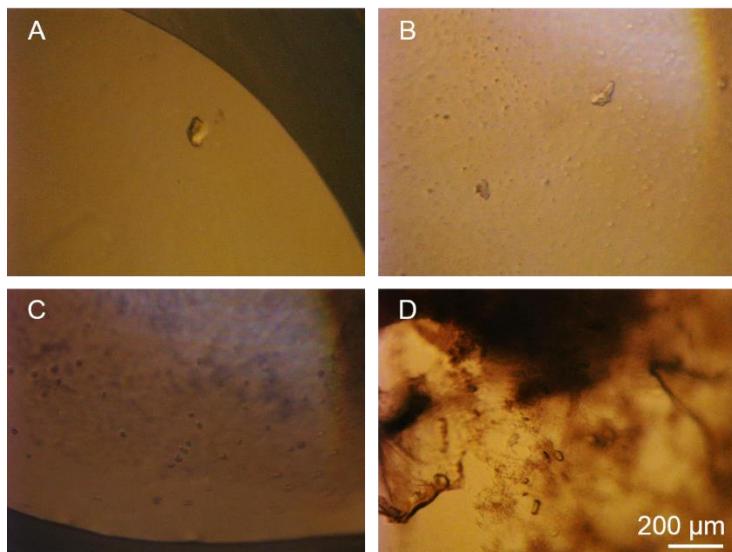


Figure 38– Micrographs of hits from manual NΔTPO crystal screening grown in **A**) 200 mM magnesium acetate, 100 mM sodium cacodylate (pH 6.5), 20% w/v PEG 8000, **B**) 200 mM potassium nitrate, 20% w/v PEG 3350, **C**) 200 mM calcium chloride, 100 mM Tris-HCl (pH 8.0), 20% w/v PEG 6000, **D**) 200 mM potassium thiocyanate, 20% w/v PEG 3350. Scale bar is approximate. These objects are thought to be small crystals not suitable for X-ray diffraction.

The NΔTPO crystals (Fig. 38) were not tested for diffraction as they were of insufficient size and often grown in drops containing lots of amorphous precipitate, making them difficult to harvest. Amorphous precipitate is typically disordered, consisting of denatured or aggregated protein³⁴¹. Depending on the kinetics of precipitation, this may redissolve given time. Crystals can still form in these drops if the concentration of protein in its native confirmation remains high enough, but it is less likely as the aggregate may provide many alternative nucleation points and the aggregate may physically interfere during crystal harvesting. In extreme cases, the precipitate forms in thick, brown clumps which cannot be resolubilised, completely preventing crystal growth.

It should also be noted that without testing for X-ray diffraction it can be very difficult to confirm that crystals are made of protein and are not formed from components of the buffer³⁴². For this reason it is not possible to confirm whether these crystals consisted of protein or not. However, this was not an issue as no further attempts to optimise these hits were made. Screens set up using a crystallography robot appeared to produce more promising results and required far less protein. As such, further efforts concentrated on this method (Figs. 40, 42 and 43 and accompanying text).

NΔTPO-10 and NΔTPO-14 behaved in a similar manner to each other in manually set sparse matrix screening. The majority of conditions produced only varying degrees of amorphous precipitate. The three sparse matrix screens were repeated with NΔTPO-10 diluted to 8 mg/mL to check if the initial protein concentration was too high. This did not improve the overall screening outcomes, but increased the frequency of clear drops where the protein remained in solution. In some conditions, most of which were shared between NΔTPO-10 and NΔTPO-14, some phase separation was observed, but no crystalline precipitate or potential crystals were observed. Phase separation is a common occurrence in crystallography, particularly when high salt concentrations or PEG containing precipitants. Under these conditions colloid droplets form containing high concentrations of protein³⁴³. If crystals are nucleated, they may grow along the interface of these droplets but this was not observed in these experiments.

NΔTPO-2 produced spherulites or possibly small crystals in one well containing calcium chloride (200 mM), Tris-HCl (100 mM, pH 8.0) and PEG 6K (20% v/v) and one containing potassium thiocyanate (200 mM) and PEG 3350 (20% v/v) from the PACT Premier screen (Fig. 39). These were too small for conventional X-ray diffraction experiments and the drops which produced them also contained some amorphous precipitate, indicating these conditions were not optimal.

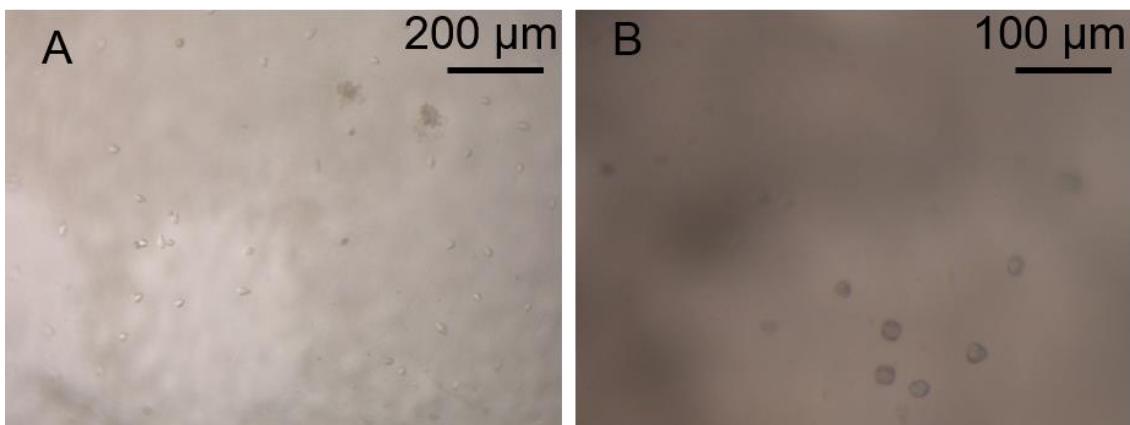


Figure 39— Micrographs of spherulites or small crystals of NΔTPO-2 grown in **A**) 200 mM calcium chloride, 100 mM Tris-HCl (pH 8.0), 20% w/v PEG 6000 or **B**) 200 mM potassium thiocyanate, 20% w/v PEG 3350. Scale bar is approximate. These were too small for X-ray diffraction experiments so further optimisation attempts were made to grow larger crystals (see below text).

Attempts were made to repeat and optimise the NΔTPO-2 hits with a manual set up without success. Screens were set calcium chloride or potassium thiocyanate (200 mM) adjusted pH values ranging between 7.6 and 10.0 and a range of PEG 3350 concentrations between 15 and 25%. These screens included direct repeats of the same conditions producing the initial hits. Lithium and sodium nitrate salts (200 mM) were also tested with the same pH and PEG conditions as those successful when screening NΔTPO and NΔTPO-2 using a crystallography robot. Each NΔTPO-2 optimisation screen was also tested at 12 mg/mL and a reduced protein concentration of 8 mg/mL to determine if this had any effect on the outcomes. The initial hits proved difficult to replicate and no improvement in growth could be made.

Crystallography conditions for each modified TPO protein were also screened in parallel by Charles River Laboratories using a crystallography robot, as this

facility was not available locally at the appropriate stage of this work. Screens set up by robot had a total drop volume of 400 nL, consisting of protein and precipitant in a 1:1 ratio in a hanging drop with a 60 µL reservoir. This was advantageous, as drop size can effect crystallization outcomes and the optimal drop size can vary between proteins³⁴⁴. Small local differences during the setup of screens can also have unpredictable effects on results. Screening with a robot uses small amounts of protein compared to what is possible with a manual set up so repeating screens on the robot was an efficient use of the available material.

Table 15– Summary of results of screening modified TPO proteins in sparse matrix crystal screens carried out by Charles River Laboratories with a crystallography robot

Protein	Screen	Outcome
NΔTPO 11 mg/mL	JCSG+	Hit in G6:200 mM sodium malonate, 20% w/v PEG 3350 (Fig. 40A)
NΔTPO 11 mg/mL	PACT Premier	Hit in E3: 200 mM sodium iodide,20% w/v PEG 3350 Hit in E5:200 mM sodium iodide, 100 mM bis-Tris-HCl propane (pH 6.5), 20% w/v PEG 3350 (Fig. 40B), Hit in E9:200 mM potassium sodium tartrate,20% w/v PEG 3350 Hit in E12:Same precipitant as JCSG+ G6 Hit in F3:200 mM sodium iodide,100 mM bis-Tris propane (pH 6.5), 20% w/v PEG 3350 Hit in F4:200 mM potassium thiocyanate, 100 mM bis-Tris propane (pH 8.5), 20% w/v PEG 3350 Hit in H5:200 mM sodium iodide, 100 mM bis-Tris propane (pH 8.5), 20% w/v PEG 3350 (Fig. 40C)
NΔTPO-2 12 mg/mL	PACT Premier	Hit in E3:200 mM sodium iodide,20% w/v PEG 3350 Hit in F4: 200 mM potassium thiocyanate, 100 mM bis-Tris propane (pH 8.5), 20% w/v PEG 3350 Hit in H5:200 mM sodium iodide, 100 mM bis-Tris propane (pH 8.5), 20% w/v PEG 3350

Screening by robot produced better results for NΔTPO and NΔTPO-2. Small crystals of NΔTPO were grown in several conditions in the PACT Premier screen (Fig. 40) that covered a range of sodium and potassium salts. Bis-Tris propane was present in some hits as a buffer with a pH of 6.5 or 8.5. This indicated a range of pH values were suitable for crystal growth. Again, PEG appeared crucial to crystal growth as all conditions which produced hits contained 20% PEG 3350. Sparse matrix screening of NΔTPO-2 by robot produced three hits in the PACT premier screen in conditions which were also successful for NΔTPO.

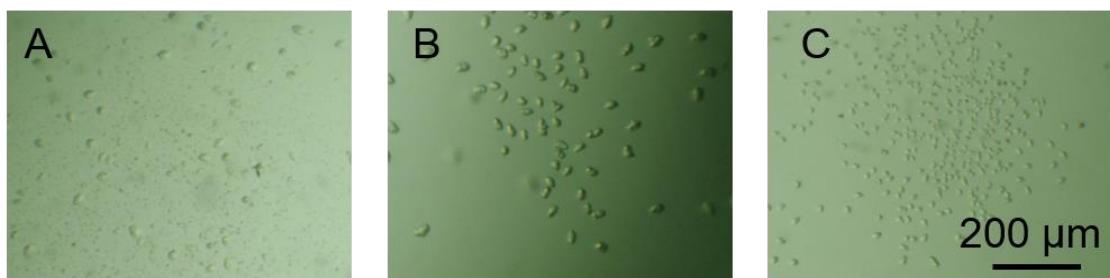


Figure 40- Micrographs of NΔTPO crystals grown at Charles River Laboratories in initial screens under various conditions: **A)** 200 mM sodium malonate, 20% w/v PEG 3350, **B)** 200 mM sodium iodide, 100 mM bis-Tris propane (pH 6.5), 20% w/v PEG 3350, **C)** 200 mM sodium nitrate, 100 mM bis-Tris propane (pH 8.5), 20% w/v PEG 3350. Scale bar is approximate. These crystals were too small to be suitable for X-ray diffraction experiments. The growth conditions which produced these crystals were used as a starting point for further optimisation.

Sparse matrix screening of NΔTPO-10 and NΔTPO-14 by a crystallography robot produced the same results as from the manual screen, no clear positive results were generated. Screens were also attempted with the reducing agent tris(hydroxypropyl)phosphine (1 mM) added to NΔTPO-10 and NΔTPO-14 prior to screening as they both appeared more prone to aggregation than NΔTPO and NΔTPO-2, but this did not improve the results.

The NΔTPO and NΔTPO-2 hits from the sparse matrix screens were optimised to produce larger crystals more suitable for diffraction studies. Optimisation screens covered a range of salts, PEG 3350 concentrations between 10 and 26% and pH values between 6.5 and 9.5, buffered with Tris-HCl or bis-Tris propane. Later screens also tested lithium and ammonium salts. Once conditions for growing large single crystals were established an additive screen (Hampton Research, USA) was carried out. The 96 additives in the screen were

tested with 200 mM lithium nitrate, 100 mM Tris-HCl pH 8.5, 15% PEG 3350 as the precipitant to determine if there was any improvement in NΔTPO crystal growth.

Screening at Charles River Laboratories was guided by results from an in-house X-ray source, allowing conditions to be selected based on the diffraction resolution of the crystals produced. Large crystals with good morphology were produced using 200 mM lithium nitrate, 100 mM Tris-HCl pH 7.0-9.0, 10-15% PEG 3350 as a precipitant (Fig. 42). The additive screen produced some hits, but these were all smaller and diffracted less well than crystals produced without the additive and so were not investigated further.

Samples of the optimised NΔTPO crystals were screened for diffraction at Diamond light source, U.K by Charles River Laboratories. The crystals did diffract, albeit to low resolution. The highest resolution data obtained contained reflections to 7 Å and was therefore no improvement on previous work²⁵⁵. It was possible to determine the probable space group and unit cell dimensions from this data to be P3, $a = b = 145 \text{ \AA}$, $c = 283 \text{ \AA}$. This is similar to the unit cell dimensions reported previously for rhTPO expressed in insect cells which crystallized in P321, P3₁21 or P3₂21 where $a = b = 100 \text{ \AA}$, $c = 215 \text{ \AA}$ ²⁵⁵.

All of these space groups belong to the trigonal crystal system, displaying threefold (120°) rotational symmetry around the c axis³⁴⁵. P3 is the simplest and has no further symmetry (Fig. 41). The other space groups proposed previously for rhTPO crystals display additional symmetry elements in addition to the rotational symmetry around the c axis. In the case of P3₁21 and P3₂21 this is in the form of a screw axis where, in addition to a rotation of 120° around the c axis, each point is translated one third of the axis length. This produces an anticlockwise (P3₁21) or clockwise (P3₂21) spiral of identical points along the axis throughout neighbouring unit cells. The space groups P321, P3₁21 and P3₂21 also display additional axes of rotational symmetry along the a and b axis.

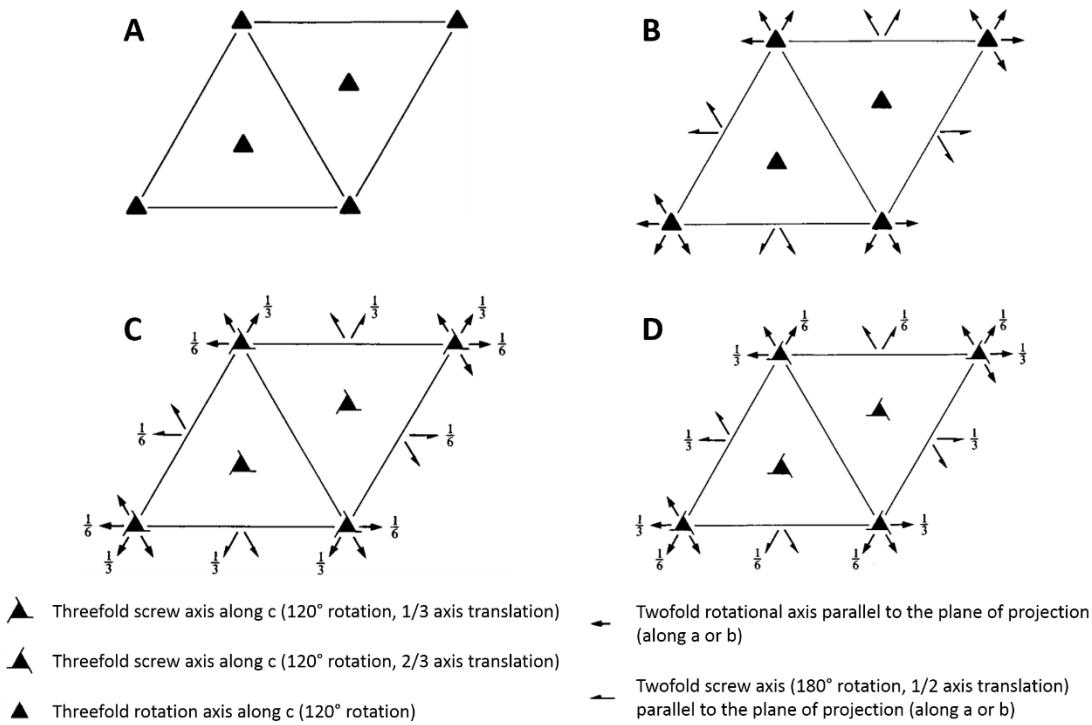


Figure 41– Symmetry elements in space group P3 (**A**), P321 (**B**), P3₁21 (**C**) and P3₂121.

Adapted from International Tables for Crystallography, Vol. A³⁴⁵.

NΔTPO-2 crystallized in similar conditions to NΔTPO (200 mM lithium nitrate, 100 mM Tris-HCl pH 7.0-9.0, 10-15% PEG 3350). Crystals of NΔTPO-2 were tested for diffraction at Diamond light source in the same manner as the NΔTPO crystals. The resolution was slightly lower than that obtained with NΔTPO, with the best crystals diffracting to 8 Å. The probable space group of the best diffracting NΔTPO-2 crystal was also P3 with unit cell dimensions of $a = b = 152$ Å, $c = 294$ Å.

Table 16– Summary of modified TPO optimisation screens set up with a crystallography robot by Charles River Laboratories

Protein	Screen	Outcome
NΔTPO 11 mg/mL	200 mM sodium nitrate/sodium malonate dibasic monohydrate/sodium iodide/potassium thiocyanate 15-20% w/v PEG 3350	Hits in multiple conditions
NΔTPO 11 mg/mL	200 mM sodium nitrate 15-26% w/v PEG 3350	Hits in multiple conditions. 200 mM sodium nitrate, 18% w/v PEG 3350 produced best morphology (Fig. 42A)
NΔTPO 11 mg/mL	200 mM sodium nitrate 100 mM Tris-HCl pH 7.5-9.0 10-20% w/v PEG 3350	Hits in multiple conditions
NΔTPO 11 mg/mL	200 mM sodium nitrate/lithium nitrate/potassium nitrate/ammonium nitrate 100 mM Tris-HCl pH 7.5-9.0 10-20% w/v PEG 3350	Hits in multiple conditions, lithium nitrate crystals diffracted to 7 Å (see Fig. 42B, C)
NΔTPO 11 mg/mL	200 mM lithium nitrate 100 mM Tris-HCl pH 8.5 15% w/v PEG 3350 + Hampton Research additive screen	Hits in multiple conditions
NΔTPO-2 12 mg/mL	200 mM sodium nitrate/sodium malonate dibasic monohydrate/sodium iodide/potassium thiocyanate 15-20% w/v PEG 3350	Hits in multiple conditions
NΔTPO-2 12 mg/mL	200 mM sodium nitrate/lithium nitrate/potassium nitrate/ammonium nitrate 100 mM Tris-HCl pH 7.5-9.0 10-20% w/v PEG 3350	Hits in multiple conditions, lithium nitrate crystals diffracted to 8 Å

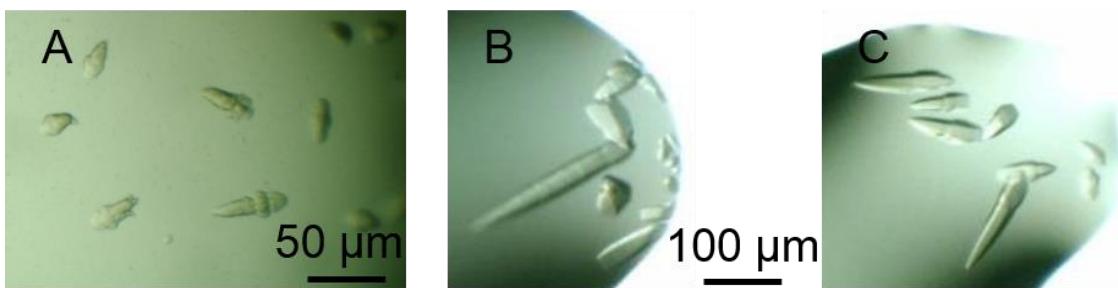


Figure 42- Micrographs of optimised crystals grown at Charles River Laboratories in **A**) 200 mM sodium nitrate, 18% w/v PEG 3350. Further optimisation produced crystals in **B**) 200 mM lithium nitrate, 0.1 M Tris-HCl pH 9.0, 14% w/v PEG 3350 and **C**) 200 mM lithium nitrate, 0.1 M Tris-HCl pH 7.5, 13% w/v PEG 3350. Crystals depicted in **B** and **C** diffracted up to 7 Å at Diamond light source. Scale bar is approximate.

There are a number of possible explanations for this disparity of results between the two methods of setting up crystal screens. Plates set up by hand used 96-well screw top hanging drop plates (Molecular Dimensions, UK). This required each drop to be dispensed before the plate is sealed. Dispensing drops and reinserting each well cover (the removable top on which the drop sits) individually was attempted but it proved difficult to reliably form an air-tight seal once the covers had been unscrewed. Dispensing drops to for the whole plate takes some time by hand and may have led to less reproducible results. Changes in local temperature or humidity can have a large effect on evaporation from a very small volume as can the amount of time between the drop being dispensed and the plate being sealed. Some drops may dehydrate quickly, particularly if the precipitant contains volatile compounds, causing rapid concentration and the formation of amorphous precipitate. There are alternative formats for screening in which each condition is set up and sealed separately. For example, crystallization drops can be set up on glass cover slips and mounted on wells individually. The lip of the well containing the reservoir is coated with grease before adding the cover slip to form an air-tight seal. This prevents individual drops being exposed to the air for extended periods of time. The drawback of this method is that it is more time consuming, requiring protein stocks to be kept on ice for extended periods or using multiple frozen aliquots if a large screen is being set up. Setting up a plate using a crystallography robot is quicker, uses smaller amounts of protein and is more accurate than by hand.

For these reasons, when possible, screening was carried out using a robot in preference to a manual set up.

Following the initial screening of the four proteins described above, a crystallography robot became available for use in this study at Cardiff University. Comparison of manual and robotically set screens showed generally favourable results produced with the robot, which also consumed less protein. For this reason robotic drop setting was the preferred method whenever possible. In some early cases when the use of a robot had to be outsourced, manually set screening was also carried out to determine if there was a significant difference between the two methods.

Further work crystallography experiments focused on NΔTPO which appeared the most promising modified TPO proteins produced so far. Using the robot the conditions for NΔTPO crystal growth identified at Charles River Laboratories were re-examined to see if they could be further optimised. The PACT Premier and JCSG+ sparse matrix screens were also repeated to search for new hits to follow up.

All hits generated in the repeat NΔTPO sparse matrix screens were in conditions similar to those already known to produce NΔTPO crystals (Fig. 43). When examining the previously optimised conditions, NΔTPO crystals could be grown reproducibly in screens containing 200 mM lithium or sodium nitrate, 100 mM Tris-HCl (pH 7.5-9.0), 10-20% w/v PEG 3350. The size and morphology of these crystals were similar to those previously observed. In many cases crystals appeared to nucleate readily, producing large numbers of very small crystals. To encourage the growth of large, single crystals experiments were carried out using the above conditions with drops set up in 1:2, 1:3, and 2:3 protein:precipitant ratios. While these screens produced crystals, there was no obvious correlation between drop ratio and crystal nucleation rate.

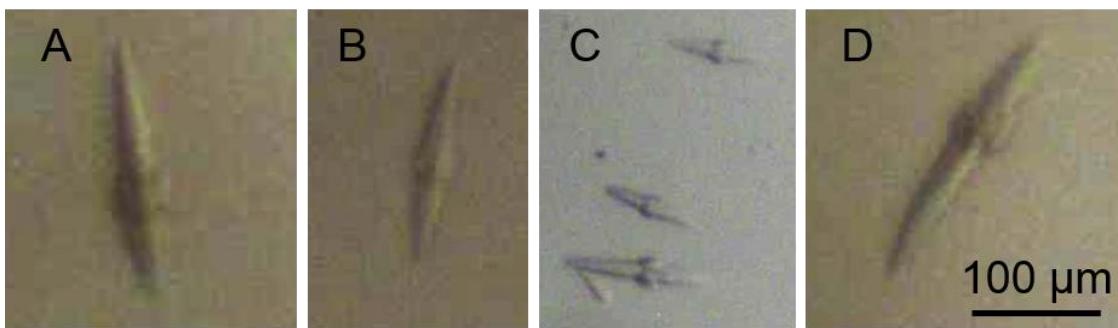


Figure 43– Micrographs of NΔTPO crystal grown from experiments set up by robot grown in 200 mM lithium or sodium nitrate, 100 mM Tris-HCl pH 7.5-9, 15-17.5% w/v PEG 3350. Scale bar is approximate. These crystals are examples of those produced at Cardiff University from which the highest resolution diffraction data was produced.

When harvesting NΔTPO crystals from the plates it was observed that the addition of 20% ethylene glycol seemed to partially dissolve the crystals. This finding was surprising as ethylene glycol is essentially the monomer of PEG which was present in the crystallization buffers (Fig. 44). Lowering the ethylene glycol concentration appeared to prevent this. This problem was not observed when using glycerol as a cryoprotectant up to 25%. As PEG can act as a cryoprotectant, some samples were frozen without further additives.



Figure 44– Structure of ethylene glycol (**1**) and polyethylene glycol (**2**).

A total of 32 single NATPO crystals, grown at Cardiff University, were screened for diffraction on the I04 beamline at Diamond light source. For each crystal, three test images were taken, 90° apart to estimate the diffracting quality of the crystal. Full data sets of 2000 images over 200° total oscillation were taken from three crystals which diffracted to ≥ 7 Å in the test images. The highest resolution dataset obtained diffracted to 5.2 Å, a slight improvement over previous work but still not sufficient to produce an accurate structure. This crystal was grown in 200 mM sodium nitrate, 100 mM Tris-HCl pH 8.6, 17% PEG 3350 in a 2:3 protein to precipitant ratio drop.

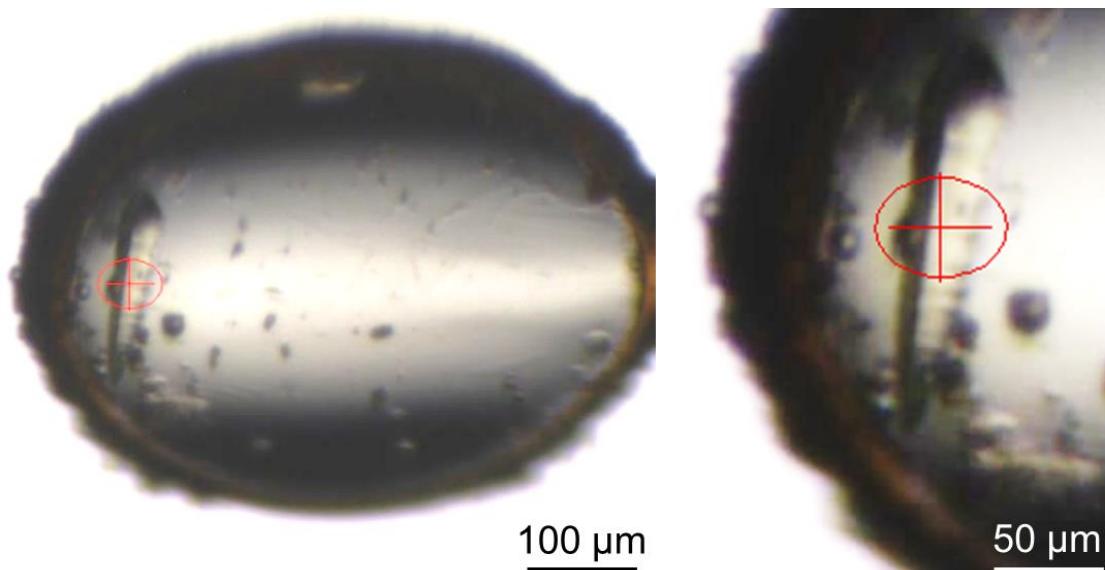


Figure 45– NATPO crystal mounted on I04 beamline. Red crosshair/circle represents position and size of X-ray beam for alignment purposes. Scale bar is approximate. This crystal diffracted to 5.2 Å.

All of the diffracting crystals produced low intensity reflections, as can be seen in Fig. 46. Many more reflections are measured in the image by the processing software than can be identified visually but a lack of visible reflections illustrates the low intensities observed. The distance of the detector from the crystal was increased in an attempt to improve reflection intensity. This limited the resolution of the data which could be collected to 3 Å but, as the crystals did not diffract to such a resolution, no data was lost as a result.

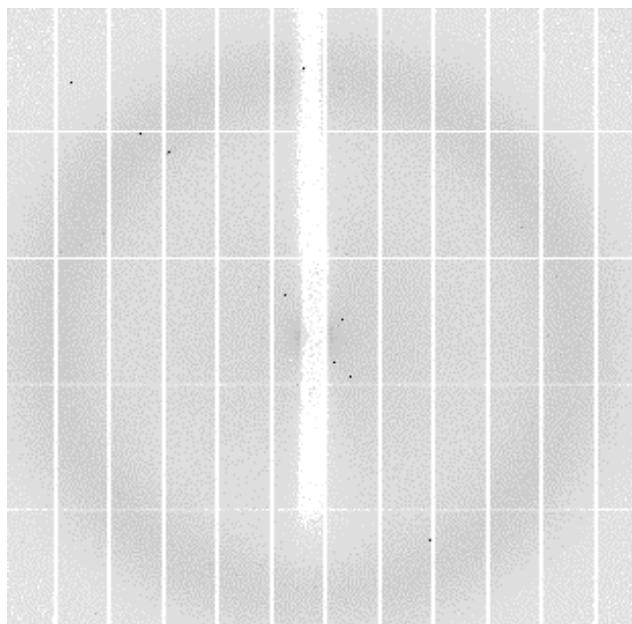


Figure 46– Example image of X-ray diffraction data collected from NΔTPO crystal. The scarcity of visible spots demonstrates the low intensity of the data gathered. Fortunately more reflections can be measured by the detector than it is possible to discern visibly.

The unit cell dimensions for this crystal were $a = b = 144 \text{ \AA}$, $c = 282 \text{ \AA}$. These corresponded very closely to those obtained by Charles River Laboratories for the same protein. The probable space group for this crystal was P6₃22. This space group has a screw axis along the c axis with sixfold (60°) rotational symmetry and a translation of half the unit cell length. In addition there are rotational symmetry axis along the a and b axis. The increased symmetry of this space group generates several additional rotational and screw axes as can be seen in Fig. 47.

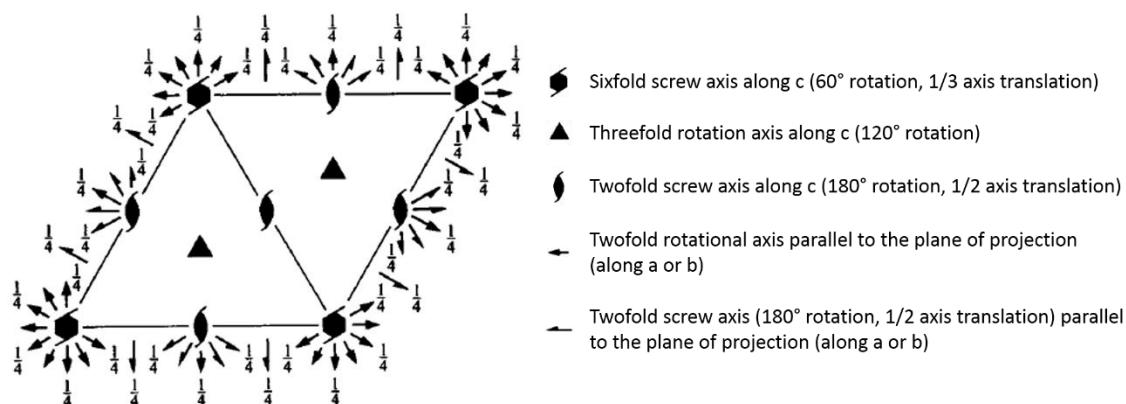


Figure 47– Symmetry elements in space group P6₃22. Adapted from International Tables for Crystallography, Vol. A³⁴⁵.

The difference in space group between the crystals grown at Cardiff University and Charles River Laboratories may be due to the difficulty of correctly determining the space group from initial data, particularly if it is of low intensity³⁴⁶. Trigonal space groups (P3, P321, P3₁21, P3₂21) have three fold rotational symmetry along the c axis where hexagonal space groups (P6₃22) have six fold rotational symmetry along the same axis³⁴⁵. Therefore, both trigonal and hexagonal space groups share the same unit cell constraints ($a = b$, $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$) and systematic absences which can be used to determine space groups.

Many of the reflection conditions used to determine the exact space group are also shared across each of the proposed space groups. Even with higher intensity and resolution the exact space group can only be determined with certainty by at least partially solving the crystal structure³⁴⁶.

Some of the NATPO crystals produced appeared to be made up of two separate forms of growth in a single crystal. These crystals had a multifaceted central region with, usually much larger, pyramidal sections protruding from one or two sides (Fig. 43, particularly C). It was hypothesised this could be caused by two different types of packing orders occurring in the same crystal. Unfortunately, it was not possible to test this theory by examining the X-ray diffraction characteristics of each region. The central square sections often broke off and became separated from the more stable, pointed end regions during harvesting. These central regions alone were also quite small and therefore difficult to align in the X-ray beam. For these reasons it was not possible to determine if the space groups of each section were different.

The Matthews coefficient was estimated for the highest resolution data set to determine if a high solvent content in the crystals may have affected diffraction. The Matthews coefficient is the volume of the unit cell per unit of protein molecular weight³⁴⁷. From this the solvent content of the crystal can be calculated. Table 17 shows the Matthews coefficients for the observed unit cell size and protein molecular weight calculated from the "Matthews_coef" component of the CCP4 software suite³⁴⁸. These results are compared to existing data from previously solved crystal structures to give a probability for

each number of molecules in the asymmetric unit³⁴⁹. The results indicate there are likely to be two molecules in the asymmetric unit and the solvent content of the crystal is around 50%, which is a reasonable value. A 2003 survey of the PDB calculated the mode average solvent content of all deposited protein crystal structures to be 47%³⁴⁹. A lower solvent content in protein crystals has been shown to correlate with increased diffraction resolution. In this case the solvent content was estimated to be close to the average value for diffracting protein crystals.

Table 17– Results output from “Matthews_coef” component of the CCP4 software suite

No. of molecules in asymmetric unit	Matthews coefficient	Solvent content %	Probability
1	4.98	75.3	0.01
2	2.49	50.6	0.98
3	1.66	25.9	0.01
4	1.24	1.2	0.00

A structure was not produced from this data due to the low resolution. A low resolution structure of rhTPO expressed in insect cells has been produced previously at 6 Å²⁵⁸. The slight increase in diffraction resolution of the data produced here would not be sufficient to make a significant improvement to this existing structure. Instead efforts were focused on producing higher resolution data.

Further crystallography experiments

Key points

- Further work was focused on NΔTPO as this seemed the most promising protein.
- Diffraction resolution of crystals could not be improved by the use of additives during growth or by dehydration.
- Testing microcrystals for diffraction proved unsuccessful.

Based on the results of the initial screening, the decision was made to focus on the NΔTPO protein for further experiments. This protein appeared the most promising during early screening, crystallising the most readily of the four proteins tested. NΔTPO-2 behaved similarly to NΔTPO during screening, but offered no clear improvements in results from crystal screening or other analysis (Chapter 2, “Protein characterization”). While the deletion made to NΔTPO-2 was small and unlikely to change the overall protein fold, as it was not obviously advantageous the less modified NΔTPO protein was selected, as this was more likely to be representative of the wild-type molecule.

As diffracting crystals of NΔTPO could be obtained fairly reliably using the conditions established, attempts were made to improve the diffracting power of these crystals. One method tested was dehydration of the crystals. Dehydrating crystals can increase the packing order of the molecules inside the crystal and therefore diffraction limits³⁵⁰. A notable example of the successful application of this technique is BSA. BSA is an extensively studied protein which also produced poorly diffracting crystals³⁵⁰. Dehydration of BSA crystals improved the diffraction resolution from 8 Å to 3.2 Å allowing the structure to be solved. Although calculating the Matthews coefficient of the data obtained indicated the solvent content of the NΔTPO crystals was only slightly above average, as the diffraction resolution was low, there was a possibility lowering the solvent content further could improve the resolution of the data.

Dehydrating crystals can be problematic as the process is difficult to control and can cause fragile crystals to crack. The method which allows the most control when dehydrating crystals is using a controlled humidity device³⁵¹. This equipment has been in use at Diamond light source but was unfortunately

unavailable during this work. Instead, a selection of NΔTPO crystals were tested ~9 months after growth. The extended incubation would cause the crystals to dehydrate, reducing the solvent content of the crystal. Fourteen crystals grown in the optimised NΔTPO screen (200 mM lithium or sodium nitrate, 100 mM Tris-HCl pH 7.5-9.0, 15-17.5% PEG 3350) and were tested for diffraction but no improvement in resolution was observed.

Annealing of the crystals *in situ* was carried out by briefly diverting the flow of cold nitrogen from the mounted crystal when it was positioned in the beamline. This has been shown to improve the diffracting power of protein crystals in some cases³⁵². No significant improvement was observed when attempting this with the NΔTPO crystals.

Sucrose was tested as an additive during crystal growth as it can act as a cryoprotectant, and its presence during growth may alter the characteristics of the resulting crystals³⁵³. NΔTPO crystals grown in established conditions plus 5% sucrose showed no significant improvement in terms of size, morphology or diffraction resolution.

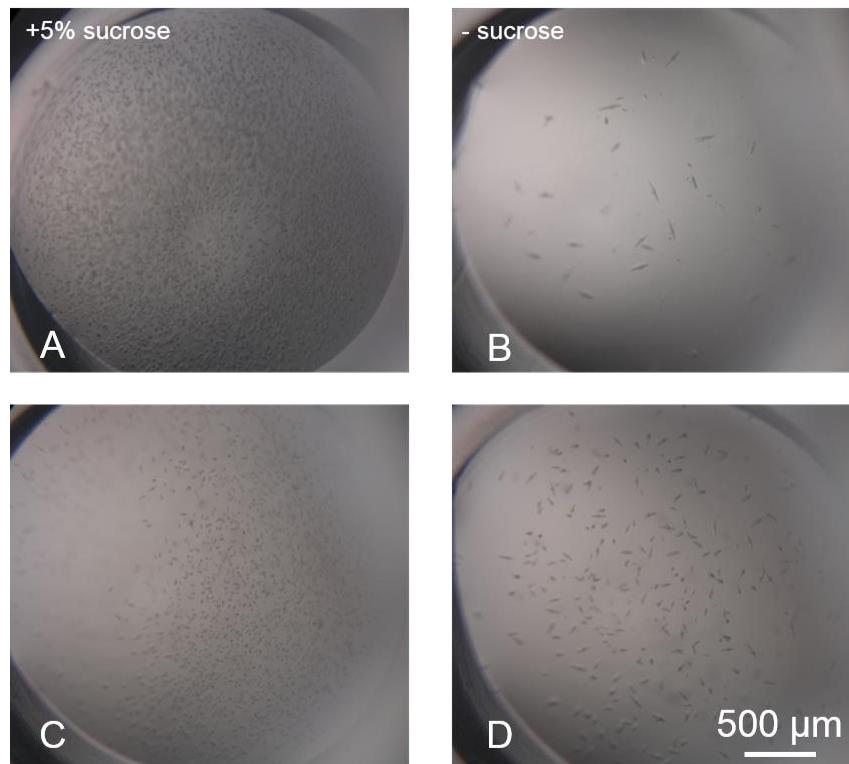


Figure 48—Micrographs of NΔTPO crystals grown in **A**) 200 mM lithium nitrate, 10 mM Tris-HCl (pH 7.8), 16% w/v PEG 3350,5% w/v sucrose, **B**) as **A** without sucrose, **C**) 200 mM lithium nitrate, 10 mM Tris-HCl (pH 8.0), 16% w/v PEG 3350,5% w/v sucrose, **D**) as **C** without 5% sucrose. The addition of sucrose was neither consistently beneficial nor detrimental to crystal growth. Scale bar is approximate.

Further NΔTPO crystals were screened on the I24 beamline, also at Diamond light source. This is a microfocus beamline which allows the focusing of the beam to $5 \times 5 \mu\text{m}$ compared to $10 \times 5 \mu\text{m}$ on the I04 beamline. The increased focus of the I24 beamline speeds up data collection and makes it possible to test microcrystals too small for traditional X-ray diffraction experiments. In some cases, it is possible to combine several partial datasets obtained from small crystals to produce a structure³⁵⁴. Both single crystals (detailed in Chapter 5) and showers of microcrystals were tested on the beamline.

The original intention was to test these crystals using *in situ* diffraction, where the crystals are exposed to X-ray radiation at room temperature whilst still in the plate³⁵⁵. This was unfortunately not possible as the plates the crystals were grown in were unsuitable for mounting in the beamline. This was due to an unforeseen incompatibility between the specific type of crystallization plate used and the mounting system for the goniometer used to position the plate in the beamline. Once this was determined, crystals were instead harvested from the

plates. While this was not ideal, it was carried out in an attempt to gain some data from the samples which could not otherwise be used.

When possible, *in situ* diffraction screening has the advantage of requiring less manipulation of the crystals. Crystal plates have been designed which produce little interference to X-ray diffraction and software tools have been developed to recognise and discard this phenomenon if it should occur³⁵⁶. This maintains the crystals in the best possible condition and avoids accidental, and often unavoidable, damage during the harvesting, cryoprotection and mounting steps³⁵⁷. Unfortunately, it was not possible to obtain further time on the I24 beamline to repeat these experiments using the correct plates.

NΔTPO microcrystals were harvested from plates using a mesh-filled loop and flash frozen in liquid N₂. However, harvesting was complicated by the formation of a sticky gel-like layer of denatured protein over the surface of the drop. This is a common occurrence in crystallography drops and may be due to the age of the plates (~3 months)³⁴¹. No diffraction data could be obtained from these samples.

Summary

The four modified TPO proteins described in Chapter 2 were tested in a number of sparse matrix screens to determine conditions for crystallization. Of the four, two proteins, NΔTPO-10 and NΔTPO-14, failed to crystallize under any of the tested conditions. NΔTPO and NΔTPO-2 crystallized under similar conditions when using a crystallography robot to set up experiments. Manually set up experiments were less successful and, as the robot allows smaller drops to be set requiring less material, a crystallography robot was used to set up subsequent screens whenever possible.

Crystals of NΔTPO and NΔTPO-2 were tested for diffraction at the Diamond light source synchrotron. The best data set obtained was produced by a NΔTPO crystal which diffracted to a resolution of 5.2 Å. While this was a slight improvement on previous work, the resolution was not sufficient to generate a reliable structure of the protein.

A number of techniques were tested to increase the diffraction resolution of crystals of this protein to generate higher resolution data. Neither annealing nor dehydrating crystals gave any improvement. Growing crystals in the presence of sucrose produced unpredictable results with no significant net benefit.

NΔTPO microcrystals were tested for diffraction on a microfocus beamline. No diffraction data could be obtained from these samples but this may be due, in part, to the less than ideal conditions when harvesting the crystals. It was not possible to investigate this further due to a lack of access to this particular facility.

Key observations

- Conditions were established for the growth of NΔTPO crystals. These conditions were sufficiently reproducible to provide ample amounts of crystals for experimentation.
- NΔTPO crystals diffracted to 5.2 Å. This resolution is an improvement on any other data published to date.
- Diffracting crystals of NΔTPO-2 were also produced.

Problems

- NΔTPO-10 and -14 failed to crystallize
- Some crystal dehydration and microcrystal diffraction experiments could not be completed due to equipment availability issues.
- The best diffraction data produced is not sufficient to produce an atomic resolution structure of TPO.

Chapter 4- Deglycosylation of Modified Thyroid Peroxidase Protein for Crystallography

Introduction

All the proteins produced in this work were expected to be N-glycosylated at 3 (N Δ TPO-10 and N Δ TPO-14) or 4 (N Δ TPO and N Δ TPO-2) sites in a similar manner to insect expressed rhTPO. N-linked glycans are covalently bound to the amide nitrogen on surface exposed asparagines in the recognition sequence asparagine-X-serine/threonine (where X is any residue except proline)³⁵⁸. One of the reasons that insect cells are widely used for producing recombinant proteins is the ability to incorporate N-linked glycans onto proteins³⁵⁹. However, the types of oligosaccharides produced by insect cells are less varied and complex than those produced by mammalian cells.

The first step of protein N-glycosylation is the transfer of an oligosaccharide precursor from a dolichol pyrophosphate carrier molecule to the protein in the endoplasmic reticulum (Fig. 49A)³⁶⁰. The four glucose residues present on this precursor are then removed by a series of enzymes to produce a high-mannose type glycan (Fig. 49B). These first steps are very similar in insect and mammalian cells, utilising equivalent enzymes. In some instances no further processing takes place and the high-mannose glycan is the end product. High-mannose glycans have been shown to appear more frequently on proteins expressed by High Five insect cells than Chinese hamster ovary cells, a mammalian cell line commonly used for recombinant protein expression³⁶¹. However in insect or mammalian cells, the high-mannose structure may be further modified in the endoplasmic reticulum and the Golgi apparatus.

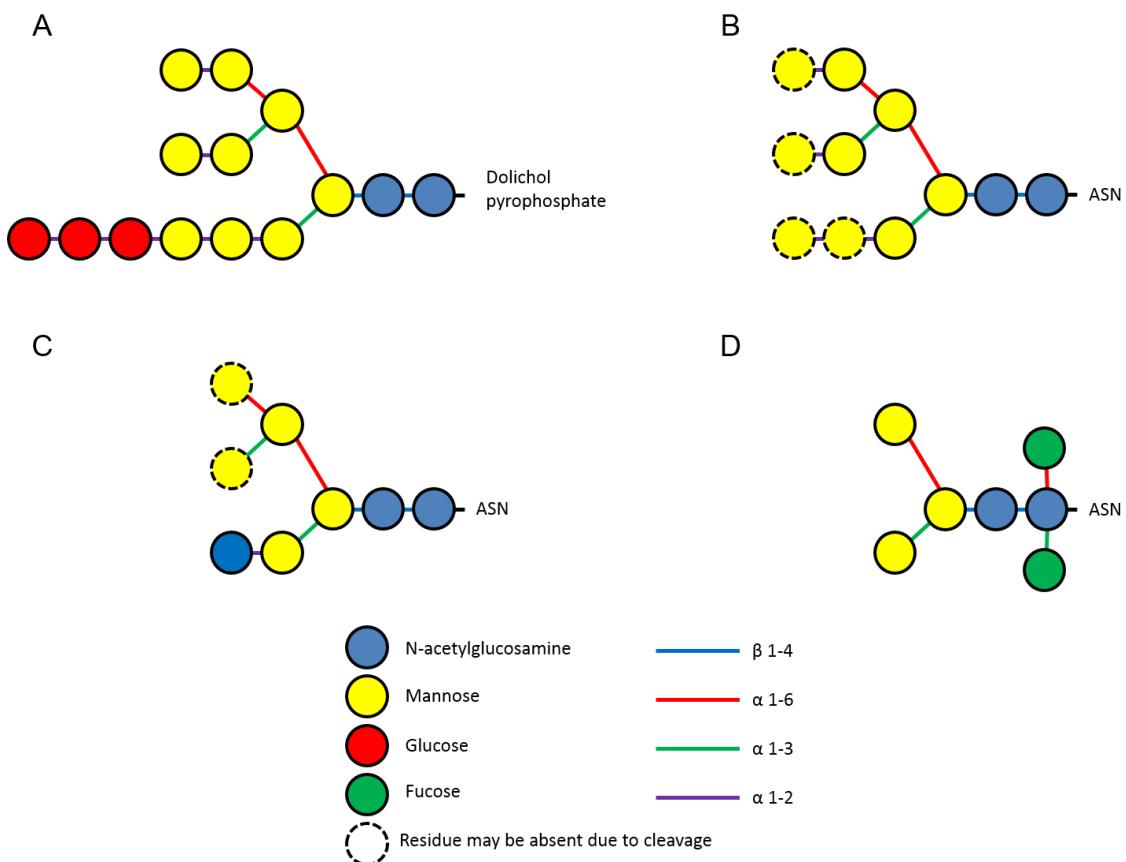


Figure 49—Diagrams of common insect N-linked glycans: **A**) Dolichol pyrophosphate linked precursor. **B**) High mannose type glycan. **C**) Hybrid type glycan, other sugars can occupy the terminal position on the 1-3 branch. **D**) Paucimannosidic type glycan, shown here with α 1-6 and 1-3 fucosylation not found in mammals. Proteins produced in this work would be expected to mostly contain glycans resembling **B**, **C** and **D** but other, more complex types are possible²⁸⁶.

High-mannose glycans can undergo further trimming to remove mannose residues from the terminal branches. Other sugars, most commonly N-acetylglucosamine, galactose, fucose or sialic acid, may be added after the removal of some of the mannose residues on the α 1-3 branch to produce a hybrid type glycan (Fig. 49C)³⁶². The branches of hybrid glycans terminate in a mannose residue at the α 1-6 branch and another sugar at the α 1-3 branch. Hybrid glycans can be further modified by fucosylation and even further trimming of the terminal branches. The addition of a α 1-6 linked fucose to asparagine linked N-acetylglucosamines occurs in both mammals and insects³⁶³. In insect cells a glycan with a α 1-6 linked fucose may then also undergo an additional α 1-3 fucosylation at the same sugar. This does not occur in mammals and the resulting glycan has been shown to be allergenic³⁶⁴.

The trimming of hybrid-type glycans in insects produces paucimannosidic glycans which may or may not be fucosylated (Fig. 49D)³⁶⁵. These consist of just the five core glycans common to all N-linked oligosaccharides, plus one or two fucoses if present. This type of glycan is not found in vertebrates³⁶⁶. Alternatively, hybrid type glycans can be elongated by the addition of a variety of sugar types to produce complex-type glycans, often terminating with a sialic acid residue³⁶⁷. This is common in mammals but appears infrequently in insects³⁶⁸. Research into the production of complex-type glycans in insect cells have shown that the ability to produce these glycans varies between species and cell lines and can be dependent on culture conditions making generalisations difficult.

The High Five (*Trichoplusia ni*) insect cell line used to express the modified TPO proteins used in this work would be expected to produce mainly high-mannose and paucimannose glycans with some hybrid types present as a minority³⁶⁹. However, there is some evidence that these cells are capable of producing complex glycans under certain circumstances. This was illustrated by High Five expressed rhTPO being recognised by sialic acid specific lectins from *Sambucus nigra* and *Maackia amurensis* as well as a lectin from *Erythrina cristagalli* which recognises terminal galactose residues²⁸⁶.

Due to the complex nature of these pathways, glycosylation of expressed proteins may vary; potential glycosylation sites on a protein may not have glycans linked in every instance³⁷⁰. N-glycosylation in insect cell culture has also been shown to vary according to culture conditions which may vary over the course of protein expression^{371,372}. This heterogeneity may be sufficient to prevent the formation of well-diffracting crystals as these large, branching structures decorating the surface of the protein may inhibit the formation of crystal contacts or adopt various different confirmations within a crystal, due to their flexible nature. For these reasons it can be desirable to remove glycans from proteins for crystallography experiments.

Results and Discussion

Protein deglycosylation method development

Key points

- Insect expressed NΔTPO can be deglycosylated with a combination of endoglycosidase H and F3.
- Results suggest the protein has paucimannosidic or complex bi/tri-antennary type glycosylation.

NΔTPO was again selected for these studies as it had so far proven the most promising of the four modified TPO proteins produced. Previous work has shown that insect expressed rhTPO can be partially deglycosylated using a combination of endoglycosidase H and F3 (RSR Ltd, unpublished). These are both endo- β -N-acetylglucosaminidases which cleave the $\beta(1\text{-}4)$ bond between the two proximal N-acetylglucosamine residues. Endoglycosidase H cleaves both high-mannose and hybrid type glycans³⁷³. It is not inhibited by $\alpha(1\text{-}6)$ fucosylation commonly produced by insect cells but does not cleave complex glycans³⁷⁴. Endoglycosidase F3 will cleave paucimannosidic and both bi- and tri-antennary complex glycans³⁷⁵. Fucosylation of the asparagine linked N-acetylglucosamine residue greatly increases the rate at which the enzyme cleaves, but it is inactive against high mannose and hybrid glycans. Together this combination of enzymes should cleave any of the glycan types expected to be present on NΔTPO.

While this strategy would not completely deglycosylate the protein, it would greatly reduce the amount of carbohydrate on the surface of the protein. Leaving a single N-acetylglucosamine residue attached at each site would still give a significant improvement in homogeneity. Peptide N-glycosidase enzymes can remove the entire glycan structure, but are generally more sensitive to protein secondary structure and therefore often require the protein to be denatured protein to work effectively³⁷⁶. This is not ideal for crystallography, as the protein then needs to be refolded following deglycosylation which can prove difficult, particularly for large, multi-domain proteins of the type studied here.

Small scale NΔTPO deglycosylation reactions were used to determine appropriate conditions for a preparative scale reaction. A range of concentrations of each endoglycosidase was tested against 50 µg NΔTPO (Table 18). Reactions were incubated in buffer (50 mM sodium acetate, pH 5.3, 15 mM sodium chloride) at 20 °C for five days. These conditions were based on the manufacturer's recommendations and previous experience with rhTPO (RSR Ltd, unpublished). When determining the amounts of enzyme to use, the units/mL value defined by a standard assay by the manufacturer was used to allow for any variability in the enzymatic activity between separate lots of enzyme.

Table 18- Trial NΔTPO deglycosylation reaction conditions

Trial reaction No.	Endoglycosidase H mUnits/mg protein	Endoglycosidase F3 mUnits/mg protein
1	86.0	57.0
2	86.0	28.5
3	86.0	14.25
4	43.0	57.0
5	43.0	28.5
6	43.0	14.25
7	21.5	57.0
8	21.5	28.5
9	21.5	14.25

Units as defined by manufacturer.

Trial reactions were analysed by lectin blotting using a biotinylated lectin from *Galanthus nivalis*. This lectin binds specifically to mannose residues, and allows visualisation using a streptavidin-horse radish peroxidase conjugate³⁷⁷. A lack of detectable signal on the blot indicates that the mannose residues have been effectively removed.

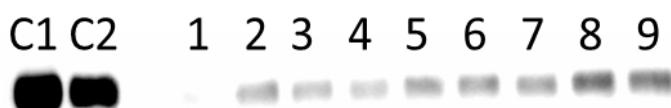


Figure 50- Lectin blot of NΔTPO trial deglycosylation samples. Control 1 (C1) untreated NΔTPO, control 2 (C2) NΔTPO incubated as trial digests without endoglycosidases, 1-9 trial deglycosylation reactions as described in Table 18. The signal becomes stronger as the amount of each enzyme decreases, indicating there are more mannose residues present on the protein.

From the lectin blot shown in Fig. 50, it is clear that both endoglycosidase H and F3 are required for complete removal of mannose residues from NΔTPO. This is demonstrated most clearly by samples 1 and 2. Sample 1 has no signal indicating the mannose has been efficiently removed where sample 2, incubated with half the amount of endoglycosidase F3, has a visible signal indicating mannose is still present. This would suggest that NΔTPO has either some paucimannosidic or complex type bi- or tri-antennary glycans present. These are likely to be similar to those reported to be found on rhTPO expressed in the same cell line, which carries complex glycans terminating in sialic acid or galactose residues²⁸⁶.

Preparative scale protein deglycosylation

Key points

- 25 mg NΔTPO was deglycosylated and purified by SEC.
- Removal of the glycans from NΔTPO appears to cause spontaneous aggregation of the protein.
- Enzymatic activity and antibody binding was not affected by deglycosylation.

Conditions for a large scale deglycosylation to prepare sufficient material for crystallography were determined from the above results (Protein deglycosylation method development). A starting weight of 25 mg protein from the NΔTPO (1) final pool (Chapter 2, “Production of recombinant TPO constructs”) was deglycosylated using the same conditions as trial reaction 1, i.e. 86 mU/mg endoglycosidase H and 57 mU/mg endoglycosidase F3. Following deglycosylation, SEC was used to remove the endoglycosidases, glycans and any aggregated or degraded TPO protein present.

The elution profile for the SEC purification of the deglycosylated NΔTPO (Fig. 51) shows a significant peak of higher molecular weight material, revealed by analytical SEC to be aggregated protein (Fig. 52). While some loss of protein was to be expected, recovery from purification was much lower than anticipated. Aggregation may have occurred during the deglycosylation reaction

as the enzymes used require a pH close to the pI of NΔTPO (5.4) to function^{375,378}.

Individual SEC fractions were analysed by lectin blot, analytical SEC, SDS-PAGE and guaiacol oxidation assay. A summary of the fractions selected for pooling is shown in Table 19. All fractions selected produced no signal when analysed by lectin blot. Concentration was determined from absorbance at 280 nm. The purity of each fraction was determined by analytical SEC.

The fractions listed in Table 19 were pooled and concentrated by centrifugal concentrator to produce a 260 µL pool of deglycosylated NΔTPO at 13.8 mg/mL (3.6 mg). Diluted samples of this pool were analysed to determine the effect of removing the glycans on the protein as described in Chapter 6. Where appropriate untreated, glycosylated NΔTPO and rhTPO were used for comparison. The deglycosylated NΔTPO appeared pure and intact by reduced SDS-PAGE. The deglycosylated protein was calculated to have a molecular weight of 81.2 kDa, 4.8 kDa less than glycosylated NΔTPO. This is a slightly larger decrease in size than previously reported for insect expressed rhTPO (2.3 ± 0.9 kDa) but, given the limited accuracy of measuring molecular mass by SDS-PAGE, is not thought to be significant²⁸⁴.

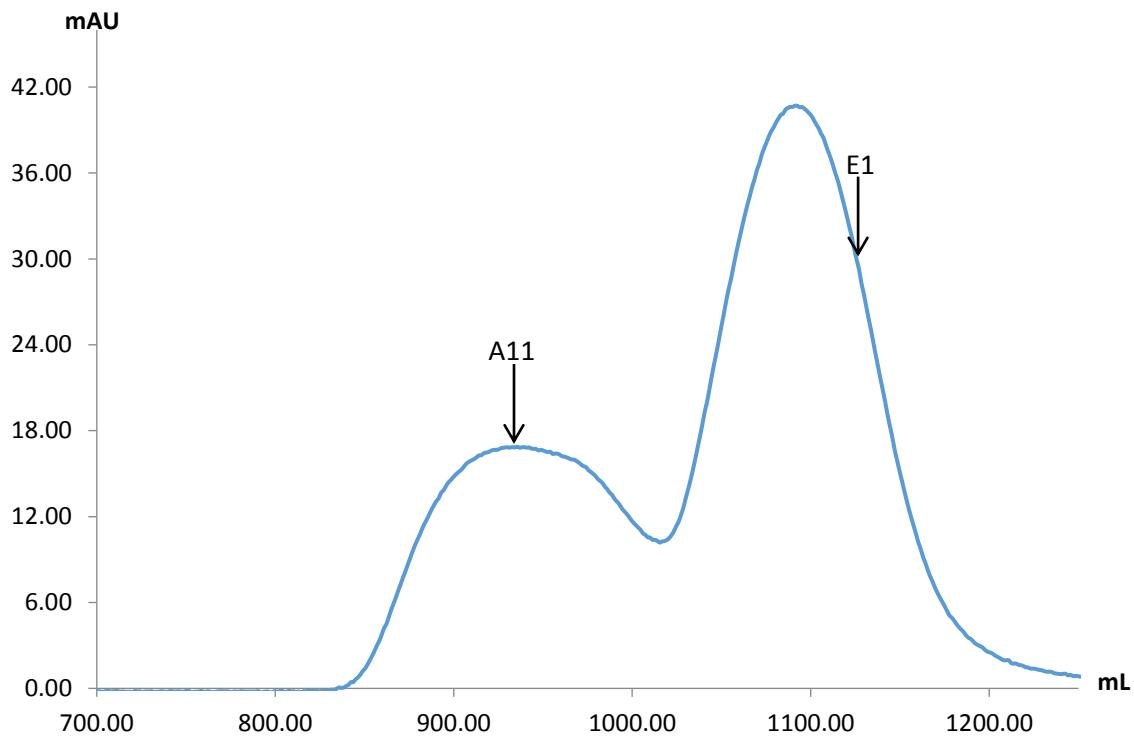


Figure 51- SEC elution profile (absorbance at 280 nm) for deglycosylated N Δ TPO. Analytical SEC profiles for the marked fractions are shown in Fig. 52. The first peak (peak approx. 925 mL) is aggregated protein whereas the second peak (peak approx. 1100mL) is monomeric protein based on elution volume.

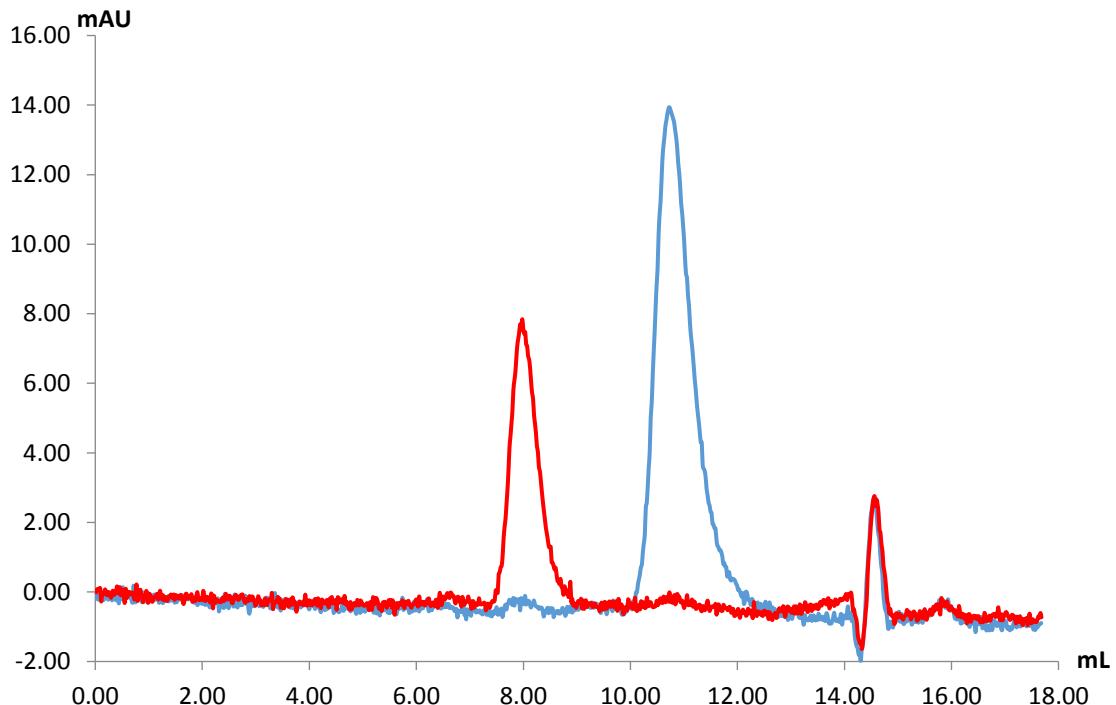


Figure 52- Analytical SEC elution profiles (absorbance at 280 nm) of deglycosylated N Δ TPO SEC fractions A11 (blue) and E1 (red). Fraction E1 elutes at the correct volume for monomeric deglycosylated N Δ TPO. Change in elution volume relative to Fig. 51 due to size differences between preparative and analytical columns.

Table 19- Fractions selected from SEC purification of deglycosylated NΔTPO for analytical SEC

Fraction	Concentration μg/mL	Weight mg	Purity by SEC %	Enzymatic activity GU/mg
D9	138.69	0.65	96.2	422
D10	134.55	0.63	97.9	606
D11	126.96	0.60	98.0	371
D12	122.82	0.58	100.0	297
E1	112.47	0.53	100.0	590
E2	102.81	0.48	100.0	461
E3	95.91	0.45	100.0	460
E4	82.11	0.39	100.0	378
E5	72.45	0.34	97.1	442
E6	62.10	0.29	97.7	119

"100%" purity by indicates there were no measurable peaks of contaminating material.

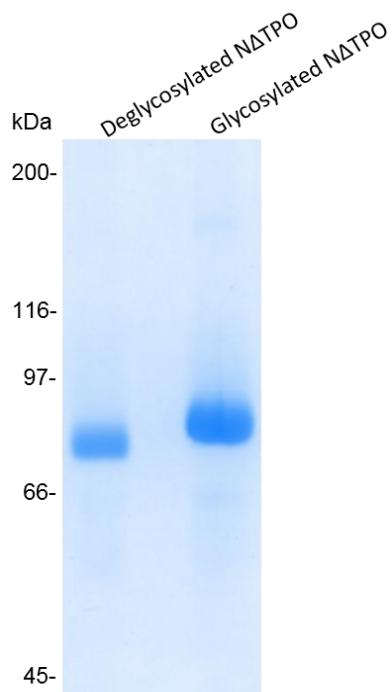


Figure 53- Stained 8% acrylamide reduced SDS-PAGE of deglycosylated NΔTPO. Glycosylated NΔTPO has an apparent mw of 86 kDa which drops to 81.2 kDa following treatment with endoglycosidase H and F3.

When analysed by analytical SEC, the pool of deglycosylated NΔTPO showed a peak of aggregated material comprising of 12.7% of the total peak area (Fig. 54). This is much higher than the level of aggregate found in the fractions, all of which had a purity of ≤96.2% (Table 19). It appears the removal of the glycans from NΔTPO made the protein prone to aggregation as the deglycosylated

protein appeared to spontaneously aggregate during any handling steps. The deglycosylated N Δ TPO was purified and treated in a similar manner to the glycosylated protein which did not have the same stability issues. Aggregation may have occurred during concentration, storage at -70 °C or during freeze-thawing, none of which could be avoided when producing this protein preparation.

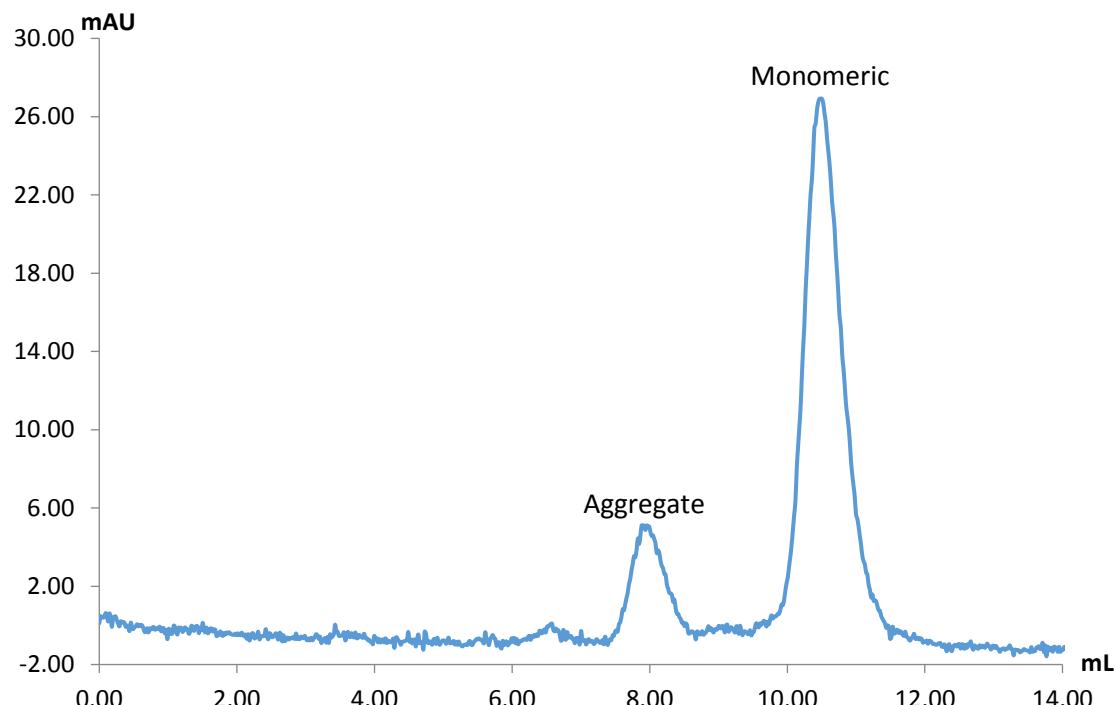


Figure 54- Analytical SEC elution profile (absorbance at 280 nm) of deglycosylated N Δ TPO pool. Protein state determined from elution volume. This pool contains a higher concentration (12.7%) of aggregated material than expected, based on the analysis of the individual fractions which contained $\leq 3.8\%$.

Enzymatic activity by guaiacol oxidation assay of deglycosylated N Δ TPO was 386 GU/mg. This is slightly lower than the 430 GU/mg value originally obtained for N Δ TPO (1) prior to deglycosylation (Chapter 2, “Characterization of modified TPO proteins”). Surface glycans are not thought to play a role in the enzymatic activity of TPO so it is more likely this drop in activity per milligram is due to the presence of inactive aggregates; indeed, the drop in activity following deglycosylation is 10.3% which is similar to the level of aggregated protein present as determined by analytical SEC (12.7%, Fig. 54). It is therefore thought that the monomeric protein present in the pool has a similar activity level to the glycosylated form.

Binding of NΔTPO to patient sera autoantibodies or 2G4 was largely unaffected by deglycosylation (Figs. 55 to 57). This result was expected as it has been previously demonstrated that both native and recombinant TPO retain their autoantibody binding characteristics following enzymatic deglycosylation^{286,379}. Insect expressed rhTPO has also been shown to bind patient autoantibodies in a similar manner to porcine and human native TPO despite differences in glycosylation³²⁴. Despite the obvious issues with aggregation, the epitopic regions of NΔTPO must still be largely intact and correctly folded to give a similar result in these assays. It is possible that the epitopes recognised by autoantibodies and 2G4 are not obstructed when the protein aggregates, allowing the IgG molecules to continue to bind.

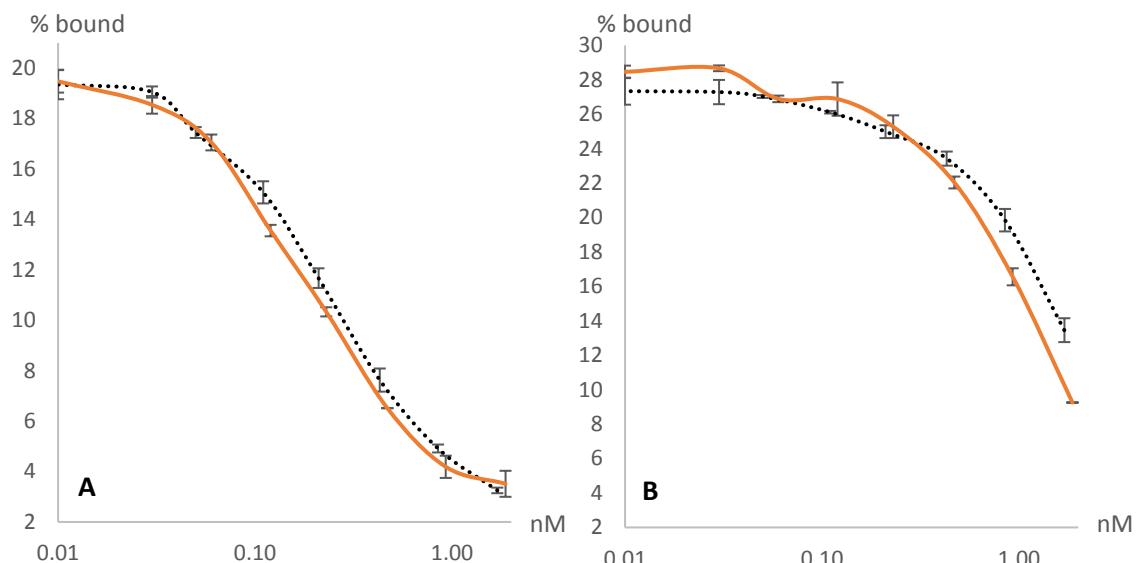


Figure 55- Pooled patient sera (**A**) and 2G4 (**B**) antibody inhibition assay for deglycosylated NΔTPO. Assay specific control (rhTPO) shown as dashed line. Error bars represent one standard deviation. This data demonstrates that removal of the glycans has not affected the antibody binding properties of the protein.

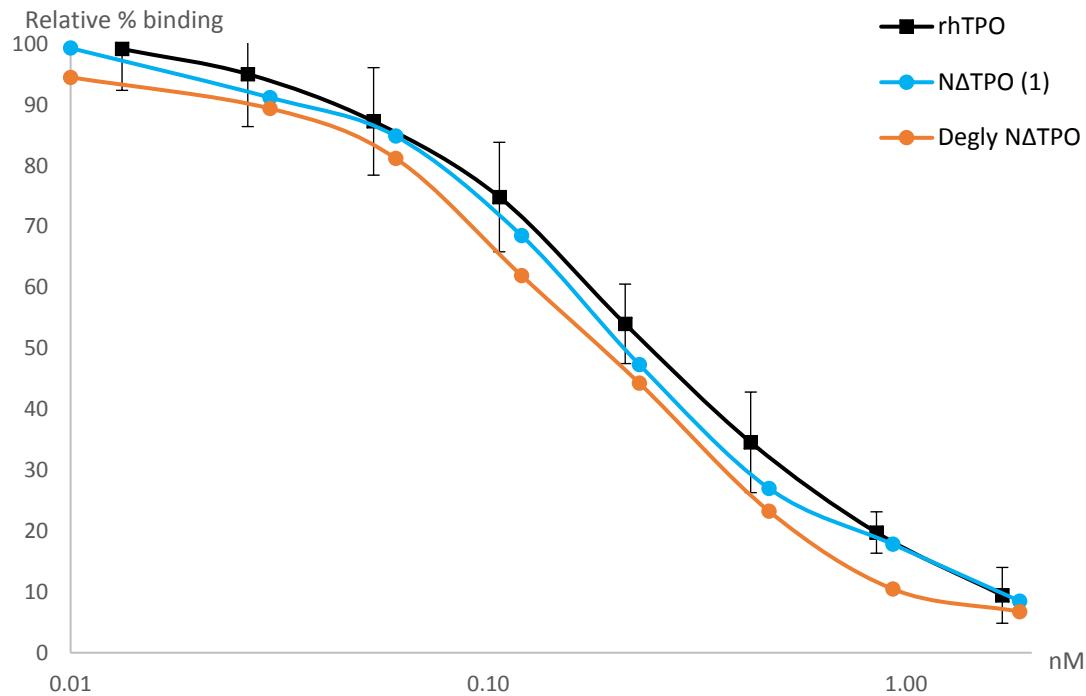


Figure 56- Pooled patient sera antibody inhibition assays for deglycosylated NΔTPO. Error bars represent one standard deviation. This shows the comparison between the glycosylated (blue) and deglycosylated protein (orange). Both proteins demonstrate inhibition levels within one standard deviation of the rhTPO control at most concentrations. Variation at 1 nM is likely due to the low signal (almost complete inhibition) at this level.

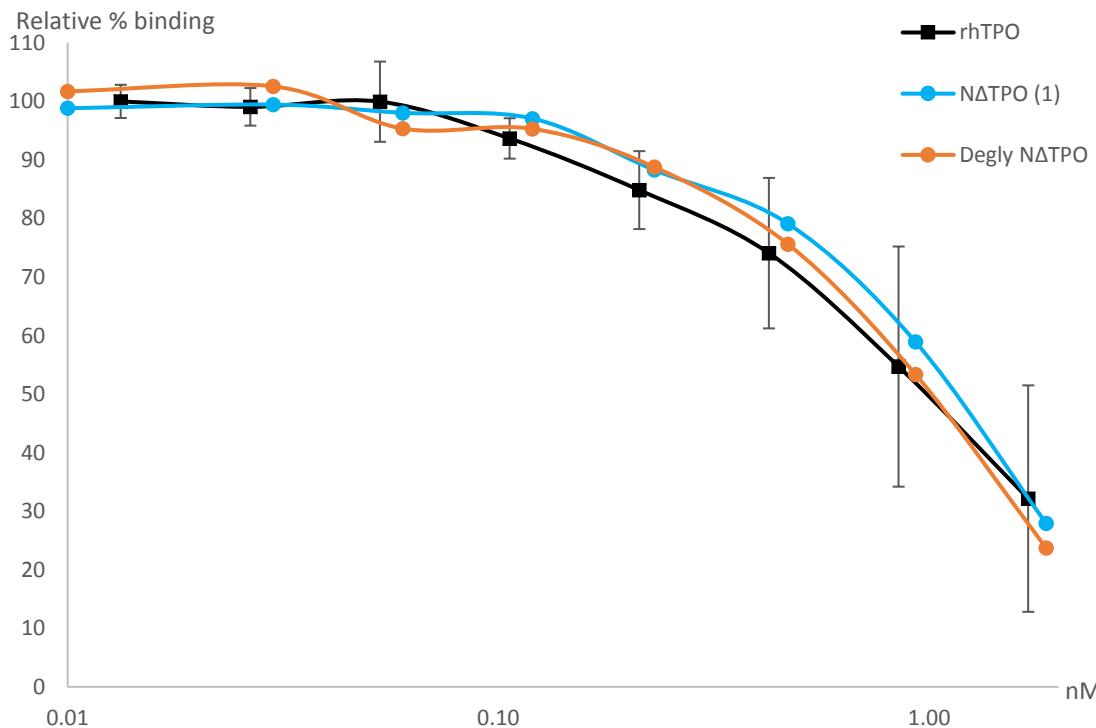


Figure 57- 2G4 antibody inhibition binding for deglycosylated NΔTPO. Error bars represent one standard deviation. This shows the comparison between the glycosylated (blue) and deglycosylated protein (orange). Both proteins demonstrate inhibition levels within one standard deviation of the rhTPO control at most concentrations.

Crystallization of deglycosylated protein

Key points

- The Final pool of deglycosylated NΔTPO was sent to Charles River Laboratories for further SEC purification and crystal screening.
- Crystals were produced which produced weak reflections to 6 Å.
- A second batch of deglycosylated NΔTPO failed to crystallize.

Despite the poor recovery and apparent instability of deglycosylated NΔTPO, efforts were made to screen the material for crystallization. Deglycosylated NΔTPO was screened by Charles River Laboratories because a crystallization robot was not available for use at the time and only very limited screening would have been possible with such a small volume of protein when setting up experiments manually due to the large minimum drop volumes. The low purity of the material (87.3% by analytical SEC) meant crystal screening would have a low chance of success. To increase the likelihood of growing crystals, a further round of SEC purification was carried out by Charles River Laboratories and screens were set up immediately with the purified material to minimise the opportunity for further aggregation (during shipping for example). Post-SEC fractions were pooled and concentrated to 8 mg/mL.

The sparse matrix screens Index and PEG/Ion (Hampton Research, UK) were set up as described previously (“Crystal screening of modified TPO proteins”). These screens are similar to those produced by Molecular Dimensions. Index covers a wide range of precipitants selected from literature reports and PEG/ion examines a range of salts in the presence of PEG 3350.

Table 20– Summary of deglycosylated NΔTPO crystal screens set up with a crystallography robot by Charles River Laboratories

Protein	Screen	Outcome
Deglycosylated NΔTPO 8 mg/mL	Index	Hit in H8:100 mM magnesium formate, 15% w/v PEG 3350 (Fig. 58A)
Deglycosylated NΔTPO 8 mg/mL	PEG/Ion	Hit in B8:200 mM magnesium formate, 20% w/v PEG 3350 (Fig. 58B)
Deglycosylated NΔTPO 8 mg/mL	100 mM magnesium formate 10.5-18% w/v PEG 3350	Hits in multiple conditions (Fig 59)
Deglycosylated NΔTPO 8 mg/mL	100 mM magnesium formate 100 mM Tris-HCl pH 7.5- 9.0 10.5-18% w/v PEG 3350	Hits in multiple conditions (Fig 59)

Crystals grew quickly in these screens, with hits obtained after 24 hours in conditions containing magnesium formate and PEG 3350 (Fig. 58). Further screens were set up with 100 mM magnesium formate and 10.5-18% w/v PEG 3350 as the precipitant. Each condition was set up both with and without 100 mM Tris-HCl pH 7.5-9.0.

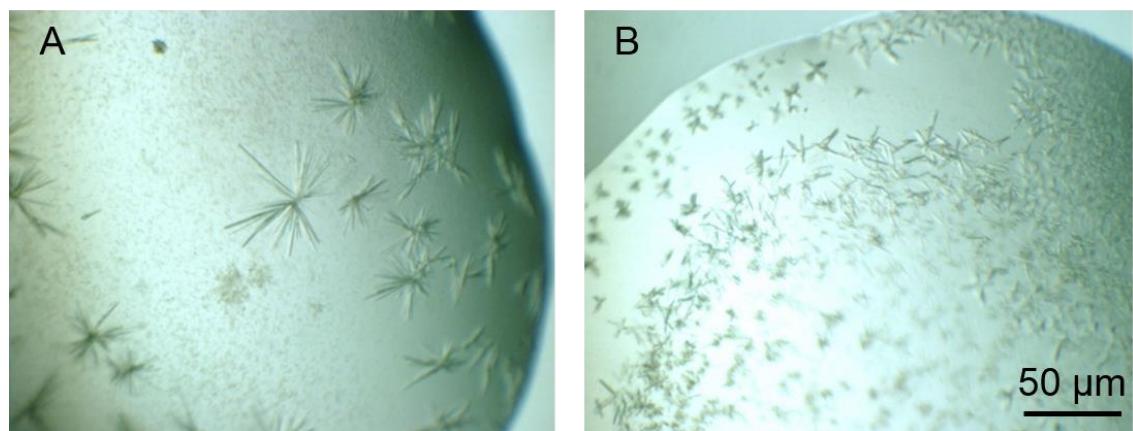


Figure 58- Micrographs of deglycosylated NΔTPO crystals grown at Charles River Laboratories in sparse matrix screens. Crystals grown in **A**) 100 mM magnesium formate, 15% w/v PEG 3350, **B**) 200 mM magnesium formate, 20% w/v PEG 3350. Scale bar is approximate. These crystals were too small to be suitable for X-ray diffraction experiments, however the growth conditions were used as a starting point for further optimisation.

Crystals from the optimisation screens were harvested and frozen in the precipitant for in-house diffraction screening by Charles River Laboratories. In some cases, 25% v/v glycerol was added to the precipitant as a cryo-protectant. Crystals grown in 100 mM magnesium formate, 100 mM Tris-HCl (pH 7.5), 12% w/v PEG 3350 produced the best diffraction, giving weak reflections up to 6 Å. The presence or absence of a cryo-protectant did not appear to have a significant effect on diffraction resolution.



Figure 59– Micrographs of deglycosylated NΔTPO crystals grown at Charles River Laboratories in optimisation screens. **A**) 100 mM magnesium formate, 15% w/v PEG 3350, **B**) 100 mM magnesium formate, 100 mM Tris-HCl (pH 7.5), 13.5% w/v PEG 3350, **C**) 100 mM magnesium formate, 100 mM Tris-HCl (pH 7.5), 12% w/v PEG 3350. Scale bar is approximate. Crystals from condition **C** produced the best diffraction up to a resolution of 6 Å.

As NΔTPO appeared to aggregate during handling and purification following deglycosylation, a further 10 mg protein was deglycosylated to determine if screening with fresher material would yield better results. Deglycosylation reactions were set up as previously, but in this instance the material was shipped to Charles River Laboratories during the incubation period. On arrival the material underwent a single step of SEC purification. Fractions eluting at the correct volume to contain monomeric, deglycosylated NΔTPO were pooled and concentrated. The same crystal screens as used for the previous batch, including the optimised conditions, were set up immediately before the protein was frozen for storage. Despite repeating the same conditions which previously produced crystals, this batch failed to crystallize.

Summary

Protein deglycosylation has been shown to be a useful technique when applied to crystallography³⁸⁰. For this reason enzymatic deglycosylation of NΔTPO was performed to make the protein more homogeneous. During the purification of deglycosylated NΔTPO it became clear that the removal of the glycans had a negative effect on protein stability, causing problems with aggregation not previously observed. Deglycosylated NΔTPO bound patient sera autoantibodies and the anti-TPO monoclonal antibody 2G4 at similar levels to glycosylated rhTPO, confirming previous observations that the glycans do not interact with autoantibodies. Some loss of enzymatic activity was observed following deglycosylation but this is likely due to the formation of inactive aggregates.

Despite the difficulties in purifying deglycosylated NΔTPO, some crystal screens were set up and crystals were successfully grown. The crystals of deglycosylated NΔTPO had a different morphology to those observed with the glycosylated protein. Some diffraction was observed from these crystals, but the reflections were of very low intensity and only to a resolution of around 6 Å, worse than the glycosylated protein. Further screening may have improved upon this however, due to the difficulty of preparing the protein and the amount of resources required, it was not feasible to explore this technique further.

Key observations

- Insect expressed NΔTPO is glycosylated with a mixture of paucimannosidic and complex type N-glycans.
- Diffracting crystals of deglycosylated NΔTPO were produced.

Problems

- Removal of the glycans using the current protocol severely destabilises the protein. This makes producing sufficient quantities for crystallography challenging.
- Diffraction from the crystals produced was weak and of low resolution (6 Å).

Chapter 5- Co-crystallization of Modified Thyroid Peroxidase Protein with Ligands

Introduction

Inhibitor screening

Following the initial crystal screening of the four modified TPO proteins, the effects of adding inhibitors to the protein were examined. Binding a ligand, such as an inhibitor, to a protein is a common method for improving the chance of success in crystallization screening³⁸¹. NΔTPO was tested with a panel of small molecule inhibitors to try to improve diffraction resolution.

As TPO is required for the production of thyroid hormones it is a drug target for the treatment of hyperthyroidism. The two main drugs in current clinical use are MMI and PTU (Chapter 1, “Treatment of autoimmune thyroid disease”)¹⁸⁷. Both drugs inhibit the enzymatic activity of TPO. Of the two, MMI is used more frequently as it requires a lower dose, can be administered less frequently and has a lower risk of producing side effects. However, PTU is recommended for the treatment of women in the early stages of pregnancy and for patients who suffer adverse side effects when treated with MMI.

Structures of LPO bound to several inhibitors have been published, including MMI and salicylhydroxamic acid^{382,383}. TPO and LPO are suspected to share high structural homogeneity around the active site, so the mode of inhibitor binding to either protein is likely to be similar. However existing data does not prove this conclusively. Solving the structure of TPO in complex with an inhibitor would allow further comparisons between the two enzymes and would be useful for determining the exact mechanism of TPO inhibition. This information could then guide the development of new inhibitors for clinical use.

Protein:F(ab) complex screening

The other major reason for scientific interest into TPO is its role as an autoantigen. A structure of TPO bound to an autoantibody would be a useful

tool in further characterising the role of this autoantigen in AITD. Obtaining such a structure was one of the original aims of this work.

Autoantibodies to TPO are predominantly of the IgG type¹¹⁰. IgG molecules can be divided into two main regions (Chapter 1, “The immune system”). These are the antibody binding F(ab) regions (two identical fragments per IgG) and the F(c) region which mediates interactions with the rest of the immune system²². Intact IgG molecules are difficult to crystallize as they possess a flexible hinge region, allowing the three main domains of the protein to move independently. However, there are numerous structures of F(ab) and F(c) fragments alone, or in complex with other molecules deposited in the PDB. F(ab)s are useful ligands for crystallography as, when in complex with other proteins, they can have a stabilising effect and encourage crystallization of otherwise resistant proteins³⁸⁴.

Papain, a cysteine protease, can be used to cleave IgG in the hinge region (Fig. 60) separating both F(ab)regions from the F(c)³⁸⁵. Alternatively, pepsin, an aspartic acid protease, can be used to cleave the IgG below the disulphide bonds to produce F(ab')₂³⁸⁶. F(ab')₂ retains the two flexible hinge regions from the intact IgG. The presence of two joined F(ab) regions allows for the cross-linking of protein antigens which may inhibit the formation of crystals. F(ab')₂ can be reduced to produce two F(ab') but part of the hinge region is retained compared to F(ab). While F(ab')₂ can be useful, generally F(ab) is preferred as a ligand for crystallography.

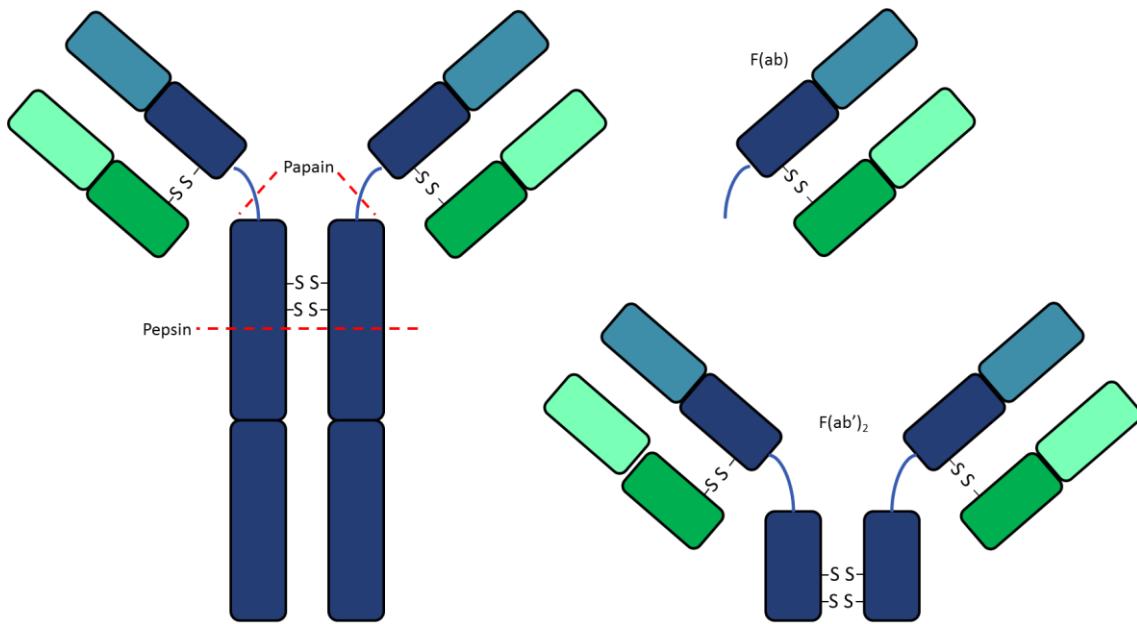


Figure 60— Diagram of IgG, F(ab) and F(ab')₂ molecules. Heavy chain-blue, light chain-green. Approximate sites of cleavage sites to produce F(ab) and F(ab')₂ indicated with red lines. F(ab) produced from 4F5 IgG was used as a ligand for crystal screening.

The mouse monoclonal IgG 4F5 was chosen to produce a complex with N Δ TPO for crystallography experiments. 4F5 is a mouse monoclonal type 1 IgG specific for TPO produced by RSR Ltd. It was produced by immunizing mice with native human TPO³⁸⁷. The crystal structure of the F(ab) region of 4F5 has been previously solved to 1.9 Å. This was one reason for selecting this antibody over 2G4, a human monoclonal IgG also used in this work, as 2G4 F(ab) has previously proven difficult to crystallize²⁵⁸. 4F5 has been shown to inhibit patient sera autoantibodies binding to TPO. This indicates a shared epitope, likely located in one of the two immunodominant regions which the majority of patient autoantibodies recognise (Chapter 1, “Thyroid peroxidase protein structure”). It does not inhibit the binding of 2G4, which also inhibits patient autoantibodies binding to TPO^{109,387}. This would suggest that these antibodies bind to separate immunodominant regions on TPO. The affinity constants of each antibody have been determined previously with 4F5 having a lower dissociation constant (K_d) than 2G4 (38 vs 400 pM)^{109,387}. 4F5 would therefore be less readily dissociated from TPO during complex production and crystallization.

Attempts have been made previously to crystallize 4F5 F(ab) in complex with insect expressed rhTPO, but no crystals were obtained (RSR Ltd, unpublished).

As NΔTPO has produced more promising diffraction data than rhTPO it was possible that the additional stabilisation of the protein with a F(ab) fragment may increase the diffraction resolution of the crystals produced to permit a 3D structure of the complex to be produced.

Results and Discussion

Initial inhibitor screening

Key points

- Methimazole co-crystallization was tested with NΔTPO.
- The presence of the inhibitor appeared to have a positive effect during crystallization.

Co-crystallization with MMI (Fig. 62) was attempted first as MMI was readily available and in current medical use. Conditions were based on those which produced NΔTPO crystals previously (200 mM lithium or sodium nitrate, 100 mM Tris-HCl pH 7.5-9, 10-20% PEG 3350). The drug was added to the protein solution prior to drop setting in a 10-fold molar excess relative to the protein. This was to ensure full occupancy of the inhibitor binding site. Otherwise the plates were set up using a crystallography robot as previously.

Adding the drug to the protein prior to setting up the experiments produced a few more hits and, in some cases, larger crystals than the same conditions without the inhibitor (Fig. 61). In the screen described above, 39 wells (of 96) contained some form of crystalline precipitate but this increased to 45 wells with MMI present. The PACT premier and JCSG+ sparse matrix screens were also repeated with MMI present to search for new crystallization conditions to follow up on. Some crystals were obtained in these screens but only in conditions similar to those already established.

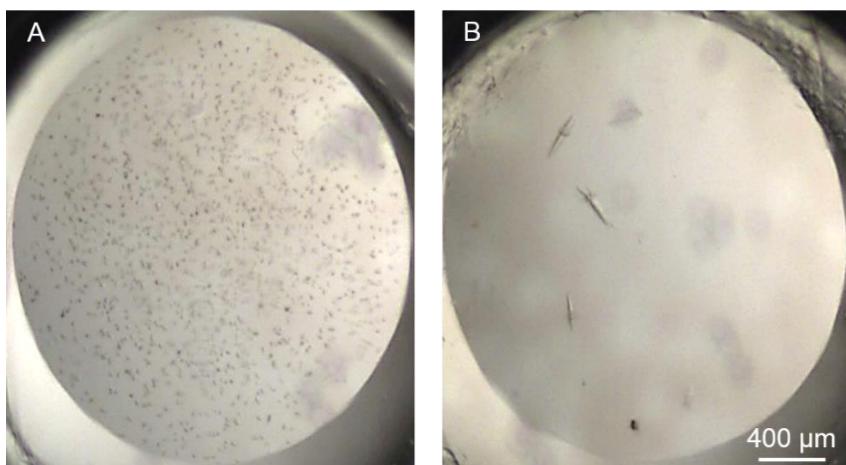


Figure 61– Micrographs of NΔTPO crystals grown in **A**) 200 mM sodium nitrate, 100 mM Tris-HCl (pH 8.53), 12% w/v PEG 3350, **B**) as **A** with MMI added. Scale bar is approximate. In many cases, such as demonstrated here, outcomes were improved when MMI was added to the protein prior to screen set up.

Selection of further inhibitors

Key points

- Further inhibitors were selected from the literature following the promising results obtained with methimazole.
- Six other compounds with TPO inhibiting properties were identified for testing.

Following the results produced with MMI, PTU and a further five TPO inhibitors from the literature (Fig. 62) were tested using a similar methodology. Thyroid inhibition by chemicals in the environment is a growing concern and a large-scale screening effort to identify some of these chemicals has been carried out by the U.S environmental protection agency³⁸⁸. This study used a combination of three high-throughput assays to test a library of over 2000 chemicals for TPO inhibition activity against porcine or rat thyroid microsomes. A fluorescence based inhibition assay was used to determine TPO activity in the presence of each chemical. Non-specific enzyme inhibition activity was assessed by a luciferase inhibition assay and cytotoxicity to a human cell line was measured using a viability assay. The results of these assays allowed chemicals which specifically inhibit TPO enzymatic activity to be identified.

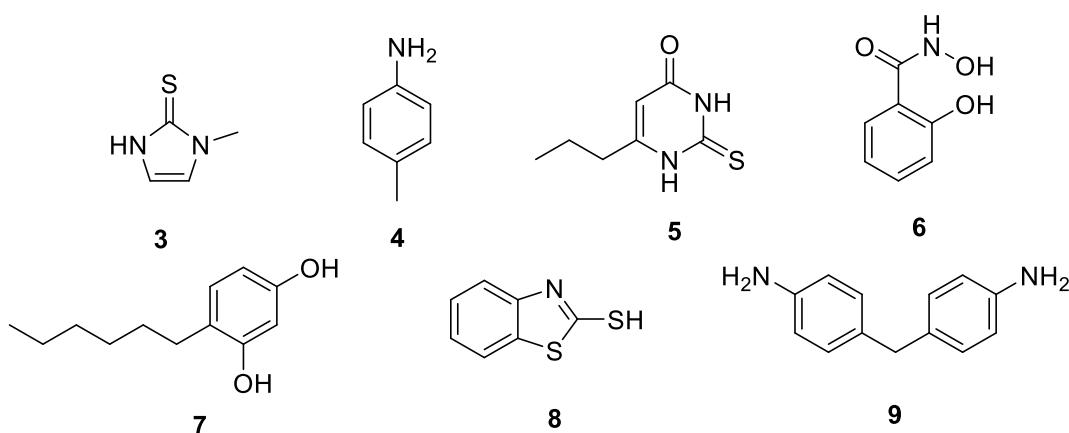


Figure 62– Inhibitors used for crystal screening: Methimazole (MMI) (**3**), *p*-toluidine (**4**), 6-propyl-2-thiouracil (PTU) (**5**), salicyhydroxamic acid (**6**), 4-hexylresorcinol (**7**), 2-methylbenzothiazole (**8**), 4,4'-methylenedianiline (**9**). These ligands were selected from the work of Paul Friedman *et al* based on their potency and selectivity as TPO inhibitors and used in crystal screens with NΔTPO³⁸⁸.

The inhibitors shown in Fig. 62 (with the exception of MMI and PTU) were selected based on the findings of Paul Friedman *et al*³⁸⁸. As well as the positioning of these chemicals near the top of the potency and selectivity ranks, availability and chemical structure was also considered. For example, resorcinol was the second most potent inhibitor of TPO identified in the study but it was not used as its chemical structure is very similar to 4-hexylresorcinol, which was the most potent inhibitor identified in the study.

Further inhibitor screening

Key points

- Inhibitors were screened for co-crystallization with NΔTPO using previously identified conditions for crystal growth.
- Crystals were produced in screens containing methimazole and *p*-toluidine.
- NΔTPO crystals grown in the presence of methimazole diffracted to 7 Å.
- NΔTPO crystals grown in the presence of *p*-toluidine diffracted to 10 Å.

Each of the six further inhibitors was tested in a screen based around previously successful conditions (200 mM lithium or sodium nitrate, 100 mM Tris-HCl pH 7.8-9.0, 15-17% w/v PEG 3350). Inhibitors were again added to the crystallization drop in a 10-fold molar excess of the protein to ensure full occupancy. Of the six other inhibitors tested, only the screens containing *p*-toluidine produced crystals (Fig. 63). The others produced mostly clear drops with some occasional light, amorphous precipitate. It is thought that in these cases, the presence of the inhibitor prevented crystallization as these conditions had been shown to reproducibly produce crystals of the protein only.

Fifteen NΔTPO+MMI crystals were screened on the I04 beamline. These were tested for diffraction initially by taking three test images in the same manner as the NΔTPO only crystals (“Crystal screening of modified TPO proteins”). No data sets were collected for these crystals as none of the test images showed diffraction beyond 7 Å. Briefly annealing the mounted crystals did not improve the diffraction limits.

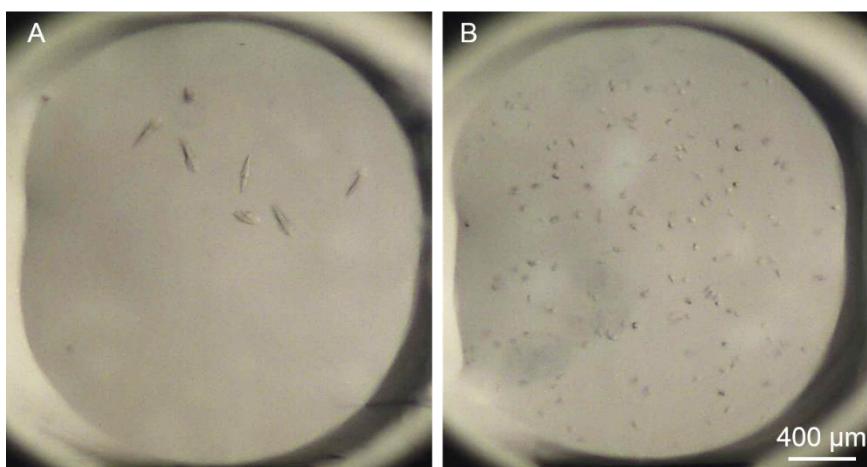


Figure 63– Micrographs of NΔTPO crystals grown in **A**) 200 mM sodium nitrate, 100 mM Tris-HCl pH 7.8, 17% w/v PEG 3350, 26.4 mM *p*-toluidine, **B**) 200 mM lithium nitrate, 100 mM Tris-HCl pH 7.8, 17% w/v PEG 3350, 26.4 mM *p*-toluidine. Scale bar is approximate. These crystals, particularly in **A** display a similar morphology to those produced with the protein only.

Four NΔTPO crystals grown in the presence of *p*-toluidine were screened on the I24 beamline. These crystals were screened at the same time as the unliganded NΔTPO microcrystals and were also originally intended for *in situ* diffraction screening (“Crystal screening of modified TPO proteins”). Due to the larger size of these crystals, harvesting them from the plate was less

problematic than the microcrystals. As with the NΔTPO crystals screened on the I04 beamline, three test images were taken, 90° apart, to estimate the diffracting quality of the crystal. The diffraction resolution of each crystal was very low, around 10 Å and again was not improved by annealing the crystal. The addition of *p*-toluidine reduced the number of hits produced during screening, and in wells where crystallisation did occur it was difficult to produce single crystals large enough for further diffraction studies.

4F5 F(ab) production

Key points

- Conditions were established for producing F(ab) from 4F5 IgG by papain digestion.
- F(ab) was purified by protein A affinity chromatography and SEC.

4F5 IgG was provided by RSR Ltd. F(ab) was prepared from 4F5 IgG by papain digestion. Initially, small scale digests were set up to determine suitable digestion conditions. Samples of 4F5 IgG were incubated at 37 °C with various papain concentrations (calculated as a weight to weight ratio) for 3 h. Samples were taken every hour for analysis by SDS-PAGE.

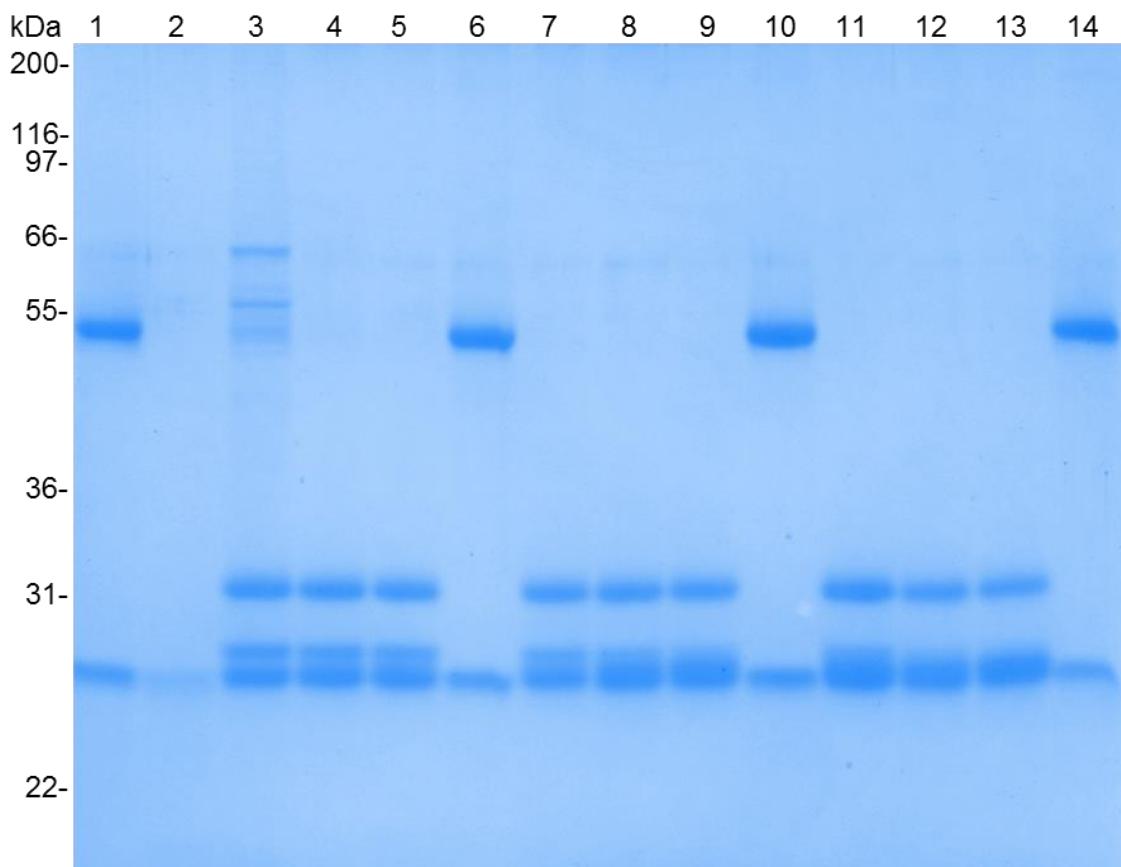


Figure 64- Stained 12% acrylamide reduced SDS-PAGE of trial papain digestions. Lane 1) undigested IgG, 2) papain, 3) 100:1 IgG:papain after 1 h, 4) 50:1 after 1 h, 5) 34:1 after 1 h, 6) negative control (no papain) after 1 h, 7) 100:1 after 2 h, 8) 50:1 after 2 h, 9) 34:1 after 2 h, 10) negative control after 2 h, 11) 100:1 after 3 h, 12) 50:1 after 3 h, 13) 34:1 after 3 h, 14) negative control after 3 h. The presence of high molecular weight bands (>50 kDa) indicate incomplete digestion (see lane 3 for example).

These results show that the IgG heavy chain is efficiently digested by the papain given at least a 50:1 IgG:papain ratio or an incubation time greater than 1 hour. As there was no evidence of protein degradation during incubation, a 3 hour incubation was selected for carrying out larger scale digests. A 34:1 IgG:papain ratio was used to ensure complete digestion. The amount of papain used could have been reduced but even at this ratio, only a small amount of the relatively inexpensive enzyme was required.

When carrying out a larger scale digestion to produce F(ab) for crystallography, a Protein A chromatography was performed following digestion to separate the F(ab) fragments from the F(c). As Protein A binds the F(c) region of IgG, undigested IgG is removed along with F(c)³⁸⁹.

Using the conditions deduced from the pilot studies described above, two separate papain digests were carried out on 50 mg samples of 4F5 IgG to produce F(ab) for crystallisation (designated 4F5 F(ab) (1) and (2)). 4F5 F(ab) preparations were analysed by SDS-PAGE, direct binding and inhibition RIAs and analytical SEC. The purpose of this was to demonstrate that the F(ab) was intact and did not contain any residual F(c).

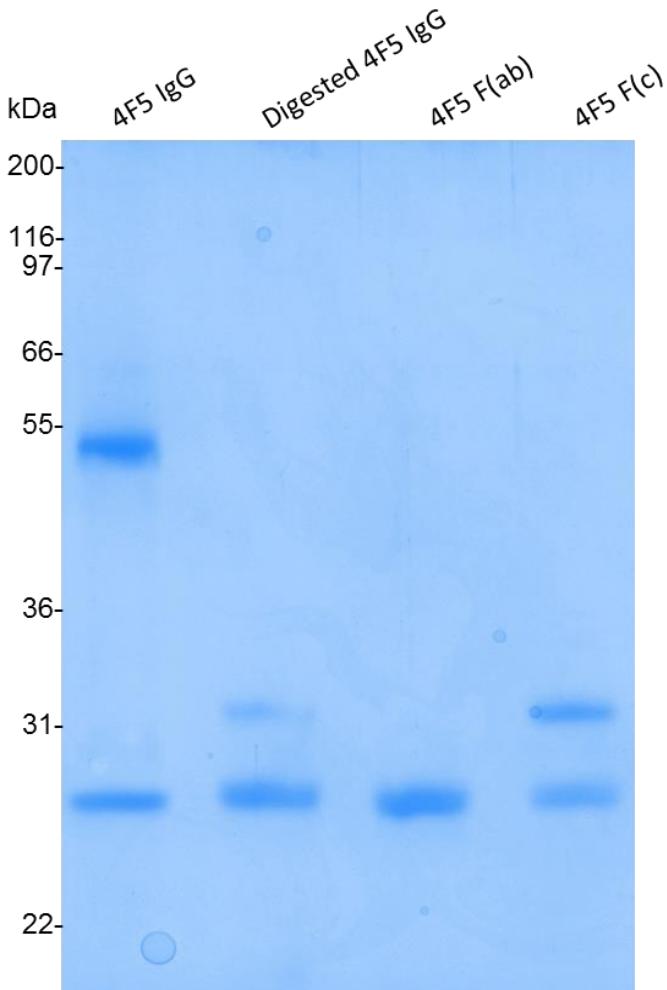


Figure 65- Stained 12% acrylamide reduced SDS-PAGE of 4F5 F(ab) (1) production samples. F(ab) and F(c) samples are post Protein A chromatography. This demonstrates the removal of F(c) from the F(ab) preparation.

The reduced SDS-PAGE gel shown in Fig. 65 shows a clear, single band at 25 kDa, the correct size for reduced 4F5 F(ab) with the second band present in the digested IgG and F(c) samples removed.

In the direct binding RIA, 4F5 F(ab) and IgG were allowed to bind ^{125}I labelled TPO, the complex precipitated with Protein A and the counts per minute (CPM)

measured. Neither 4F5 F(ab) preparation gave a signal (CPM similar or less than negative control) at concentrations up to 1 mg/mL. In contrast, 4F5 IgG which produced a positive signal at concentrations of 100 ng/mL and above (Fig. 66A). In the inhibition assay, a range of 4F5 F(ab) concentrations were allowed to bind ¹²⁵I labelled TPO before a fixed concentration of 4F5 IgG was introduced. The complex was then precipitated with Protein A and the CPM measured as for the direct binding assay. The results (Fig. 66B) demonstrate that 4F5 F(ab) inhibited the binding of 4F5 IgG to ¹²⁵I-TPO in a dose dependent manner. Taken together, these results indicate that both 4F5 F(ab) preparations bind TPO but not Protein A, indicating the F(c) region has been successfully removed.

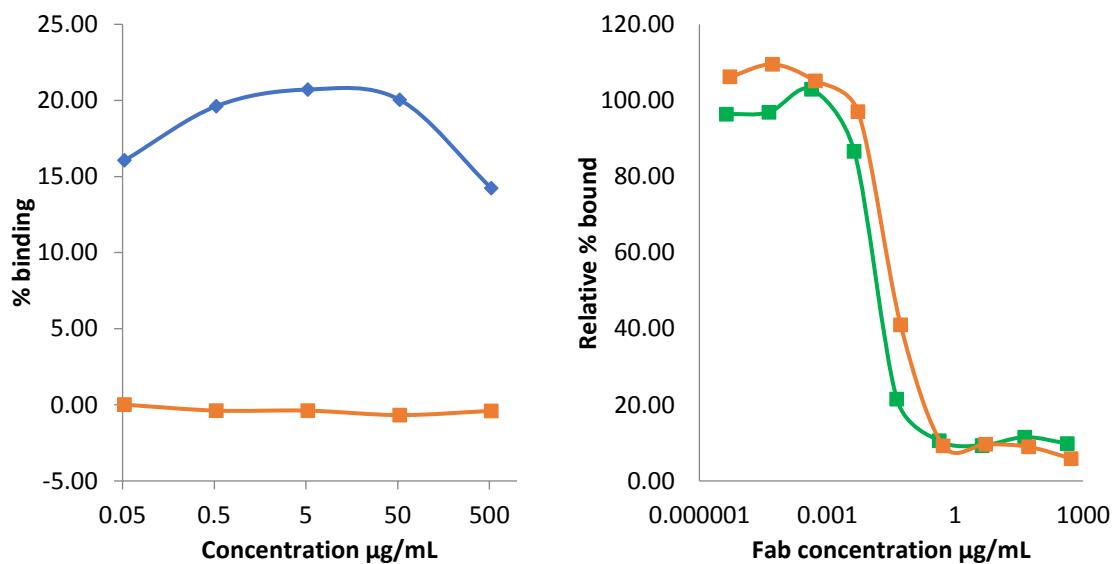


Figure 66- A) Direct binding assay results for 4F5 F(ab) (2) (orange) and 4F5 IgG (blue). Results for 4F5 F(ab) (1) not shown but largely similar. This shows that IgG binds protein A, F(ab) does not due to the removal of the f(c) region. **B)** Inhibition assay results for 4F5 F(ab)(1) (green) and 4F5 F(ab) (2) (orange). Values for relative % bound are a percentage of the negative (no F(ab) present) control. This demonstrates that both 4F5 F(ab) and IgG inhibit antibody binding to radiolabelled TPO at similar levels.

When analysing 4F5 F(ab) (1) by analytical SEC it was noted that the preparation contained a second peak eluting before the main F(ab) peak (Fig. 67). This unidentified impurity was of similar MW to the complex, and would therefore be impossible to separate by SEC if it were not removed first. To avoid this impurity being included in any preparations of the complex, 4F5 F(ab) (1) was purified by SEC before using it for the production of NΔTPO:4F5 F(ab)

complex. 4F5 F(ab) (2) contained only a single peak when analysed by analytical SEC. A SEC purification step was therefore not necessary for this material.

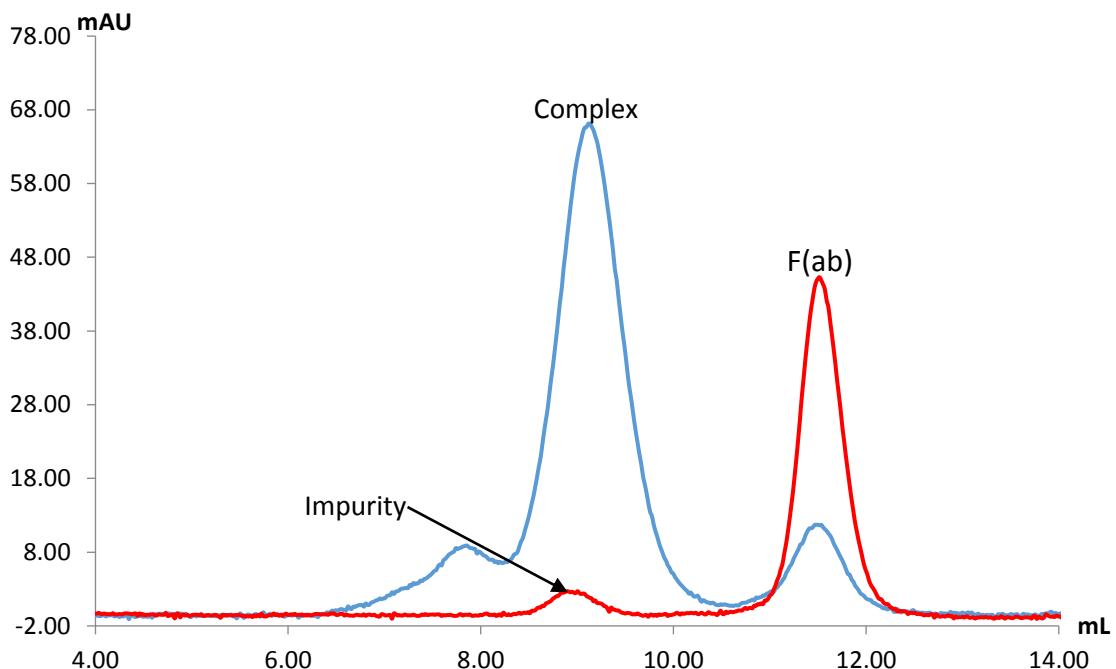


Figure 67- Overlaid analytical SEC elution profiles (absorbance at 280 nm) of unpurified NΔTPO:4F5 F(ab) complex (blue) and 4F5 F(ab) (1) (red). Due to its MW, the impurity in 4F5 F(ab) (1) would elute at the same volume as the complex.

Post-SEC fractions were analysed by SDS-PAGE and analytical SEC. Fractions appearing pure by both analysis methods were pooled and concentrated to produce 4F5 F(ab) (1). Analysis of this preparation was repeated and similar results obtained with the exception of the analytical SEC elution profile now showing only a single peak at the correct elution volume for 4F5 F(ab).

NΔTPO:4F5 F(ab) complex production

Key points

- Pilot studies were carried out to determine the most suitable protein:F(ab) ratio for complex production.
- Two pools of NΔTPO:4F5 F(ab) complex were produced, purified and analysed.

Δ TPO:4F5 F(ab) complex was produced by mixing post-IMAC Δ TPO and 4F5 F(ab) and allowing them to bind. Theoretically, a 1:1 molar ratio would give complete binding of F(ab) to TPO however in practise, it is desirable to have a slight excess of F(ab) to ensure this. As 4F5 F(ab) has a much lower MW than Δ TPO (50 vs 86 kDa) unbound F(ab) is easier to separate from the complex (136 kDa) by SEC than unbound Δ TPO. To determine a suitable ratio of F(ab) to TPO to use, a series of ratios between 1:1 and 1:1.75 (TPO:F(ab) molar ratio) were tested on a small scale. After allowing the F(ab) to bind, each condition was analysed by analytical SEC. Based on the results of this experiment (Fig. 69) it was determined that a 1:1.25 Δ TPO:4F5 F(ab) molar ratio produced a slight excess of F(ab) and ensured complete occupancy.

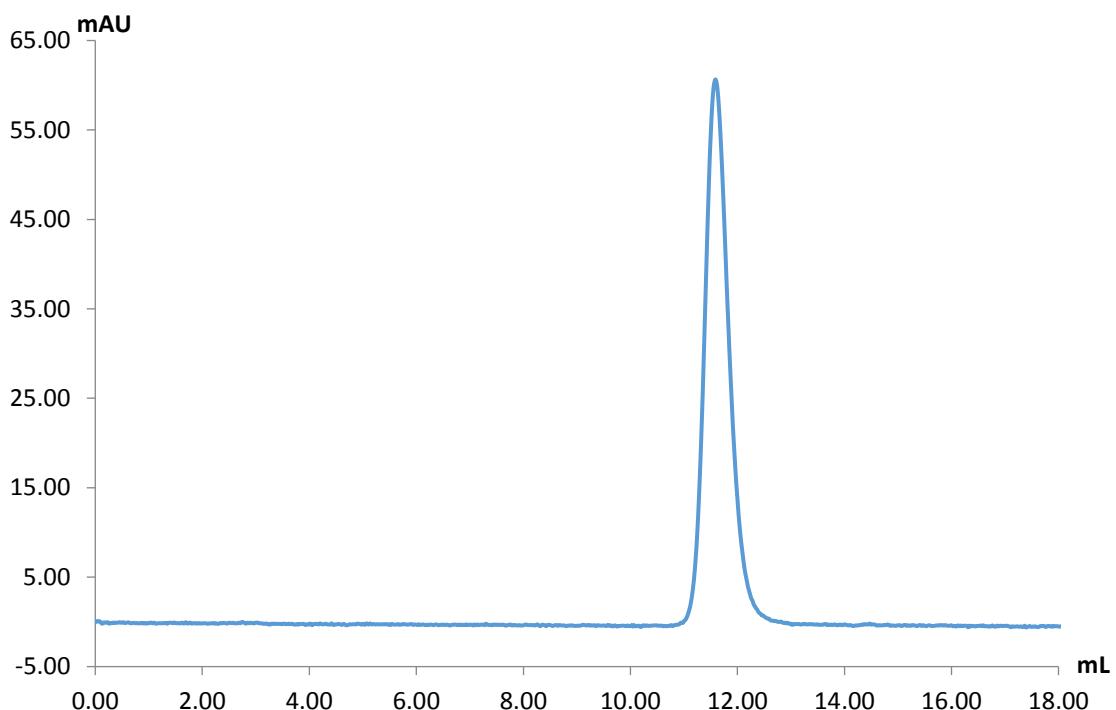


Figure 68- Analytical SEC elution profile (absorbance at 280 nm) of 4F5 F(ab) post preparative SEC. All trace of the high molecular weight impurity was removed by SEC.

To produce the Δ TPO:4F5 F(ab) complex, 13 mg post-IMAC Δ TPO (4) (Chapter 2, “Production of modified TPO proteins”) was mixed with 10 mg 4F5 F(ab) to give a TPO:F(ab) molar ratio of 1:1.25. The mixture was incubated to allow the complex to form before SEC purification. Fractions were selected immediately based on the absorbance at 280 nm profile.

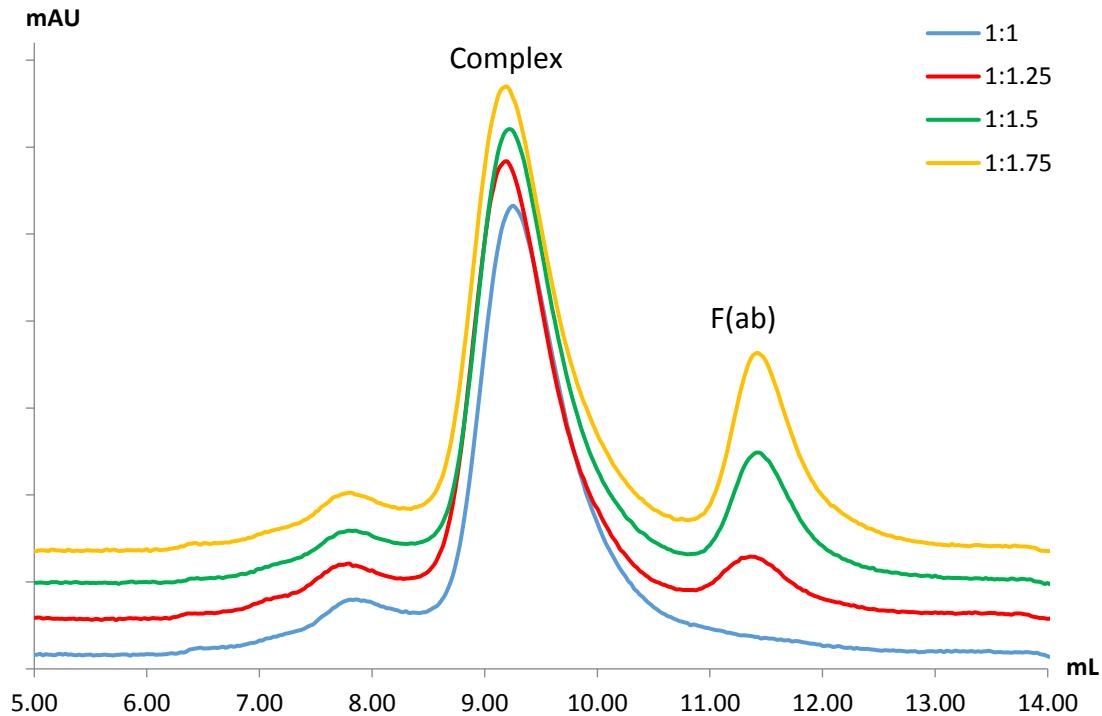


Figure 69- Analytical SEC elution profiles (absorbance at 280 nm) of NΔTPO:4F5 F(ab) (2) complex ratio trials (profiles offset for clarity). Ratios in legend are NΔTPO to 4F5 F(ab) molar ratios. As the amount of F(ab) present increases, the intensity of the unbound F(ab) peak relative to the complex peak increases.

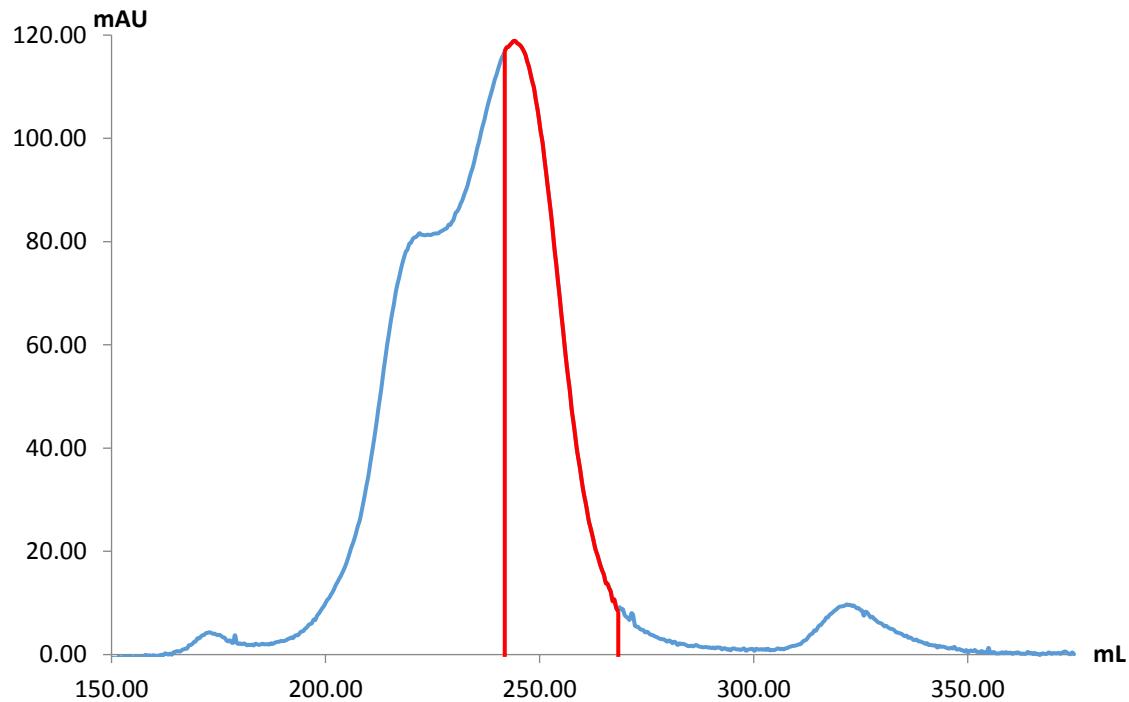


Figure 70– Preparative SEC elution profile (absorbance at 280 nm) for NΔTPO:4F5 F(ab) (1). Fractions selected for pooling highlighted (red). Fractions were selected from the trailing edge of the peak to avoid contamination with aggregated material which may have not been completely separated.

Table 21- NΔTPO:4F5 F(ab) complex final pool weights and volumes

NΔTPO:4F5 F(ab) complex (run)	Number of fractions pooled	Volume (mL)	Post-concentration volume (μ L)	NΔTPO27:4F5 F(ab) complex concentration (mg/mL)	Weight of NΔTPO27:4F5 F(ab) complex (mg)
NΔTPO:4F5 F(ab) (1)	15	26	482	10	4.8
NΔTPO:4F5 F(ab) (2)	17	31	548	10	5.5

The NΔTPO:4F5 F(ab) complex preparations were analysed by reduced SDS-PAGE and analytical SEC. SDS-PAGE analysis of both complex preparations showed bands corresponding to NΔTPO and 4F5 F(ab) were present in each preparation with no contaminating bands present (Fig. 71). NΔTPO and 4F5 F(ab) appeared as separate bands due to the denaturing effects of SDS.

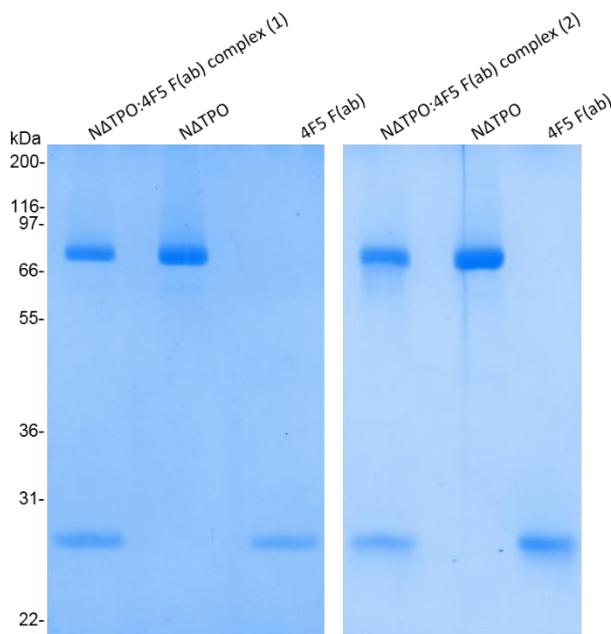


Figure 71- Stained 12% acrylamide reduced SDS-PAGE gels of NΔTPO:4F5 F(ab) (1) and (2). NΔTPO and 4F5 F(ab) appear as separate bands due to the denaturing conditions used for SDS-PAGE.

Analytical SEC showed both preparations of the NΔTPO:4F5 F(ab) complex to be highly pure and therefore suitable for crystallography experiments (Fig. 72). NΔTPO:4F5 F(ab) (1) was slightly purer (98.4% by analytical SEC) than NΔTPO:4F5 F(ab) (2) (96.1%), which contained more higher molecular weight

material, possibly aggregated protein, but further purification was not deemed necessary. No unbound F(ab) was detected in either preparation.

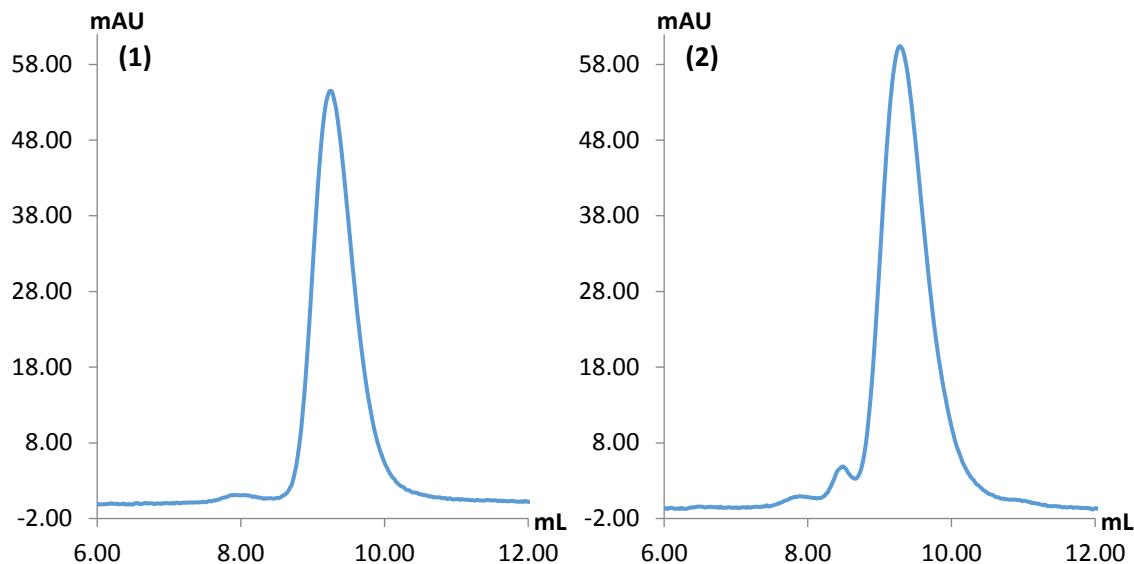


Figure 72- Analytical SEC elution profiles (absorbance at 280 nm) of NΔTPO:4F5 F(ab) complex final pools. Pool displayed slightly higher purity (98.4% Vs 96.1%) but both were acceptable for crystal screening.

NΔTPO:4F5 F(ab) complex crystallography

Key points

- Sparse matrix screening of NΔTPO:4F5 F(ab) complex produced a single hit which could not be replicated.
- A single crystal of NΔTPO:4F5 F(ab) complex was tested for diffraction but did not produce any measurable reflections.

Both NΔTPO:4F5 F(ab) complex preparations were screened for crystallization in a similar manner to the initial sparse matrix screening of the modified TPO proteins. Structure screens 1 and 2, JCSG+ and PACT premier were set up by hand and with a crystallography robot at Cardiff University. No hits for either NΔTPO:4F5 F(ab) preparation were observed in any of the sparse matrix or optimisation screens set up with the crystallography robot. Other than a single condition, all other experiments were unsuccessful, generating only amorphous precipitate or clear drops. Crystals of NΔTPO:4F5 F(ab) (1) formed in a single

well of the manually set up JCSG+ screen containing 100 mM bicine (pH 9.0), 10% w/v PEG 6000 as the precipitating agent (Fig. 73).

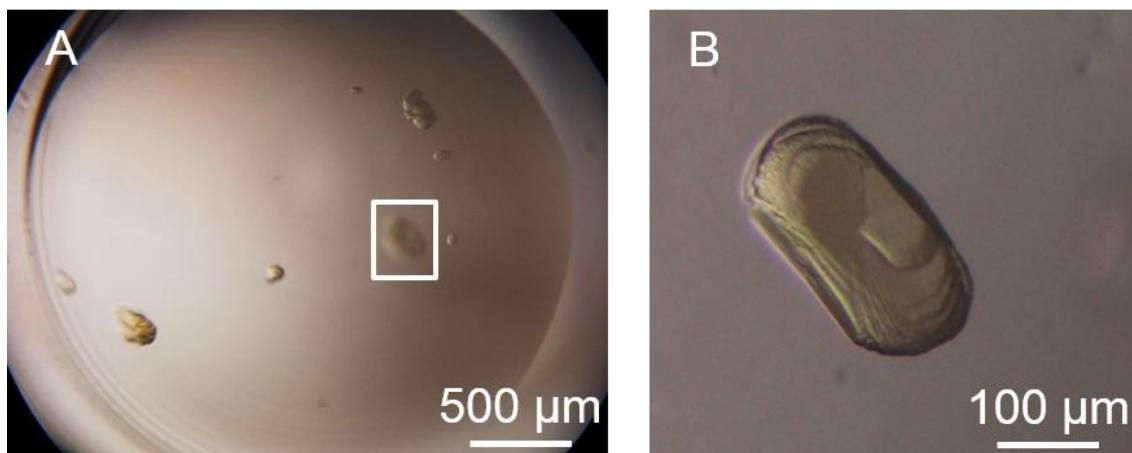


Figure 73– **A)** Micrograph of NΔTPO:4F5 F(ab)(1) crystals grown in 100mM bicine pH 9.0, 10% w/v PEG 6K. **B)** Magnification of boxed area of **A**. Scale bar is approximate. The crystal in **B** did not diffract. No further crystals of the NΔTPO:4F5 F(ab) could be produced.

Crystals of NΔTPO complexed with 4F5 F(ab) could not be reproduced despite several attempts. Optimisation screens were set up around the conditions which produced the initial hit (100 mM bicine pH 7.6-9.0, 10-24% PEG 6000) by hand with both NΔTPO:4F5 F(ab) preparations. The optimisation screens were also tested with the crystallography robot and at a 2:1 drop ratio. Repeats of identical conditions which produced the initial hit were included in these optimisation screens in an effort to replicate the hit. This also proved unsuccessful for both complex preparations using manual or robotic drop setting. A single crystal of NΔTPO:4F5 F(ab) (1) was harvested from the original hit and frozen in liquid N₂ with glycerol added to 25% as a cryoprotectant. This crystal was screened on the I04 beamline at Diamond Light Source but did not produce any measurable reflections.

Summary

Inhibitor screening

As NΔTPO crystallized the most readily of the four modified TPO proteins and produced the best diffracting crystals, further efforts focused on this protein. Both small molecule ligands and larger proteins such as F(ab)s can have a stabilising influence during crystallography improving both crystal growth and diffraction³⁸¹.

Small molecule ligands were selected from a recent publication which examined thousands of chemicals for TPO inhibiting properties³⁸⁸. Five inhibitors identified in this work were screened along with two compounds used clinically to treat hyperthyroidism. Despite one compound, the TPO inhibiting drug MMI showing some promise during the screening of crystal growth conditions, no improvement in diffraction was produced. Of the seven compounds tested five completely prevented crystallization of NΔTPO. While these inhibitors may have merely shifted crystallization conditions away from those previously established, repeating the screening from the sparse matrix stage while including the inhibitors would have been time consuming, costly and not considered a good use of resources. While a structure of any of these inhibitors bound to TPO, particularly the two in clinical use would be of great interest, this may have to wait until high resolution diffracting crystals of TPO or a TPO analogue can be produced.

F(ab) complex screening

Stabilising NΔTPO with a F(ab) fragment from an anti-TPO mouse monoclonal IgG was also attempted. This technique has proven successful for other proteins and would produce interesting data regarding the nature of autoantibody binding to TPO if successful. A NΔTPO:4F5 F(ab) complex was produced and purified. This complex was tested in sparse matrix crystal screens in a similar manner to the modified TPO proteins. Crystals of the NΔTPO:4F5 F(ab) complex were obtained from a single condition, but could not be replicated by subsequent experiments. When a crystal from this experiment

was tested, it failed to produce any significant diffraction. Overall it was shown that the modified TPO proteins produced here, particularly NΔTPO, have some potential for crystallography, but further work is required to produce a ≤ 3 Å structure.

Key observations

- NΔTPO was co-crystallized with the inhibitors methimazole and *p*-toluidine.
- NΔTPO was crystallized in complex with 4F5 F(ab).

Problems

- Despite producing NΔTPO crystals with a number of ligands, no improvement in diffraction resolution was observed.
- A number of the small molecule inhibitors tested for co-crystallization with NΔTPO inhibited crystallization.

Final Summary- Thyroid Peroxidase Crystallography

The overall goal of this project was to further understand the protein structure of human TPO and how that relates to its role in autoimmunity. Many biochemical studies of TPO have been carried out since it was conclusively shown to be the previously unidentified, “thyroid microsomal antigen” in 1987¹⁰³. A high resolution crystal structure of TPO has been long sought after as part of these investigations but, despite diffracting crystals of the protein first being produced in 1999, this still remains elusive²⁵⁵.

For this project the findings of a number of studies on TPO were combined with the wealth of knowledge available on crystallizing various reluctant proteins, in attempt to produce a modified form of TPO which would produce high resolution diffracting crystals. While improved diffraction data was produced, this still did not reach the 3 Å resolution required for a reliable structure (Chapter 3). The highest resolution data obtained was from the protein designated “NΔTPO”, consisting of the TPO amino acids 1-14,109-796 plus a 6-histidine tag. Further modifications made to NΔTPO did not result in any significant improvement in protein stability or crystal screening outcomes.

While the main aim of the crystallography experiments was not achieved this is not to say that the work itself is not without merit or use. The diffraction data produced from crystals of NΔTPO was an improvement on anything previously published for TPO. What is also interesting is that this protein could be crystallized fairly reliably using the conditions established here. Crystals of native or recombinant TPO produced elsewhere have proven difficult to repeat. This would indicate that NΔTPO is a good starting point for future crystallography based studies as opposed to the full length protein, especially given that NΔTPO has been demonstrated to share many key features with native TPO.

The slow progress towards an atomic level structure of TPO is disappointing particularly when related peroxidases such as MPO and LPO are readily crystallisable. While these proteins are the closest relatives to TPO it is

important not to forget about the large differences between these proteins and how they may effect crystallography. Although TPO is treated as a soluble protein in this work by removing the transmembrane domains, this is an inherent difference. The presence of domains such as the CCP and N-terminal domains which are required for a fully informative structure further complicate the crystallization of TPO. It should also be noted that the most homologous regions shared between the peroxidases are around the active site and therefore buried within the protein structure. Again this allows for much variation at the surface of the protein which effects protein crystallization. This is demonstrated by the marked differences in the pI of each protein. MPO, LPO and EPO are all cationic at physiological pH where TPO, with a pI of 5.4 would be anionic. While this may not affect features such as enzymatic function *in vivo*, it will completely change suitable conditions for the crystallization of each protein. This is not to say that the structures of MPO and LPO cannot be related to TPO in a meaningful manner, but it is clear that when attempting to crystalize TPO, the amount of information which can be gained from these structures is limited.

Chapter 6- Thyroid Peroxidase Epitope Mapping Studies

Introduction

While TPO plays a key role in the regulation of human growth and metabolism, much of the interest in this particular enzyme stems from its role as a key autoantigen in AITD. TPO has been shown to be a target for autoantibodies in the majority of AITD patients¹⁰³. Previous work has shown these autoantibodies recognise two main IDRs on the surface of TPO (Chapter 1, “Thyroid peroxidase protein structure”).

A number of different methods have been used to determine the nature of these IDRs, in some cases identifying individual residues which play a role in autoantibody binding. In lieu of a 3D structure of TPO bound to an autoantibody, the new data on TPO IDRs obtained by other biochemical methods could provide new insights and improve understanding of the nature of autoantibody binding to TPO.

The most successful and widely used techniques for determining the TPO IDRs along with a summary of their findings are discussed in Chapter 1. It is generally accepted that there are two non-linear IDR's on the surface of TPO, one at the C-terminal end of the peroxidase and another overlapping the peroxidase and CCP domains. Many of the key residues or sequences which make up the IDRs have been deduced by comparing the results of studies utilising methods such as competitive binding assays and site-directed mutagenesis.

While the current literature on this subject is extensive, there are some techniques which have not been applied to TPO. One of these techniques, epitope excision, could provide new information regarding the IDRs and further corroborate past work³⁹⁰. This technique involves enzymatically digesting an antigen in complex with an IgG molecule. The IgG protects the epitope from digestion, allowing it to be separated from the rest of the antigen. This isolates short peptides representing the epitope, which can then be identified.

The advantage of this technique is that it requires little prior knowledge of the antigen structure, provided an antibody to it can be obtained³⁹¹. Monoclonal antibodies are preferable for this type of study as they provide clearer results, although in some cases polyclonal patient sera has been used successfully³⁹². The resolution obtained is dependent on the location of protease sites for the enzyme used and may be improved by repeated digestion with different enzymes³⁹³.

For TPO there is an added benefit, a short peptide which bound to a monoclonal antibody could be useful for crystallographic studies. A small TPO peptide in complex with a F(ab) fragment could provide an easier target for crystallography than a complex with the entire protein, while still giving insight into the mechanism of binding.

The two TPO monoclonal antibodies, 4F5 and 2G4, used elsewhere in this work were used for this study. Previous work has shown that 4F5 and 2G4 both inhibit patient sera autoantibody binding to TPO, but not each other³⁸⁷. This indicates that each antibody recognises a different one of the two IDRs on TPO. While the binding kinetics and genetic structure of these two antibodies have been previously studied it is not currently known which IDR either antibody is directed against^{109,258}.

The protease selected for this study was bovine pancreatic trypsin. Trypsin is a well characterised, 24 kDa serine protease³⁹⁴. It hydrolyses proteins on the C-terminal side of accessible lysine and arginine residues³⁹⁵. It cleaves at a slower rate if the residue C-terminal to the lysine or arginine is acidic, and is inhibited by the presence of proline in this position. Trypsin has a long history of use in protein characterization experiments as it is readily available and efficiently cleaves proteins in a predictable manner³⁹⁶. It is particularly favourable for mass spectrometry (MS) based studies as the peptides produced have a positively charged residue at the C-terminus, which aids positive ionization.

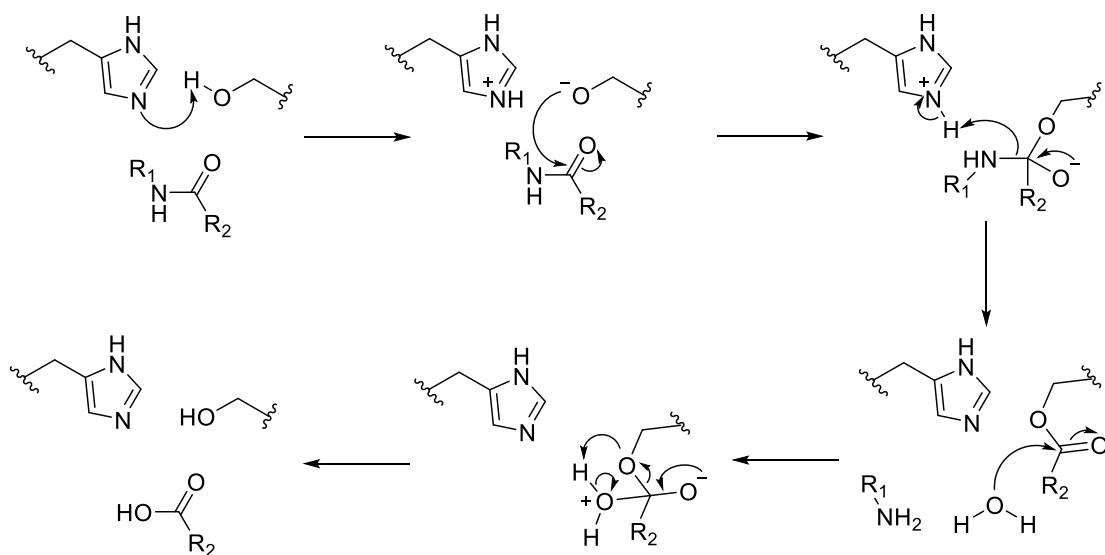


Figure 74- Trypsin reaction mechanism³⁹⁷. The histidine residue accepts a proton from serine which then attacks the carbonyl carbon of the substrate. The peptide bond is broken and water enters the active site. The water attacks the acyl-enzyme intermediate releasing the C-terminal region of the peptide and regenerating the active site.

To ensure the enzymatic cleavage sites could be reliably predicted trypsin stocks with tosyl phenylalanyl chloromethyl ketone (TPCK) added were used. TPCK has no effect on trypsin activity but inhibits chymotrypsin by forming a covalent bond with the active site histidine³⁹⁸. Chymotrypsin is a common contaminant in trypsin preparations even following repeated purification by crystallization. TPCK does not inhibit trypsin as the phenylalanine moiety is not accepted into the active site. By inhibiting chymotrypsin activity, the number of potential cleavage sites is reduced, simplifying later analysis.

Results and Discussion

Antibody selection

Key points

- The anti-TPO monoclonal antibodies 4F5 and 2G4 were tested for resistance to trypsin digestion.
- 4F5 showed a reasonable level of resistance to trypsin.
- 2G4 was readily digested by trypsin and therefore was not suitable for epitope excision studies.

For antibody excision to be successful it is vital that the antibody used is resistant to digestion by proteases. The first step of this study was to determine if the two TPO antibodies were resistant to digestion and therefore suitable for use. To test this a small amount of each IgG was incubated with the enzyme in various IgG:trypsin weight ratios. After digestion samples were compared to a negative control incubated without trypsin by analytical SEC and reduced SDS-PAGE. A significant change in the shape of the analytical SEC elution profile or the appearance of additional bands by SDS-PAGE was taken as evidence of IgG digestion. The trypsin stock was also analysed by analytical SEC to confirm that, at the concentration used, the addition of trypsin did not produce any significant new absorbance peaks in the analytical SEC elution profile.

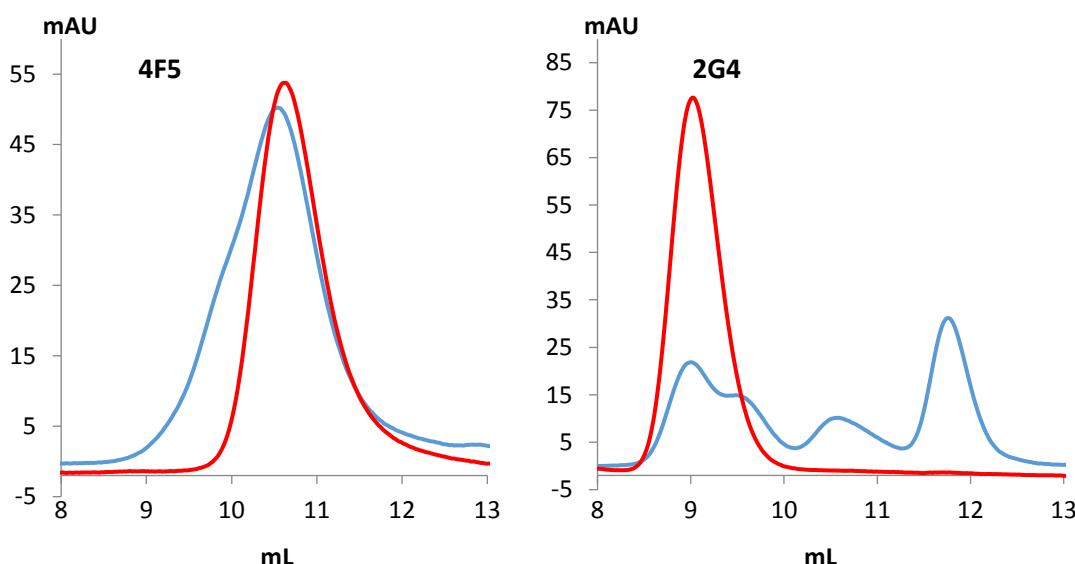


Figure 77- SEC elution profiles (absorbance at 280 nm, peaks only for clarity) for 100:1 protein:trypsin digested 4F5 and 2G4 IgG (blue), and respective IgG samples incubated under same buffer conditions without trypsin (red).

The analytical SEC elution profiles in Fig. 77 show that 4F5 maintains a largely similar elution profile following an overnight trypsin digestion when a 100:1 IgG:trypsin ratio is used. Beyond this ratio the main peak begins to appear less well defined, indicating some digestion of the IgG is occurring. This data indicates 4F5 was suitable for epitope excision studies provided a high protein:trypsin ratio was not required.

The 2G4 profiles in Fig. 75 show this IgG was much more readily digested by trypsin than 4F5. Even at a 100:1 ratio there are several significant peaks present in the profile following digestion and the peak corresponding to intact 2G4 IgG is much less intense. This suggests the protein is being cleaved into several fragments. As they are eluting well before the void volume of the column, the molecular weight of these fragments must be large enough for them to represent sizable fragments of the molecule.

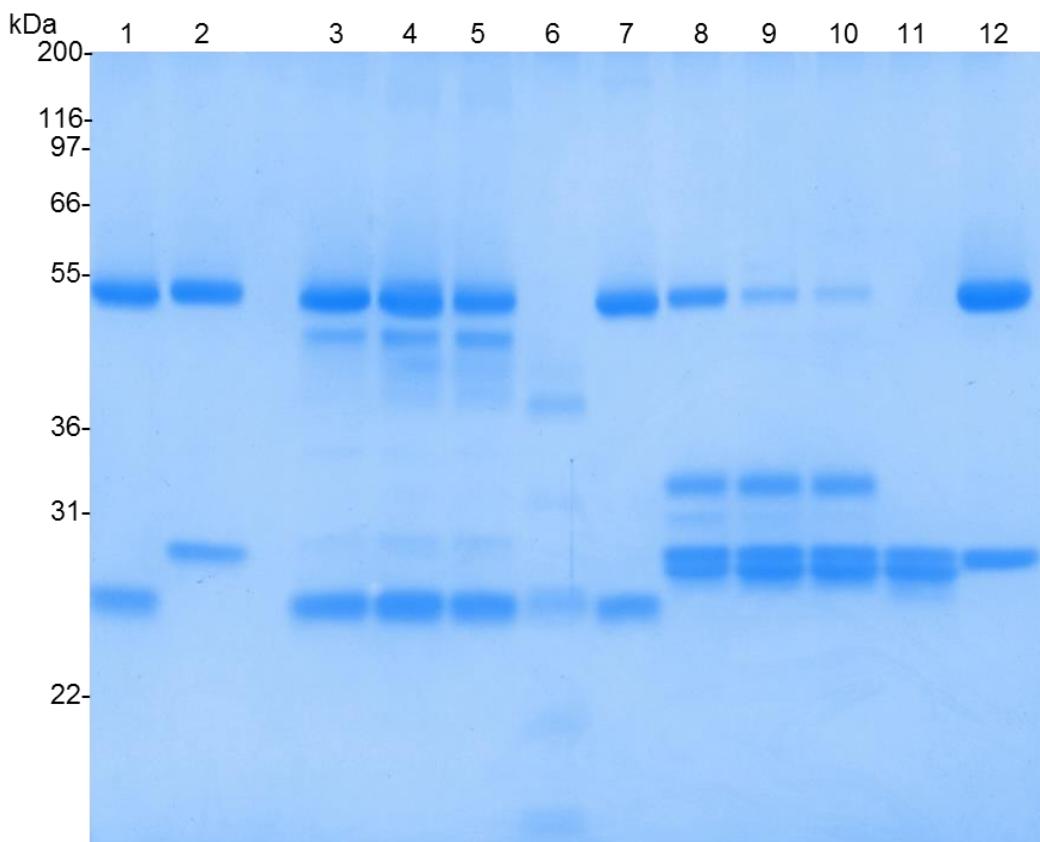


Figure 76- Stained 8% acrylamide reduced SDS-PAGE of trypsin-digested 4F5 and 2G4. Lane 1: undigested 4F5. Lane 2: undigested 2G4. Lanes 3 to 6: 4F5 digested with a 100:1, 50:1, 40:1, 4:1 trypsin ratios. Lane 7: 4F5 incubated without trypsin. Lanes 8 to 11: 2G4 digested with a 100:1, 50:1, 40:1, 4:1 trypsin ratios. Lane 12: 2G4 incubated without trypsin. The presence of additional bands relative to the controls (lane 7 and 12) may indicate digestion of the IgG.

The SDS-PAGE analysis results (Fig. 76) agree with those produced by analytical SEC. 4F5 appears mostly intact, although a band at 50 kDa is visible which is absent in the negative control (Fig. 76, lanes 3-5). This indicates limited digestion of the heavy chain occurred, but the majority of molecules have remained undigested, as evidenced by the intense bands corresponding to the heavy and light chains. 2G4 shows more significant signs of digestion. Extra bands are visible at 34 and 24 kDa in the test samples and the intensity of the heavy chain band at 55 kDa is noticeably decreased.

The epitope excision method used for in this work (described later) involves the digestion of TPO whilst in complex with Sepharose bound IgG. The Sepharose may introduce some steric hindrance which would prevent significant digestion of the 2G4 IgG however, based on the results of this experiment, the decision was made to concentrate on 4F5 as it was clearly more resistant to trypsin digestion. Given the existing knowledge of these antibodies, if it can be shown that 4F5 binds to a specific TPO IDR, it could be inferred that 2G4 recognises the other.

While 2G4 may have resisted lower trypsin concentrations than those used here, overnight trypsin digests of rhTPO showed the protein to be fairly resistant to digestion whilst in its native conformation (Fig. 77). Epitope excision only works in cases where the IgG is less readily digested than the bound protein.

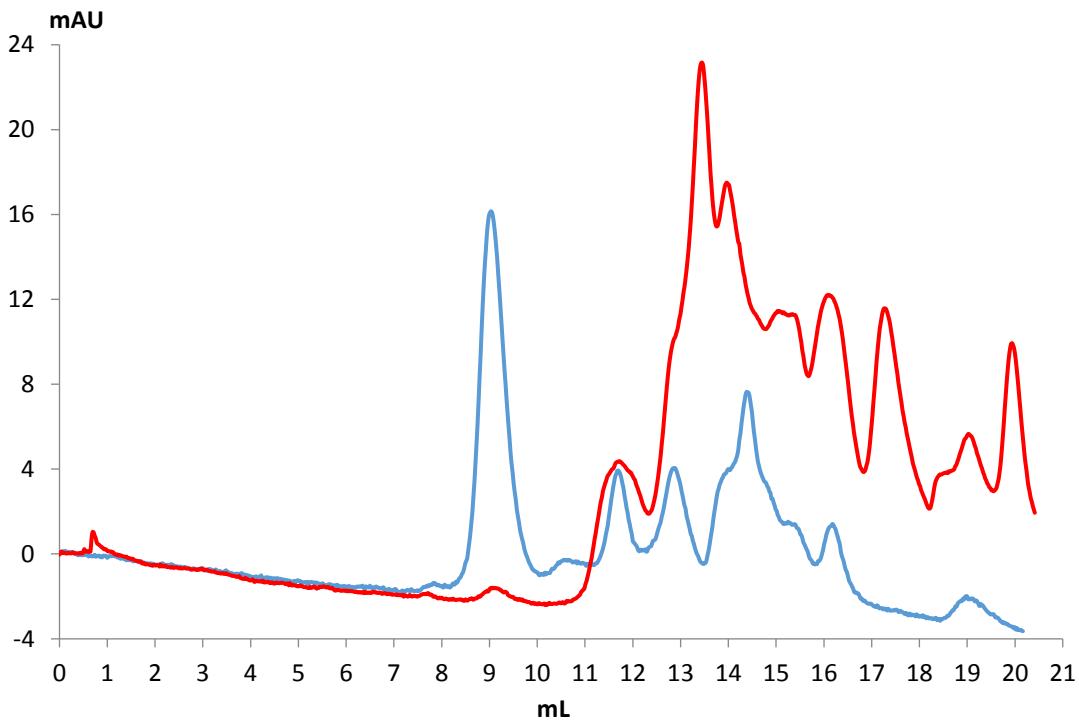


Figure 77- Analytical SEC elution profiles (absorbance at 280 nm) of rhTPO digested with trypsin at a 100:1 w/w ratio (blue) and 5:1 w/w (red). Intact rhTPO elutes at ~9 mL, any peaks eluting after this will be of lower molecular weight and therefore present due to digestion.

4F5 epitope excision method development

Key points

- A method for epitope excision was developed utilizing sepharose bound 4F5 IgG.
- The final result is obtained by LC-MS/MS analysis of peptides produced by the epitope excision process.

While the general principals remain the same in many studies, there is no single common method of epitope excision used throughout the literature³⁹⁰. The exact methodology used differs depending on the study requirements and the properties of the antibody in question. Therefore, while a general method for epitope excision using 4F5 IgG could be developed using ideas from the literature, some experimentation was required to fully develop a working technique.

An outline of the procedure used in this work is shown in Fig. 78;

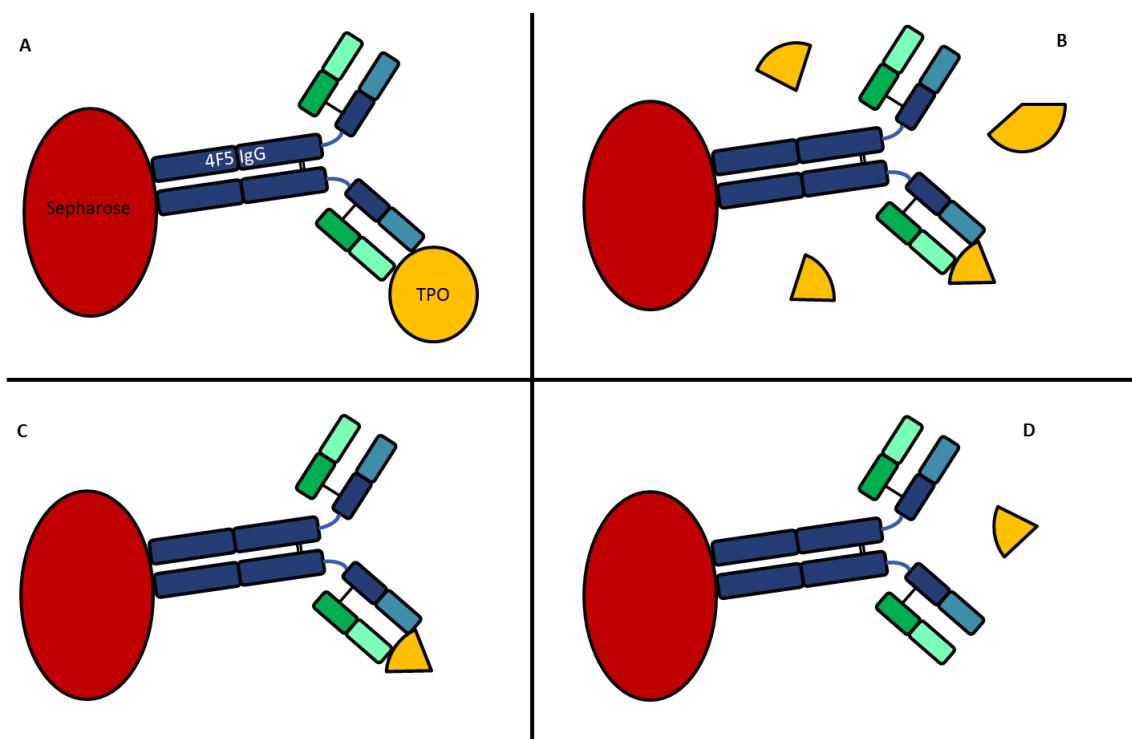


Figure 78- Overview of epitope excision procedure used. **A** TPO is bound to 4F5 IgG Sephadex. **B** IgG bound TPO is digested with trypsin, the epitope region is protected from digestion by the IgG and remains bound. **C** Unbound TPO peptides are removed leaving only the epitope peptide in complex with the IgG. **D** The peptide is eluted from the IgG and identified.

One issue in developing this method was how to efficiently isolate the IgG and any bound peptides. Trypsin and unbound protein fragments would need to be removed from the samples, prior to analysis, to ensure the correct peptides were identified. For this work 4F5 IgG was coupled to cyanogen bromide activated Sephadex (G.E, UK). This allowed the IgG, along with any bound peptides, to be removed from suspension by centrifugation. While effective, this method was not without limitations. IgG bound Sephadex can act as an ion exchanger giving some potential for non-specific binding³⁹⁹. It is also difficult to completely remove the liquid the matrix is suspended in following centrifugation. Complete drying of the matrix is undesirable as it may damage the bound molecules and prove difficult to rehydrate. To ensure only the desired peptides were isolated, the matrix was washed repeatedly with a high ionic strength buffer following the digestion of 4F5 bound TPO. This would remove any peptides bound through a non-specific ionic bond to the matrix. Previous experiments have shown that this does not remove TPO bound to 4F5 (RSR Ltd, unpublished observations).

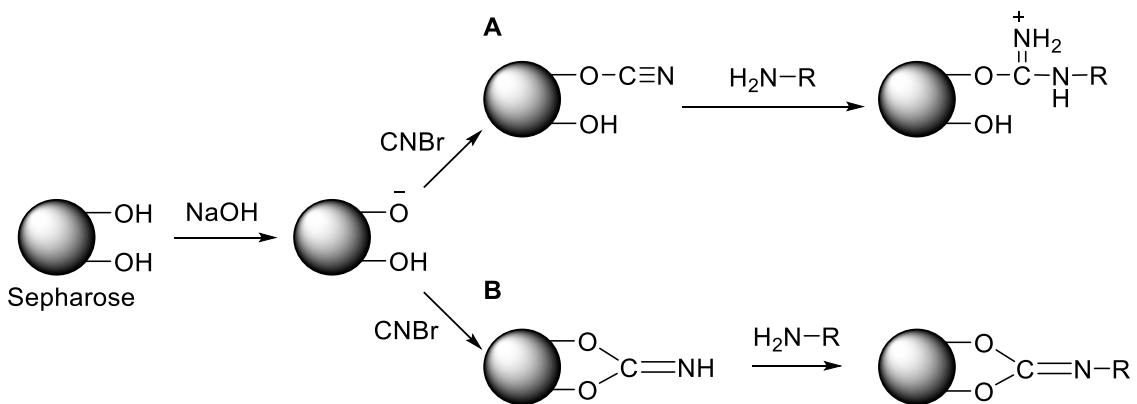


Figure 79- Mechanism of IgG coupling to CNBr Sepharose. Cyanogen bromide reacts with the oxygen anion groups on the surface of agarose beads to produce either **A)** a cyanate ether which reacts with an amine group to produce an isourea derivative or **B)** a cyclic imidocarbonate which is substituted with an amine group from the ligand.

Trypsin digestion conditions for the TPO:IgG Sepharose complex were deduced from the findings of the antibody selection experiments described previously. A protein:trypsin ratio of 100:1 was used as this did not appear to significantly digest the 4F5 IgG. When rhTPO was digested alone under these conditions digestion appeared to be incomplete (Fig. 77). Denaturing or reducing the protein may produce a more complete digestion as many trypsin sites would be expected to be buried in the folded protein and inaccessible to the enzyme. This was unfortunately not a viable option as previous work with TPO has shown that reducing or denaturing the protein has a negative effect on antibody binding²⁹³. A cautious approach was therefore taken by using a 100:1 trypsin ratio in the digestion step as IgG digestion could produce unclear or false results, complicating the analysis of the data produced.

Once the epitope peptide was isolated a method for eluting and identifying the peptide was required. Liquid chromatography coupled to MS analysis (LC-MS) allowed any eluted peptides to be separated on a reverse phase (RP) column prior MS analysis. This gave an insight into the number of peptides and other components present in the eluent as well as the mass of each. Electrospray ionization was used as this produces unfragmented, multiply charged peptide ions⁴⁰⁰.

Individual peptides were identified by MS/MS. This technique uses a quadrupole to selects individual mass/charge ratios of ions which are then fragmented by acceleration through a gas filled collision chamber⁴⁰¹. At low kinetic energy levels, peptides fragment along the peptide backbone at the amide bond, producing ions of varying length depending on the amino acid composition⁴⁰². This allows the protein sequence to be determined, *de novo*, from the mass intervals between the ions present in the MS/MS spectra⁴⁰³. As peptides in the eluent will have originated from one of three known proteins, TPO, 4F5 or trypsin any peptide sequences observed by mass spectrometry could be checked against the sequences of these proteins to determine if they matched a feasible digestion product.

The Edman degradation is an alternative method of peptide sequencing and was commonly used prior to the development of MS based methods⁴⁰⁴. This process involves labelling the amino acid at the N-terminal of a peptide with phenyl isothiocyanate. The labelled amino acid can then be cleaved under mildly acidic conditions and identified by RP chromatography⁴⁰⁵. Successive rounds of the degradation reaction allow the sequence to be built up. This technique had a potential use in this work as having a confirmed N-terminal starting point from Edman degradation would be useful when identifying the epitope containing peptide.

Peptide sequencing by Edman degradation for this work was carried out by Alta Bioscience (UK). The results produced were unclear as multiple peptides of different sequences were present in the sample. Edman degradation requires a sample of mostly pure peptide to be successful and in this case, it appears the sample purity was not sufficient. This would be less problematic for MS based approaches so sequencing by Edman degradation was not explored further.

Previous studies have shown that protein conformation is important for patient sera autoantibody recognition of TPO⁴⁰⁶. However, the residues assigned to the TPO IDR-B consist of some linear sequences of amino acids (Fig. 84). If 4F5 was directed to this epitope then these peptides, when free of the protein structure, may still be able to adopt a conformation which allows them to bind to the antibody. To test this, an experiment was carried out using a similar method

to the epitope excision process. TPO was digested with trypsin before binding to the antibody. Once complete the digestion was stopped and Sepharose bound 4F5 was added to the mixture. Following incubation, the Sepharose bound 4F5 along with any peptides bound to the antibody, were separated from the mixture, washed and eluted. The samples taken throughout the process were analysed by analytical SEC with the eluted material also analysed by LC-MS. Analysis by analytical SEC showed that peptides were still present in the supernatant following incubation with the IgG and only a very small amount of material was removed by a high ionic strength buffer wash following binding. Both the analytical SEC and LC-MS results showed no peptides had been eluted at the end of the process. The elution method had previously been validated by other experiments so this would suggest that no peptides bound to the IgG when TPO was digested prior to binding. This would indicate that the 4F5 epitope is conformationally dependent. While this result does not provide any evidence regarding which TPO IDR 4F5 recognises, it would agree with previous data which suggests that the nature of 4F5 binding is similar to that of patient sera autoantibodies.

As the above method produced a negative result with 4F5, it was not tested with 2G4. While this method would resolve the 2G4 digestion issue, both TPO IDRs are thought to be conformationally dependent and therefore similar results would be expected^{293,294}.

Epitope excision results and discussion

Key points

- A peptide (designated “EP1”) covering the 4F5 epitope was isolated and identified.
- EP1 consists of two disulphide bonded, tryptic peptides.
- EP1 contains most of the amino acids assigned to the TPO IDR-B in previous studies, indicating that this is the IDR 4F5 IgG recognises.
- Based on these results and previous work comparing the binding of 4F5 and 2G4, 2G4 must therefore recognise the TPO IDR-A.

The epitope excision procedure described above allowed a peptide representing the 4F5 epitope to be isolated. The large peak eluting at 20.38 min in Fig. 80 is a peptide of mass 4690.34 Da (Fig. 81). This peptide, referred to as “EP1” was observed as the most abundant peptide peak in nine repeats of the process.

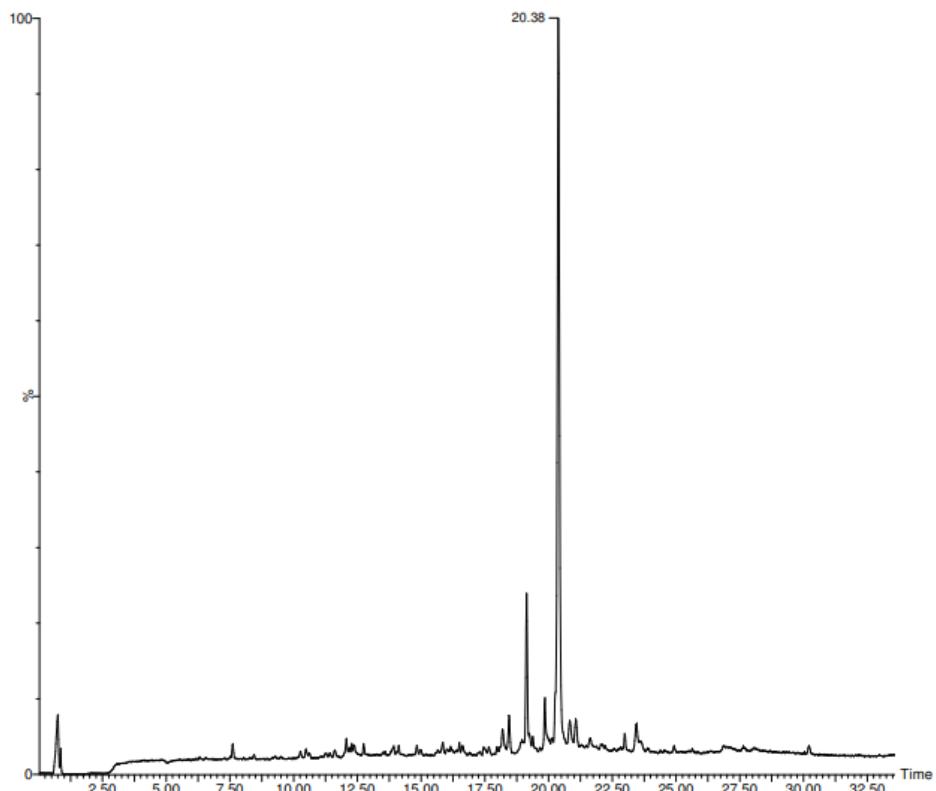


Figure 80- Epitope excision elution LC-MS total ion count (TIC) chromatogram. EP1 elutes at 20.38 minutes when using a 3-60% acetonitrile:water gradient over 31 minutes over an Acquity CSH C18 column (Waters, UK). The total ion current was 4.48×10^6 .

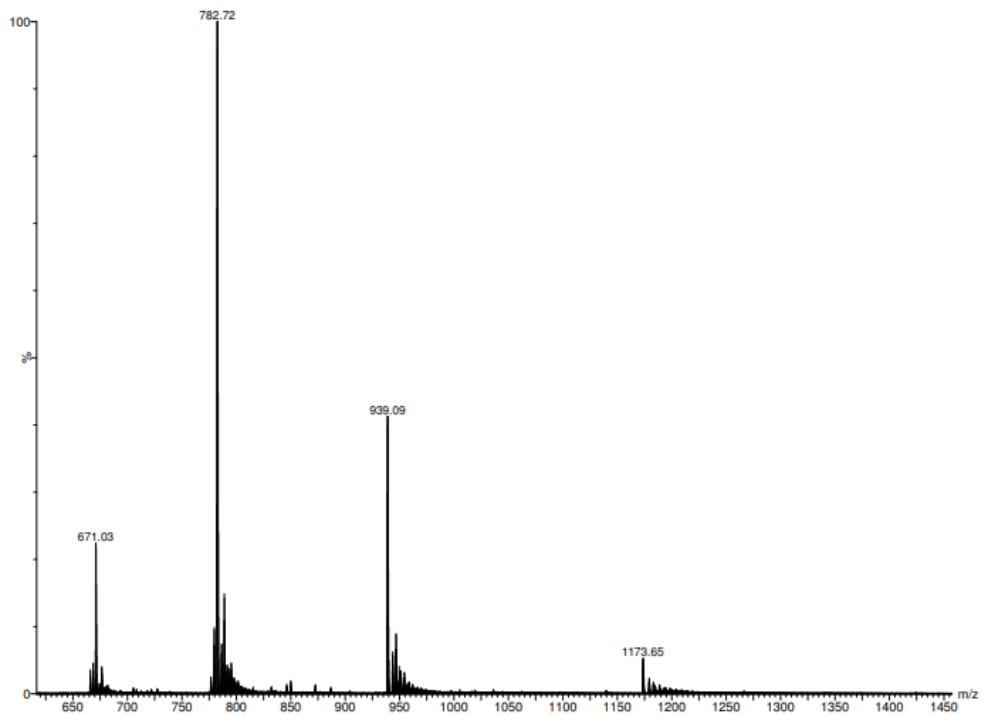


Figure 81- Deconvoluted spectra of peak eluted at 20.38 min in Fig. 77 (EP1). Intact peptide mass = 4689.34 ± 0.16 Da. The total ion current was 8.27×10^5 .

To aid in identifying EP1, simulated trypsin digests of the rhTPO protein sequence were carried out using the PeptideMass program⁴⁰⁷. This program uses the known conditions required for trypsin cleavage to generate all the possible peptides produced from a given protein sequence. The EP1 peptide is of a greater mass than the largest expected fragment from the simulated digest. There are several explanations for this including post-translational modifications or missed trypsin cleavage sites within the peptide. The latter would be unsurprising when digesting a protein in its native confirmation and may also have occurred due to steric hindrance from the 4F5 IgG. However, even when allowing for up to five missed cleavage sites, EP1 does not clearly match the mass of any predicted peptide from the sequence.

To further clarify the identity of EP1 MS/MS analysis was carried out. The N and C terminal sequences could be identified by manual curation of the CID fragmentation data (Fig 82).

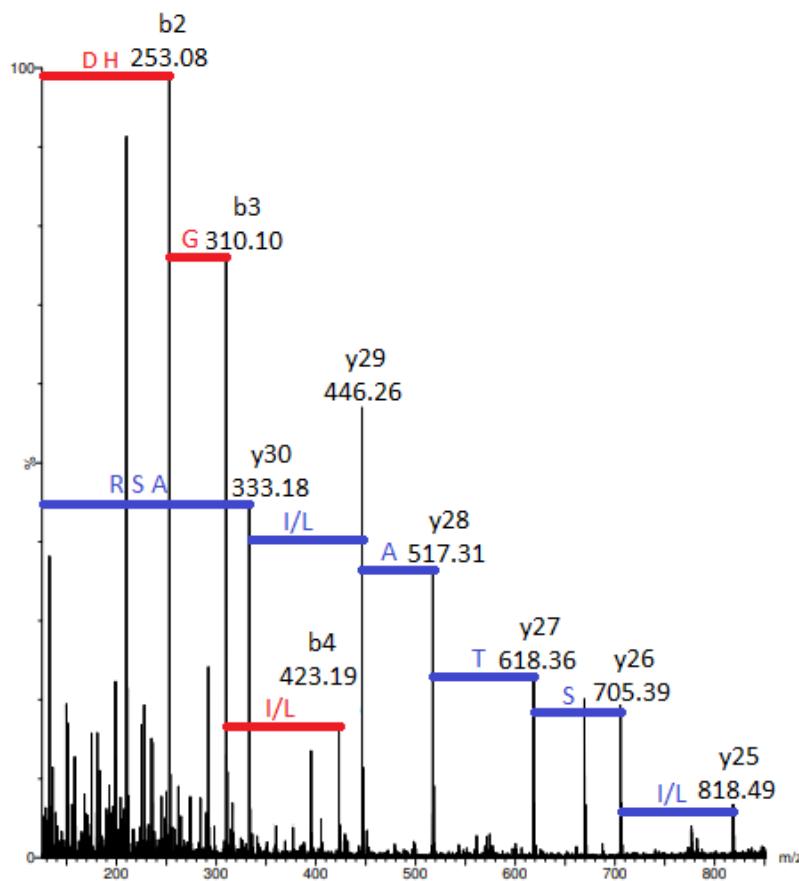


Figure 82- MS/MS spectra of EP1 peptide fragmented by CID showing b ion series (red) and y ion series (blue). Intact mass 4689.34 Da. The total ion current was 3.53×10^5 .

The MS/MS data indicates EP1 is a tryptic peptide spanning the TPO amino acids 585-616. The calculated monoisotopic mass of this peptide (3570.73 Da) is far below the observed mass of EP1 obtained from the epitope excision experiments. However, there is a cysteine present in the sequence identified (amino acid 598). All experiments were carried out under oxidising conditions and therefore this residue would be expected to form a disulphide bond with another cysteine. Based on sequence alignments with related peroxidases (Chapter 1, “Thyroid peroxidase protein structure”) this would be expected to be cysteine 655. A tryptic peptide spanning amino acids 649-659 would contain this residue and have a calculated monoisotopic mass of 1118.62 Da. These two peptides, joined by a disulphide bond, would have an average mass of 4689.35 which closely matches the value observed here for EP1 (2.13 parts per million (ppm) error) and is therefore thought to be the identity of EP1 (Fig. 83).

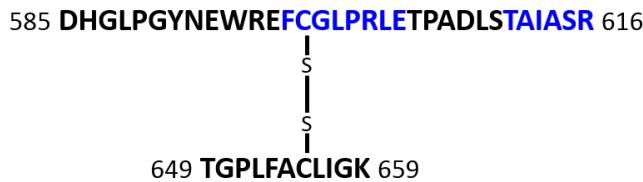


Figure 83- EP1 peptide structure in single letter amino acid code. Residues contributing to TPO IDR-B highlighted (blue).

The sequence between amino acids 585-616 covers most of the residues assigned previously to the TPO IDR-B. While this would indicate that 4F5 is directed to this IDR, it is assuming that the 585-616 peptide is the site of binding. An argument could be made that the 649-659 peptide is bound to the antibody and the larger peptide (between amino acids 585-616) is present due to the disulphide bond. While this can't be directly disproved by the data presented here, it is unlikely when considering results from the existing literature³⁸⁷. 4F5 has been shown to inhibit patient sera autoantibodies and would therefore be expected to overlap with some of the known IDR residues.

```

109 TQQSQHPTDALSEDLLSIIANMSGCLPYMLPPKCPNTCLANKYRPITGACNNRDHPRWGA
169 SNTALARWLPPVYEDGFSQPRGWNPFLYNGFPLPPVREVTRHVIQVSNEVVTDDDRYSD
229 LLMAWGQYIDHDIAFPQSTSAAFGGGADCQMTCEQNQPCFPIQLPEEARPAAGTACLP
289 FYRSSAACGTGDQGALFGNLSTANPRQQMNGLTSFLDASTVYGSPPALERQLRNWTSAEG
349 LLRVHARLRDSGRAYLPFVPPRAPSACAPEPGIPGETRGPCFLAGDGRATEVPSLTALHT
409 LWLREHNRLAALKALNAHWSADAVYQEARKVVGALHQIITLRDYIPRILGPEAFQQYVG
469 PYEGYDSTANPTVSNVFSTAARFRGHATIHPLVRRLDASFQEHPDLPGLWLHQAFFSPWT
529 LLRGGGLDPLIRGLLARPAKLQVQDQLMNEELTERLFVLSNSSTLDLASINLQRGRDHGL
589 PGYNEWREFCGLPRLTPADLSTAIASRSVADKILDLYKHPDNIDVWLGGLAENFLPRA
649 TGPLFACLIGKQMKALRGDWFWWENSHVFTDAQRELEKHSSLRVICDNTGLTRVPMDA
709 FQVGKFPEDFESCDISIPGMNLEAWRETFPQDDKGFPESVENGDFVHCEESGRRLVLYSC
769 RHGYELOGREQLTCTQEGWDFQPPLC

```

Figure 84- Human TPO amino acids 109-796 (N-terminal, peroxidase and CCP domains). █ Immunodominant region A █ Immunodominant region B. Amino acids 585-616 and 649-659 underlined.

The amino acid sequences 597-604 and 611-618 were assigned to the TPO IDR-B by Bresson *et al* by probing short TPO peptides with a polyclonal rabbit antisera raised against the TPO amino acid sequence 599-617^{256,300}. The peptide fragment identified here covers most of the same amino acids but lacks amino acids 617 and 618 at the C-terminal. Bresson *et al* tested the influence of each amino acid from 597 to 618 by synthesising peptides with alanine substituted for each individual residue. At some positions, including amino acids 617 and 618, the alanine mutation greatly decreased the level of antibody binding suggesting these residues are important for the binding of antibodies

from the rabbit sera. Given this information it is surprising the peptide isolated by this technique terminated at amino acid 616, particularly as there is another potential cleavage site just slightly further along the sequence at lysine 621. Cleavage at lysine 621 would also preserve aspartic acid 620 which was identified by Dub ska *et al* as another key residue for autoantibody binding³⁰⁴. This data was produced by producing single residue TPO mutants and testing the binding of these mutants against two human IDR-B specific F(ab)s originating from a HT patient. The data presented here shows that this cleavage site is not blocked by 4F5 and is still accessible to trypsin. This may suggest that there are some slight differences in the exact residues involved in binding when comparing 4F5 to these particular F(ab)s.

Given the major differences in the techniques used to generate and characterise TPO antibodies for both works some slight discrepancies are unsurprising. The rabbit sera used by Bresson *et al* bound short (eight amino acids in length) TPO peptides where 4F5 would not bind trypsinized TPO. This indicates there are some inherent differences between the antibodies which is likely reflective of the different methods used to produce them. 4F5 was raised in mice against purified TPO from human thyroid tissue and as such recognises a conformational epitope, unlike the rabbit sera used by Bresson *et al* which was raised against a short, linear sequence.

The work of Godlewska *et al* adds some weight to this theory; they expressed human TPO with single amino acid mutations in CHO cells to probe the specificity of three IDR-B mouse monoclonal antibodies in a similar manner to Dub ska *et al*^{304,408}. The antibodies used in this work were originally produced by injecting mice with TPO purified from human thyroid tissue, a similar method to that used to produce 4F5²⁹². Godlewska *et al* observed that a mutation at glutamic acid 604 inhibited the binding of all three antibodies. However, mutations at aspartic acid 620, aspartic acid 624 and lysine 627 did not significantly reduce the binding of all three antibodies tested. This would indicate that even antibodies produced in the same manner may show some slight variation in the contact forming residues when binding.

Overall, this work is in agreement with the existing literature and confirms 4F5 binds a TPO IDR at the same region as AITD patient sera antibodies. Further clarification of the exact residues involved in antibody binding would likely require a structure of TPO, preferably in complex with 4F5 or another antibody, a project which is receiving on-going attention.

Conclusions

The epitope of the anti-TPO IgG 4F5 (EP1) has been isolated and identified as a pair of disulphide linked, tryptic peptides. The peptide sequence between amino acids 585-616 was confirmed to be present from MS/MS data. This sequence does not match the observed mass of the intact fragment but this can be explained by the presence of a second peptide, representing amino acids 649-659, disulphide bonded to the confirmed sequence. While this second peptide was not observed directly, the predicted average mass of this proposed fragment matches the observed mass closely enough to give a reasonable level of certainty that it has been correctly identified.

The 585-616 sequence would overlap with residues previously assigned to the TPO IDR-B³⁰⁰. Previous work has shown 4F5 inhibits patient sera binding which suggests it binds either at or near, one of the two TPO IDRs³⁸⁷. As the observations from this work agree with this data, it can be concluded that 4F5 is directed to the TPO IDR-B, although the exact nature of this epitope cannot be further defined from these results. Experiments involving binding pre-digested TPO to 4F5 provided further confirmation that the 4F5 epitope is conformationally dependent.

While not studied directly here, 2G4, another TPO antibody used in the course of this project, has also been shown to inhibit patient sera autoantibodies from binding to TPO while not inhibiting 4F5³⁸⁷. This would suggest that each of these antibodies are directed to separate TPO IDRs. 2G4 was not suitable for epitope excision using the above method so this could not be confirmed directly. However having demonstrated that 4F5 binds to TPO IDR-B and considering previous work which has shown 2G4 binds to one of the IDRs while not

inhibiting 4F5, it can be stated with some confidence that 2G4 binds to, or in close proximity to, IDR-A.

In the case of TPO, where antibodies mostly bind to two fairly well-defined areas of the protein, this method could be useful for quickly determining which of the two IDRs an antibody binds without the need for a previously defined antibody as a reference. This is of course assuming that the antibody to be tested is resistant to trypsin digestion, which this work has shown is not always the case. While in this instance the IDR was not further defined based on the data produced, future work could utilise repeated digestion steps with other enzymes to narrow down the critical residues involved in antibody binding. Alternatively, applying this method to more TPO antibodies could also produce interesting data by comparing the peptides produced and establishing if there are any common patterns.

Summary

The data produced here demonstrates for the first time that the TPO monoclonal antibody 4F5 is directed to the IDR-B. This information, combined with previous work, can be used to conclude that the monoclonal antibody 2G4, must therefore be directed to the IDR-A. While no new information regarding the key residues of the TPO IDRs was produced here, an extension or further application of this technique does have the potential to do so in the future.

Key observations

- Epitope excision is a viable technique for the study of some TPO antibodies.
- The monoclonal antibody 4F5 has been shown to recognise the TPO IDR-B. This information may prove useful when re-examining the previously produced crystal structure of 4F5 F(ab).
- From this data it can be inferred that the monoclonal antibody 2G4 is directed against IDR-A.

Problems

- Not all antibodies are trypsin resistant and therefore a different protease would be required to study them by this method.
- While this work has produced some novel data further work would be required to extend the observations beyond what has previously been shown with other antibodies.

Chapter 7- Future work and outlook

While the crystallography experiments carried out in this work were extensive, particularly with the promising NΔTPO protein, they were not exhaustive. The NΔTPO protein could prove to be a useful starting point for future work. There are several techniques which could still be examined with this protein or other modified TPO proteins across the whole crystallography process. Many of these could be combined and may, in conjunction, produce incremental improvements which could finally allow the 3 Å diffraction target to be reached. In the longer term, there are emerging techniques which may prove more suited to producing a structure of a challenging target such as TPO.

Future crystallography work: short term

There are several techniques for improving crystallography outcomes which could be tested with NΔTPO or the other proteins produced here. Using this work as a starting point, future work could explore techniques such as chemical or enzymatic modification. Modifying NΔTPO by reductive alkylation is one example of this. This process involves modifying primary amines of surface exposed lysine residues and at the N-terminus to a tertiary amine. Reductive alkylation has been shown to be useful for difficult proteins which resist crystallization despite being highly pure and homogeneous⁴⁰⁹. Modifying the protein by this technique could improve the packing order of the crystals sufficiently to reach the required resolution for an accurate structure⁴¹⁰. A similar effect can be achieved by rational mutagenesis of the protein sequence⁴¹¹. However, when prior structural knowledge is limited, a rational approach requires some trial and error which can be arduous, particularly for a large protein such as TPO. The advantages of reductive alkylation are that no prior knowledge is required and, assuming the protein of interest has already been obtained, the process is much quicker. The exact mechanisms by which this improves crystal diffraction quality are not known but it is suspected to be due to a reduction in the protein surface entropy, increasing the number of sites where crystal contacts could form on the protein surface⁴¹². All four of the modified

TPO proteins produced in this work could potentially be suitable for this technique.

Another modification which could be tested is enzymatic digestion. Very gentle digestion of NΔTPO in its native conformation with a protease such as trypsin may trim some of the surface exposed loops while leaving the more densely folded core of the protein intact⁴¹³. This would alter the surface of the protein, possibly allowing previously unavailable contacts to form during crystallization and also reducing heterogeneity in the positioning of surface loops. This could be carried out directly in the crystallization drop and has the benefit of not requiring rational design unlike the approach taken here with the NΔTPO-2, -10 and -14 proteins. While this is the simplest approach it may be difficult to reproduce any successful conditions. Alternatively, the stable proteolysis product could be identified by mass spectrometry and expressed recombinantly⁴¹⁴. This would be more time consuming but should improve the reproducibility of the process and simplify later crystallization optimisation steps.

Once crystals are obtained there are several methods which may improve the resolution of data obtained from them. Further optimisation of the cryoprotectants used when freezing the crystals is one area which could receive further attention in future studies. This was not investigated more extensively in the course of this work due to limitations on time and resources. While glycol based cryoprotectants, as used in this work, have a long history of successful application they can enhance the solubility of proteins which may explain the observed melting of some crystals during cryoprotection⁴¹⁵. More complex cryoprotectant mixtures containing alcohols and dimethyl sulfoxide have been developed in recent years which may better preserve NΔTPO crystals⁴¹⁶. These have shown to be less likely to dissolve or disrupt the packing order of crystals of several proteins.

It is thought that any handling or manipulation of protein crystals will introduce slight imperfections which will go on to decrease the overall diffraction resolution⁴¹⁷. In a case like this, where there is low resolution diffraction, this could be enough to prevent the structure being solved. *In situ* diffraction is an emerging technique which is becoming more widely used as its benefits

become more widely known^{418,419}. It was not possible to attempt *in situ* diffraction of any of the crystals produced during the course of this work. Equipment to carry out these experiments is improving and becoming more common making this an interesting avenue for further study.

Longer term studies: crystallography and beyond

If the techniques discussed above do not successful with the proteins produced in this work, then further experiments could look at other modifications of the TPO protein. Two of the TPO proteins produced with loop truncations, NΔTPO-10 and NΔTPO-14, completely failed to crystallize despite appearing very similar to NΔTPO by the analysis carried out. NΔTPO-2 fared better, which may reflect the less significant change to the protein, but the crystals produced still did not diffract as well as those of NΔTPO. It is possible further analysis of the structure of these proteins may reveal some clues as to why this occurred, but this will not necessarily aid in producing high quality TPO crystals.

The modifications made to all three of these proteins were based on information from the structures of other peroxidase enzymes (Chapter 2). The structural significance of these changes is difficult to predict *in silico* without a reliable structure of TPO and therefore needed to be deduced experimentally. In this case the effects of the truncations were mostly negative, but this does not mean this approach is not worth further consideration. Due to the large amounts of protein required and the necessity of using a eukaryotic expression system, it was not possible to test more loop truncations in the course of this work. However, this could be explored further in the future, using the findings of this study as a starting point. If the optimal truncation for each of the two loops examined here could be determined then both these modifications could be combined in a single protein to further limit heterogeneity at the protein surface.

A single insect cell line was used for expression of all the modified TPO proteins used in this work. This was the well-established, High Five cell line from *Trichoplusia ni*. This represents another opportunity for study as the same proteins expressed in another cell line may behave differently. The production of correctly folded TPO requires a eukaryotic expression system but previous

work has shown it is difficult to obtain good yields from yeast cells, which are faster growing and cheaper to maintain³⁰⁹. This is unlikely to be different for the truncated forms produced here but mammalian cells such as CHO cells or insect cells from an alternative species may prove successful. TPO has been previously expressed in mammalian and alternative insect cell lines but, judging from the published literature, not with an intent to use the expressed protein for crystallography^{289,328,420}. The glycosylation pathways differ even between insect cell lines from different species and in many cases are poorly understood³⁵⁹. As variation in N-glycosylation would alter the surface of the protein produced this could have a large effect on crystal packing.

Unfortunately, despite an increasing interest in X-ray crystallography for determining protein structures, predicting crystallization outcomes is still difficult. The “scatter gun” approach of sparse matrix screening is still widely used to determine conditions for crystal growth and the only real test of crystal diffracting power comes when exposing the crystal to X-rays. More targeted approaches have been suggested but have yet to be widely implemented due to technical complexity⁴²¹. Presently, the sparse matrix screening approach used here would be difficult to improve upon.

Another potential avenue for exploration is serial crystallography. Serial crystallography involves collecting data from many individual microcrystals to produce a full data set⁴²². This type of experiment is difficult to carry out using traditional synchrotron beamlines because radiation damage occurs quickly in small crystals, particularly at room temperature⁴²³. This can be avoided by using recently developed, X-ray free electron lasers (XFELs)⁴²⁴. XFEL beamlines produce high intensity X-ray pulses on a femtosecond timescale. The intensity and speed at which X-ray pulses are produced allows diffraction data to be collected from a crystal before radiation damage occurs and limits the effects of atomic vibrations due to temperature. One issue with this approach is that, due to the intensity of the beam, microcrystals are destroyed almost instantly so that only a single frame of data can be obtained each individual crystal. This problem is solved by delivering a constant supply of new crystals into the beam while the detector captures diffraction images as quickly as is possible⁴²⁵. While many of the images captured will be blank, each time a crystal passes through

the beam, a frame of diffraction data in a random orientation will be captured⁴²⁶. Given enough time and an ample supply of microcrystals, a sufficient number of images can be captured to produce a full data set. Software has been developed to identify images containing diffraction data and then scale and index the data so that it can be used to solve the protein structure in a similar manner to data from a single crystal⁴²⁶.

Showers of microcrystals were commonly produced in the NATPO crystal screens. These were too small for single crystal diffraction experiments and therefore not used in this work. These crystals could provide a good starting material for serial crystallography experiments. It would also be possible to re-examine the screening results generated here and optimise conditions towards producing more microcrystals, rather than large, single crystals. These smaller crystals may have different growth habits than the larger crystals and therefore produce better diffraction data. XFEL facilities require major investment and there are currently only five operating worldwide⁴²⁷. However, it has been recognised that XFELs allow unique experiments to be performed in several fields besides structural biology and as such there are plans for the expansion of some existing facilities, along with plans to commission new XFELs in the future.

While X-ray crystallography was judged to currently be the technique most likely to succeed, this may change with future technological developments. The reasons justifying the use of X-ray crystallography are discussed in Chapter 3 with the conclusion that, at the present, this was the most viable method. Cryo-EM was one of the alternative techniques discussed. This method of protein imaging is developing rapidly and producing a structure of TPO by this technique in the near future could be feasible, if further crystallography continues to prove unsuccessful.

Currently, cryo-EM is only applicable to molecules larger than 100 kDa, although the lower size limit of what can be resolved to atomic resolution is decreasing⁴²⁸. It has been theorised that it could be possible to image molecules as small as 38 kDa at a resolution of 3 Å using this technique⁴²⁹. An exciting recent example is the 3.2 Å structure of haemoglobin, a 64 kDa protein,

published in 2017⁴³⁰. In this case the inherent symmetry of haemoglobin was utilised when refining the structure, although a 3.4 Å structure could be produced by treating the same data as asymmetric.

The binding of a F(ab) molecule to TPO could be helpful for cryo-EM studies as the complex produced would be larger than the protein alone. This would increase the likelihood of obtaining atomic resolution data⁴³¹. This approach would also help further characterise the nature of autoantibody binding to TPO. However, during cryo-EM, as with X-ray crystallography, the final image is produced from an average of thousands of images of many individual molecules⁴³². Therefore, some of the features of TPO which are suspected of limiting the diffraction resolution (variable glycosylation and loops or domains in multiple conformations) could also have a detrimental effect on the resolution of a cryo-EM structure. This also means an anti-TPO IgG, which would produce a larger complex than a F(ab), is unlikely to be a suitable ligand due to the inherent flexibility of these molecules. As is the case with X-ray crystallography, it is not currently possible to predict the success of cryo-EM so experimentation would be needed to determine the suitability of this technique for use.

Outlook

While a high-resolution structure of human TPO was not generated during the course of this work, progress was made towards obtaining one. Using this work as a starting point and applying the techniques discussed above, could allow higher resolution data to be generated and the structure of the protein to be solved. While there are still many promising methods left to explore, it cannot be said with any certainty that a structure of TPO can or will be generated using X-ray crystallography. If this proves to be the case then encouraging results such as the previously mentioned cryo-EM structure of haemoglobin, may indicate that the structure of TPO could be solved in the near future using this alternative technique⁴³⁰.

Despite the difficulty in obtaining a structure of the TPO protein this is still a subject worthy of further attention. An accurate structure of the TPO domains involved in autoimmunity could be used to gain a greater understanding of this

aspect of AITD at a molecular level. This could also act as a starting point for developing a structure of the intact protein and answer questions such as those raised by Le *et al* who produced two models of TPO, each with an alternative arrangement of the domains²⁵⁷. Both of these models were plausible and there is currently insufficient structural or biochemical knowledge of TPO to disprove either possibility. A TPO protein structure also has potential to improve the treatment of AITD. Current AITD drug treatments focus on easing the disease symptoms (Chapter 1, “Treatment of autoimmune disease”). An accurate structure of the protein would further the current knowledge of both the enzymatic mechanism and the nature of antibody binding, information which could be useful when developing improved treatments for both GD and HT which do not require surgical intervention.

Drug based therapies for GD involve inhibiting TPO activity with thiourea compounds. These treatments can cause a range of side-effects, many of which are serious⁸⁴. While the TPO inhibiting effect of these drugs can be easily demonstrated *in vitro*, the exact mechanism of inhibition and wider effects of these drugs are not fully understood⁴³³. A readily crystallisable form of the TPO protein would be a useful tool for further characterising existing drugs. A crystal structure of caprine LPO in complex with MMI, one of the commonly used TPO inhibiting drugs, has been produced but without a structure of TPO for comparison, conclusions must be drawn with care³⁸². Once current treatments are better understood this information could then be used when designing new compounds, which may be more effective or carry less adverse side effects.

As well as being markers for AITD, TPO autoantibodies have been shown to play a role in thyroid destruction by antibody dependent cell cytotoxicity and complement fixation^{111,434}. While this is not the only mechanism of thyroid damage observed in HT it is one potential route by which disease symptoms could be alleviated. Recently, work into several separate diseases have examined the use of small molecules such as peptides to block protein-protein interactions *in vivo*, and the possible use of these as therapeutics^{435–438}. A similar approach may yield new treatments for HT. The EP1 peptide identified here (Chapter 4) could be a good starting point for studies into this. Further work would focus on more closely determining the key amino acids required to

bind patient sera antibodies effectively and the reasons behind the small discrepancies observed between studies. Once a better understanding was gained then the use of small sections of the TPO protein sequence as autoantibody inhibitors could be investigated.

This is another area of research where a TPO protein structure would hasten progress. Despite the multitude of investigations into TPO autoantibody binding the pertinent amino acids are still not clearly established. A crystal structure of TPO in complex with an autoantibody would help clarify this issue particularly if multiple structures with different antibodies could be generated.

Chapter 8 - Materials and Methods

DNA cloning

The cDNA for TPO amino acids 1-833 in pUC18 was obtained from RSR Ltd, U.K. Further modifications were introduced by site directed mutagenesis using a Quikchange II site-directed mutagenesis kit (Agilent Technologies, U.K) following the manufacturer's instructions. All primers were obtained from Sigma-Aldrich, U.K. Briefly, a PCR reaction was run containing 10x reaction buffer (5 µl), template DNA (the gene of interest in a suitable vector plasmid, 50 ng), forward and reverse primer (125 ng each), deoxyribose nucleoside triphosphate (dNTP) mix (1 µl) and sterile H₂O (to 50 µl). PfuUltra HF DNA polymerase (2.5 U) was added last prior to thermocycling. Reactions were heated to 95 °C for 1 min followed by 18 cycles of 95 °C for 30 seconds, 55 °C for 1 minute and 68 °C for 1 min/kbp plasmid length. Following PCR cycling, reactions were digested with *Dpn*I (10 U) at 37 °C for 1 hour.

XL 1-Blue *E. coli* supercompetent cells were incubated with digested PCR product (1 µL) on ice for 30 minutes. Cells were transformed by heat shocking in a water filled heating block at 42 °C for 45 seconds. Cells were removed and incubated on wet ice at 0 °C for 2 minutes. Room temperature (18-25 °C) LB media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride) was added to the transformed cells to a final volume of 0.5 mL and the suspension was incubated at 37 °C for 4 hours with shaking at 220 rpm. Transformed cells were plated out on LB agar (LB media + 20 g/L agar) containing ampicillin (100 µg/mL) and incubated 37 °C, overnight.

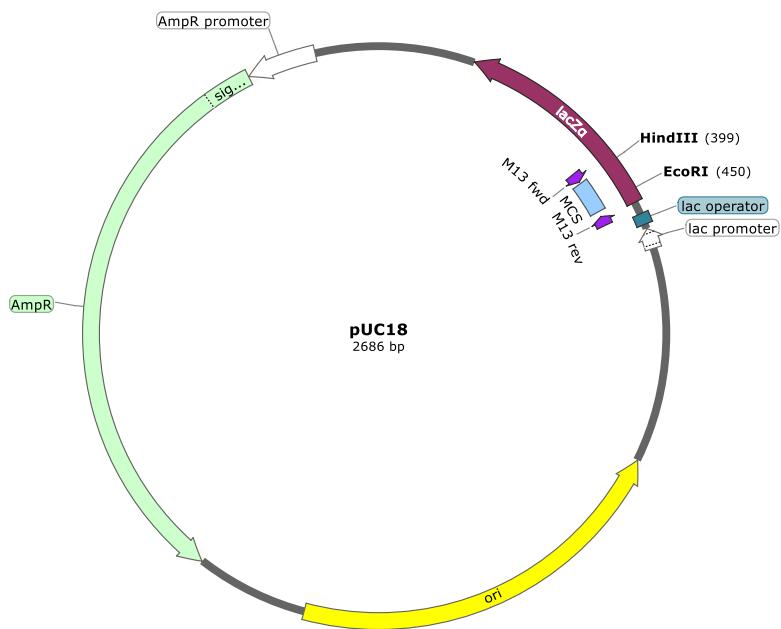


Figure 85- Map of pUC18 plasmid. TPO aa 1-833 sequence positioned between *Eco*RI and *Hind*III RE recognition sequences.

Other cloning work was carried out in *E. coli* strains NM522 or JM110. Chemically competent cells from these strains were freshly prepared for each use. A 100 mL culture of the required strain in LB media was grown until absorbance at 600 nm was ~0.35 (all absorbance measurements were made in a UV-1280 spectrophotometer (Shimadzu, Japan) with a 1 cm path length unless stated). Cells were pelleted by centrifugation at 2000 x g for 30 minutes at 4 °C and washed with 60 mL cold magnesium chloride (80 mM) and calcium chloride (20 mM) solution. Cells were resuspended in cold calcium chloride solution (4 mL, 20 mM) and incubated on ice for 3 hours. Heat shock transformations were carried out as above (200 µL competent cell suspension with 1-5 ng DNA).

Single, well defined colonies of transformed cells were picked from selective agar plates and used to seed LB media (3 mL) containing the appropriate selective antibiotics. These starter cultures were grown 37 °C for ~7 hours with shaking at 220 rpm and were then used to seed midi cultures (100 mL) which were grown overnight under the same conditions.

DNA was purified from *E. coli* cells using a plasmid midi kit (Qiagen, UK) following the manufacturer's instructions. Briefly, cells were harvested from overnight midi cultures by centrifugation at 2000 x g for 30 minutes at 4 °C.

Pellets were resuspended in buffer P1 (4 mL) and lysed with buffer P2 (4 mL) for 5 min. Genomic DNA and other cellular material was precipitated by adding buffer P3 (4 mL) and incubating on ice for 15 minutes. Samples were centrifuged at 20217 x g for 45 minutes at 4 °C and the supernatant applied to a gravity fed Qiagen-tip 100 column. Columns were washed with buffer QC (2 x 10 mL) and eluted with buffer QF (5 mL) into glass centrifuge tubes. Isopropanol (3.5 mL) was added to the eluate and the resulting mixture was centrifuged 9690 x g for 30 minutes at 4 °C. The supernatant was removed and pelleted DNA was washed with ethanol (70% v/v, 2mL). DNA was centrifuged at 9690 x g for 10 minutes at 4 °C and supernatant removed. DNA pellet was air dried and resuspended in sterile H₂O (500 µL). DNA concentrations were determined from absorbance at 260 nm of a 1 in 20 dilution in HPLC grade H₂O (water purified by a Labstar Hypol system, Wychwood water systems, UK) using the equation;

$$\text{Concentration } (\mu\text{g/mL}) = \text{absorbance at } 260 \text{ nm (1/20 dilution)} \times 1000$$

Purity was determined by the ratio absorbance at 260 and 280 nm ratio with 1.7-2 deemed acceptable.

RE enzymes were obtained from Roche, UK or New England Biolabs, UK. All RE digests were carried out in 20 µL volumes for analytical digests or 40 µL for preparative digests using supplied buffers. RE digests were incubated at 37 °C for 3 hours. Agarose gels were prepared by adding agarose (3 g) to TAE buffer (40 mM Tris-HCl, 20 mM acetic acid, 1 mM ethylenediaminetetraacetic acid, 300 mL) and heating in a microwave oven to dissolve the solids. Once cooled to below ~50 °C, GelRed nucleic acid stain (30 µL, Cambridge Bioscience, UK) was added and the solution was poured into a suitable mould. RE digests were mixed with 1/6 total volume of 6x DNA loading buffer (10 mM Tris-HCl pH 7.6, 60% (v/v) glycerol, 0.5% (w/v) Orange G) and loaded into gel. Gels were run at a constant 150 V for 1 hour at RT and visualised on a ChemiDoc XRS+ imaging system (Bio-Rad, USA).

DNA bands from preparative gels were excised with a sterile scalpel and DNA was recovered from the gel using a Qiaquick gel extraction kit (Qiagen, UK) following the manufacturer's instructions. Briefly, the gel slice was dissolved in

buffer QG (3x w/v) at 50 °C. Isopropanol (equal w/v) was added and mixed before applying to a QIAquick spin column. Columns were centrifuged at 12470x g for 1 minute to load sample and for each wash. Column were then washed with buffer QG (500 µL) followed by buffer PE (750 µL). DNA was eluted with sterile H₂O (50 µL).

DNA ligations were carried out by incubating equimolar concentrations of insert and vector DNA with T4 DNA ligase (1 µL, Roche, UK), 10 x ligation buffer (2 µL, supplied with enzyme) in sterile H₂O (to 20 µL). Ligation reactions were incubated in the dark at 16 °C overnight.

PCR was used to amplify some sequences for further cloning. Reactions contained template DNA (656 ng), Pfx DNA polymerase (2 U, Invitrogen, UK), 10 x reaction buffer (10 µL), 10 x enhancer solution (10 µL, reaction buffer and enhancer solution supplied with enzyme), forward and reverse primers (0.3 µM each), dNTP mixture (0.3 mM), magnesium sulfate (1.25 mM) and sterile H₂O (to 100 µL). Reactions were incubated for 5 minutes at 94 °C followed by 30 cycles of 94 °C for 15 seconds, 55 °C for 30 seconds and 68 °C for 2.5 minutes in a Geneamp PCR system 2700 (Applied Biosystems, UK).

DNA linker sequences were formed by annealing equimolar concentrations of appropriate oligonucleotides. Oligonucleotides were mixed and heated to 95 °C for 5 minutes. The mixture was then allowed to cool for 1 hour then stored on ice before use the same day.

Synthetic genes were obtained from Life Technologies, UK. All DNA constructs were sequenced by automated Sanger sequencing carried out by Source Bioscience, UK⁴³⁹.

Recombinant baculovirus production

The Bac-to-Bac system (Invitrogen, UK) was used to produce recombinant baculoviruses for the expression of each protein. Bacmid DNA was produced by transforming DH10Bac *E. coli* with a pFastBac1 vector containing the gene of interest. Chemically competent DH10Bac were incubated with pFastBac DNA (1 ng) and transformed by heat shock as above. LB media was added to 1 mL

after heat shocking and the cells incubated at 37 °C for 4 hours with shaking at 220 rpm. Transformed cells were serially diluted to 10⁻³ in LB media and plated out on LB agar containing kanamycin (50 µg/mL), gentamicin (7 µg/mL), tetracycline (10 µg/mL), isopropyl-β-D-thiogalactoside (40 µg/mL) and X-gal (100 µg/mL). Blue-white selection was used to determine colonies in which the gene of interest has been successfully transposed into the bacmid using the mini-Tn7 transposon. After 48 hours single, well defined, white colonies were amplified and bacmid DNA purified using the protocols described previously.

Recombinant baculoviruses carrying the gene of interest under the polyhedrin promoter were produced and amplified in SF9 (*Spodoptera frugiperda*) cells. Logarithmic phase cells in TC100 media (Gibco, UK) supplemented with foetal bovine serum (10% v/v), gentamicin (7 µg/mL), potassium iodide (0.1 mM) were used to seed 6-well culture plates. Plates were incubated for 1 hour at 27 °C to allow cells to attach. Cellfectin (Invitrogen, UK) was mixed with bacmid DNA (1 µg) in unsupplemented TC100 media (200 µL). A separate mixture was prepared for each well transfected. Cells were washed twice with unsupplemented TC100, then overlayed with the transfection mix plus unsupplemented TC100 media (800 µL) and incubated at 27 °C for 5 hours at which point the transfection mix was removed and replaced with supplemented TC100 (2 mL). Cells were incubated at 27 °C for 72 hours. P1 viruses were harvested by removing the media from the plates with a sterile pipette. Each well was considered a separate P1 virus.

P1 viruses were amplified in SF9 cells to produce P2 and, if required, P3 viruses. P2 viruses were generated in spinner flask cultures (50 mL). Cultures were set up at 2x10⁶ cells/mL in supplemented TC100 and seeded with an entire P1 (2 mL) virus sample. P2 cultures were incubated at 27 °C for 14 days with stirring at 60 rpm. Amplified virus was harvested by centrifugation 491 x g at R.T for 5 min. The supernatant was removed and used for protein expression and P3 virus production. All virus preparations were stored 4 °C, protected from light.

If further amplification was required, P3 virus cultures were set up and grown as P2 cultures on a 500 mL scale. Cultures were inoculated with P2 viruses to give

a multiplicity of infectivity (MOI) of 0.1 and grown for 14 days as above. P3 viruses were harvested in the same manner as P2 virus.

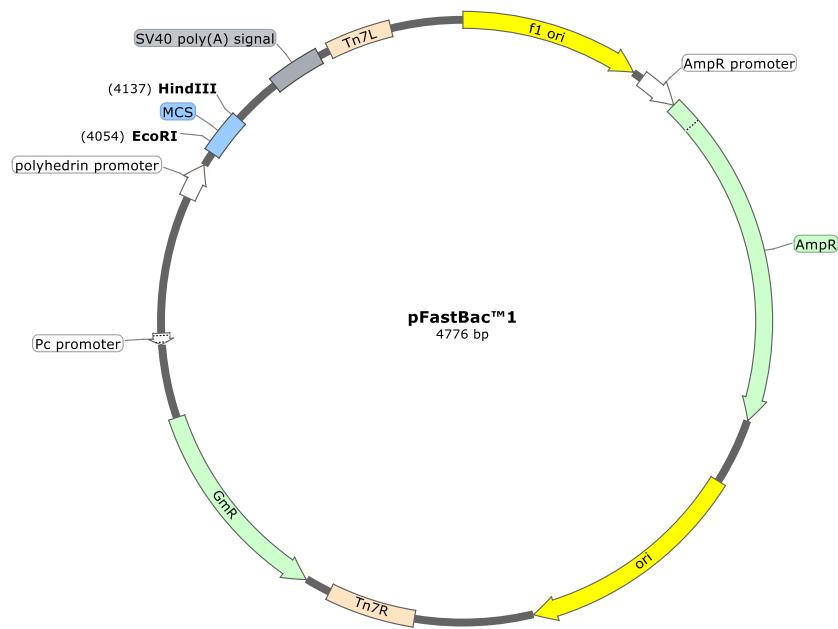


Figure 86- Map of pFastBac1 plasmid. Genes for expression were positioned between *EcoRI* and *HindIII* RE recognition sequences.

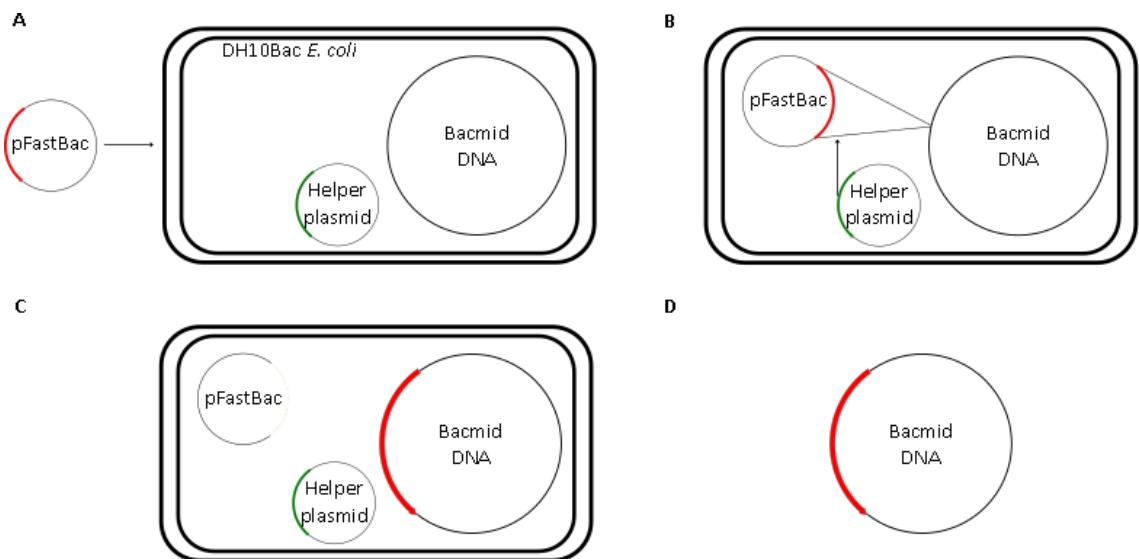


Figure 87- Diagram of bacmid DNA production process.⁴⁴⁰ **A)** DH10Bac *E. coli* are transformed with a pFastBac vector containing the gene of interest (red). **B)** The helper plasmid expresses recombinases (green) which transpose gene of interest into the bacmid DNA. **D)** Transformed DH10Bac cells are grown up in selective media and bacmid DNA containing the gene of interest purified.

Virus titration was carried out by a quantitative PCR (qPCR) based assay developed for in-house use by RSR Ltd, UK. DNA was extracted from a virus sample (200 µL) using a QIAamp MinElute virus spin kit (Qiagen, UK) following

the manufacturer's instructions. Briefly, the virus was mixed with an equal volume buffer AL and protease (25 µL) and mixed by vortexing for 15 seconds. Samples were incubated at 56 °C for 30 min before the addition of ethanol (250 µL) and further incubation for 5 min at R.T. Samples were applied to a QIAamp MinElute column by centrifugation (6000 x g for 1 minute at RT. The column was washed with buffer AW1, buffer AW2 and finally ethanol (each 500 µL) before drying at 56 °C for 3 minutes with the lid open. DNA was eluted from column with sterile H₂O (50 µL).

Extracted DNA was diluted in a 4-fold series to 1 in 256 in sterile H₂O. Dilutions were mixed with primer stock (5 µL, 800 nM, containing forward and reverse virus titration primers) and power SYBR green mastermix (10 µL, Applied Biosystems, UK) in a 96-well V-bottomed q-PCR plate. Each dilution was tested in duplicate. PCR was carried out in a StepOnePlus qPCR machine (Applied biosystems, UK). PCR reactions were incubated for 95 °C 10 minutes followed by 40 cycles of 95 °C 15 seconds, 60 °C 1 minute with fluorescence measured after each cycle. Virus titres were determined using a standard curve produced with DNA from a virus of known titre (determined previously by plaque assay, supplied by RSR Ltd) extracted and amplified alongside test samples.

Protein expression

All protein expression was carried out using High Five (*Trichoplusia ni*) cells. Stock cultures for maintaining the cell line were grown in Insect Express media (400 mL, Lonza, UK) supplemented with gentamicin (7 µg/mL), potassium iodide (0.1 mM) in Erlenmeyer flasks (1 L capacity). Cultures were seeded with 2 x 10⁸ cells and incubated at 27 °C with shaking at 110 rpm. Cultures were passaged every 3/4 days to maintain logarithmically growing stocks.

Expression cultures were seeded as above and grown for 3 days before being resuspended in fresh media (2 L) in Fernbach flasks (3 L capacity). The following day, the media was replaced with fresh media and the cell density was measured by counting cells manually using a haemocytometer. Cultures were infected with a MOI of 0.096 and incubated at 23 °C for 137 hours with shaking at 90 rpm. Cultures for protein expression were supplemented with δ-

aminolevulinic acid (50 µg/mL) prior to infection, and again 70 hours post infection to facilitate haem production.

Expressed recombinant protein was harvested from culture supernatant by centrifugation at 558 x g for 10 min at RT. Protease inhibitor cocktail tablets (Roche, UK) were dissolved in H₂O (50 mL per tablet) and added (10 mL/L culture) prior to storage at -80 °C.

Protein purification

Proteins harvested from culture supernatants were purified by three rounds of chromatography. Initial capture was by fluidized bed ion exchange chromatography carried out in a Streamline 100 column packed with streamline DEAE matrix (GE healthcare, UK). Culture supernatant was diluted in reverse osmosis water (1 in 3, purified by a Labstar 90 reverse osmosis system, Wychwood water systems, UK) and the pH adjusted to 7.5 with Tris base (2 M). Diluted supernatant was loaded onto the column at 400 mL/min. During loading the flow rate was adjusted to maintain a matrix expansion level between 2 and 3 x the settled matrix height. Unbound material was washed off with streamline buffer (50 mM sodium chloride, 5 mM Tris-HCl pH 7.5, 0.1 mM potassium iodide) until absorbance at 280 nm was negligible. The matrix was then collapsed for further washing and elution as a packed bed. Protein was eluted in a single step with streamline elution buffer (400 mM sodium chloride, 20 mM Tris-HCl pH 7.6, 0.1 mM potassium iodide). Up to 24 400 mL fractions were collected by hand. A guaiacol oxidation assay was preformed to identify fractions containing the protein of interest (“Protein analysis”). All fractions with enzymatic activity were pooled for further purification.

IMAC purification was carried out on an Äkta chromatography system (GE Healthcare, UK). Chelating Sepharose Fast Flow (58 ml, GE Healthcare, UK) was packed in an adjustable length column adapter (26 mm diameter) and charged with nickel sulfate (200 mM, 1.5 column volumes (CV)). Excess nickel sulfate was washed off with HPLC grade water and IMAC running buffers (IMAC buffer A- 500 mM sodium chloride, 20 mM Tris-HCl pH 7.6, IMAC buffer B- 500 mM sodium chloride, 20 mM Tris-HCl pH 7.6, 500 mM imidazole).

Streamline elution pools were adjusted with imidazole (to 20 mM) and sodium chloride (to approximately 500 mM) and passed through a 0.2 µm filter. The column was loaded at a suitable flow rate for overnight loading at R.T. Once loaded, the column was washed with 4% IMAC buffer B (20 mM imidazole) at 15 ml/min until absorbance at 280 nm was negligible. Proteins were eluted by applying a gradient of increasing IMAC buffer B concentration. The first step increased buffer B concentration to 10% (50mM imidazole) over 3CV. This concentration was held until absorbance at 280 nm was negligible at which point buffer B concentration was increased to 30% (150 mM imidazole) over 3 CV. Fractions (10 mL) were collected automatically when absorbance at 280 nm reached \leq 20 mAU (over a 0.2 cm path length). Peak fractions were immediately pooled and dialysed into TPO buffer (50 mM sodium chloride, 20 mM Tris-HCl pH 7.6, 0.1 mM potassium iodide).

SEC was carried out on an Äkta chromatography system using Superdex 200 prep grade matrix (GE Healthcare, UK) packed in an adjustable length column adaptor (50 mm diameter). Total column volume was 1870 mL. Post-IMAC pools were concentrated to \leq 2% of a column volume in a nitrogen pressurised ultrafiltration cell at 4 °C and allowed to de-gas at RT prior to loading. Material was loaded onto the column at 5 mL/min in TPO buffer. Fractions (5 mL) were collected automatically when absorbance at 280 nm reached \leq 10 mAU (over a 0.2 cm path length). Each individual fraction was sampled for analysis using techniques described in “Protein analysis”.

Final protein preparations were produced by pooling and concentrating selected SEC fractions. Selected fractions contained \leq 2.5% aggregate by analytical SEC, were enzymatically active by guaiacol assay and appeared intact by SDS-PAGE. Pooled fractions were concentrated in a nitrogen pressurised ultrafiltration cell at 4 °C until the concentration of protein was \geq 10 mg/mL.

Post-IMAC NΔTPO was used to test the effect of incubating the protein with hydrogen peroxide and hemin (both Sigma-Aldrich, UK). Hydrogen peroxide (to 10 µM), hemin (to 600 µM) or both were added to NΔTPO (200 µg) in TPO buffer and incubated at 37 °C for 3 hours with shaking. After incubation samples were dialyzed back into TPO buffer in a Vivaspin 6 centrifugal concentrator

(Sigma-Aldrich, UK). Analysis was carried out as detailed below for final pool samples by guaiacol oxidation assay and measurement of the optical absorbance at 412/280nm.

Protein analysis

Peroxidase activity was assessed using a guaiacol oxidation assay. A protein sample (100 µL) and guaiacol solution, (2 mL, 40 mM guaiacol, 10mM Tris-HCl pH 8.0) were added to a 1 cm path length plastic cuvette and incubated at 37 °C for 20 minutes. The cuvette was placed in a spectrophotometer, blanked at 470 nm and hydrogen peroxide (10 µL, 60 mM) added. The test sample was mixed and a 1 minute timer started immediately. The absorbance was recorded at 15, 30 and 60 seconds. Samples were diluted in guaiacol solution or TPO buffer to give an absorbance at 470 nm of between 0.3-0.8 at 30 seconds. Absorbance values were converted to guaiacol units (GU) for comparison. The absorbance at 470 nm after 15 seconds was multiplied by 4, to give a 60 seconds value free from any reaction quenching effects. This value was then multiplied by 10 to give the value for 1 mL and finally multiplied by the dilution factor, if applicable, to convert to guaiacol units/mL (GU/mL). This was divided by sample concentration to give the specific enzymatic activity (GU/mg).

Protein concentration for impure (pre-IMAC) samples was assessed by inhibition RIA utilizing [¹²⁵I]-labelled TPO (RSR Ltd, UK). Samples were diluted to a suitable level in RIA buffer (150 mM sodium chloride, 10 mM Tris-HCl pH 7.5, 5 mg/mL BSA, 7.7 mM sodium azide, 0.1% (v/v) Tween 20). Test samples (100 µL) along with standards consisting of known concentrations of rhTPO (3, 10, 30, 100, 300, 1000 ng/mL) were incubated with a rabbit polyclonal anti-TPO sera (100 µL) at 37 °C for 15 minutes. [¹²⁵I]-TPO (100 µL) was added and incubated for 30 minutes at R.T followed by Protein A suspension (50 µL) and further incubation for 1 hour at R.T. An excess of cold RIA buffer (1 mL) was added to each sample followed by centrifugation at 2000 x g for 30 minutes at 4 °C. Supernatant was aspirated from each sample and total counts over 1 minute measured in a LB2111 γ-counter (Berthold Technologies, UK). CPM were converted to % bound using the equation:

$$\% \text{ bound} = (\text{CPM test sample}/\text{CPM } [^{125}\text{I}]\text{-tracer only +ve control}) \times 100$$

Results for standards in % bound were plotted on a linear-logarithmic scale from which the concentration of test samples was determined. Concentrations were multiplied by the dilution factor to give the concentration of the neat samples. At least two different dilutions of each sample were tested in the same assay. If more than one was within the assay effective range (30-130 ng/mL) then an average was taken across each sample.

Protein concentration for semi-pure (post-IMAC) material was calculated from absorbance at 280 nm using the equation:

$$\text{Concentration} = \text{absorption at 280 nm} \times \text{extinction coefficient} \times \text{path length}$$

Extinction coefficients were estimated from the amino acid sequence using ProtParam³²⁰. A value for haem incorporation was determined by dividing absorbance at 412 nm by absorbance at 280 nm. This 412/280 ratio was used to compare haem incorporation between proteins.

Western blotting followed probing with a rabbit polyclonal anti-TPO sera (RSR Ltd, UK) was used to assess protein integrity/degradation in impure (pre-IMAC) samples. Samples were run overnight at a constant 45 V at R.T on an 8% acrylamide SDS-PAGE gel. SDS-PAGE gels were cast in 18 x 16.5 x 0.75 cm moulds as follows:

Resolving gel- 6.9 mL HPLC grade H₂O, 4 mL 30% (w/v) acrylamide (37.5:1 acrylamide:bis-acrylamide), 3.8 mL 1.5 M Tris-HCl pH 8.8, 150 µL 10% (w/v) sodium dodecyl sulfate, 150 µL 10% (w/v) ammonium persulfate, 9 µL tetramethylethylenediamine.

Stacking gel- 3.4 mL HPLC grade H₂O, 1.66 mL 30% (w/v) acrylamide (37.5:1 acrylamide:bis-acrylamide), 1.26 mL 1 M Tris-HCl pH 6.8, 50 µL 10% (w/v) sodium dodecyl sulfate, 50 µL 10% (w/v) ammonium persulfate, 5 µL tetramethylethylenediamine.

Up to 1 µg protein was loaded per well. Prior to loading SDS-PAGE loading buffer (1/3 sample volume, 17.4 % (v/v) glycerol, 4% (w/v) sodium dodecyl sulfate, 10 mM Tris-HCl pH 6.8, bromophenol blue added to achieve suitable

colour) and dithiothreitol (100 mM in 2 M Tris-HCl pH 8.3 added to 11.1 mM) were added and the sample was incubated at 100 °C for 4 minutes. After separation, proteins were transferred to a nitrocellulose membrane by blotting for 3 hours at a constant 55 V at 4 °C in blotting buffer (192 mM glycine, 25 mM Tris-HCl pH 8.3). The membrane was washed with phosphate buffered saline plus 0.5% tween 20 (PBS+T) and blocked with polyvinyl alcohol (100 µg/mL). The membrane was then incubated with polyclonal anti-TPO rabbit sera for 1 hour at 37 °C, followed by another washing step. A horseradish peroxidase conjugated goat anti-rabbit IgG (Thermo, UK) was used as the secondary antibody. Incubation was for 1 hour at RT and was followed by a final wash step. The blot was visualised using a chemiluminescent substrate for horseradish peroxidase on a ChemiDoc XRS+ imaging system.

Post-IMAC samples were visualised using the same SDS-PAGE protocol followed by staining with SimplyBlue SafeStain (Invitrogen, UK). For stained gels, 5 µg protein was loaded per well, prepared in the same manner as samples for Western blotting. Gels were washed with water for 3x 5 minutes. Gels were stained for 1 hour and destained for 2x 1 hour with water at RT with gentle shaking. Gel images were captured on a ChemiDoc XRS+ imaging system.

Isoelectric focusing (IEF) was carried out on pre-cast Novex IEF gels (Gels and all buffers supplied by Invitrogen, UK) with an effective pH range of 3-10. Final pool samples were diluted in HPLC water (to 2 mg/mL, 10 µg protein total) and loaded onto the gel in an equal volume of sample buffer. Gels were run for 1 hour at a constant 100 V, 1 hour at a constant 200 V then for 30 minutes at a constant 500 V. Gels were stained for 30 min with IEF stain (27% (v/v) isopropanol, 10% (v/v) acetic acid, 0.4 g/L Coomassie brilliant blue R, 0.5 g/L crocein scarlet 7B) and destained with IEF destaining solution (40% (v/v) ethanol, 10% (v/v) acetic acid) until sufficiently clear. Gel images were captured on a ChemiDoc XRS+ imaging system.

Thermal shift assays were carried out in a StepOnePlus qPCR machine using a protein thermal shift dye kit (both Applied Biosystems, UK). All proteins tested including controls were from final pools prepared from post-SEC fractions.

Samples were diluted to 0.6 mg/mL in TPO buffer and mixed with thermal shift dye in a 96-well V-bottomed q-PCR plate. Each well contained the test sample (12.5 µL), and thermal shift dye (7.5 µL). A melt curve program was run from 25 °C to 95 °C with a ramp rate of 1.2 °C/min. The first derivative of the raw fluorescence data was calculated using the supplied StepOne software with the temperature of the first derivative peak taken as the melting temperature for comparisons between proteins.

Patient sera and monoclonal antibody binding was assessed by RIA. For each assay a dilution series of the test protein was prepared from 160 ng/mL to 1.25 ng/mL in RIA buffer. 2G4 IgG (50 µL, 21 ng/mL) or a pool of TPO Ab positive patient sera (50 µL) was added to each sample or standard (50 µL) and incubated for 1 hour at R.T. [¹²⁵I]-labelled TPO (50 µL) was added and incubated for 1 hour at R.T. The assay was then completed as the inhibition RIA described previously. CPM data was converted to % bound as described previously and plotted on a linear-logarithmic scale to assess binding activity. Relative % binding was calculated by setting the % bound value of the assay negative control (buffer only) as 100% and calculating the percentage of this value each sample represents.

Crystallization screening

The sparse matrix crystal screens Structure Screens 1&2, JCSG+ and PACT Premier were purchased from Molecular Dimensions, U.K. Optimisation screens were prepared from concentrated stock solutions. Concentrated (10x) stock solutions of pH adjusted buffers or salts were prepared in HPLC grade water (water purified by a Labstar Hypol system, Wychwood Water Systems, UK). Tris buffers were made by adjusting Tris base to the appropriate pH with hydrochloric acid. PEG stock solutions were prepared at 30% w/v in HPLC grade water and diluted to give the required concentrations. All stock solutions were 0.2 µM filtered prior to use.

Crystal screens were set up by hand or using a Crystal Phoenix robot (Art Robbins Instruments, USA). Some screening (as specified in Chapter 3

“Results and discussion”), was carried out by Charles River Laboratories, UK using a Mosquito crystallography robot (Art Robbins Instruments, USA).

Crystal screens set up by hand were in the hanging drop format in 96-well screw top crystallization plates (Molecular Dimensions, UK). Plates were incubated in a temperature-controlled room at 22 °C. Hanging drops were generally set at 2 µL sizes, but in some cases up to 4 µL was used. Drops consisted of 1:1 protein to precipitant unless otherwise stated, and were suspended over 100 µL precipitant.

Crystal screens set up by a robot used 96-well sitting drop or hanging drop format. Drops were 400 nL with a 60 µL reservoir of precipitant. The protein to precipitant ratio was 1:1 unless stated. Plates set up by robot were incubated at 19 °C in a dedicated incubator.

Crystal plates were monitored by microscope regularly for at least 3 months after set up. Plates were still checked after this time, albeit less frequently. No crystal plates were discarded until all conditions on the plate were completely desiccated.

Inhibitors used for co-crystallization experiments were obtained from Sigma-Aldrich, UK. With the exception of methimazole (which was soluble in water), all inhibitors were dissolved in ethanol to produce a concentrated stock solution. This was added to the precipitant to give a final concentration of 20 molar equivalents of quantity of protein used. When a 1:1 ratio crystallography drop was set up this gave a 10x molar excess of inhibitor relative to the protein in the drop.

X-ray diffraction experiments

Crystals for X-ray diffraction studies were harvested from plates in SPINE standard pins and flash frozen in liquid nitrogen. Prior to harvesting, ethylene glycol or glycerol was added as a cryoprotectant if required. When used, the cryoprotectant was added to the precipitant reservoir, then the drop, to give a final concentration in the drop of 25-13%.

X-ray diffraction data collection was carried out on the I04 and I24 beamlines at Diamond light source, UK. When screening crystals, three test images were taken 90° apart to estimate the diffracting potential of each crystal. Full data sets of 2000 images over 200° total oscillation were taken from diffracting crystals selected based on results of the initial screening images. Exposure time was 0.2 seconds per image at a wavelength 0.9795 Å.

All processing of X-ray diffraction data was carried out using programs contained within the Collaborative Computational Project Number 4 (CCP4) software suite³⁴⁸.

Deglycosylation

Trial protein deglycosylation reactions were carried out using endoglycosidase H (21.5-86.0 mU/mg, Roche, Germany) and endoglycosidase F3 (14.25-57.00 mU/mg, Calbiochem, USA). Units were as defined by the manufacturers and each endoglycosidase was obtained from a single source throughout this work. Reactions were incubated at 20 °C for 120 hours in deglycosylation buffer (15 mM sodium chloride, 50 mM sodium acetate pH 5.3). Preparative protein deglycosylation reactions were carried out using endoglycosidase H (86 mU/mg) and endoglycosidase F3 (57 mU/mg). Reactions were incubated at 20 °C for 120 hours in deglycosylation buffer.

Post-deglycosylation samples were analysed by SDS-PAGE and lectin blotting. SDS-PAGE protocol and transfer to a nitrocellulose membrane was as described previously. Following transfer, the membrane was washed with Tris buffered saline plus tween 20 (0.5% v/v, TBS+T) and blocked with gelatine (30 mg/mL) in TBS+T. The membrane was then incubated with biotinylated *Galathus nivalis* lectin (80 µg) (Vector laboratories, U.K) with gelatine (10 mg/mL), and calcium chloride (1 mM) in TBS+T for 1 hour at 37 °C, followed by another washing step. A streptavidin-horseradish peroxidase conjugate (RSR Ltd, U.K) was used to facilitate visualisation. Incubation was 1 hour at R.T and was followed by a final wash step. The blot was visualised using a chemiluminescent substrate on a ChemiDoc XRS+ imaging system.

Deglycosylated material was purified by SEC as described previously (“Protein purification”). Protocols for determining protein concentration by optical absorbance, analytical SEC, guaiacol oxidation assay, pooled patient sera binding assay and 2G4 binding assay of deglycosylated proteins were also as described above (“Protein analysis”).

F(ab) complex production

4F5 IgG was obtained from RSR Ltd, U.K. IgG and F(ab) concentrations were determined from absorbance at 280 nm using the same method as for protein quantification (“Protein analysis”). 4F5 F(ab) was produced from 4F5 IgG by papain digestion. Trial digests were set up with IgG (850 µg per test condition) to determine suitable conditions for digestion. Papain (Sigma-Aldrich, UK) from Papaya latex was added to 4F5 IgG in a 100:1, 50:1 and 34:1 w/w IgG:papain ratio in a total volume of 130 µL. Digestion reactions were incubated at 37 °C in PBS plus L-cysteine (7.7 mM) and ethylenediaminetetraacetic acid (1.5 mM) with occasional mixing by inversion. Samples for analysis were taken every 1 hour for 3 hours. The reactions were stopped by the addition of iodoacetamide (to 45.5 mM) and incubation at R.T for 30 minutes. Trial digests were analysed by reduced and non-reduced SDS-PAGE on a 12% (v/v) gel as described previously (“Protein analysis”).

Preparative digests of 4F5 IgG (50 mg) were digested with papain (1.47 mg, 34:1 w/w ratio) for 3 hours as the trial reactions. Reactions were stopped by the addition of iodoacetamide as with the trial reactions.

4F5 F(ab) was purified by affinity chromatography using MabSelect matrix (GE, UK). Digested IgG was loaded onto a 2.2 mL volume column equilibrated with a sodium chloride (150 mM), glycine (1 M, pH 8.6) buffer. The flow-through containing the F(ab) was collected at the outlet. To ensure maximum recovery, 1 CV of buffer was run through the column following sample loading. During this phase the outlet flow through was also collected and pooled with the F(ab) fraction.

Preparative scale digests were also analysed by reduced and non-reduced SDS-PAGE as described previously (“Protein analysis”). Direct binding and inhibition assays were carried out using [¹²⁵I]-TPO, Protein A and RIA buffer as described previously (“Protein analysis”). For the direct binding assay, 4F5 F(ab) and IgG were diluted to a suitable level in RIA buffer. Diluted sample (50 µL) was added to 5 mL tubes containing a fixed amount of [¹²⁵I]-TPO and incubated for 1 hour at room temperature. The assay was then completed as the inhibition RIA and CPM data converted to % bound as described previously (“Protein analysis”). The % bound values were corrected for background signal by subtracting the % bound value of a buffer only negative control from the % bound value for each sample.

For the inhibition assay 4F5 F(ab) was diluted as above. Diluted samples (25 µL) were added to tubes (5 mL) with a fixed amount of [¹²⁵I]-TPO and incubated for 1 hour at room temperature. 4F5 F(ab) (25 µL, 263 ng/mL) was added and incubated for 1 hour at room temperature. The assay was then completed as the inhibition RIA and CPM data converted to % bound as described previously (“Protein analysis”). Relative % bound was calculated by assigning a buffer only negative control (no F(ab) present) as 100% binding and calculating all other samples as a percentage of this value.

When required, 4F5 F(ab) was purified by SEC using a scaled down version of the protocol described previously (“Protein purification”). Superdex 200 preparative grade matrix (GE Healthcare, UK) was packed in a 26 mm diameter column adaptor to give a column volume of 472 mL. Material was loaded onto the column at 2 mL/min in TPO buffer. Fractions were collected automatically as described previously (“Protein purification”).

The optimal TPO:F(ab) ratio for producing a complex was determined by setting up a series of mixtures at various ratios and incubating at 21 °C for 1 hour in TPO buffer with occasional mixing by inversion. IMAC purified NΔTPO was added to 4F5 F(ab) in a 1:1, 1:1.25, 1:1.5 and 1:1.75 TPO:F(ab) molar ratio to give a total volume of 250 µL. Following incubation, samples were analysed immediately by analytical SEC as described previously (“Protein analysis”).

NΔTPO:4F5 F(ab) complex was formed on a preparative scale by mixing post-IMAC NΔTPO and 4F5 F(ab) in a 1:1.25 molar ratio and incubating at 21 °C for 1 hour with occasional mixing by inversion. NΔTPO:4F5 F(ab) complex was purified by SEC using the same protocol as 4F5 F(ab). Fractions were selected based on the elution profile (absorbance at 280 nm), pooled immediately, and concentrated in a 10 kDa molecular weight cut-off centrifugal concentrator.

Analysis of NΔTPO:4F5 F(ab) complex by analytical SEC and SDS-PAGE was as described previously (“Protein analysis”). NΔTPO:4F5 F(ab) complex crystal screening was carried out using the same protocols as for the modified TPO proteins described previously.

Antibody selection

TPCK treated trypsin from bovine pancreas was purchased from Sigma-Aldrich, U.K. Trypsin (500 µg/mL) was resuspended in calcium chloride (20 mM), Tris-HCl (50 mM, pH 8). 4F5 and 2G4 IgG were incubated with trypsin in 100:1, 50:1 40:1 and 4:1 protein:trypsin weight ratios at 37 °C, overnight with end over end mixing. Samples of each test condition were analysed by analytical SEC and reduced SDS-PAGE as described previously (“Protein analysis”).

Antibody coupling to Sepharose

4F5 IgG (RSR Ltd, UK) was coupled to cyanogen bromide (CNBr) activated Sepharose matrix (G.E, UK) following the manufactures instructions. Briefly, matrix was prepared by washing the matrix (2 g) with hydrochloric acid (2 L, 1 mM). Washed matrix was incubated with 4F5 IgG (48.6 mg) for 1 hour in sodium chloride (500 mM), sodium carbonate/sodium hydrogen carbonate (100 mM, pH 8.3) buffer with end over end mixing. IgG coupled matrix was packed into an 16 mm diameter column adapter and washed with alternating acid (5 CV, 500 mM sodium chloride, 100 mM sodium acetate, pH 4) and alkali (5 CV, 500 mM sodium chloride, 100 mM Tris-HCl pH 8) buffers, three times each. Matrix was then washed with guanidium HCl (1 CV, 6 M) before being washed

back into alkali buffer plus sodium azide (0.2 g/L) for storage and unpacking from the column.

Epitope excision

4F5 Sepharose slurry was resuspended by gentle, end over end mixing. Once homogeneous, matrix slurry (1 mL) was removed from the main stock and washed into high ionic strength buffer (500 mM sodium chloride, 20 mM Tris-HCl pH 7.6) by diluting with buffer (to 5 mL), settling by centrifugation at 340 x g for 6 minutes at R.T and removing the supernatant with a pipette. This was repeated three times before adding rhTPO (800 µg) plus high ionic strength buffer (to 1 mL) and incubating at 37 °C for 2 hour with end over end mixing. The matrix was then washed three times, as above, with TPO buffer. Trypsin (80 µg) prepared as previously (“Antibody selection”) and TPO buffer (to 1 mL) were added and the matrix was incubated at 37 °C overnight (~16 hours) with end over end mixing. Following digestion, the matrix was washed three times as above with high ionic strength buffer to stop the trypsin digestion and remove any unbound peptides. Peptides were eluted with guanidinium HCl (3.5 M) or sodium citrate (100 mM, pH 3) by adding elution buffer to the matrix (to 1 mL) and incubating at 37 °C for 1 hour with end over end mixing before settling the matrix and removing the supernatant as previously. If sodium citrate was used for elution the supernatant was adjusted to approximately pH 7.5 with Tris-HCl (2 M, pH 8.3). Elution supernatants were concentrated by freeze drying and resuspension in HPLC grade water.

Liquid chromatography-mass spectrometry analysis

LC-MS was carried out on an Acquity H class ultra-performance liquid chromatography (UPLC) system coupled to a Synapt G2-Si mass spectrometer (both Waters, UK). An Acquity UPLC CSH C18 RP column (2.1 x 100 mm, 1.7 µm pore size, Waters, UK) was used for peptide separation in an oven maintained at 40 °C. Peptides were eluted with an increasing gradient of acetonitrile in water containing formic acid (0.1%). The gradient consisted of 1

minute held at 3% acetonitrile followed by a linear increase to 60% acetonitrile over 30 minutes. Eluting molecules were ionised by electrospray ionisation and analysed by tandem time-of-flight mass analysers. Data was collected in data independent acquisition mode with a scan range of 100-2500 m/z. Collision induced dissociation energy was ramped from 25 to 55 V. Analysis of data was carried out using the MassLynx and BioLynx software (Waters, UK). Monoisotopic masses were calculated using PeptideMass⁴⁰⁷. PPM error was calculated using the equation;

$$\text{Error (ppm)} = ((\text{Observed mass} - \text{calculated monoisotopic mass}) / \text{observed mass}) \times 10^6$$

References

1. R. Medzhitov and C. Janeway, *Immunol. Rev.*, 2000, **173**, 89–97.
2. J. A. Sanford and R. L. Gallo, *Semin. Immunol.*, 2013, **25**, 370–377.
3. A. Aderem and D. M. Underhill, *Annu. Rev. Immunol.*, 1999, **17**, 593–623.
4. J. L. Gowans and E. J. Knight, *Proc. R. Soc. Lond. B Biol. Sci.*, 1964, **159**, 257–282.
5. A. D. Luster, *Curr. Opin. Immunol.*, 2002, **14**, 129–135.
6. N. S. Merle, R. Noe, L. Halbwachs-Mecarelli, V. Fremeaux-Bacchi and L. T. Roumenina, *Front. Immunol.*, 2015, **6**, 257.
7. T. E. Hugli, *Springer Semin. Immunopathol.*, 1984, **7**, 193–219.
8. D. Bubeck, *Biochemistry*, 2014, **53**, 1908–1915.
9. H. Veiga-Fernandes, U. Walter, C. Bourgeois, A. McLean and B. Rocha, *Nat. Immunol.*, 2000, **1**, 47–53.
10. Q. Zhang, R. Iida, T. Yokota and P. W. Kincade, *Curr. Opin. Hematol.*, 2013, **20**, 265–272.
11. N. Hozumi and S. Tonegawa, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 3628–3632.
12. J. Neefjes, M. L. M. Jongsma, P. Paul and O. Bakke, *Nat. Rev. Immunol.*, 2011, **11**, 823–836.
13. V. Vasilevko, A. Ghochikyan, M. J. Holterman and M. G. Agadjanyan, *DNA Cell Biol.*, 2002, **21**, 137–149.
14. P. S. Linsley, J. L. Greene, P. Tan, J. Bradshaw, J. A. Ledbetter, C. Anasetti and N. K. Damle, *J. Exp. Med.*, 1992, **176**, 1595–1604.
15. W. Liao, J.-X. Lin and W. J. Leonard, *Curr. Opin. Immunol.*, 2011, **23**, 598–604.
16. N. Anikeeva and Y. Sykulev, *Immunol. Res.*, 2011, **51**, 183–194.
17. T. D. Randall, L. B. King and R. B. Corley, *Eur. J. Immunol.*, 1990, **20**, 1971–1979.
18. R. Maki, W. Roeder, A. Traunecker, C. Sidman, M. Wabl, W. Raschke and S. Tonegawa, *Cell*, 1981, **24**, 353–365.
19. P. K. Mongini, W. E. Paul and E. S. Metcalf, *J. Exp. Med.*, 1982, **155**, 884–902.
20. D. L. Delacroix, C. Dive, J. C. Rambaud and J. P. Vaerman, *Immunology*, 1982, **47**, 383–385.
21. W. E. Winter, N. S. Hardt and S. Fuhrman, *Arch. Pathol. Lab. Med.*, 2000, **124**, 1382–1385.
22. G. Vidarsson, G. Dekkers and T. Rispens, *Front. Immunol.*, 2014, **5**, 520.
23. A. Ferrante, L. J. Beard and R. G. Feldman, *Pediatr. Infect. Dis. J.*, 1990, **9**, 516–524.
24. T. Damelang, S. J. Rogerson, S. J. Kent and A. W. Chung, *Trends Immunol.*, 2019, **40**, 197–211.
25. R. C. Aalberse, S. O. Stapel, J. Schuurman and T. Rispens, *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.*, 2009, **39**, 469–477.

- 26.** R. L. Stanfield and I. A. Wilson, *Microbiol. Spectr.*, DOI:10.1128/microbiolspec.AID-0012-2013.
- 27.** D. R. Davies and G. H. Cohen, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 7–12.
- 28.** J. M. Dal Porto, S. B. Gauld, K. T. Merrell, D. Mills, A. E. Pugh-Bernard and J. Cambier, *Mol. Immunol.*, 2004, **41**, 599–613.
- 29.** Z. K. Indik, J. G. Park, S. Hunter and A. D. Schreiber, *Blood*, 1995, **86**, 4389–4399.
- 30.** T. R. Mosmann, H. Cherwinski, M. W. Bond, M. A. Giedlin and R. L. Coffman, *J. Immunol.*, 1986, **136**, 2348–2357.
- 31.** A. Lanzavecchia, *Nature*, 1985, **314**, 537–539.
- 32.** P. Garside, E. Ingulli, R. R. Merica, J. G. Johnson, R. J. Noelle and M. K. Jenkins, *Science*, 1998, **281**, 96–99.
- 33.** J. B. Hay, M. J. Murphy, B. Morris and M. C. Bessis, *Am. J. Pathol.*, 1972, **66**, 1–24.
- 34.** R. F. Coico, B. S. Bhogal and G. J. Thorbecke, *J. Immunol.*, 1983, **131**, 2254–2257.
- 35.** G. Teng and F. N. Papavasiliou, *Annu. Rev. Genet.*, 2007, **41**, 107–120.
- 36.** N. Maizels, *Annu. Rev. Genet.*, 2005, **39**, 23–46.
- 37.** M. Larijani and A. Martin, *Semin. Immunol.*, 2012, **24**, 255–263.
- 38.** M. Ziegner, G. Steinhauser and C. Berek, *Eur. J. Immunol.*, 1994, **24**, 2393–2400.
- 39.** P. Silacci, A. Mottet, V. Steimle, W. Reith and B. Mach, *J. Exp. Med.*, 1994, **180**, 1329–1336.
- 40.** I. Dogan, B. Bertocci, V. Vilmont, F. Delbos, J. Mégret, S. Storck, C.-A. Reynaud and J.-C. Weill, *Nat. Immunol.*, 2009, **10**, 1292–1299.
- 41.** R. E. Billingham, L. Brent and P. B. Medawar, *J. Immunol. Baltim. Md 1950*, 2010, **184**, 5–8.
- 42.** J. W. Kappler, N. Roehm and P. Marrack, *Cell*, 1987, **49**, 273–280.
- 43.** H. S. Teh, P. Kisielow, B. Scott, H. Kishi, Y. Uematsu, H. Blüthmann and H. von Boehmer, *Nature*, 1988, **335**, 229–233.
- 44.** K. Nagamine, P. Peterson, H. S. Scott, J. Kudoh, S. Minoshima, M. Heino, K. J. Krohn, M. D. Lalioti, P. E. Mullis, S. E. Antonarakis, K. Kawasaki, S. Asakawa, F. Ito and N. Shimizu, *Nat. Genet.*, 1997, **17**, 393–398.
- 45.** M. S. Anderson, E. S. Venanzi, L. Klein, Z. Chen, S. P. Berzins, S. J. Turley, H. von Boehmer, R. Bronson, A. Dierich, C. Benoist and D. Mathis, *Science*, 2002, **298**, 1395–1401.
- 46.** H. Kishimoto and J. Sprent, *J. Exp. Med.*, 1997, **185**, 263–271.
- 47.** G. Iezzi, E. Scotet, D. Scheidegger and A. Lanzavecchia, *Eur. J. Immunol.*, 1999, **29**, 4092–4101.
- 48.** J. R. Lamb, B. J. Skidmore, N. Green, J. M. Chiller and M. Feldmann, *J. Exp. Med.*, 1983, **157**, 1434–1447.

49. S. B. Hartley, M. P. Cooke, D. A. Fulcher, A. W. Harris, S. Cory, A. Basten and C. C. Goodnow, *Cell*, 1993, **72**, 325–335.
50. D. Nemazee, *Nat. Rev. Immunol.*, 2017, **17**, 281–294.
51. A. S. Wiener and E. Russow, *Ann. Intern. Med.*, 1957, **47**, 1–9.
52. E. P. Nagele, M. Han, N. K. Acharya, C. DeMarshall, M. C. Kosciuk and R. G. Nagele, *PLoS One*, 2013, **8**, e60726.
53. K. Hayakawa, R. R. Hardy, M. Honda, L. A. Herzenberg, A. D. Steinberg and L. A. Herzenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 2494–2498.
54. B. Cukrowska, J. Sinkora, L. Mandel, I. Splíchal, A. T. Bianchi, F. Kovářů and H. Tlaskalová-Hogenová, *Immunology*, 1996, **87**, 487–492.
55. K. Elkon and P. Casali, *Nat. Clin. Pract. Rheumatol.*, 2008, **4**, 491–498.
56. S. Duquerroy, E. A. Stura, S. Bressanelli, S. M. Fabiane, M. C. Vaney, D. Beale, M. Hamon, P. Casali, F. A. Rey, B. J. Sutton and M. J. Taussig, *J. Mol. Biol.*, 2007, **368**, 1321–1331.
57. B. Sutton, A. Corper, V. Bonagura and M. Taussig, *Immunol. Today*, 2000, **21**, 177–183.
58. Y. Ichiyoshi, M. Zhou and P. Casali, *J. Immunol.*, 1995, **154**, 226–238.
59. P. D. Burbelo, J. Keller and J. B. Kopp, *Discov. Med.*, 2015, **20**, 17–25.
60. J. J. Goronzy and C. M. Weyand, *Arthritis Res. Ther.*, 2008, **10 Suppl 1**, S3.
61. A. B. Glick, A. Wodzinski, P. Fu, A. D. Levine and D. N. Wald, *Thyroid*, 2013, **23**, 871–878.
62. W. Chen and J. E. Konkel, *J. Mol. Cell Biol.*, 2010, **2**, 30–36.
63. B. Akinci, A. Comlekci, S. Yener, F. Bayraktar, T. Demir, M. A. Ozcan, F. Yuksel and S. Yesil, *Arch. Med. Res.*, 2008, **39**, 397–401.
64. I. Manolova, J. Gerenova and M. Ivanova, *Eur. Cytokine Netw.*, 2013, **24**, 69–74.
65. A. Saxena, S. Khosraviani, S. Noel, D. Mohan, T. Donner and A. R. A. Hamad, *Cytokine*, 2015, **74**, 27–34.
66. H. Groux and F. Cottrez, *J. Autoimmun.*, 2003, **20**, 281–285.
67. B. Koppelman, J. J. Neefjes, J. E. de Vries and R. de Waal Malefyt, *Immunity*, 1997, **7**, 861–871.
68. A. Joss, M. Akdis, A. Faith, K. Blaser and C. A. Akdis, *Eur. J. Immunol.*, 2000, **30**, 1683–1690.
69. N. F. Go, B. E. Castle, R. Barrett, R. Kastelein, W. Dang, T. R. Mosmann, K. W. Moore and M. Howard, *J. Exp. Med.*, 1990, **172**, 1625–1631.
70. A. Rahman and D. A. Isenberg, *N. Engl. J. Med.*, 2008, **358**, 929–939.
71. R. K. Mallavarapu and E. W. Grimsley, *South. Med. J.*, 2007, **100**, 896–898.
72. G. J. Friou, S. C. Finch and K. D. Detre, *J. Immunol.*, 1958, **80**, 324–329.
73. C. C. Mok and C. S. Lau, *J. Clin. Pathol.*, 2003, **56**, 481–490.

- 74.** L. E. Muñoz, K. Lauber, M. Schiller, A. A. Manfredi and M. Herrmann, *Nat. Rev. Rheumatol.*, 2010, **6**, 280–289.
- 75.** H.-J. Anders, *Trends Mol. Med.*, 2009, **15**, 553–561.
- 76.** K. Ohl and K. Tenbrock, *Eur. J. Immunol.*, 2015, **45**, 344–355.
- 77.** C. T. Sawin, *Endocr. J.*, 2002, **49**, 399–403.
- 78.** C. G. P. Roberts and P. W. Ladenson, *Lancet*, 2004, **363**, 793–803.
- 79.** P. Caturegli, A. De Remigis and N. R. Rose, *Autoimmun. Rev.*, 2014, **13**, 391–397.
- 80.** Y. Li, Y. Bai, Z. Liu, T. Ozaki, E. Taniguchi, I. Mori, K. Nagayama, H. Nakamura and K. Kakudo, *Pathol. Int.*, 2009, **59**, 636–641.
- 81.** D. Kottahachchi and D. J. Topliss, *Eur. Thyroid J.*, 2016, **5**, 231–239.
- 82.** A. P. Weetman, *Horm. Res.*, 2003, **59 Suppl 1**, 114–118.
- 83.** P. Laurberg, K. M. Pedersen, H. Vestergaard and G. Sigurdsson, *J. Intern. Med.*, 1991, **229**, 415–420.
- 84.** S. De Leo, S. Y. Lee and L. E. Braverman, *Lancet*, 2016, **388**, 906–918.
- 85.** D. D. Adams, *J. Clin. Endocrinol. Metab.*, 1958, **18**, 699–712.
- 86.** J. P. Kriss, V. Pleshakov and J. R. Chien, *J. Clin. Endocrinol. Metab.*, 1964, **24**, 1005–1028.
- 87.** B. R. Smith and R. Hall, *Lancet*, 1974, **2**, 427–431.
- 88.** R. S. Bahn, *N. Engl. J. Med.*, 2010, **362**, 726–738.
- 89.** W. M. Wiersinga and L. Bartalena, *Thyroid*, 2002, **12**, 855–860.
- 90.** S. Moshkelgosha, P.-W. So, N. Deasy, S. Diaz-Cano and J. P. Banga, *Endocrinology*, 2013, **154**, 3008–3015.
- 91.** U. Berchner-Pfannschmidt, S. Moshkelgosha, S. Diaz-Cano, B. Edelmann, G.-E. Görtz, M. Horstmann, A. Noble, W. Hansen, A. Eckstein and J. P. Banga, *Endocrinology*, 2016, **157**, 1673–1682.
- 92.** A. Eckstein, M. Schittkowski and J. Esser, *Best Pract. Res. Clin. Endocrinol. Metab.*, 2012, **26**, 339–358.
- 93.** A. Stagnaro-Green, M. Abalovich, E. Alexander, F. Azizi, J. Mestman, R. Negro, A. Nixon, E. N. Pearce, O. P. Soldin, S. Sullivan, W. Wiersinga and American Thyroid Association Taskforce on Thyroid Disease During Pregnancy and Postpartum, *Thyroid*, 2011, **21**, 1081–1125.
- 94.** A. Stagnaro-Green, S. H. Roman, R. H. Cobin, E. el-Harazy, S. Wallenstein and T. F. Davies, *J. Clin. Endocrinol. Metab.*, 1992, **74**, 645–653.
- 95.** W. K. Nicholson, K. A. Robinson, R. C. Smallridge, P. W. Ladenson and N. R. Powe, *Thyroid*, 2006, **16**, 573–582.
- 96.** H. E. W. Robertson, *Br. Med. J.*, 1948, **2**, 93.
- 97.** A. Stagnaro-Green, *Best Pract. Res. Clin. Endocrinol. Metab.*, 2004, **18**, 303–316.

98. S. Othman, D. I. Phillips, A. B. Parkes, C. J. Richards, B. Harris, H. Fung, C. Darke, R. John, R. Hall and J. H. Lazarus, *Clin. Endocrinol. (Oxf.)*, 1990, **32**, 559–564.
99. S. De Leo and E. N. Pearce, *Lancet Diabetes Endocrinol.*, 2018, **6**, 575–586.
100. R. Negro, G. Formoso, T. Mangieri, A. Pezzarossa, D. Dazzi and H. Hassan, *J. Clin. Endocrinol. Metab.*, 2006, **91**, 2587–2591.
101. Y. Li, Z. Shan, W. Teng, X. Yu, Y. Li, C. Fan, X. Teng, R. Guo, H. Wang, J. Li, Y. Chen, W. Wang, M. Chawinga, L. Zhang, L. Yang, Y. Zhao and T. Hua, *Clin. Endocrinol. (Oxf.)*, 2010, **72**, 825–829.
102. E. E. Wasserman, J. P. Pillion, A. Duggan, K. Nelson, C. Rohde, E. C. Seaberg, M. V. Talor, R. H. Yolken and N. R. Rose, *Pediatr. Res.*, 2012, **72**, 525–530.
103. J. Ruf, B. Czarnocka, C. De Micco, C. Dutoit, M. Ferrand and P. Carayon, *Acta Endocrinol. Suppl. (Copenh.)*, 1987, **281**, 49–56.
104. F. Gentile, M. Conte and S. Formisano, *Immunology*, 2004, **112**, 13–25.
105. B. Rapoport, G. D. Chazenbalk, J. C. Jaume and S. M. McLachlan, *Endocr. Rev.*, 1998, **19**, 673–716.
106. T. Chardès, N. Chapal, D. Bresson, C. Bès, V. Giudicelli, M.-P. Lefranc and S. Péraldi-Roux, *Immunogenetics*, 2002, **54**, 141–157.
107. G. D. Chazenbalk, S. Portolano, D. Russo, J. S. Hutchison, B. Rapoport and S. McLachlan, *J. Clin. Invest.*, 1993, **92**, 62–74.
108. N. Chapal, T. Chardès, D. Bresson, M. Pugnière, J. C. Mani, B. Pau, M. Bouanani and S. Péraldi-Roux, *Endocrinology*, 2001, **142**, 4740–4750.
109. M. Horimoto, V. S. Petersen, C. A. Pegg, N. Fukuma, N. Wakabayashi, Y. Kiso, J. Furmaniak and B. Rees Smith, *Autoimmunity*, 1992, **14**, 1–7.
110. A. B. Parkes, S. M. McLachlan, P. Bird and B. Rees Smith, *Clin. Exp. Immunol.*, 1984, **57**, 239–243.
111. L. Chiavato, P. Bassi, F. Santini, C. Mammoli, P. Lapi, P. Carayon and A. Pinchera, *J. Clin. Endocrinol. Metab.*, 1993, **77**, 1700–1705.
112. M. W. Szkudlinski, V. Fremont, C. Ronin and B. D. Weintraub, *Physiol. Rev.*, 2002, **82**, 473–502.
113. B. R. Smith, J. Sanders and J. Furmaniak, *Thyroid*, 2007, **17**, 923–938.
114. J. Furmaniak, J. Sanders, R. Núñez Miguel and B. Rees Smith, *Horm. Metab. Res.*, 2015, **47**, 735–752.
115. S. A. Morshed and T. F. Davies, *Horm. Metab. Res.*, 2015, **47**, 727–734.
116. A. P. Weetman, M. E. Yateman, P. A. Ealey, C. M. Black, C. B. Reimer, R. C. Williams, B. Shine and N. J. Marshall, *J. Clin. Invest.*, 1990, **86**, 723–727.
117. J. Furmaniak, J. Sanders and B. Rees Smith, *Auto- Immun. Highlights*, 2013, **4**, 11–26.
118. J. Sanders, D. Y. Chirgadze, P. Sanders, S. Baker, A. Sullivan, A. Bhardwaja, J. Bolton, M. Reeve, N. Nakatake, M. Evans, T. Richards, M. Powell, R. N. Miguel, T. L. Blundell, J. Furmaniak and B. R. Smith, *Thyroid*, 2007, **17**, 395–410.

- 119.** P. Sanders, S. Young, J. Sanders, K. Kabelis, S. Baker, A. Sullivan, M. Evans, J. Clark, J. Wilmot, X. Hu, E. Roberts, M. Powell, R. Nunez Miguel, J. Furmaniak and B. Rees Smith, *J. Mol. Endocrinol.*, DOI:10.1530/JME-10-0127.
- 120.** S. A. Morshed, T. Ando, R. Latif and T. F. Davies, *Endocrinology*, 2010, **151**, 5537–5549.
- 121.** M. Evans, J. Sanders, T. Tagami, P. Sanders, S. Young, E. Roberts, J. Wilmot, X. Hu, K. Kabelis, J. Clark, S. Holl, T. Richards, A. Collyer, J. Furmaniak and B. R. Smith, *Clin. Endocrinol. (Oxf.)*, 2010, **73**, 404–412.
- 122.** B. Di Jeso and P. Arvan, *Endocr. Rev.*, 2016, **37**, 2–36.
- 123.** P. Caturegli, R. C. Kuppers, S. Mariotti, C. L. Burek, A. Pinchera, P. W. Ladenson and N. R. Rose, *Clin. Exp. Immunol.*, 1994, **98**, 464–469.
- 124.** S. M. McLachlan and B. Rapoport, *Thyroid*, 2004, **14**, 510–520.
- 125.** L. Muixí, M. Carrascal, I. Alvarez, X. Daura, M. Martí, M. P. Armengol, C. Pinilla, J. Abian, R. Pujol-Borrell and D. Jaraquemada, *J. Immunol.*, 2008, **181**, 795–807.
- 126.** A. McGrogan, H. E. Seaman, J. W. Wright and C. S. de Vries, *Clin. Endocrinol. (Oxf.)*, 2008, **69**, 687–696.
- 127.** W. M. Tunbridge, D. C. Evered, R. Hall, D. Appleton, M. Brewis, F. Clark, J. G. Evans, E. Young, T. Bird and P. A. Smith, *Clin. Endocrinol. (Oxf.)*, 1977, **7**, 481–493.
- 128.** M. P. Vanderpump, W. M. Tunbridge, J. M. French, D. Appleton, D. Bates, F. Clark, J. Grimley Evans, D. M. Hasan, H. Rodgers and F. Tunbridge, *Clin. Endocrinol. (Oxf.)*, 1995, **43**, 55–68.
- 129.** J. G. Hollowell, N. W. Staehling, W. D. Flanders, W. H. Hannon, E. W. Gunter, C. A. Spencer and L. E. Braverman, *J. Clin. Endocrinol. Metab.*, 2002, **87**, 489–499.
- 130.** S. Ansar Ahmed, W. J. Penhale and N. Talal, *Am. J. Pathol.*, 1985, **121**, 531–551.
- 131.** I. F. Cook, *Vaccine*, 2008, **26**, 3551–3555.
- 132.** M. Butterworth, B. McClellan and M. Allansmith, *Nature*, 1967, **214**, 1224–1225.
- 133.** J. E. Castro, *Proc. R. Soc. Lond. B Biol. Sci.*, 1974, **185**, 437–451.
- 134.** J. A. Da Silva, *Ann. Rheum. Dis.*, 1995, **54**, 6–16.
- 135.** X. Yin, R. Latif, Y. Tomer and T. F. Davies, *Ann. N. Y. Acad. Sci.*, 2007, **1110**, 193–200.
- 136.** L. Martin, *Q. J. Med.*, 1945, **14**, 207–219.
- 137.** T. H. Brix, K. Christensen, N. V. Holm, B. Harvald and L. Hegedüs, *Clin. Endocrinol. (Oxf.)*, 1998, **48**, 397–400.
- 138.** T. H. Brix, K. O. Kyvik, K. Christensen and L. Hegedüs, *J. Clin. Endocrinol. Metab.*, 2001, **86**, 930–934.
- 139.** D. A. Ringold, J. T. Nicoloff, M. Kesler, H. Davis, A. Hamilton and T. Mack, *Thyroid*, 2002, **12**, 647–653.
- 140.** R. Villanueva, D. A. Greenberg, T. F. Davies and Y. Tomer, *Thyroid*, 2003, **13**, 761–764.
- 141.** S. M. Pulst, *Arch. Neurol.*, 1999, **56**, 667–672.

- 142.** M. J. Simmonds and S. C. L. Gough, *Brief. Funct. Genomics*, 2011, **10**, 77–90.
- 143.** M. T. Pirro, V. De Filippis, A. Di Cerbo, A. Scillitani, A. Liuzzi and V. Tassi, *Thyroid*, 1995, **5**, 461–464.
- 144.** Y. Tomer, G. Barbesino, M. Keddache, D. A. Greenberg and T. F. Davies, *J. Clin. Endocrinol. Metab.*, 1997, **82**, 1645–1648.
- 145.** M. Stefan and L. C. Faustino, *Front. Endocrinol.*, 2017, **8**, 57.
- 146.** O. J. Brand, J. C. Barrett, M. J. Simmonds, P. R. Newby, C. J. McCabe, C. K. Bruce, B. Kysela, J. D. Carr-Smith, T. Brix, P. J. Hunt, W. M. Wiersinga, L. Hegedüs, J. Connell, J. A. H. Wass, J. A. Franklyn, A. P. Weetman, J. M. Heward and S. C. L. Gough, *Hum. Mol. Genet.*, 2009, **18**, 1704–1713.
- 147.** R. Colobran, M. D. P. Armengol, R. Faner, M. Gärtner, L.-O. Tykocinski, A. Lucas, M. Ruiz, M. Juan, B. Kyewski and R. Pujol-Borrell, *Hum. Mol. Genet.*, 2011, **20**, 3415–3423.
- 148.** K. Sakai, S. Shirasawa, N. Ishikawa, K. Ito, H. Tamai, K. Kuma, T. Akamizu, M. Tanimura, K. Furugaki, K. Yamamoto and T. Sasazuki, *Hum. Mol. Genet.*, 2001, **10**, 1379–1386.
- 149.** Y. Tomer, Y. Ban, E. Concepcion, G. Barbesino, R. Villanueva, D. A. Greenberg and T. F. Davies, *Am. J. Hum. Genet.*, 2003, **73**, 736–747.
- 150.** Y. Ban, D. A. Greenberg, E. Concepcion, L. Skrabaneck, R. Villanueva and Y. Tomer, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15119–15124.
- 151.** Y. Ban, T. F. Davies, D. A. Greenberg, E. S. Concepcion, R. Osman, T. Oashi and Y. Tomer, *Genes Immun.*, 2004, **5**, 203–208.
- 152.** F. Menconi, M. C. Monti, D. A. Greenberg, T. Oashi, R. Osman, T. F. Davies, Y. Ban, E. M. Jacobson, E. S. Concepcion, C. W. Li and Y. Tomer, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 14034–14039.
- 153.** S. E. Hodge, Y. Ban, L. J. Strug, D. A. Greenberg, T. F. Davies, E. S. Concepcion, R. Villanueva and Y. Tomer, *Thyroid*, 2006, **16**, 351–355.
- 154.** D. Pastuszak-Lewandoska, E. Sewerynek, D. Domańska, A. Gładys, R. Skrzypczak and E. Brzezińska, *Arch. Med. Sci.*, 2012, **3**, 415–421.
- 155.** M. R. Velaga, V. Wilson, C. E. Jennings, C. J. Owen, S. Herington, P. T. Donaldson, S. G. Ball, R. A. James, R. Quinton, P. Perros and S. H. S. Pearce, *J. Clin. Endocrinol. Metab.*, 2004, **89**, 5862–5865.
- 156.** Y. Tomer, *Thyroid*, 2010, **20**, 715–725.
- 157.** G. A. Bishop and B. S. Hostager, *Cytokine Growth Factor Rev.*, 2003, **14**, 297–309.
- 158.** R. Khattri, T. Cox, S.-A. Yasayko and F. Ramsdell, *Nat. Immunol.*, 2003, **4**, 337–342.
- 159.** A. Drutel, F. Archambeaud and P. Caron, *Clin. Endocrinol. (Oxf.)*, 2013, **78**, 155–164.
- 160.** E. Guastamacchia, V. A. Giagulli, B. Licchelli and V. Triggiani, *Endocr. Metab. Immune Disord. Drug Targets*, 2015, **15**, 288–292.
- 161.** A. Taurog, M. L. Dorris and D. R. Doerge, *Arch. Biochem. Biophys.*, 1996, **330**, 24–32.
- 162.** M. B. Zimmermann, *Endocr. Rev.*, 2009, **30**, 376–408.

- 163.** E. N. Pearce, M. Andersson and M. B. Zimmermann, *Thyroid*, 2013, **23**, 523–528.
- 164.** G. J. Kahaly, H. P. Dienes, J. Beyer and G. Hommel, *Eur. J. Endocrinol.*, 1998, **139**, 290–297.
- 165.** I. B. Pedersen, N. Knudsen, A. Carlé, P. Vejbjerg, T. Jørgensen, H. Perrild, L. Ovesen, L. B. Rasmussen and P. Laurberg, *Clin. Endocrinol. (Oxf.)*, 2011, **75**, 120–126.
- 166.** M. B. Zimmermann and K. Boelaert, *Lancet Diabetes Endocrinol.*, 2015, **3**, 286–295.
- 167.** F. Latrofa, E. Fiore, T. Rago, L. Antonangeli, L. Montanelli, D. Ricci, M. A. Provenzale, M. Scutari, M. Frigeri, M. Tonacchera and P. Vitti, *J. Clin. Endocrinol. Metab.*, 2013, **98**, E1768–1774.
- 168.** G. Carayanniotis, *Autoimmunity*, 2003, **36**, 423–428.
- 169.** M. Li and S. C. Boyages, *Autoimmunity*, 1994, **18**, 31–40.
- 170.** V. Triggiani, E. Tafaro, V. A. Giagulli, C. Sabbà, F. Resta, B. Licchelli and E. Guastamacchia, *Endocr. Metab. Immune Disord. Drug Targets*, 2009, **9**, 277–294.
- 171.** I. Bülow Pedersen, N. Knudsen, A. Carlé, L. Schomburg, J. Köhrle, T. Jørgensen, L. B. Rasmussen, L. Ovesen and P. Laurberg, *Clin. Endocrinol. (Oxf.)*, 2013, **79**, 584–590.
- 172.** I. Wimmer, T. Hartmann, R. Brustbauer, G. Minear and K. Dam, *Horm. Metab. Res.*, 2014, **46**, 707–709.
- 173.** K. H. Winther, S. J. Bonnema and L. Hegedüs, *Curr. Opin. Endocrinol. Diabetes Obes.*, 2017, **24**, 348–355.
- 174.** P. Vestergaard, *Eur. J. Endocrinol.*, 2002, **146**, 153–161.
- 175.** W. M. Wiersinga, *Clin. Endocrinol. (Oxf.)*, 2013, **79**, 145–151.
- 176.** R. M. Belin, B. C. Astor, N. R. Powe and P. W. Ladenson, *J. Clin. Endocrinol. Metab.*, 2004, **89**, 6077–6086.
- 177.** B. O. Asvold, T. Bjøro, T. I. L. Nilsen and L. J. Vatten, *Arch. Intern. Med.*, 2007, **167**, 1428–1432.
- 178.** I. B. Pedersen, P. Laurberg, N. Knudsen, T. Jørgensen, H. Perrild, L. Ovesen and L. B. Rasmussen, *Eur. J. Endocrinol.*, 2008, **158**, 367–373.
- 179.** T. G. A. Strieder, M. F. Prummel, J. G. P. Tijssen, E. Endert and W. M. Wiersinga, *Clin. Endocrinol. (Oxf.)*, 2003, **59**, 396–401.
- 180.** R. Talhout, T. Schulz, E. Florek, J. van Benthem, P. Wester and A. Opperhuizen, *Int. J. Environ. Res. Public. Health*, 2011, **8**, 613–628.
- 181.** T. J. Cawood, P. Moriarty, C. O'Farrelly and D. O'Shea, *J. Clin. Endocrinol. Metab.*, 2007, **92**, 59–64.
- 182.** M. F. Cusick, J. E. Libbey and R. S. Fujinami, *Clin. Rev. Allergy Immunol.*, 2012, **42**, 102–111.
- 183.** S. Benvenga and F. Guarneri, *Rev. Endocr. Metab. Disord.*, DOI:10.1007/s11154-016-9363-2.
- 184.** R. Dalan and M. K. S. Leow, *Eur. J. Intern. Med.*, 2012, **23**, 682–691.

- 185.** H. B. Burch and D. S. Cooper, *JAMA*, 2015, **314**, 2544–2554.
- 186.** I. Kravets, *Am. Fam. Physician*, 2016, **93**, 363–370.
- 187.** D. S. Ross, H. B. Burch, D. S. Cooper, M. C. Greenlee, P. Laurberg, A. L. Maia, S. A. Rivkees, M. Samuels, J. A. Sosa, M. N. Stan and M. A. Walter, *Thyroid Off. J. Am. Thyroid Assoc.*, 2016, **26**, 1343–1421.
- 188.** L. Delbridge, *ANZ J. Surg.*, 2003, **73**, 761–768.
- 189.** C. Camenzuli, P. Schembri Wismayer and J. Calleja Agius, *JSLS*, , DOI:10.4293/JSLS.2018.00026.
- 190.** B. Corvilain, A. Hamy, L. Brunaud, F. Borson-Chazot, J. Orgiazzi, L. Bensalem Hachmi, M. Semrouni, P. Rodien and C. Lussey-Lepoutre, *Ann. Endocrinol.*, 2018, **79**, 618–635.
- 191.** M. Luster, A. Pfestroff, H. Hänscheid and F. A. Verburg, *Semin. Nucl. Med.*, 2017, **47**, 126–134.
- 192.** E. K. Alexander and P. R. Larsen, *J. Clin. Endocrinol. Metab.*, 2002, **87**, 1073–1077.
- 193.** H. F. Escobar-Morreale, J. I. Botella-Carretero and G. Morreale de Escobar, *Best Pract. Res. Clin. Endocrinol. Metab.*, 2015, **29**, 57–75.
- 194.** A. Mohebati and A. R. Shaha, *Clin. Anat. N. Y. N.*, 2012, **25**, 19–31.
- 195.** L. E. Braverman, S. H. Ingbar and K. Sterling, *J. Clin. Invest.*, 1970, **49**, 855–864.
- 196.** P. M. Yen, *Physiol. Rev.*, 2001, **81**, 1097–1142.
- 197.** A. J. Forhead and A. L. Fowden, *J. Endocrinol.*, 2014, **221**, R87–R103.
- 198.** P. R. Larsen, J. E. Silva and M. M. Kaplan, *Endocr. Rev.*, 1981, **2**, 87–102.
- 199.** A. Marsili, A. M. Zavacki, J. W. Harney and P. R. Larsen, *J. Endocrinol. Invest.*, 2011, **34**, 395–407.
- 200.** M. Marinò and R. T. McCluskey, *Am. J. Physiol. Cell Physiol.*, 2000, **279**, C1295-1306.
- 201.** A. D. Dunn, H. E. Crutchfield and J. T. Dunn, *Endocrinology*, 1991, **128**, 3073–3080.
- 202.** R. Ekholm and S. H. Wollman, *Endocrinology*, 1975, **97**, 1432–1444.
- 203.** M. L. Coval and A. Taurog, *J. Biol. Chem.*, 1967, **242**, 5510–5523.
- 204.** P. G. Furtmüller, M. Zederbauer, W. Jantschko, J. Helm, M. Bogner, C. Jakopitsch and C. Obinger, *Arch. Biochem. Biophys.*, 2006, **445**, 199–213.
- 205.** L. Fayadat, P. Niccoli-Sire, J. Lanet and J. L. Franc, *J. Biol. Chem.*, 1999, **274**, 10533–10538.
- 206.** A. Taurog and M. Wall, *Thyroid*, 1998, **8**, 185–191.
- 207.** T. L. Poulos and J. Kraut, *J. Biol. Chem.*, 1980, **255**, 8199–8205.
- 208.** J. N. Rodríguez-López, D. J. Lowe, J. Hernández-Ruiz, A. N. Hiner, F. García-Cánovas and R. N. Thorneley, *J. Am. Chem. Soc.*, 2001, **123**, 11838–11847.
- 209.** A. Virion, F. Courtin, D. Dème, J. L. Michot, J. Kaniewski and J. Pommier, *Arch. Biochem. Biophys.*, 1985, **242**, 41–47.

- 210.** O. M. Lardinois and P. R. Ortiz de Montellano, *Biochem. Biophys. Res. Commun.*, 2000, **270**, 199–202.
- 211.** O. M. Lardinois, K. F. Medzihradzky and P. R. Ortiz de Montellano, *J. Biol. Chem.*, 1999, **274**, 35441–35448.
- 212.** B. Corvilain, J. van Sande, E. Laurent and J. E. Dumont, *Endocrinology*, 1991, **128**, 779–785.
- 213.** H. Kohler, A. Taurog and H. B. Dunford, *Arch. Biochem. Biophys.*, 1988, **264**, 438–449.
- 214.** H. B. Dunford, *Redox Rep.*, 2000, **5**, 169–171.
- 215.** J. Kessler, C. Obinger and G. Eales, *Thyroid*, 2008, **18**, 769–774.
- 216.** D. R. Doerge, A. Taurog and M. L. Dorris, *Arch. Biochem. Biophys.*, 1994, **315**, 90–99.
- 217.** T. J. Fiedler, C. A. Davey and R. E. Fenna, *J. Biol. Chem.*, 2000, **275**, 11964–11971.
- 218.** A. K. Singh, N. Singh, S. Sharma, S. B. Singh, P. Kaur, A. Bhushan, A. Srinivasan and T. P. Singh, *J. Mol. Biol.*, 2008, **376**, 1060–1075.
- 219.** L. Lamas, P. C. Anderson, J. W. Fox and J. T. Dunn, *J. Biol. Chem.*, 1989, **264**, 13541–13545.
- 220.** P. Lemansky, G. M. Popp, J. Tietz and V. Herzog, *Endocrinology*, 1994, **135**, 1566–1575.
- 221.** F. Rodesch and J. E. Dumont, *Exp. Cell Res.*, 1967, **47**, 386–396.
- 222.** J. T. Dunn and A. D. Dunn, *Biochimie*, 1999, **81**, 505–509.
- 223.** U. Berndorfer, H. Wilms and V. Herzog, *J. Clin. Endocrinol. Metab.*, 1996, **81**, 1918–1926.
- 224.** J. M. Gavaret, J. Nunez and H. J. Cahnmann, *J. Biol. Chem.*, 1980, **255**, 5281–5285.
- 225.** A. Taurog, M. Dorris and D. R. Doerge, *Arch. Biochem. Biophys.*, 1994, **315**, 82–89.
- 226.** F. Courtin, D. Deme, A. Virion, J. L. Michot, J. Pommier and J. Nunez, *Eur. J. Biochem.*, 1982, **124**, 603–609.
- 227.** A. Virion, D. Deme, J. Pommier and J. Nunez, *Eur. J. Biochem.*, 1981, **118**, 239–245.
- 228.** S. Kimura, T. Kotani, O. W. McBride, K. Umeki, K. Hirai, T. Nakayama and S. Ohtaki, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 5555–5559.
- 229.** S. Kimura, Y. S. Hong, T. Kotani, S. Ohtaki and F. Kikkawa, *Biochemistry*, 1989, **28**, 4481–4489.
- 230.** R. Elisei, G. Vassart and M. Ludgate, *J. Clin. Endocrinol. Metab.*, 1991, **72**, 700–702.
- 231.** P. Niccoli, L. Fayadat, V. Panneels, J. Lanet and J. L. Franc, *J. Biol. Chem.*, 1997, **272**, 29487–29492.
- 232.** A. Taurog, *Biochimie*, 1999, **81**, 557–562.
- 233.** E. Zanelli, M. Henry, B. Charvet and Y. Malthièry, *Biochem. Biophys. Res. Commun.*, 1990, **170**, 735–741.
- 234.** M. Ferrand, V. Le Fourn and J.-L. Franc, *J. Biol. Chem.*, 2003, **278**, 3793–3800.

- 235.** T. Ueda, K. Sakamaki, T. Kuroki, I. Yano and S. Nagata, *Eur. J. Biochem. FEBS*, 1997, **243**, 32–41.
- 236.** M. Zámocký, S. Hofbauer, I. Schaffner, B. Gasselhuber, A. Nicolussi, M. Soudi, K. F. Pirker, P. G. Furtmüller and C. Obinger, *Arch. Biochem. Biophys.*, 2015, **574**, 108–119.
- 237.** M. Zamocký, C. Jakopitsch, P. G. Furtmüller, C. Dunand and C. Obinger, *Proteins*, 2008, **72**, 589–605.
- 238.** A. L. Mitchell, T. K. Attwood, P. C. Babbitt, M. Blum, P. Bork, A. Bridge, S. D. Brown, H.-Y. Chang, S. El-Gebali, M. I. Fraser, J. Gough, D. R. Haft, H. Huang, I. Letunic, R. Lopez, A. Luciani, F. Madeira, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, G. Nuka, C. Orengo, A. P. Pandurangan, T. Paysan-Lafosse, S. Pesseat, S. C. Potter, M. A. Qureshi, N. D. Rawlings, N. Redaschi, L. J. Richardson, C. Rivoire, G. A. Salazar, A. Sangrador-Vegas, C. J. A. Sigrist, I. Sillitoe, G. G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, S.-Y. Yong and R. D. Finn, *Nucleic Acids Res.*, 2019, **47**, D351–D360.
- 239.** S. J. Klebanoff, *Science*, 1970, **169**, 1095–1097.
- 240.** B. D. Polis and H. W. Shmukler, *J. Biol. Chem.*, 1953, **201**, 475–500.
- 241.** G. T. Archer, G. Air, M. Jackas and D. B. Morell, *Biochim. Biophys. Acta*, 1965, **99**, 96–101.
- 242.** J. Zeng and R. E. Fenna, *J. Mol. Biol.*, 1992, **226**, 185–207.
- 243.** M. Yamada, *J. Biol. Chem.*, 1982, **257**, 5980–5982.
- 244.** T. Ravnsborg, G. Houen and P. Højrup, *Biochim. Biophys. Acta*, 2010, **1804**, 2046–2053.
- 245.** P. Van Antwerpen, M.-C. Slomianny, K. Z. Boudjeltia, C. Delporte, V. Faid, D. Calay, A. Rousseau, N. Mogilevsky, M. Raes, L. Vanhamme, P. G. Furtmüller, C. Obinger, M. Vanhaeverbeek, J. Nève and J.-C. Michalski, *J. Biol. Chem.*, 2010, **285**, 16351–16359.
- 246.** K. Shin, H. Hayasawa and B. Lönnerdal, *Biochem. Biophys. Res. Commun.*, 2001, **281**, 1024–1029.
- 247.** S. Sharma, A. K. Singh, S. Kaushik, M. Sinha, R. P. Singh, P. Sharma, H. Sirohi, P. Kaur and T. P. Singh, *Int. J. Biochem. Mol. Biol.*, 2013, **4**, 108–128.
- 248.** G. Sievers, *FEBS Lett.*, 1981, **127**, 253–256.
- 249.** S. M. Wolf, R. P. Ferrari, S. Traversa and K. Biemann, *J. Mass Spectrom. JMS*, 2000, **35**, 210–217.
- 250.** M. G. Carlson, C. G. Peterson and P. Venge, *J. Immunol.*, 1985, **134**, 1875–1879.
- 251.** I. Olsson, A. M. Persson, K. Strömberg, I. Winqvist, P. C. Tai and C. J. Spry, *Blood*, 1985, **66**, 1143–1148.
- 252.** A. R. Thomsen, L. Sottrup-Jensen, G. J. Gleich and C. Oxvig, *Arch. Biochem. Biophys.*, 2000, **379**, 147–152.
- 253.** C. Oxvig, A. R. Thomsen, M. T. Overgaard, E. S. Sorensen, P. Højrup, M. J. Bjerrum, G. J. Gleich and L. Sottrup-Jensen, *J. Biol. Chem.*, 1999, **274**, 16953–16958.
- 254.** A. Gardas, M. K. Sohi, B. J. Sutton, A. M. McGregor and J. P. Banga, *Biochem. Biophys. Res. Commun.*, 1997, **234**, 366–370.

- 255.** E. Hendry, G. Taylor, K. Ziemnicka, F. Grennan Jones, J. Furmaniak and B. Rees Smith, *J. Endocrinol.*, 1999, **160**, R13–15.
- 256.** P. Hobby, A. Gardas, R. Radomski, A. M. McGregor, J. P. Banga and B. J. Sutton, *Endocrinology*, 2000, **141**, 2018–2026.
- 257.** S. N. Le, B. T. Porebski, J. McCoey, J. Fodor, B. Riley, M. Godlewska, M. Góra, B. Czarnocka, J. P. Banga, D. E. Hoke, I. Kass and A. M. Buckle, *PLoS One*, 2015, **10**, e0142615.
- 258.** E. Hendry, PhD Thesis, University of St. Andrews, 2000.
- 259.** V. Le Fourn, M. Ferrand and J.-L. Franc, *J. Biol. Chem.*, 2005, **280**, 4568–4577.
- 260.** H. Owji, N. Nezafat, M. Negahdaripour, A. Hajiebrahimi and Y. Ghasemi, *Eur. J. Cell Biol.*, 2018, **97**, 422–441.
- 261.** M. Godlewska, M. Góra, A. M. Buckle, B. T. Porebski, E. H. Kemp, B. J. Sutton, B. Czarnocka and J. P. Banga, *Thyroid*, 2014, **24**, 371–382.
- 262.** M. Zederbauer, P. G. Furtmüller, S. Brogioni, C. Jakopitsch, G. Smulevich and C. Obinger, *Nat. Prod. Rep.*, 2007, **24**, 571–584.
- 263.** I. M. Kooter, N. Moguilevsky, A. Bollen, L. A. van der Veen, C. Otto, H. L. Dekker and R. Wever, *J. Biol. Chem.*, 1999, **274**, 26794–26802.
- 264.** A. Jacquet, L. Garcia-Quintana, V. Deleersnyder, R. Fenna, A. Bollen and N. Moguilevsky, *Biochem. Biophys. Res. Commun.*, 1994, **202**, 73–81.
- 265.** L. B. Klickstein, W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson and D. T. Fearon, *J. Exp. Med.*, 1987, **165**, 1095–1112.
- 266.** K. B. Reid and A. J. Day, *Immunol. Today*, 1989, **10**, 177–180.
- 267.** M. D. Kirkitadze and P. N. Barlow, *Immunol. Rev.*, 2001, **180**, 146–161.
- 268.** J. M. Weiler, M. R. Daha, K. F. Austen and D. T. Fearon, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 3268–3272.
- 269.** G. J. Kotwal, S. N. Isaacs, R. McKenzie, M. M. Frank and B. Moss, *Science*, 1990, **250**, 827–830.
- 270.** D. G. Norman, P. N. Barlow, M. Baron, A. J. Day, R. B. Sim and I. D. Campbell, *J. Mol. Biol.*, 1991, **219**, 717–725.
- 271.** A. P. Wiles, G. Shaw, J. Bright, A. Perczel, I. D. Campbell and P. N. Barlow, *J. Mol. Biol.*, 1997, **272**, 253–265.
- 272.** I. D. Campbell and P. Bork, *Curr. Opin. Struct. Biol.*, 1993, **3**, 385–392.
- 273.** C. R. Savage, T. Inagami and S. Cohen, *J. Biol. Chem.*, 1972, **247**, 7612–7621.
- 274.** R. F. Doolittle, D. F. Feng and M. S. Johnson, *Nature*, 1984, **307**, 558–560.
- 275.** U. Hommel, T. S. Harvey, P. C. Driscoll and I. D. Campbell, *J. Mol. Biol.*, 1992, **227**, 271–282.
- 276.** H. S. Lu, J. J. Chai, M. Li, B. R. Huang, C. H. He and R. C. Bi, *J. Biol. Chem.*, 2001, **276**, 34913–34917.

277. P. A. Handford, M. Mayhew, M. Baron, P. R. Winship, I. D. Campbell and G. G. Brownlee, *Nature*, 1991, **351**, 164–167.
278. P. R. Winship and A. C. Dragon, *Br. J. Haematol.*, 1991, **77**, 102–109.
279. G. Nijbroek, S. Sood, I. McIntosh, C. A. Francomano, E. Bull, L. Pereira, F. Ramirez, R. E. Pyeritz and H. C. Dietz, *Am. J. Hum. Genet.*, 1995, **57**, 8–21.
280. N. Bolar, L. Van Laer and B. L. Loeys, *Curr. Opin. Pediatr.*, 2012, **24**, 498–504.
281. W. P. Russ and D. M. Engelman, *J. Mol. Biol.*, 2000, **296**, 911–919.
282. J. R. Baker, P. Arscott and J. Johnson, *Thyroid*, 1994, **4**, 173–178.
283. D. Foti, K. D. Kaufman, G. D. Chazenbalk and B. Rapoport, *Mol. Endocrinol.*, 1990, **4**, 786–791.
284. F. Grennan Jones, A. Wolstenholme, S. Fowler, S. Smith, K. Ziernicka, J. Bradbury, J. Furmaniak and B. Rees Smith, *J. Mol. Endocrinol.*, 1996, **17**, 165–174.
285. T. M. Yi and E. S. Lander, *J. Mol. Biol.*, 1993, **232**, 1117–1129.
286. Y. Kiso, J. Furmaniak, C. Morteo and B. R. Smith, *Autoimmunity*, 1992, **12**, 259–269.
287. A. B. Rawitch, G. Pollock, S. X. Yang and A. Taurog, *Arch. Biochem. Biophys.*, 1992, **297**, 321–327.
288. B. Czarnocka, J. Ruf, M. Ferrand, P. Carayon and S. Lissitzky, *FEBS Lett.*, 1985, **190**, 147–152.
289. K. D. Kaufman, B. Rapoport, P. Seto, G. D. Chazenbalk and R. P. Magnusson, *J. Clin. Invest.*, 1989, **84**, 394–403.
290. L. Fayadat, P. Niccoli-Sire, J. Lanet and J. L. Franc, *Endocrinology*, 1998, **139**, 4277–4285.
291. J. P. Banga, *Curr. Opin. Endocrinol. Diabetes Obes.*
292. J. Ruf, M. E. Toubert, B. Czarnocka, J. M. Durand-Gorde, M. Ferrand and P. Carayon, *Endocrinology*, 1989, **125**, 1211–1218.
293. Y. Nakajima, R. D. Howells, C. Pegg, E. D. Jones and B. R. Smith, *Mol. Cell. Endocrinol.*, 1987, **53**, 15–23.
294. V. Estienne, C. Dutheoit, L. Vinet, J. M. Durand-Gorde, P. Carayon and J. Ruf, *J. Biol. Chem.*, 1998, **273**, 8056–8062.
295. B. Czarnocka, D. Pastuszko, P. Carayon, J. Ruf and A. Gardas, *Autoimmunity*, 1996, **23**, 145–154.
296. T. Nishikawa, J. C. Jaume, S. M. McLachlan and B. Rapoport, *J. Clin. Endocrinol. Metab.*, 1995, **80**, 1461–1466.
297. E. Zanelli, M. Henry and Y. Malthiery, *Clin. Exp. Immunol.*, 1992, **87**, 80–86.
298. F. Grennan Jones, K. Ziernicka, J. Sanders, A. Wolstenholme, R. Fiera, J. Furmaniak and B. Rees Smith, *Autoimmunity*, 1999, **30**, 157–169.
299. T. Nishikawa, Y. Nagayama, P. Seto and B. Rapoport, *Endocrinology*, 1993, **133**, 2496–2501.

- 300.** D. Bresson, S. A. Rebuffat, B. Nguyen, J. P. Banga, A. Gardas and S. Peraldi-Roux, *Endocrinology*, 2005, **146**, 2834–2844.
- 301.** F. Libert, M. Ludgate, C. Dinsart and G. Vassart, *J. Clin. Endocrinol. Metab.*, 1991, **73**, 857–860.
- 302.** D. Bresson, M. Cerutti, G. Devauchelle, M. Pugnière, F. Roquet, C. Bes, C. Bossard, T. Chardès and S. Péraldi-Roux, *J. Biol. Chem.*, 2003, **278**, 9560–9569.
- 303.** W. Ebrahimizadeh and M. Rajabibazl, *Curr. Microbiol.*, 2014, **69**, 109–120.
- 304.** M. Dub ska, J. P. Banga, D. Plochocka, G. Hoser, E. H. Kemp, B. J. Sutton, A. Gardas and M. Gora, *Endocrinology*, 2006, **147**, 5995–6003.
- 305.** S. A. Rebuffat, *Int. Immunol.*, 2006, **18**, 1091–1099.
- 306.** V. Estienne, C. Dut hoit, S. Blanchin, R. Montserret, J.-M. Durand-Gorde, M. Chartier, D. Baty, P. Crayon and J. Ruf, *Int. Immunol.*, 2002, **14**, 359–366.
- 307.** M. Gora, A. Gardas, P. F. Watson, P. Hobby, A. P. Weetman, B. J. Sutton and J. P. Banga, *Biochem. Biophys. Res. Commun.*, 2004, **320**, 795–801.
- 308.** J. Guo, X. M. Yan, S. M. McLachlan and B. Rapoport, *J. Immunol.*, 2001, **166**, 1327–1333.
- 309.** N. Wedlock, J. Furmaniak, S. Fowler, Y. Kiso, J. Bednarek, A. Baumann-Antczak, C. Morteo, P. Sudbery, A. Hinchcliff and B. Rees Smith, *J. Mol. Endocrinol.*, 1993, **10**, 325–336.
- 310.** J. Hata, S. Yamashita, S. Yagihashi, H. Kato, S. Kabeno, K. Hirai, K. Kuma, S. Kimura, K. Umeki and T. Kotani, *Biochem. Biophys. Res. Commun.*, 1989, **164**, 1268–1273.
- 311.** M. Benvenuti and S. Mangani, *Nat. Protoc.*, 2007, **2**, 1633–1651.
- 312.** D. Bresson, S. A. Rebuffat and S. Péraldi-Roux, *J. Autoimmune Dis.*, 2005, **2**, 2.
- 313.** S. M. McLachlan and B. Rapoport, *Thyroid Off. J. Am. Thyroid Assoc.*, 2007, **17**, 939–948.
- 314.** S. M. Marino and V. N. Gladyshev, *J. Biol. Chem.*, 2012, **287**, 4419–4425.
- 315.** T. Nakashima and A. Taurog, *Clin. Chim. Acta Int. J. Clin. Chem.*, 1978, **83**, 129–140.
- 316.** D. R. Doerge, R. L. Divi and M. I. Churchwell, *Anal. Biochem.*, 1997, **250**, 10–17.
- 317.** D. L. Jarvis, *Methods Enzymol.*, 2009, **463**, 191–222.
- 318.** S. Ohtaki, T. Kotani and Y. Nakamura, *J. Clin. Endocrinol. Metab.*, 1986, **63**, 570–576.
- 319.** A. Gardas, A. Lewartowska, B. J. Sutton, Z. Pasieka, A. M. McGregor and J. P. Banga, *J. Clin. Endocrinol. Metab.*, 1997, **82**, 3752–3757.
- 320.** M. R. Wilkins, E. Gasteiger, A. Bairoch, J. C. Sanchez, K. L. Williams, R. D. Appel and D. F. Hochstrasser, *Methods Mol. Biol.*, 1999, **112**, 531–552.
- 321.** A. K. Dunker and R. R. Rueckert, *J. Biol. Chem.*, 1969, **244**, 5074–5080.
- 322.** R. Pitt-Rivers and F. S. Impiombato, *Biochem. J.*, 1968, **109**, 825–830.
- 323.** A. R. Williamson, M. R. Salaman and H. W. Kreth, *Ann. N. Y. Acad. Sci.*, 1973, **209**, 210–

224.

324. P. Gut, F. Grennan Jones, A. Sullivan, K. Ziemnicka, S. Smith, D. Jaskólski, J. Furmaniak and B. Rees Smith, *Thyroid*, 2000, **10**, 543–550.
325. A. B. Rawitch, A. Taurog, S. B. Chernoff and M. L. Dorris, *Arch. Biochem. Biophys.*, 1979, **194**, 244–257.
326. A. Taurog, M. L. Dorris, N. Yokoyama and C. Slaughter, *Arch. Biochem. Biophys.*, 1990, **278**, 333–341.
327. G. D. DePillis, S. i Ozaki, J. M. Kuo, D. A. Maltby and P. R. Ortiz de Montellano, *J. Biol. Chem.*, 1997, **272**, 8857–8860.
328. P. Kaur, H. Patil, P. B. Bhanushali, S. B. Badgugar and A. K. Gupta, *Int. J. Biol. Macromol.*, 2018, **106**, 87–94.
329. U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker and P. Nordlund, *Anal. Biochem.*, 2006, **357**, 289–298.
330. Z. Diwu, Y. Lu, C. Zhang, D. H. Klaubert and R. P. Haugland, *Photochem. Photobiol.*, 1997, **66**, 424–431.
331. M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane and F. R. Salemme, *J. Biomol. Screen.*, 2001, **6**, 429–440.
332. M. Vedadi, F. H. Niesen, A. Allali-Hassani, O. Y. Fedorov, P. J. Finerty, G. A. Wasney, R. Yeung, C. Arrowsmith, L. J. Ball, H. Berglund, R. Hui, B. D. Marsden, P. Nordlund, M. Sundstrom, J. Weigelt and A. M. Edwards, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15835–15840.
333. D. P. Frueh, A. C. Goodrich, S. H. Mishra and S. R. Nichols, *Curr. Opin. Struct. Biol.*, 2013, **23**, 734–739.
334. L. Skora, B. Shrestha and A. D. Gossert, *Methods Enzymol.*, 2015, **565**, 245–288.
335. R. Henderson and P. N. Unwin, *Nature*, 1975, **257**, 28–32.
336. R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann and K. H. Downing, *J. Mol. Biol.*, 1990, **213**, 899–929.
337. Y. Cheng, N. Grigorieff, P. A. Penczek and T. Walz, *Cell*, 2015, **161**, 438–449.
338. J. Jancarik and S. H. Kim, *J. Appl. Crystallogr.*, 1991, **24**, 409–411.
339. J. Newman, D. Egan, T. S. Walter, R. Meged, I. Berry, M. Ben Jelloul, J. L. Sussman, D. I. Stuart and A. Perrakis, *Acta Crystallogr. D Biol. Crystallogr.*, 2005, **61**, 1426–1431.
340. A. Kulkarni and C. Zukoski, *J. Cryst. Growth*, 2001, **232**, 156–164.
341. J. R. Luft, J. R. Wolfley and E. H. Snell, *Cryst. Growth Des.*, 2011, **11**, 651–663.
342. A. McPherson and J. A. Gavira, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, 2014, **70**, 2–20.
343. Y. G. Kuznetsov, A. J. Malkin and A. McPherson, *J. Cryst. Growth*, 2001, **232**, 30–39.
344. J. Newman, J. Xu and M. C. Willis, *Acta Crystallogr. D Biol. Crystallogr.*, 2007, **63**, 826–

- 345.** M. I. Aroyo and Internationale Union für Kristallographie, Eds., *International tables for crystallography. Volume A: Space-group symmetry*, Wiley, West Sussex, Sixth, revised edition., 2016.
- 346.** P. R. Evans, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 282–292.
- 347.** B. W. Matthews, *J. Mol. Biol.*, 1968, **33**, 491–497.
- 348.** M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 235–242.
- 349.** K. A. Kantardjieff and B. Rupp, *Protein Sci. Publ. Protein Soc.*, 2003, **12**, 1865–1871.
- 350.** I. Russo Krauss, F. Sica, C. A. Mattia and A. Merlini, *Int. J. Mol. Sci.*, 2012, **13**, 3782–3800.
- 351.** C. M. C. Loble, J. Sandy, J. Sanchez-Weatherby, M. Mazzorana, T. Krojer, R. P. Nowak and T. L. Sorensen, *Acta Crystallogr. Sect. Struct. Biol.*, 2016, **72**, 629–640.
- 352.** B. L. Hanson and G. J. Bunick, *Methods Mol. Biol.*, 2007, **364**, 31–42.
- 353.** J. W. Pflugrath, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, 2015, **71**, 622–642.
- 354.** J. Foadi, P. Aller, Y. Alguel, A. Cameron, D. Axford, R. L. Owen, W. Armour, D. G. Waterman, S. Iwata and G. Evans, *Acta Crystallogr. D Biol. Crystallogr.*, 2013, **69**, 1617–1632.
- 355.** D. Axford, R. L. Owen, J. Aishima, J. Foadi, A. W. Morgan, J. I. Robinson, J. E. Nettleship, R. J. Owens, I. Moraes, E. E. Fry, J. M. Grimes, K. Harlos, A. Kotecha, J. Ren, G. Sutton, T. S. Walter, D. I. Stuart and G. Evans, *Acta Crystallogr. D Biol. Crystallogr.*, 2012, **68**, 592–600.
- 356.** R. Bingel-Erlenmeyer, V. Olieric, J. P. A. Grimshaw, J. Gabadinho, X. Wang, S. G. Ebner, A. Isenegger, R. Schneider, J. Schneider, W. Glettig, C. Pradervand, E. H. Panepucci, T. Tomizaki, M. Wang and C. Schulze-Briese, *Cryst. Growth Des.*, 2011, **11**, 916–923.
- 357.** A. McPherson, *J. Appl. Crystallogr.*, 2000, **33**, 397–400.
- 358.** E. Bause and G. Legler, *Biochem. J.*, 1981, **195**, 639–644.
- 359.** X. Shi and D. L. Jarvis, *Curr. Drug Targets*, 2007, **8**, 1116–1125.
- 360.** M. Aebi, *Biochim. Biophys. Acta*, 2013, **1833**, 2430–2437.
- 361.** L. A. Palomares, I. K. Srivastava, O. T. Ramírez and M. M. J. Cox, *Adv. Biochem. Eng. Biotechnol.*, DOI:10.1007/10_2018_61.
- 362.** R. Kornfeld and S. Kornfeld, *Annu. Rev. Biochem.*, 1985, **54**, 631–664.
- 363.** E. Staudacher, F. Altmann, I. B. Wilson and L. März, *Biochim. Biophys. Acta*, 1999, **1473**, 216–236.
- 364.** C. Prenner, L. Mach, J. Glössl and L. März, *Biochem. J.*, 1992, **284 (Pt 2)**, 377–380.
- 365.** F. Altmann, H. Schwihla, E. Staudacher, J. Glössl and L. März, *J. Biol. Chem.*, 1995, **270**,

17344–17349.

366. H. Schachter, *Carbohydr. Res.*, 2009, **344**, 1391–1396.
367. P. Stanley, *Cold Spring Harb. Perspect. Biol.*, DOI:10.1101/cshperspect.a005199.
368. I. Marchal, D. L. Jarvis, R. Cacan and A. Verbert, *Biol. Chem.*, 2001, **382**, 151–159.
369. E. Ailor, N. Takahashi, Y. Tsukamoto, K. Masuda, B. A. Rahman, D. L. Jarvis, Y. C. Lee and M. J. Betenbaugh, *Glycobiology*, 2000, **10**, 837–847.
370. O. Gornik, T. Pavić and G. Lauc, *Biochim. Biophys. Acta*, 2012, **1820**, 1318–1326.
371. M. Donaldson, H. A. Wood, P. C. Kulakosky and M. L. Shuler, *Biotechnol. Bioeng.*, 1999, **63**, 255–262.
372. L. Joshi, T. R. Davis, T. S. Mattu, P. M. Rudd, R. A. Dwek, M. L. Shuler and H. A. Wood, *Biotechnol. Prog.*, 2000, **16**, 650–656.
373. A. L. Tarentino and F. Maley, *Biochem. Biophys. Res. Commun.*, 1975, **67**, 455–462.
374. R. B. Trimble and A. L. Tarentino, *J. Biol. Chem.*, 1991, **266**, 1646–1651.
375. A. L. Tarentino and T. H. Plummer, *Methods Enzymol.*, 1994, **230**, 44–57.
376. H. Hirayama, A. Hosomi and T. Suzuki, *Semin. Cell Dev. Biol.*, 2015, **41**, 110–120.
377. N. Shibuya, I. J. Goldstein, E. J. Van Damme and W. J. Peumans, *J. Biol. Chem.*, 1988, **263**, 728–734.
378. A. L. Tarentino and F. Maley, *J. Biol. Chem.*, 1974, **249**, 811–817.
379. D. Foti and B. Rapoport, *Endocrinology*, 1990, **126**, 2983–2988.
380. H. M. Baker, C. L. Day, G. E. Norris and E. N. Baker, *Acta Crystallogr. D Biol. Crystallogr.*, 1994, **50**, 380–384.
381. A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, S.-J. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu and L. M. Shewchuk, *Acta Crystallogr. D Biol. Crystallogr.*, 2007, **63**, 72–79.
382. R. P. Singh, A. Singh, H. V. Sirohi, A. K. Singh, P. Kaur, S. Sharma and T. P. Singh, *FEBS Open Bio*, 2016, **6**, 640–650.
383. A. K. Singh, N. Singh, M. Sinha, A. Bhushan, P. Kaur, A. Srinivasan, S. Sharma and T. P. Singh, *J. Biol. Chem.*, 2009, **284**, 20311–20318.
384. L. C. Kovari, C. Momany and M. G. Rossmann, *Struct. Lond. Engl.* 1993, 1995, **3**, 1291–1293.
385. Y. Yamaguchi, H. Kim, K. Kato, K. Masuda, I. Shimada and Y. Arata, *J. Immunol. Methods*, 1995, **181**, 259–267.
386. R. G. A. Jones and J. Landon, *J. Immunol. Methods*, 2002, **263**, 57–74.
387. E. Hendry, G. Taylor, F. Grennan-Jones, A. Sullivan, N. Liddy, J. Godfrey, N. Hayakawa, M. Powell, J. Sanders, J. Furmaniak and B. R. Smith, *Thyroid Off. J. Am. Thyroid Assoc.*, 2001, **11**, 1091–1099.

- 388.** K. Paul Friedman, E. D. Watt, M. W. Hornung, J. M. Hedge, R. S. Judson, K. M. Crofton, K. A. Houck and S. O. Simmons, *Toxicol. Sci. Off. J. Soc. Toxicol.*, 2016, **151**, 160–180.
- 389.** H. Hjelm, K. Hjelm and J. Sjöquist, *FEBS Lett.*, 1972, **28**, 73–76.
- 390.** K. F. M. Opuni, M. Al-Majdoub, Y. Yefremova, R. F. El-Kased, C. Koy and M. O. Glocker, *Mass Spectrom. Rev.*, DOI:10.1002/mas.21516.
- 391.** L. Yu, S. J. Gaskell and J. L. Brookman, *J. Am. Soc. Mass Spectrom.*, 1998, **9**, 208–215.
- 392.** J. A. Coffey, K. R. Jennings and H. Dalton, *Eur. J. Biochem.*, 2001, **268**, 5215–5221.
- 393.** C. E. Parker and K. B. Tomer, *Methods Mol. Biol.*, 2000, **146**, 185–201.
- 394.** W. Bode and P. Schwager, *J. Mol. Biol.*, 1975, **98**, 693–717.
- 395.** J. J. Perona and C. S. Craik, *Protein Sci.*, 1995, **4**, 337–360.
- 396.** E. Vandermarliere, M. Mueller and L. Martens, *Mass Spectrom. Rev.*, 2013, **32**, 453–465.
- 397.** L. Hedstrom, *Chem. Rev.*, 2002, **102**, 4501–4524.
- 398.** V. Kostka and F. H. Carpenter, *J. Biol. Chem.*, 1964, **239**, 1799–1803.
- 399.** J. Kohn and M. Wilchek, *Enzyme Microb. Technol.*, 1982, **4**, 161–163.
- 400.** J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Science*, 1989, **246**, 64–71.
- 401.** J. M. Wells and S. A. McLuckey, *Methods Enzymol.*, 2005, **402**, 148–185.
- 402.** I. A. Papayannopoulos, *Mass Spectrom. Rev.*, 1995, **14**, 49–73.
- 403.** J. Seidler, N. Zinn, M. E. Boehm and W. D. Lehmann, *Proteomics*, 2010, **10**, 634–649.
- 404.** P. Edman, *Arch. Biochem.*, 1949, **22**, 475.
- 405.** H. D. Niall, *J. Agric. Food Chem.*, 1971, **19**, 638–644.
- 406.** R. Finke, P. Seto and B. Rapoport, *J. Clin. Endocrinol. Metab.*, 1990, **71**, 53–59.
- 407.** M. R. Wilkins, I. Lindskog, E. Gasteiger, A. Bairoch, J. C. Sanchez, D. F. Hochstrasser and R. D. Appel, *Electrophoresis*, 1997, **18**, 403–408.
- 408.** M. Godlewska, B. Czarnocka and M. Gora, *Autoimmunity*, 2012, **45**, 476–484.
- 409.** Y. Kim, P. Quartey, H. Li, L. Volkart, C. Hatzos, C. Chang, B. Nocek, M. Cuff, J. Osipiuk, K. Tan, Y. Fan, L. Bigelow, N. Maltseva, R. Wu, M. Borovilos, E. Duggan, M. Zhou, T. A. Binkowski, R. Zhang and A. Joachimiak, *Nat. Methods*, 2008, **5**, 853–854.
- 410.** K. Tan, Y. Kim, C. Hatzos-Skitges, C. Chang, M. Cuff, G. Chhor, J. Osipiuk, K. Michalska, B. Nocek, H. An, G. Babnigg, L. Bigelow, G. Joachimiak, H. Li, J. Mack, M. Makowska-Grzyska, N. Maltseva, R. Mulligan, C. Tesar, M. Zhou and A. Joachimiak, *Methods Mol. Biol.*, 2014, **1140**, 189–200.
- 411.** Z. S. Derewenda and P. G. Vekilov, *Acta Crystallogr. D Biol. Crystallogr.*, 2006, **62**, 116–124.
- 412.** P. Sledz, H. Zheng, K. Murzyn, M. Chruszcz, M. D. Zimmerman, M. D. Chordia, A. Joachimiak and W. Minor, *Protein Sci.*, 2010, **19**, 1395–1404.

- 413.** A. Dong, X. Xu, A. M. Edwards, Midwest Center for Structural Genomics, Structural Genomics Consortium, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Vvakumov, A. Bochkarev, L.-G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, S. Flodin, A. Flores, S. Gräslund, M. Hammerström, M. D. Herman, B.-S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng and H. Zhu, *Nat. Methods*, 2007, **4**, 1019–1021.
- 414.** X. Gao, K. Bain, J. B. Bonanno, M. Buchanan, D. Henderson, D. Lorimer, C. Marsh, J. A. Reynes, J. M. Sauder, K. Schwinn, C. Thai and S. K. Burley, *J. Struct. Funct. Genomics*, 2005, **6**, 129–134.
- 415.** L. Vera, B. Czarny, D. Georgiadis, V. Dive and E. A. Stura, *Cryst. Growth Des.*, 2011, **11**, 2755–2762.
- 416.** L. Vera and E. A. Stura, *Cryst. Growth Des.*, 2014, **14**, 427–435.
- 417.** K. Michalska, K. Tan, C. Chang, H. Li, C. Hatzos-Skitges, M. Molitsky, R. Alkire and A. Joachimiak, *J. Synchrotron Radiat.*, 2015, **22**, 1386–1395.
- 418.** L. Jacquemet, J. Ohana, J. Joly, F. Borel, M. Pirocchi, P. Charrault, A. Bertoni, P. Israel-Gouy, P. Carpentier, F. Kozielski, D. Blot and J.-L. Ferrer, *Structure*, 2004, **12**, 1219–1225.
- 419.** A. le Maire, M. Gelin, S. Pochet, F. Hoh, M. Pirocchi, J. F. Guichou, J. L. Ferrer and G. Labesse, *Acta Crystallogr. D Biol. Crystallogr.*, 2011, **67**, 747–755.
- 420.** S. Jaeger, B. Fernandez and P. Ferrier, *Immunology*, 2013, **139**, 141–150.
- 421.** L. J. Delucas, D. Hamrick, L. Cosenza, L. Nagy, D. McCombs, T. Bray, A. Chait, B. Stoops, A. Belgovskiy, W. William Wilson, M. Parham and N. Chernov, *Prog. Biophys. Mol. Biol.*, 2005, **88**, 285–309.
- 422.** H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hämke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. H. Spence, *Nature*, 2011, **470**, 73–77.
- 423.** R. Neutze, R. Wouts, D. van der Spoel, E. Weckert and J. Hajdu, *Nature*, 2000, **406**, 752–757.

- 424.** L. C. Johansson, B. Stauch, A. Ishchenko and V. Cherezov, *Trends Biochem. Sci.*, 2017, **42**, 749–762.
- 425.** D. Oberthuer, J. Knoška, M. O. Wiedorn, K. R. Beyerlein, D. A. Bushnell, E. G. Kovaleva, M. Heymann, L. Gumprecht, R. A. Kirian, A. Barty, V. Mariani, A. Tolstikova, L. Adriano, S. Awel, M. Barthelmess, K. Dörner, P. L. Xavier, O. Yefanov, D. R. James, G. Nelson, D. Wang, G. Calvey, Y. Chen, A. Schmidt, M. Szczepak, S. Frielingsdorf, O. Lenz, E. Snell, P. J. Robinson, B. Šarler, G. Belšak, M. Maček, F. Wilde, A. Aquila, S. Boutet, M. Liang, M. S. Hunter, P. Scheerer, J. D. Lipscomb, U. Weierstall, R. D. Kornberg, J. C. H. Spence, L. Pollack, H. N. Chapman and S. Bajt, *Sci. Rep.*, 2017, **7**, 44628.
- 426.** T. A. White, *Methods Mol. Biol.*, 2017, **1607**, 325–347.
- 427.** P. Ball, *Nature*, 2017, **548**, 507–508.
- 428.** A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. S. Milne and S. Subramaniam, *Cell*, 2016, **165**, 1698–1707.
- 429.** R. M. Glaeser, *J. Struct. Biol.*, 1999, **128**, 3–14.
- 430.** M. Khoshouei, M. Radjainia, W. Baumeister and R. Danev, *Nat. Commun.*, 2017, **8**, 16099.
- 431.** R. Henderson, S. Chen, J. Z. Chen, N. Grigorieff, L. A. Passmore, L. Ciccarelli, J. L. Rubinstein, R. A. Crowther, P. L. Stewart and P. B. Rosenthal, *J. Mol. Biol.*, 2011, **413**, 1028–1046.
- 432.** F. J. Sigworth, *Microscopy*, 2016, **65**, 57–67.
- 433.** G. Roy and G. Mugesh, *Chem. Biodivers.*, 2008, **5**, 414–439.
- 434.** P. Rodien, A. M. Madec, J. Ruf, F. Rajas, H. Bornet, P. Carayon and J. Orgiazzi, *J. Clin. Endocrinol. Metab.*, 1996, **81**, 2595–2600.
- 435.** S. Ortiz-Urda, A. Elbe-Burger, J. Smolle, Y. Marquart, Y. Chudnovsky, T. W. Ridky, P. Bernstein, K. Wolff and K. Rappersberger, *J. Immunol.*, 2003, **171**, 6244–6250.
- 436.** Y. Xia, E. Eryilmaz, E. Der, R. D. Pawar, X. Guo, D. Cowburn and C. Puttermann, *Clin. Exp. Immunol.*, 2016, **183**, 369–379.
- 437.** R. Herrendorff, P. Hänggi, H. Pfister, F. Yang, D. Demeestere, F. Hunziker, S. Frey, N. Schaeren-Wiemers, A. J. Steck and B. Ernst, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E3689–E3698.
- 438.** A. Havasi, W. Lu, H. T. Cohen, L. Beck, Z. Wang, C. Igwebuike and S. C. Borkan, *Am. J. Physiol. Renal Physiol.*, 2017, **312**, F1016–F1025.
- 439.** F. Sanger, S. Nicklen and A. R. Coulson, *Proc. Natl. Acad. Sci. U. S. A.*, 1977, **74**, 5463–5467.
- 440.** Invitrogen, *Bac-to-Bac® Baculovirus Expression System user guide*, U.S.A, A.O., 2015, vol. MAN0000414.