Structural Studies on Determinants of Receptor/Ligand Binding in the Tumour Necrosis Factor and T Cell Receptor Protein Families

Jon Marles-Wright

Division of Structural Biology, the Wellcome Trust Centre for Human Genetics and Oriel College, Oxford



A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Summer 2005

Abstract

Structural Studies on Determinants of Receptor/Ligand Binding in the Tumour Necrosis Factor and T Cell Receptor Protein Families

Jon Marles-Wright

Division of Structural Biology, the Wellcome Trust Centre for Human Genetics and

Oriel College, Oxford

Thesis submitted for the degree of Doctor of Philosophy

Summer 2005

Protein-protein recognition plays a central role in the surveillance of self and non-self in the mammalian immune system and ultimately in cellular survival within the organism. Two systems of fundamental importance to the immune system are the Tumour Necrosis Factor (TNF) and the T cell receptor (TCR) families.

High-throughput methods developed within the Oxford Protein Production Facility have been successfully applied to the production of members of the TNF receptor and ligand superfamilies for structural characterisation. The TNF receptor DR6 was successfully refolded from *E.coli* inclusion bodies using a rapid-dilution technique and yielded diffraction quality crystals. Data collected from these crystals will be used to obtain an x-ray crystallographic model of DR6. Vascular Endothelial Growth Inhibitor (VEGI) was produced as a soluble recombinant protein in *E.coli,* and formed a number of poorly diffracting crystals, it is hoped that further trials and optimization of conditions will lead to improved data quality. Lymphotoxin β receptor was produced in a Eukaryotic system. This has shed light on the complications posed by signal peptide cleavage and glycosylation on the production of protein for crystallization trials.

TNF superfamily proteins are ideal targets for the design of novel therapeutic agents due to their involvement in a number of disease pathologies. Various methods of molecular docking and small molecule design were applied to the search for potential inhibitors of receptor binding for the TNF ligand proteins TRAIL and BAFF. A number of potential drug leads were identified from the National Cancer Institute drug database.

The Natural Killer (NK) T cell restricted TCRs recognise CD1d-presented glycolipid. Determination of the crystal structures of the invariant NK TCR and the NK restricted TCRs 5E and 5B shows that these proteins adopt the canonical structures of class I MHC restricted TCRs. This suggests that the binding of CD1d-glycolipid by these receptors will conform to the same model of binding seen for the class I MHC restricted TCRs.

Acknowledgements

My D.Phil could not have been completed without the help and encouragement of many people and I am grateful to them all. Firstly, I would like to thank Prof. Yvonne Jones for the opportunity to work within her group on some challenging projects. Prof. David Stuart and members of the Division of Structural biology have provided an excellent laboratory to work in.

I would particularly like thank James Brown for his help and good humour as he helped me find my feet within the laboratory, his involvement in the TNF project and for his comments and advice on the preparation of this thesis. I owe thanks to Michael Koch for his help and collaboration on the TCR project and to Stefan Gadola for the protein material for this project.

Thanks to all those involved in various aspects of the TNF project, Jo Nettleship, Anil Verma, Ray Owens, Ju Mongkolsapaya and Prof. Gavin Screaton. For making my synchrotron trips relatively painless and productive, I would like to thank Karl Harlos, Guillaume Stewart-Jones, Nicola Abrescia and Geoff Sutton. To all who offered advice, or help over the course of my studies and work in the laboratory, Margaret Jones, Ross Robinson, Christian Siebold, Radu Aricescu, Rishi Matadeen. To Robert Esnouf, Jon Diprose and Jun Dong for their first rate computing facilities and rescuing my data. To Anna Stamp, Masayo Kotaka and Jeff Ishuzuka for providing amusing lunches over the last year. Thanks to everyone who made my time outside of the laboratory in Oxford enjoyable and enlightening, particularly Tom Akehurst, Matthew Batchelor, Anita Isalska and Kristin Carideo for her superb grammar.

For My Family

Diane, Bill and Ann

Table of Contents

Introduction to Thesis **13**

General Introduction **13**

Aims of the project **14**

Chapter 1 OPPF TNF receptor and ligand project **15**

1.1 Introduction **16**
1.1.1 The Tumour Necrosis Factor Family 16
1.1.2 TNF Biology 17
1.1.3 TNF Signalling 20
1.1.4 Structural biology of the TNF receptors 25
1.1.5 Ligand/receptor binding 30
1.1.6 TNF receptor signalling 32
1.1.7 Aims 37

1.2 Materials and methods **38**
1.2.0 Enzymes and chemicals 38
1.2.1 Tumour Necrosis Factor Superfamily Ligands and Receptors 39
1.2.2 Construct design 40
1.2.3 Primer design 43
1.2.4 Cloning 44
1.2.A Use of T7 based vectors 44
1.2.5 OPPF Gateway™ cloning system 45
1.2.6 Protein expression in a prokaryotic system 49
1.2.B SDS-PAGE analysis 50
1.2.7 Refolding of insoluble prokaryotic expressed protein products 52
1.2.8 Purification of prokaryotic expressed proteins 53
1.2.9 Eukaryotic protein expression 54
1.2.10 Purification of Eukaryotic expressed proteins 55
1.2.C Light Scattering by proteins 57
1.2.11 Quality control 58
1.2.12 Crystallization 59

1.3 OPPF TNF project **60**
1.3.1 Overview of OPPF TNF project results 62
1.3.2 Cloning Results 63
1.3.3 Protein Expression Results 63
1.3.4 Purification of recombinant TRAIL 65
1.3.5 Purification of Recombinant DR6 67
1.3.6 Purification of Osteoprotegerin 70
1.3.7 Purification of soluble VEGI 73
1.3.8 Prokaryotic Protein refolding 76
1.3.9 Purification of Eukaryotic expressed Fc-tagged LTβR 81
1.3.10 Purification of Histidine tagged LTβR 84
1.3.11 Eukaryotic production of TNF receptor targets 86
1.3.12 Crystallization of TNF targets 88

1.4 Analysis of X-ray data from TNF protein crystals **89**
1.4.1 Preliminary X-ray analysis of DR6 crystals 89
1.4.A Molecular Replacement 92
1.4.2 Initial screening of VEGI crystals 98

1.5 Conclusions 100
1.5.1 OPPF pipeline 100
1.5.2 Refolding of protein material from E.coli inclusion bodies 101
1.5.3 Crystallization of TNF superfamily targets 101
1.5.4 Future Perspectives 102

Chapter 2 TNF/TNF receptor binding and potential small molecule inhibition 103

2.1 Introduction 104
2.1.1 Protein Recognition 104
2.1.2 TNF therapies 105
2.1.3 High-throughput and in silico drug discovery 105
2.1.4 Aims 106

2.2 Materials and methods 107
2.2.1 Analysis of TNF-receptor binding sites on TNF ligands 107
2.2.2 Small molecule generation by Ligbuilder 107
2.2.3 NCI database search using generated molecules 109
2.2.4 NCI Diversity set docking 109

2.3 Analysis of TNF ligands and their receptor binding sites 112
2.3.1 Suitability for small molecule binding 112
2.3.2 How to hit a moving target 112
2.3.3 Receptor/ligand complexes 113
2.3.4 Receptor binding surface analysis of TNF ligands 116
2.3.5 Determinants of receptor binding 119

2.4 Design of small molecule ligands for TNF family members 121
2.4.1 3D Pharmacophores and Database mining 121
2.4.A Pharmacophores 121
2.4.2 Docking 122
2.4.3 Pharmacophore modelling for TNF ligands 124
2.4.4 Small molecule design using Ligbuilder 127
2.4.B The Ligbuilder methodology 127
2.4.C The Lipinski rules 128
2.4.5 Database search with generated ligands 133
2.4.D SMILES 133
2.4.6 In Silico screening with the NCI diversity set 137

2.5 Conclusions 144
2.5.1 Pockets, Avidity and active sites 144
2.5.2 Ligbuilder 145
2.5.3 Automated docking in drug screening 145

Chapter 3 T cell receptor recognition of CD1d 146

3.1 Introduction 147
3.1.1 The Cellular Immune Response 147
3.1.2 T cell development 148
3.1.3 TCR recombination and specificity for MHC 151
3.1.4 The T cell receptor complex 151
3.1.5 Structural biology of the T cell Receptor 153
3.1.6 TCR/MHC binding 154
3.1.7 TCR/CD1 binding 156
3.1.8 CD1 ligand binding 158
3.1.9 Recognition of CD1d bound α-GalCer by NKT cell receptors 160
3.1.10 TCR/CD1 binding scheme 161
3.1.11 Aims 161

3.2 Determination of the structure of three NKT cell receptors 163
3.2.1 Crystallization and data collection 163
3.2.2 Structure determination and refinement 167
3.2.3 Structure of the *i*NKTCR, 5E and 5B TCRs 176
3.2.4 CDR loop architecture 177
3.A Canonical structures in the TCR Complementarity Determining Regions 180
3.2.5 Structural comparison of human T cell receptors 186

3.3 Conclusions 194
3.3.1 NKTCR Structures 194
3.3.2 CD1d binding mechanism 194
3.3.3 Future perspectives 195

Summary of Thesis 196

Appendices 198
Appendix A: The TNF ligand superfamily 199
Appendix B: The TNF receptor superfamily 200
Appendix Reagents 202
Appendix TNFR: sequences and primer design for TNF-receptor constructs 204
Appendix TNF Ligands: sequences and primer design for TNF-ligand constructs 221
Appendix XTAL: published structures of TNF ligand and receptors 234
Appendix PCR 237
Appendix Vectors 238
Appendix DLS 241
Appendix MS: Mass spectroscopy data for TNF targets 245
Appendix TCR: Published structures of T-Cell Receptors and T-Cell Receptor/MHC complexes. 249

Bibliography 252

List of Figures

Figure 1.1.1 Interactions of TNF ligands and receptors. 19
Figure 1.1.2 Variation between TNF ligand subunit structures 22
Figure 1.1.3 Trimeric arrangement of TNF ligands 23
Figure 1.1.4 TNF receptor schematic 26
Figure 1.1.5 Basic modules of TNF receptors 28
Figure 1.1.6 TNF/TNF receptor binding scheme 30
Figure 1.1.7 The TNFR1 signalling pathway 33
Figure 1.2.1 Workflow for the TNF family project 39
Figure 1.2.2 Sequence alignment of the tumour necrosis factor receptor targets 41
Figure 1.2.3 Sequence alignment of the tumour necrosis factor ligand targets 42
Figure 1.2.4 OPPF amplification and cloning protocol 46
Figure 1.2.5 The BP reaction 47
Figure 1.2.6 The LR reaction 48
Figure 1.3.1 purification of recombinant TRAIL 65
Figure 1.3.2 Characterisation of purified TRAIL by SDS-PAGE 66
Figure 1.3.3 Purification of refolded DR6 67
Figure 1.3.4 Characterisation of DR6 by SDS-PAGE 68
Figure 1.3.5 Example DR6 crystals 69
Figure 1.3.6 purification of refolded Osteoprotegerin 70
Figure 1.3.7 SDS-PAGE analysis of Osteoprotegerin 71
Figure 1.3.8 Purification of VEGI 73
Figure 1.3.9 SDS-PAGE analysis of VEGI 74
Figure 1.3.10 VEGI crystals. 75
Figure 1.3.11 purification of Fc-tagged LTβR 81
Figure 1.3.12 SDS-PAGE analysis of LTβR 82
Figure 1.3.13 Example crystals of LTβR 83
Figure 1.3.14 purification of histidine tagged LTβR 84
Figure 1.3.15 SDS-PAGE analysis of LTβR 85
Figure 1.4.1 Diffraction pattern from DR6 crystals. 90
Figure 1.4.2 Crystals of DR6 grown in the Salt RX screen 97
Figure 1.4.3 Crystals of VEGI and x-ray diffraction pattern produced by the crystals 99
Figure 2.2.1 Conceptual workflow of AutoDock and NCI database search 111
Figure 2.3.1a Electrostatic interaction surface of TNF Ligands 117
Figure 2.3.1b Electrostatic interaction surface of TNF Ligands 118
Figure 2.4.1 Receptor ligand contacts for TRAIL/DR5 and BAFF/BAFFR 126
Figure 2.4.2 Small molecule products of Ligbuilder 131
Figure 2.4.3 NCI molecules binding within the TRAIL and BAFF receptor binding grooves from AutoDock screening. 139
Figure 3.1.1 T cell Receptor gene rearrangements 150
Figure 3.1.2 The T cell receptor signalling complex 152
Figure 3.1.3 T cell Receptor structure 153
Figure 3.1.4 TCR/MHC I binding 156
Figure 3.1.5 The structure of human CD1d with α-GalCer 159
Figure 3.1.6 Alignment of three NKT cell receptor sequences 162
Figure 3.2.1 X-ray diffraction images for 5E TCR 164
Figure 3.2.2 X-ray diffraction image for 5B TCR 166

Figure 3.2.3 Ramachandran plot for stereochemistry observed in the 5E TCR model 169
Figure 3.2.4 Ramachandran plot for stereochemistry observed in the 5B TCR model 171
Figure 3.2.5 Ramachandran plot for stereochemistry observed in the *i*NKTCR model. 174
Figure 3.2.6 Structure of The 5E, 5B and NKT T cell receptors 176
Figure 3.2.7 $2F_o$-F_c Electron density of 5E TCR CDR loops 182
Figure 3.2.8 $2F_o$-F_c Electron density of 5B TCR CDR loops 183
Figure 3.2.9 $2F_o$-F_c Electron density of TCR NKT CDR loops 184
Figure 3.2.10 Omit maps for CDR3 regions of The NKT cell receptors 185
Figure 3.2.11 Structural alignment of human T cell receptors 186
Figure 3.2.12 Alignment of the structures of human T cell receptor CDR loops 188
Figure 3.2.13 Alignment of T cell receptor CDR alpha regions 190
Figure 3.2.14 Alignment of T cell receptor CDR beta regions 191
Figure 3.2.15 CD1 recognition surfaces of T cell receptors 193

List of Tables

Table 1.3.1 Overview of OPPF TNF project results 62
Table 1.4.1 processing statistics for DR6 90
Table 1.4.2 processing statistics for DR6 96
Table 1.4.3 Anomalous signal analysis from SHELXC for DR6 96
Table 2.3.1 Solvent accessible surface area and contact areas of TNF receptor/ligand complexes 115
Table 2.4.1 Representative molecules generated by Ligbuilder against BAFF and TRAIL using the growing and linking strategies of the programme. 130
Table 2.4.2 Molecules identified by NCI similarity searching using molecules generated against the TRAIL receptor-binding site by LigBuilder. 135
Table 2.4.3 Molecules identified by NCI similarity searching using molecules generated against the BAFF receptor-binding site by LigBuilder. 136
Table 2.4.4 Potential TRAIL receptor antagonists from the NCI diversity set identified by virtual screening 140
Table 2.4.5 Potential BAFF receptor antagonists from the NCI diversity set identified by virtual screening 141
Table 3.2.1 Statistics for crystallographic data collection and refinement 175
Table 3.2.2 CDR loop architecture of human T cell receptors 189

All protein crystallographic models are referred to using their PDB accession codes and references for these can be found in the Appendices.

Abbreviations

α-GalCer	α-galactosylceramide
A_{260}	Absorbance at 280 nm
A_{280}	Absorbance at 280 nm
APRIL	A Proliferation Inducing Ligand
β2m	β2-microglobulin
BAFF	B Cell Activating Factor
BAFFR	BAFF Receptor
BCMA	B Cell Maturation Antigen
C1q	Complement 1q
CCD	Charge Coupled Device
CCDC	Cambridge Crystallographic Data Centre
CCP4	Collaborative Computational Project Number 4
CD	Cluster of Differentiation
CDR	Complementarity Determining Region
CRD	Cysteine Rich Domain
DcR	Decoy Receptor
DD	Death Domain
DED	Death Effector Domain
DLS	Dynamic Light Scattering
DN	$CD4^-$, $CD8^-$ Double Negative
DNA	Deoxyribonucleic Acid
DR	Death Receptor
DTT	Dithiothreitol
E.coli	Escherichia coli
EDA	Ectodysplasin A
EDTA	Ethylene Diamine Tetra Acetate
EPO-R	Erythropoeietin Receptor
ESRF	European Synchrotron Radiation Facility
FADD	Fas Associated Death Domain
FCS	Foetal Calf Serum
FLIP	FLICE-like Inhibitory Protein
FPLC	Fast Performance Liquid Chromatography
GWF	Gateway™ Forward primer sequence
GWR	Gateway™ Reverse primer sequence
HLA	Human Leukocyte Antigen
HPLC	High Performance Liquid Chromatography
HVEM	Herpes Virus Entry Mediator
ICAM-1	Intercellular Adhesion Molecule 1
iGb3	Isoglobotrihexosylceramide
IMAC	Immobilised Metal-ion Affinity Chromatography
IPTG	Iso-βD-Thiogalactoside
JAK	Janus Kinase
LB	Luria-Bertani
LFA-1	Lymphocyte Function-Associated 1
LIGHT	Herpes Virus Entry Mediator Ligand
LTα	Lymphotoxin α
LTβ	Lymphotoxin β
LTβR	Lymphotoxin β Receptor
MGC	Mammalian Genome Collection
MHC	Major Histocompatibility Complex
MPD	2-Methyl-2-4-pentanediol
MRC	Medical Research Council
MS	Mass Spectroscopy
MWCO	Molecular Weight Cut-off
NCI	National Cancer Institute
NF-κB	Nuclear Factor κB
NK	Natural Killer
NKTCR	Natural Killer T cell receptor

OD	Opdical Density
OPG	Osteoprotegerin
OPPF	Oxford Protein Production Facility
PAGE	Poly Acrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PEI	poly(ethylenimine)
PLAD	Pre-Ligand Association Domain
POI	Protein of Interest
POV	Persistence of Vision
Q-TOF	Quadrupole Time Of Flight
RANKL	Receptor Activator of NF-κB
RBS	Ribosome Binding Site
RELT	Receptor Expressed In Lymphoid Tissues
RIP	Receptor-Interacting Protein
rmsd	Root Mean Square Deviation
RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulphate
SLE	Systemic Lupus Erythematosus
SMILES	Simple MolecularInput Line Entry System
SODD	Silencer Of Death Domain
SPDBV	Swiss PDB Viewer
TACI	Trans membrane Activator and CAML Interactor
TCR	T Cell Receptor
THD	TNF Homology Domain
TNF	Tumour Necrosis Factor
TNFR	Tumour Necrosis Factor Receptor
TRADD	TNFR Associated Death Domain
TRAF	TNFR Associated Factor
TRAIL	TNF Related Apoptosis Inducing Ligand
TRAPS	TNF Associated Periodic Syndrome
TROY	Toxicity and JNK inducer
TWEAK	TNF-related Weak Incuder of Apoptosis
VEGI	Vascular Endothelial Growth Inhibitor
XLHED	X-Linked Hypohidrotic Ectodermal Dysplasia

Amino Acid Single and three letter codes

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

Introduction to Thesis

General Introduction

Cell proliferation and organisation must be tightly regulated in the multi-cellular organism to ensure the correct action of tissues and organs. Interactions occurring at the cell surface between surface proteins, and those in the stroma, act to control and regulate these processes. Protein-protein recognition plays a central role in the mammalian immune system in the surveillance of self and non-self and ultimately, in cellular survival within the organism. Two systems of fundamental importance to the immune system are the Tumour Necrosis Factor (TNF) and the T cell receptor (TCR) families. The TNF family proteins are involved in the signalling of growth and proliferation, and apoptosis, across a wide range of cell types and milieu. T cell receptor proteins act in the recognition of antigens presented by the Major Histocompatibility Complex (MHC) and CD1 proteins as part of the immune response by T cells.

Interest in these systems has grown out of their role in disease and the subsequent discovery of related proteins through genomics initiatives. Structural studies on these molecules have led to a greater understanding of the manner in which they fulfil their roles. These families of proteins use a core-scaffold common to members of each family, to facilitate recognition of different molecules. Mechanisms and determinants of recognition can be elucidated from structural studies, with obvious medical benefits coming from the potential for drug and small molecule interference in these events.

Aims of the project

The aims of this project can be summarised as three main strands, which are described as follows.

- To express and purify a number of TNF receptor and ligand constructs in collaboration with the Oxford Protein Production Facility for crystallographic studies of receptor/ligand binding in the TNF super families.
- To carry out an analysis of determinants of receptor/ligand binding from those structures of TNF receptors and ligands that exist in the protein database and assess these for potential small-molecule inhibition using ligand design software and molecular docking.
- To solve the crystal structures of a number of novel, CD1-glycolipid recognising T cell receptors and assess determinants of ligand specificity in these proteins.

Chapter 1 OPPF TNF receptor and ligand project

This chapter describes the implementation of the OPPF high throughput structural genomics pathway to the TNF receptor and ligand families. The introduction provides an overview of the interactions between the extracellular regions of the TNF receptor and ligand superfamily members and the role of these proteins in humans. The aim of this chapter is the recombinant expression and purification of a number of TNF receptor and ligand targets for structural studies.

1.1 Introduction

1.1.1 The Tumour Necrosis Factor Family

Historically, the existence of Tumour Necrosis Factor (TNF) was first postulated during the Crimean war as an agent responsible for the shrinkage of tumours in patients suffering from severe bacterial infection; this was termed cachectin [1]. It was not until the 1970's that TNF was first isolated and the TNF family was described as a group of proteins in mammals whose levels increase subsequent to stimulation of the immune system [2,3].

TNF receptor family members are known to be involved in a number of diverse pathways, including the apoptotic events activated by caspases within the cell and the NF-κb proliferation pathway. The involvement of TNFs in these systems has been widely reviewed [4,5]. Downstream signalling pathways activated by the family are well known and are understood to a large degree. However, the selectivity mechanisms by which cells respond to different TNF ligand signals, in state-dependent situations, are currently unclear and represent a challenge in the growing field of systems biology as we seek to integrate results of the numerous signals that cells receive and act on.

A substantial amount of research has gone into the dissection of the structural and biochemical pathways in which the growing family of TNF ligands and receptors are implicated. As such, a wealth of protein-structures and biochemical data exists from which to piece together the pathways for the apoptosis inducing members of the family [6].

1.1.2 TNF Biology

a. Immunity

Members of the TNF and TNF Receptor (TNFR) family are primarily involved in immune system responses, although a number of proteins are implicated in the bone formation and developmental processes[7]. The position of these proteins within cellular signalling couples them to the basic processes of cell proliferation, survival and differentiation. The role of these proteins in immunity is starkly highlighted in diseases that are directly linked to the breakdown of correct signalling response, such as Systematic Lupus Erythematosus (SLE) and rheumatoid psoriasis [8]. A polymorphism in the TNFR2 gene leads to susceptibility to SLE, and in rheumatoid psoriasis over expression of TNF within the joints causes inappropriate inflammation and leads to the further progression of the disease [9,10].

b. Expression

The expression of TNF ligands varies considerably, from the highly localized expression of Ectodysplasin A (EDA), which is restricted to foetal ectodermal tissue, to the systematic expression of TNF Related Apoptosis Inducing Ligand (TRAIL). Similarly, TNFRs have a wide range of expression levels; many of the receptors localize to the haematopoietic organs, while TRAIL receptors are found more broadly distributed.

c. Interaction Patterns

The current map of TNF/TNFR interaction is reviewed in Bodmer et al. [11] (Figure 1.1.1, after Bodmer). Those TNF ligands with a wide tissue distribution, such as TRAIL and Lymphotoxin, have a number of interaction partners. This could be

ascribed to their involvement in a number of distinct pathways or, in the case of TRAIL, a need to regulate its systemic activation by using a complement of decoy receptors. A number of the ligands have only one known receptor and the signalling pathways in which they are involved are likely to be tightly controlled at the level of expression and cell surface presentation.

No binding partners for the TNF receptors, Death Receptor 6 (DR6), Toxicity and JNK Inducer (TROY) and Receptor Expressed in Lymphoid Tissues (RELT) have yet been found.

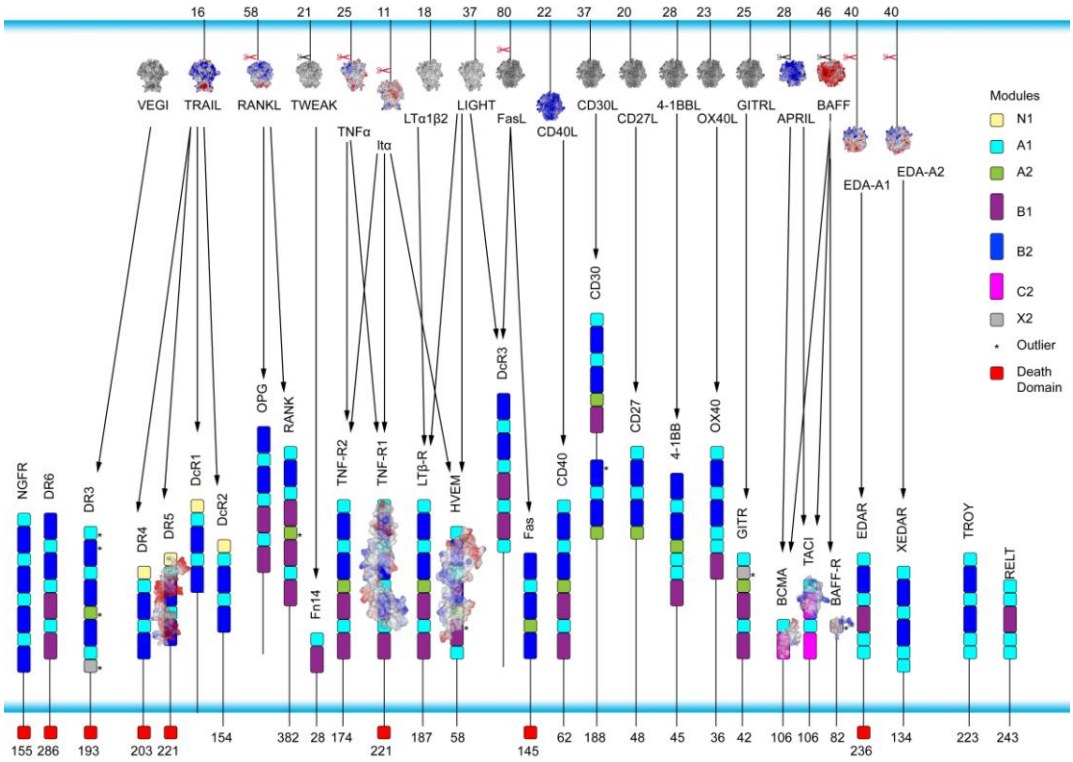


Figure 1.1.1 Interactions of TNF ligands and receptors.
TNF ligands are shown at the top and TNF receptors at the bottom of the figure. The TNF ligands are represented as cartoon structures; coloured for those that have crystallographic models and in grey for those that have yet to be solved. Red scissors represent processing by furin proteases, while black scissors represent processing by other proteases.

The TNF receptors are depicted schematically with N1, A1, A2, B1, B2, C2 and X2 modules, stars denote modules whose cysteine connectivity does not conform to the canonical pattern. Those proteins for which a model has been solved are shown against their schematic. The lengths of intracellular domains, in number of amino acids, are shown for both receptors and ligands and the presence of a death domain is shown as a red box for receptors.

Receptors and ligands are not drawn to scale.

1.1.3 TNF Signalling

The selectivity and signalling mechanisms of the TNF receptor family members and their ligands present an interesting and exciting challenge to the structural biologist and come with the baggage of more than a small amount of controversy. Well-established evidence shows that TNF ligands exist as trimeric complexes, either tethered to the cell surface as Type II membrane proteins or free in the inter-cellular medium after metalloprotease cleavage. These ligands engage their receptor partners within three binding grooves that form at the monomer interfaces [12,13]. Compared to the TNF ligands, TNFRs display a more complex array of pre- and post-ligand binding events which are suggested to ensure maximal effectiveness and fidelity of signalling [14].

The TNF ligand family consists of nineteen type II membrane proteins each with a homologous extracellular C-terminal domain, termed the TNF-homology domain (THD) (Appendix A). TNF like ligands are found as trimers, which is the functional arrangement of these proteins. Four members of the family can form hetero-trimers (BAFF/APRIL, LTα/LTβ). Family members have a sequence identity of 20-30% and share a conserved hydrophobic core that is rich in aromatic residues [11,15].

The activity of TNF ligands can be modulated by their solubilisation: a number of the proteins can exist freely in solution after limited proteolytic cleavage by metalloproteases. Solubilisation by the metalloprotease furin is necessary for the function of EDA, where a mutation in the furin recognition sequence can cause the disorder X-linked hypohidrotic ectodermal dysplasia (XLHED) [16]. Conversely, the soluble form of FAS is dramatically less cytotoxic than the membrane bound form [17].

a. Structure

The structure of TNF was solved by Jones et al. [18]; this crystallized as a trimeric beta-jellyroll structure with a suggested structural homology to viral capsid proteins (Figure 1.1.2a). Two stacked beta-sheets consisting of five strands each form the core of the protein. The inner sheet (strands A, A', H, C and F) forms the trimer interface and the outer sheet (strands B, B', D, E and G) is exposed to solvent. Structures of a number of other TNF ligands have been solved (TRAIL, LTα, RANKL, CD40L, BAFF, murine APRIL and EDA) and are shown structurally superposed in Figure 1.2b (A list of TNF ligands for which a crystal structure exists is found in Appendix Xtal).

Each protein conforms strongly to the same basic structure as the original TNF-ligand, with a trimer formed from beta-jellyroll monomers, interacting through one beta-sheet with a highly conserved pattern of aromatic residues and forming receptor-binding surfaces at the interface between subunits. TNF and CD40 ligand contain a disulphide link between C–D and E–F loops; similar features are predicted in Fas ligand, LIGHT, CD30L and CD27. A disulphide bridge is found between strands E and F of BAFF and from sequence homology, this feature is also suggested to exist for TWEAK, EDA and APRIL.

b. Family heterogeneity

Major differences between TNF-ligands are found at the solvent exposed surfaces, which make up the binding sites for their cognate receptors. Surface charge and shape differences appear to confer receptor specificity. The most striking example of this is seen in the structure of BAFF, which deviates from the canonical TNF structure in two regions: the D–E loop (Figure 1.1.2a orange region between strands D and E) and the elbow region of loop A''–A''' [19] (Figure 1.1.2a purple region between strands A'' and A''').

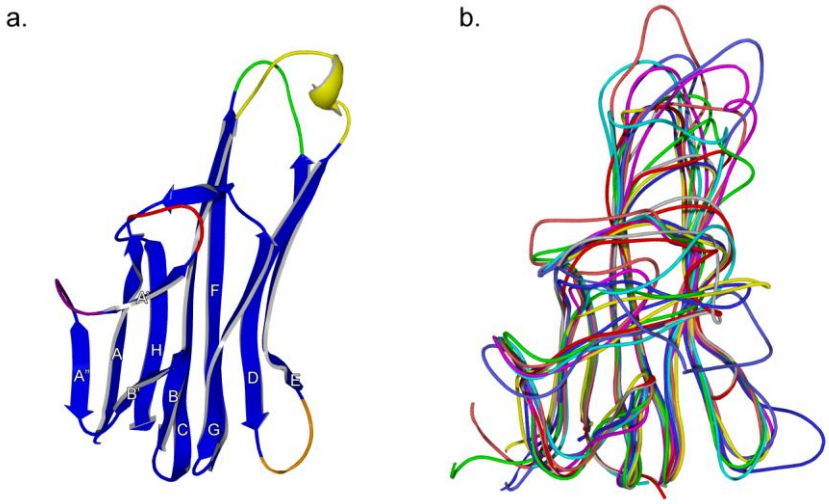


Figure 1.1.2 Variation between TNF ligand subunit structures
a. A cartoon schematic of the monomeric TNF ligand structure is depicted with the beta jellyroll strands labelled A-H and regions of high variation between the published structures are shown in yellow, green, red, pink and orange.

b. Structural superposition of TNF ligand monomer subunits, showing conserved regions and variation. The TNF ligand monomers are coloured accordingly: APRIL, yellow; BAFF, blue; CD40L, green; EDA-A1, red; EDA-A2, grey; Lymphotoxin α, pink; RANKL, cyan; TNFα, salmon; TRAIL, violet (references for the structure of each of these are in Appendix Xtal).

c. Trimeric complexes

The trimeric structure of the TNF ligands presents a compact and highly stable scaffold for the recruitment of a diverse range of receptors (Figure 1.1.3). The proteins do not display particularly high sequence homology, but the central core of the trimer is highly homologous. For both TRAIL and BAFF, this trimer is further stabilized by metal ion co-ordination by cysteine residues. In the TRAIL trimer, incomplete Zinc ion co-ordination causes a disulphide linked hepatotoxic species to form, whereas BAFF does not form trimers in the absence of magnesium [20,21].

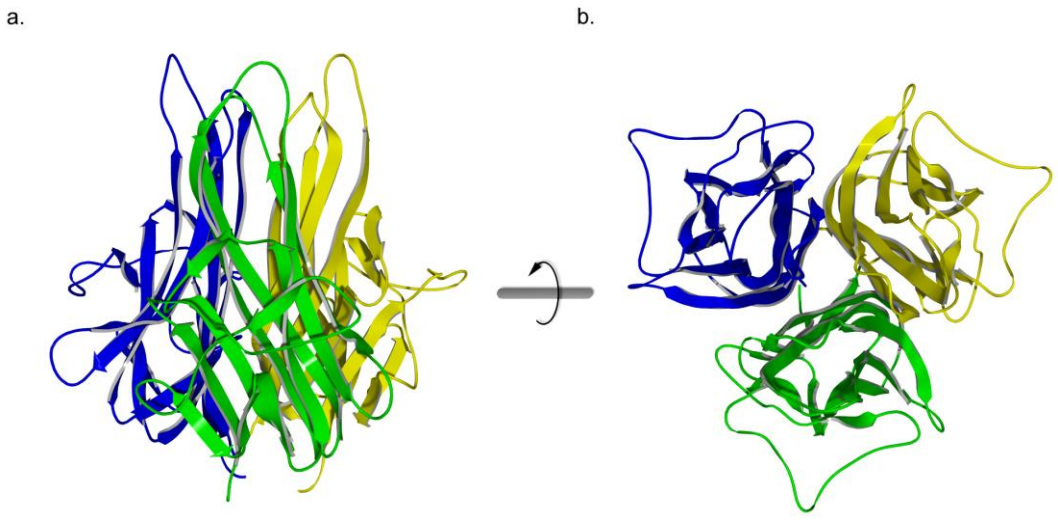


Figure 1.1.3 Trimeric arrangement of TNF ligands
a. secondary structure cartoon of the active trimeric structure of TRAIL with the three protein chains coloured blue, green and yellow, with the cell membrane at the bottom of the figure.
b. Orthogonal view of a, viewed down the pseudo-threefold axis of the trimer.

Controversial evidence for the formation of higher-order TNF-ligand structures arising from the solution of the structure of BAFF has led to heated debate on whether this is a crystallographic artefact caused by pH dependent multimerisation of histidine tags or is a truly biological state for the molecules [22,23]. The structure presented seems

to exist in a conformation that would preclude receptor interaction with many of the sixty BAFF trimers that make up the assembly and thus this is most likely to be an interesting quirk of the crystallization conditions of the protein rather than being biologically relevant.

d. Structural Homology

The crystal structure of the globular domain of the mouse complement-1Q family protein ACRP30 revealed a highly homologous structure to the THD (Shapiro and Scherer, 1998). The C1q family proteins recognize immune complexes and trigger the classical complement pathway [24]. Although the proteins share limited sequence homology with TNF (four residues in the hydrophobic core) the overall β-sheet structure is highly conserved. This points to the evolutionary conservation of structure over sequence, or convergent evolution [25]. To date there is no evidence that the receptor interaction scheme of the TNF ligands is seen in the C1q proteins.

The overall fold of the THD has been noted to look similar to viral capsid proteins of small plant viruses (Satellite Tobacco Necrosis Virus) and picornaviruses (including Foot and Mouth Disease Virus and the common Rhinoviruses) [18]. Where the THD associates along a threefold axis and the viral proteins form a fivefold axis the connectivity of the β-strands is almost identical (excepting strands A and B in the viral proteins, which have no intervening loops). Rather than suggesting an evolutionary relationship, this oligomerisation most likely shows the propensity of the jellyroll structure to form higher order structures.

1.1.4 Structural biology of the TNF receptors

The TNF receptors are a group of modular proteins consisting of a variable number of structurally conserved cysteine rich domains (CRD), which make up the N-terminal extra-cellular portion of the protein, a membrane-spanning domain and an intracellular region that often contains a Death Domain (DD) that is conserved across a number of members of the family. To date, twenty-nine TNF receptors have been identified in humans (Appendix B). These are mainly type I transmembrane proteins, although a number (BAFFR, BCMA, TACI and XEDAR) are type III proteins with no signal peptide. The architecture and structure/function relationships of the TNF receptors has been reviewed extensively [11,26].

a. Classification

TNF receptors have been classified based on the presence of an intracellular Death Domain region that functions to recruit TNF Related Death Domain (TRADD) and Fas Associated Death Domain (FADD) family proteins. Receptors with a DD are responsible for the activation of the caspase dependent apoptosis pathway. A number of TNF receptor proteins exist which lack a functional cytoplasmic domain and these have been termed decoy receptors for their proposed role in modulating functional receptor activation. Decoy receptors can be membrane anchored, as is the case for DcR1 and DcR2, or soluble in the case of OPG and DcR3, which is a decoy for FasL and is up-regulated in a number of tumours [27]. This increased expression protects against cytotoxic effects of FasL and thus confers an advantage to survival of tumour cells that possess this adaptation.

b. TNF receptor structure

TNF receptors typically possess three cysteine rich domains, but can have between one and four of these modules (although CD30 has a number of duplicate domains). These domains form a twisted disulphide 'ladder' that gives the proteins an elongated, slender appearance as shown in Figure 1.1.4. The BAFF receptors [19,28-30] are interesting in that they present a minimal TNF binding region. An engineered version of the BAFF receptor possesses only half of a CRD, yet ligand binding experiments have shown that this minimal BAFF-binding domain has a nanomolar-affinity for its ligand, comparable to that of the full two-domain ligand-binding motif found in TNFR1 and DR5.

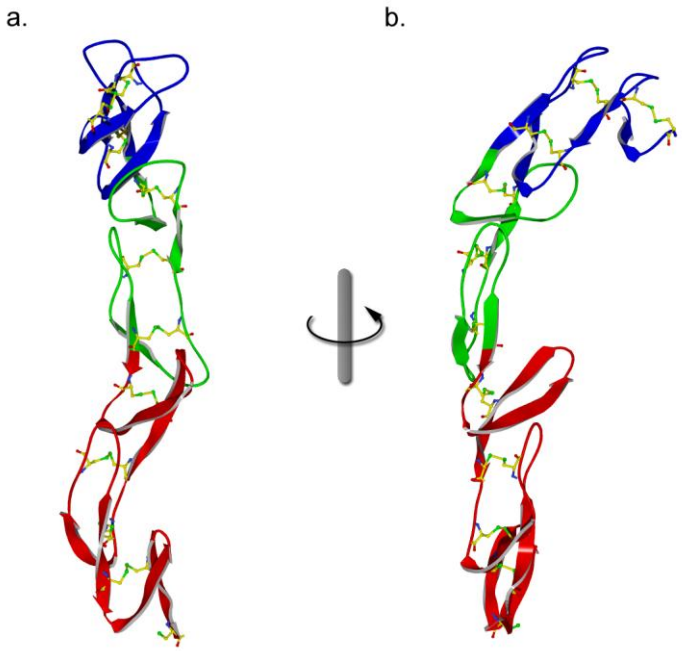


Figure 1.1.4 TNF receptor schematic
a. A schematic of TNFR1 is shown with cysteine residues shown in CPK representation and the domain boundaries coloured blue for CRD1, green for CRD2 and red for CRD3.
b. Rotation 90 º into the plane of the paper of a.

c. CRD modularity

Receptor CRD modules usually contain six cysteine residues, forming three disulphide linkages; because these repeats are not strictly conserved the sequence alignment of TNF receptors has in the past proven difficult. Naismith and Sprang [31] developed a structural classification of TNF-modules based on distinct motifs from members of the family for which crystal structures have been solved. These are termed A, B, C and N for repeats that have crystal structures and X for those for which no structural information was available, the number of disulphide linkages found is indicated with a number. Typically, a CRD is formed by an A1-B2 or an A2-B1 module (Figure 1.1.5). The conservation of these modules has allowed sequence database mining techniques to identify many new TNF receptors [32].

A-modules

A1 modules have a consensus sequence of Cys1-X_2-Gly-X-Tyr/Phe-X_{5-10}-Cys2 (where X_n is a number 'n' of intervening amino acids). This module has a distinctive C shape consisting of three short β-strands linked by turns, with strands one and three connected by the disulphide bond (Figure 1.1.5a). A2 are similar to A1 modules, but have an additional disulphide linkage (Cys3-Cys4). The consensus sequence is Cys1-X_2-Gly-X-Tyr/Phe-Cys3-X_{4-9}-Cys4-X_2-Cys2.

B-Modules

The B2 module adopts a meat-hook-like S-shaped fold of three anti-parallel strands. The consensus sequence is Cys1-X_2-Cys2-X_{3-6}-X_5-Cys3-Thr-X_{2-5}-Asn-Thr-Val-Cys4, with disulphide bonds between Cys1-Cys3 and Cys2-Cys4 (Figure 1.1.5b). The Asn-Thr-Val motif stabilizes the structure, with the conserved threonine forming a

hydrogen bond to a backbone amide in the third strand [12]. The B1 module lacks the Cys1-Cys3 disulphide but otherwise forms an almost identical structure.

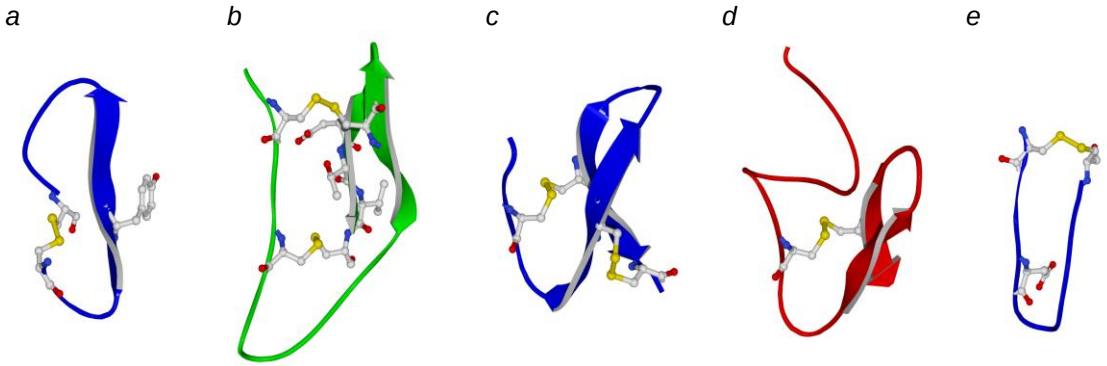


Figure 1.1.5 Basic modules of TNF receptors
All panels are shown as cartoon representation with cysteine and conserved residues are shown in ball and stick representation with oxygen in red, nitrogen in blue, carbon in white and sulphur in yellow.
a) A1 with the conserved tyrosine shown
b) B2 with conserved Asn-Thr-Val motif
c) C2 showing different disulphide connectivity to the B2 module
d) N1 with extended N-terminal loop
e) The minimal BAFF receptor module

Other modules

The connectivity of a number of CRD modules differs to that of the A and B modules. The C2 module shown in figure 1.3c is similar to a B2 module, but has differing disulphide connectivity. DR5 possesses an N1, module at its N-terminal end, which is again a unique module (Figure 1.1.5d).

BAFFR module

The key region of the BAFF receptor is a very small, non-canonical, nineteen amino-acid structure stabilized by a disulphide link that differs from that seen in other TNF receptor family members; it presents an interesting minimal binding loop structure that has been suggested could be used therapeutically (Figure 1.1.5e).

d. Structural homology

Cysteine rich TNF-like repeats also form the structural basis of the Epidermal Growth Factor (EGF) like laminin proteins [33]. These proteins also contain unique short-linker domains, which connect CRD modules. Laminins bind a connecting protein that forms networks within the basement membrane; the γ-1 chain of laminin binds to Nidogen-1, but this protein is structurally unrelated to the TNF ligands.

Structural superposition of a number of Low-density Lipoprotein Complement repeat domains onto TNFR1 shows that the proteins have A-module-like repeats at their N-termini [34,35].

Both insulin receptor proteins and the small T-knot proteins have been suggested to share some structural homology with the TNF receptor family [36,37]. Closer inspection of their structures reveals that they show too little sequence homology to say whether they are indeed related; the more plausible explanation is that some form of convergent evolution is acting on their structures.

1.1.5 Ligand/receptor binding

a. Structural studies

The structure of TNF in complex with its receptor TNF Receptor 1 (TNFR1) was solved as the trimeric TNF ligand with three copies of TNFR1 binding at the interfaces between TNF monomers (Figure 1.1.6) [12]. Six years later, the structure of TRAIL/DR5 was solved independently by a number of groups and the arrangement of this receptor-ligand pair was highly similar to that of the TNF/TNFR1 structure [13,38,39]. Of note is the fact that the geometry of this arrangement matches that of the intracellular binding partner of TNFR2 and CD40 - TRAF-2 [40]; and thus a scheme for activation of downstream signalling pathways presents itself. There are no conformational changes of the trimeric TNF type molecule on receptor-ligand binding.

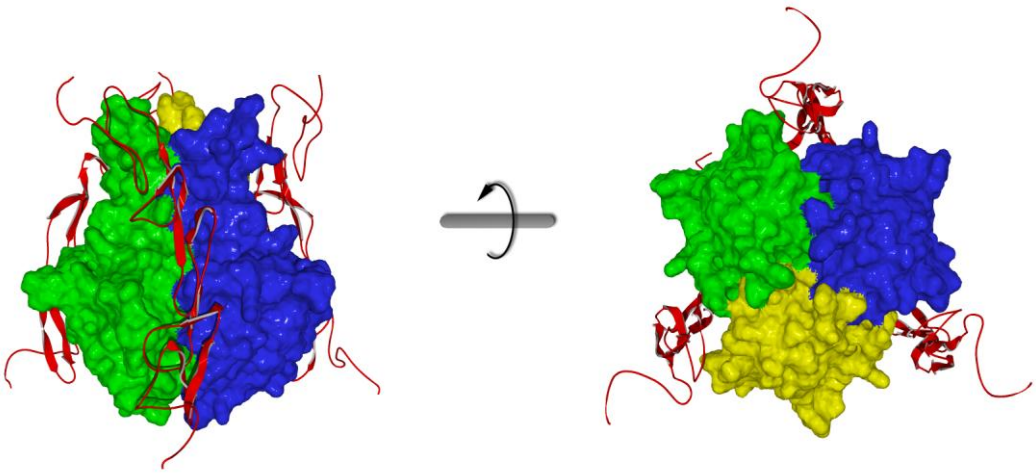


Figure 1.1.6 TNF/TNF receptor binding scheme
Binding of TNF Receptors occurs along the subunit interfaces of the TNF trimer. The TNF ligand is shown in surface representation with subunits coloured green, yellow and blue and the TNF Receptors are shown as red secondary structure cartoons.

b. Specificity determinants

Aside from the gross structural differences found in the TNF-ligand and receptor families, such as loop regions in TRAIL and B cell Activating Factor (BAFF) and the CRD arrangement of TNF receptors, the main determinants of ligand-receptor binding appear to be specific electrostatic and steric interactions between receptors and ligands. The surface charge distribution for each member of the family is highly divergent (this is discussed further in Chapter 2).

Residue conservation in the receptor-ligand interface between TRAIL/DR5 and TNF/TNFR1 is limited to a single tyrosine in both TRAIL and TNFβ (Tyr216 and 142 respectively); the remainder of this interface is highly variable and would contribute to the receptor specificity of the ligands.

In TRAIL the A-A" loop forms a specific electrostatic interaction between Arg 149 of TRAIL and Glu147 of DR5. TNFβ has a much shorter A-A" loop than TRAIL and thus this interaction is not found with its receptors.

The receptors Transmembrane Activator and CAML Interactor (TACI) and B-Cell Maturation Antigen (BCMA) bind promiscuously to A Proliferation Inducing Ligand (APRIL) and BAFF [41,42]. The core structures of both are almost identical, while the BAFF specific Receptor (BAFFR) has a unique disulphide link between its constituent modules. It is speculated that this structural feature prevents the conformational flexibility necessary to bind APRIL.

1.1.6 TNF receptor signalling

The ultimate outcome of the engagement of TNF-ligands by TNF receptors is signalling through a network of protein-protein interactions to the cellular machinery, to either activate the caspase dependent apoptotic pathway, or the NF-κB/JNK pathways that promote gene expression [4]. The TNFR death domain is a sixty amino acid ankyrin like domain with a six helix bundle that is responsible for homotypic interaction with other death-domain containing proteins [43]. A number of TNFRs that lack a death domain can still signal apoptosis through interaction with adaptor proteins that possess the death domain, although the non-functioning decoy receptors can not act in this way [44].

The control of signalling is of great importance in any system, and the TNF receptors have a number of mechanisms whereby constitutive signalling is prevented and the system is kept 'poised' for activation.

a. TNFR1 signalling paradigm

The mechanism of signalling by TNFR1 has been taken as a paradigm for signalling by death domain containing TNFRs [45] and serves well to illustrate the multiple proteins and outcomes that are involved in the TNF signalling pathway (Figure 1.1.7).

Through death-domain interactions, the activated TNFR1 recruits trimeric TNFR Associated Death Domain (TRADD), which in turn acts as a scaffold for a number of further proteins including Receptor-Interacting Protein (RIP) and TNFR Associated Factor 2 (TRAF2). These final two proteins are responsible for the recruitment of specific mitogen dependent kinases that are responsible for the activation of NF-κB.

Fas-Associated Death Domain (FADD) acts by recruiting caspase-8 through Death Effector Domain (DED) interactions, which then self-cleaves and initiates an apoptotic protease cascade.

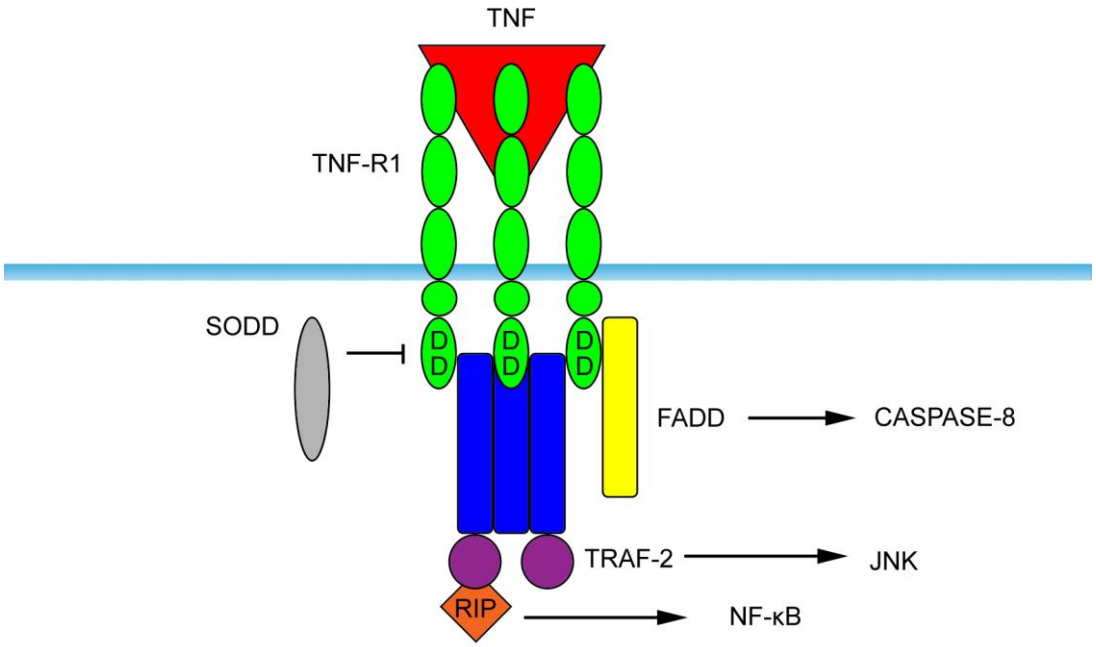


Figure 1.1.7 The TNFR1 signalling pathway
Engagement of TNFR1 by TNF results in the release of the suppressor of death domain (SODD) protein; this in turn allows the formation of the intracellular-death domain complex containing the adaptor proteins TRADD, TRAF-2 and RIP, which initiate the NF-κB pathway through specific activating kinases and the activation of JNK through TRAF2. Signalling through FADD initiates caspase dependent apoptotic pathways.

TNFRs lacking the death domain can still initiate apoptosis in certain circumstances, but do so through the recruitment primarily of TRAF molecules. The actual pathways of apoptotic induction by non death domain receptors are less well understood than those with death domains [46].

The potential for the activation of conflicting pathways of cell proliferation/differentiation, and apoptosis, by TNFRs, points to the control that is exerted by the cell over the outcome of TNF signalling. In common with much of cell

biology, the wider state of the cellular system is as important as the fine detail of the signalling events in determining signalling outcomes.

b. Control of receptor activation

There are currently two major extra-cellular mechanisms by which the activation of TNFRs are controlled. The first is competition with non-signalling decoy receptors and the second is a pre-association of receptors, which poises the proteins in a 'ready' state and prevents the interaction of intracellular domains with downstream signalling partners.

Decoy receptors

A number of non-signalling receptors that lack an intracellular domain exist and it has been suggested that these proteins act as decoy receptors, preventing signalling by a mechanism of competition with signalling death receptors. The decoy receptors for TRAIL, Decoy Receptor 1 and 2 (DcR1, DcR2) are membrane anchored, and DcR2 possesses a vestigial death domain [47]. Those decoy receptors that have been found for Receptor Activator of NF-κB (RANK) and FAS/LIGHT, Osteoprotegerin (OPG) and DcR3 respectively, exist as soluble decoys [27,48]. The action of DcR3 has been shown to have a role in tumour evasion of the immune system. A number of pseudo-decoy receptors can be formed by alternative splicing of FAS and Death Receptor 3 (DR3) [49].

Pre-ligand association

A second method by which receptor activation is controlled would appear to be focussed on the prevention of constitutive and inappropriate signalling. Ligand-induced receptor trimerisation has been accepted as the method by which TNFRs are activated and there is evidence that the structure of TRADD is trimeric [50,51]. Hence, there is potential for the activation of TNF receptor signalling in the absence of TNF-ligand by purely random encounters between receptors.

The solution of the crystal structure of TNFR1 [52] led to the proposal of a mechanism by which signalling could be silenced. This mechanism is in part based on observations of an unrelated class of protein, the erythropoietin receptors [53,54]. In the ligand free state, the erythropoietin receptor adopts a conformation very much like the blades of a pair of scissors. This structure aligns the cytoplasmic domains in such a manner that they are unable to trans-activate the JAK2 protein that is needed for the action of the signalling pathway. On interaction with erythropoietin, a domain rearrangement takes place that brings the cytoplasmic domains into close contact and allows JAK2 activation.

Chan and co-workers [14,55] have shown that a domain in the N-terminus of TNFR1 is responsible for receptor interaction and is necessary for correct signalling, although this is not involved in the interaction with TNF. Removal of this domain of residues 10-54 eliminates self-association and ligand binding. Further mutagenesis identified three residues, which were directly involved in the receptor association. This domain is physically distinct from the ligand-binding domain, yet is essential for proper receptor function and has been termed the Pre-Ligand Association Domain (PLAD).

The requirement for the PLAD in receptor function has also been shown for FAS [56] and it is notable in that a defect in this region causes the disease Autoimmune Lympho Proliferative Syndrome [57].

This model differs somewhat from that proposed by Naismith *et al* in that when endogenous TNFR1 is cross-linked it seems to form a trimeric complex. This seems logical given that the activation complex is trimeric, but the mechanism by which the complex rearranges to one that is signalling has yet to be described. Further research and structural studies would be expected to support one model above the other.

Despite these being very different signalling systems, work on TNFR1 and FAS has shown that the EPO-R model would be an appropriate starting point for a conceptual framework for TNFR silencing.

Intracellular regulation of TNF receptor signalling

Two main mechanisms by which TNFR signalling can be silenced exist on the intracellular side of the signalling pathway. The first of these is the interaction of SODD proteins with the TNF death domain [58]. This protein is thought to act as a negative regulator of TNF signalling by interacting with the un-complexed TNF receptor. On TNFR engagement by its ligand, the silencing protein is disengaged and signalling can take place. In essence, this protein acts to curtail the tendency of TNFR death domains to self-associate and thus prevent constitutive signalling. This mechanism of silencing has much that ties it to the mechanism of pre-ligand association and it may be the case that the pre-association complex is 'silenced' by the

SODD proteins rather than for the geometric reasons that have been suggested in common with EPO-R pre-association.

The second silencing mechanism involves the FLICE-like inhibitory proteins (FLIPs), a class of death effector domain containing molecules that are thought to act as inactive decoys for caspase-8. The expression of these proteins is reduced in activated T- and B-cells, corresponding to an increased sensitivity to death signalling [59,60]. These proteins act to inhibit of the correct formation of the signalling complex [61]. The initial discovery of these proteins came from work on the survival of cells infected by adenoviruses [62] and their ability to subvert FAS dependent apoptosis; one can see how the subversion of the death signalling machinery would be in the interests of the survival of a virus. This strategy has also been seen in tumour cells [63] and underlines the importance of immune surveillance in the effective response to tumours.

1.1.7 Aims

Because of the level of interest in the biology of the TNF family, the ligand and receptor proteins present an ideal set of targets for a concerted high-throughput structural genomics initiative to add to the knowledge base of X-ray crystal structures for the family. Thus, those members of the TNF receptor and ligand super families that have not yet been solved by X-ray crystallographic means were chosen as targets for the Oxford Protein Production Facility (OPPF) protein production pipeline and their progress through this from initial construct design to crystallization and x-ray diffraction studies are discussed in this chapter.

1.2 Materials and methods

1.2.0 Enzymes and chemicals

Chemicals were purchased from Sigma, Fisher, Melford Laboratories, or MP fine chemicals, reagents for crystallization screens were from Hampton Research or Emerald Biostructures. Crystallization plates and consumables were purchased from Greiner Bio-one. The formulations for all buffers and solutions can be found in Appendix Reagents. All solutions were stored at 20 ºC; apart from refold buffers and inclusion body wash buffers, which were stored at 4 ºC. Distilled water was taken from a Millipore Milli-Q reverse osmosis system and had a resistivity of 18.2 mΩ.

IMAGE clones were acquired from the MRC gene service. A number of TNF ligand and receptor DNA clones were the gift of Professor Gavin Screaton (Imperial College). All primers were purchased from Invitrogen in unpurified, desalt-grade, 96-sample block format. The KOD HiFi DNA polymerase from Novagen was used for all PCR reactions. Cloning was carried out using the Invitrogen Gateway^{TM} system and all reagents for this were sourced from Invitrogen.

All gel filtration was carried out on GE Healthcare ÄKTA systems and columns, using solutions that were filtered using a 0.22 μM Millipore filter and degassed using a vacuum pump filtration system. SDS-PAGE was performed using the Invitrogen X-cell system with 1.0 mm cartridges.

1.2.1 Tumour Necrosis Factor Superfamily Ligands and Receptors

In collaboration with the OPPF, the TNF receptor and ligand super families were chosen as targets for a pilot, high-throughput protein expression and crystallization programme. TNF ligands and receptors represented in the MGC collection of genomic clones [64], or in the collection of Professor Gavin Screaton (Department of Medicine, Imperial College, London), were selected as targets for this pipeline, the workflow of which is shown in Figure 1.2.1. All construct design was carried out *in silico* prior to cloning and expression trials, which were carried out by OPPF staff in the OPPF. Large-scale expression, refolding, purification and crystallization of targets were carried out within Strubi and the OPPF.

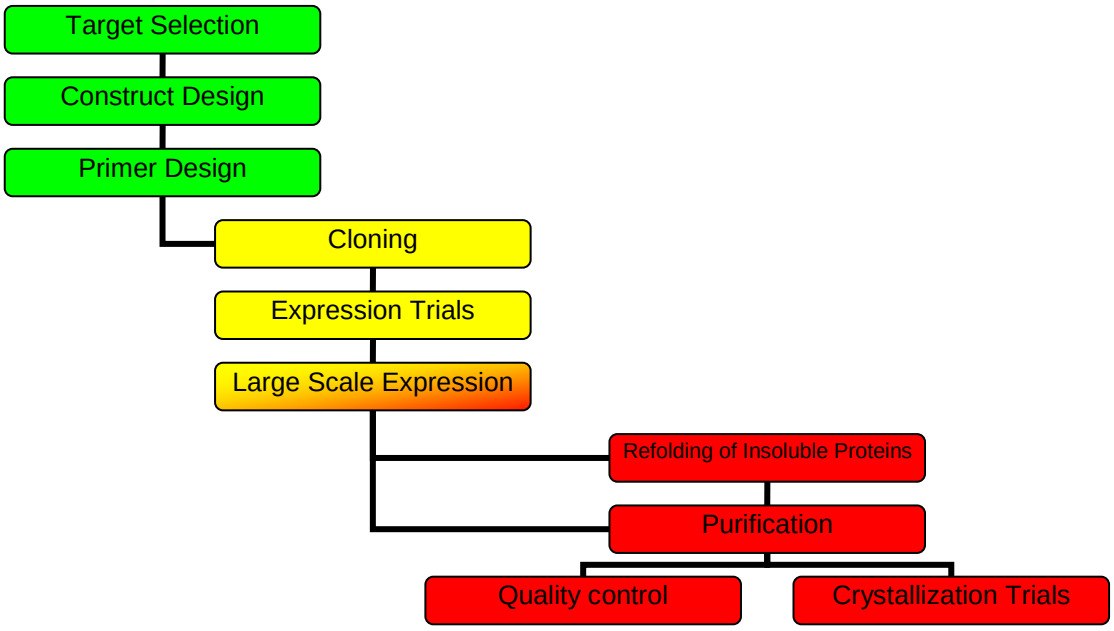


Figure 1.2.1 Workflow for the TNF family project
Production of TNF receptor and ligand family members for protein crystallization experiments was carried out to the following scheme. Steps in green were carried out *in silico;* those in yellow in the OPPF, by OPPF staff; red steps were carried out both in Strubi and with OPPF staff.

1.2.2 Construct design

The focus of the TNF project was on the determinants of ligand/receptor binding and thus constructs were designed to contain the extra-cellular protein interaction domain for both TNF receptor and TNF-ligand Families.

TNF receptors

Using the OPPF OPTIC viewer in concert with the SRS7 sequence search and viewing tool [65] and the Uniprot protein database [66], constructs for sixteen TNF receptors were designed. Each construct was designed to include the N-terminal CRDs of the protein, with the N-terminus of the protein truncated to remove the signal sequence and the C-terminus after the final CRD region (Figure 1.2.2). The single exception was with the CD30 construct, where the repeated domains were also omitted. (See Appendix TNFR for sequences of all TNF receptor constructs and primers).

TNF ligands

A sequence alignment of the extracellular TNF THDs [11] (Figure 1.2.3) was used as the basis for the design of the constructs for the thirteen chosen TNF ligands. The constructs were N-terminally truncated to remove the membrane spanning region of the protein and included the full C-terminal region of the protein encompassing the full extent of the extra-cellular beta-jellyroll domain of the protein. (See Appendix TNF Ligands for sequences of all TNF-ligand constructs).

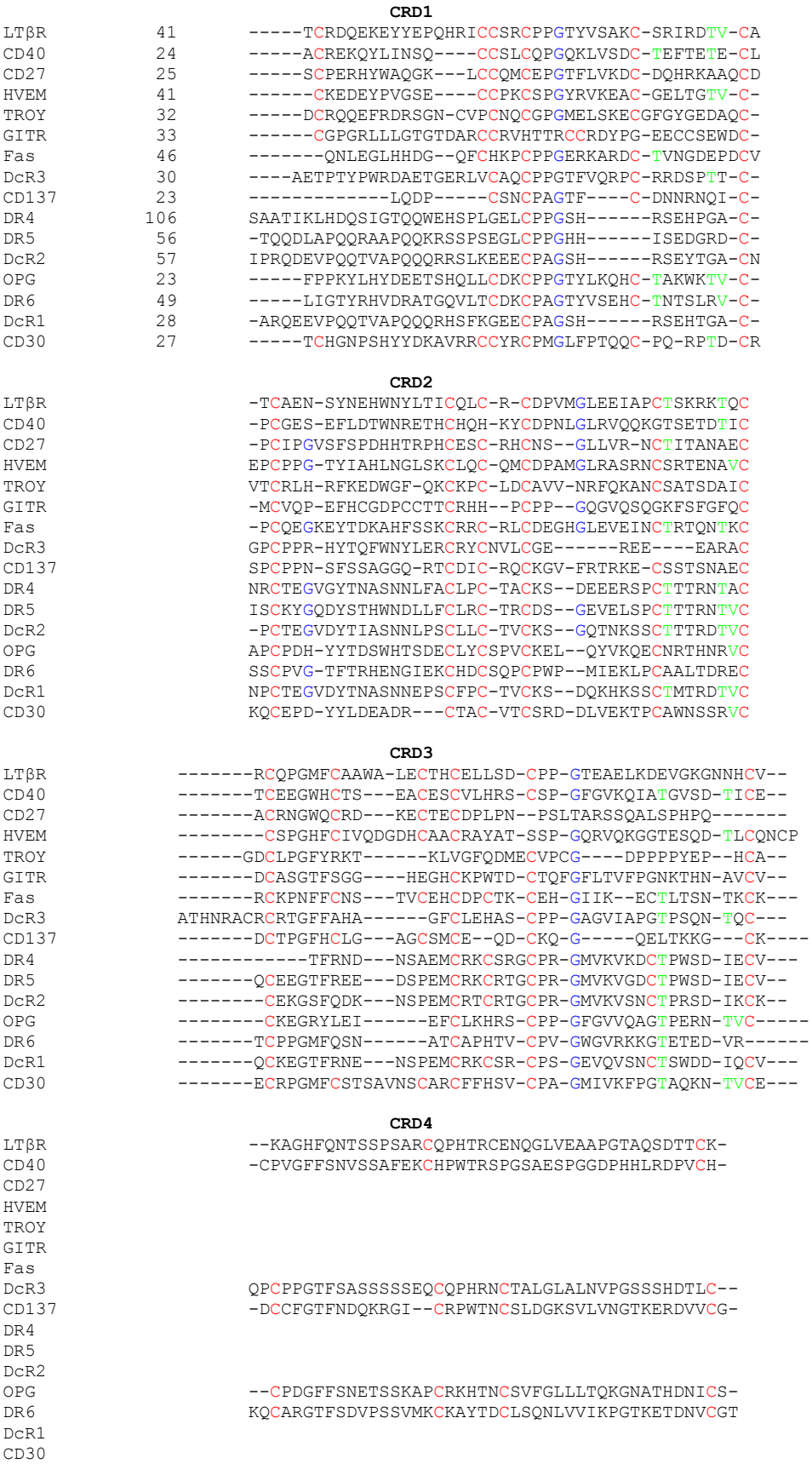


Figure 1.2.2 Sequence alignment of the tumour necrosis factor receptor targets
Primary sequence alignments of the sixteen TNF receptors to show the extent of the constructs used and the arrangement of cysteine rich domains. The sequence alignment was performed on a domain-by-domain basis using ClustalW, with conserved cysteine residues labelled in red, concerved glycines in blue and other structurally conserved residues in green. The numbers of N-terminal residues preceding the CRDs are noted before the CRD1 alignment.

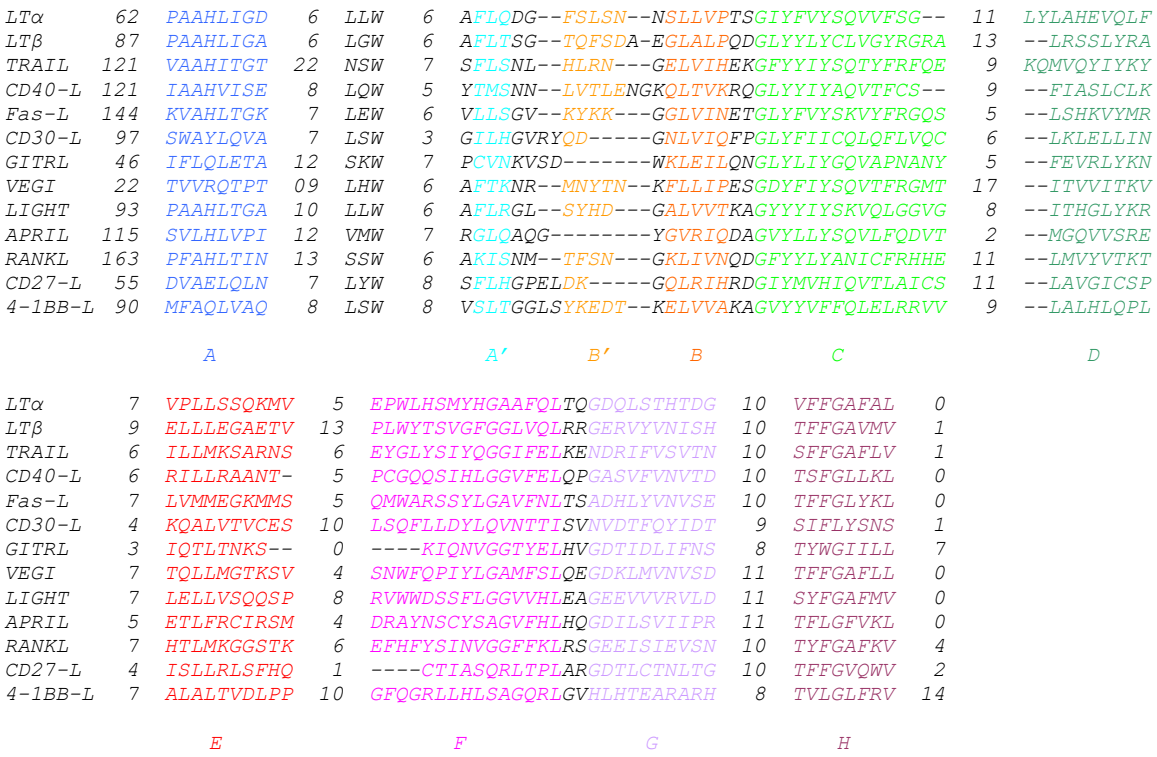


Figure 1.2.3 Sequence alignment of the tumour necrosis factor ligand targets
Primary sequence alignments of the thirteen TNF THDs to show the extent of the constructs used and the position of the structurally conserved beta strand regions. The sequence alignment has been compacted to show only areas of high-homology, with numbers of intervening residues indicated between regions. The ten beta strands are coloured sequentially and named beneath the sequences.

1.2.3 Primer design

Primers for TNF ligand and receptor constructs were designed using the OPINE programme (which ensures complementarity of primer binding and the correct reading frame) to amplify the protein sequence selected during construct design. Primers were designed to include all elements that are needed for Gateway^{TM} cloning.

Forward primers contained the following sequences:

- Four guanine residues at the 5' end
- The 25 base pair *attB1* site
- 18-25 base pairs of template specific sequence

Because the *attB* site ends with a thymidine nucleotide, primers for proteins with N-terminal tags had an additional two nucleotides added to ensure the correct reading frame (not AA, AG, or GA because these code for a translation termination codon).

Reverse primers contained the following sequences:

- Four guanine residues at the 5' end
- The 25 base pair *attB2* site
- 18-25 base pairs of template specific sequence

Again, if a C-terminal tag is required a single extra nucleotide must be added to ensure the proper reading frame.

1.2.4 Cloning

All cloning was carried out within the OPPF using a semi-automated implementation of Gateway™ ligation independent cloning.

a. Prokaryotic vectors

The majority of TNF receptor and ligand constructs were inserted into T7 polymerase based pDEST14 vector with an N-terminal His_6 tag (Appendix Vectors). The sequence of each construct is described in either Appendix TNF Ligands, or Appendix TNFR. A TRAIL construct using a pET9c vector was a gift of J. Mongkolsapaya.

1.2.A Use of T7 based vectors

In the T7 based vector system, the gene of interest is under the control of the bacteriophage T7 promoter and requires the T7 RNA polymerase for transcription and thus expression. The vector also contains the *lac* repressor gene. Strains of *E.coli* used for expression commonly contain a chromosomal copy of the T7 RNA polymerase controlled by the *lac* repressor and thus the expression of that protein from the vector prevents constitutive expression of the T7 polymerase, hence under normal conditions there is no significant expression of the target protein. This is particularly important if the target protein is toxic to the cell.

When isopropyl-β-D-thiogalactopyranoside (IPTG) is added to the growth medium, typically at a concentration of 1 mM, the lac repressor is inactivated by binding of this molecule, allowing the transcription of the T7 polymerase and thus expression of the target protein. This system allows the expression of target protein to be tightly regulated.

b. Eukaryotic vectors

The TNF receptor like molecule Lymphotoxin Beta Receptor LTβR was inserted into a pHLsec vector with a simple six-histidine tag. A construct with an Fc tag in a cDM7 vector was a gift of J. Mongkolsapaya/A. Cowper

1.2.5 OPPF Gateway™ cloning system

The Invitrogen Gateway™ cloning system [67] is used within the OPPF, for its benefits of ligation free insertion and the ability to transfer the gene of interest between different vectors while maintaining correct orientation and reading frame. The OPPF cloning protocol is outlined in Figure 1.2.4.

Primers for all OPPF targets were sourced un-purified (de-salt grade, Invitrogen) in a 96-well format and inserts were initially amplified using standard PCR conditions. Quality control was carried out by agarose gel electrophoresis and final sequencing of expression vectors. Targets were amplified using the KOD HiFi DNA polymerase (Novagen) following standard PCR conditions as detailed in the Novagen protocol (see Appendix PCR).

The Gateway™ cloning system is based on the site-specific recombination properties of bacteriophage λ [68]. The integration of bacteriophage λ into the *E.coli* chromosome involves site-specific recombination events mediated by bacteriophage and *E.coli* recombination enzymes that recognise specific DNA sequences and catalyse the exchange of genetic material. The sites of recombination are known as attachment sites and are named *attB* on the E.coli chromosome and *attP* on the bacteriophage

chromosome. The integration of bacteriophage DNA into the *E.coli* chromosome at fifteen base pair sites, recognised by recombination proteins gives rise to *attL* and *attR* sites, which, in the lytic cycle, are recognised by bacteriophage excision factors.

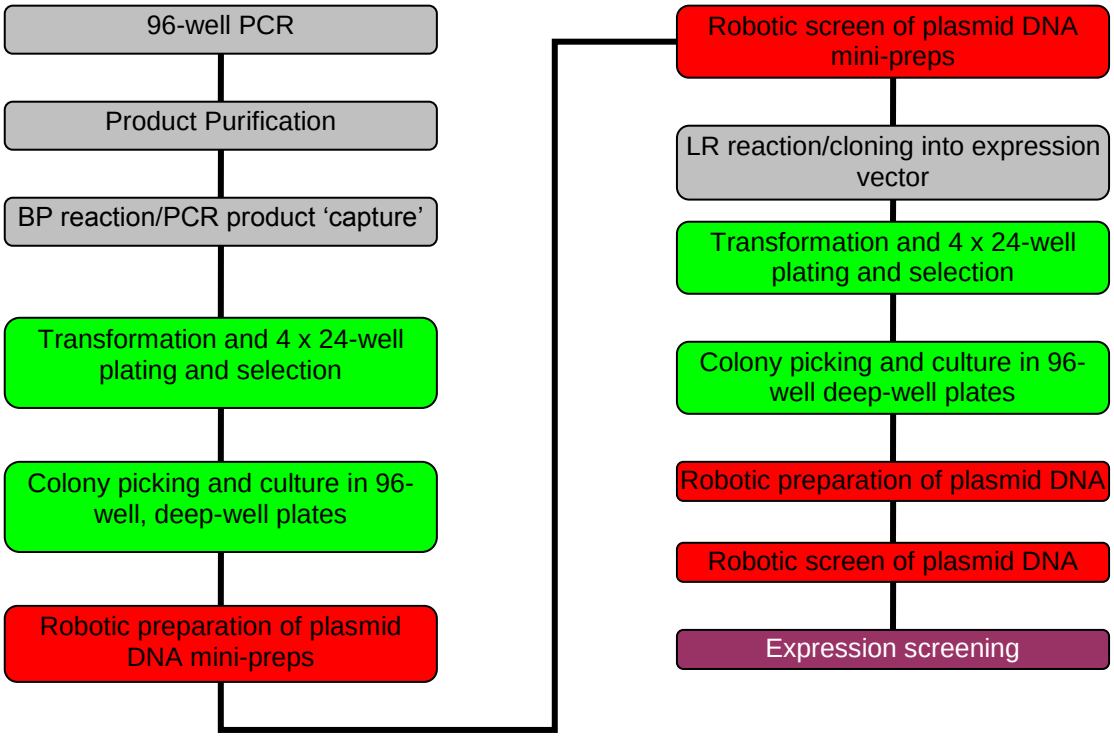


Figure 1.2.4 OPPF amplification and cloning protocol
Initial clean-up, amplification and gateway steps are coloured in grey, selection steps are coloured in green, semi-automated steps are red and downstream steps are coloured in purple. The BP and LR reactions are described further in the text.

The Gateway™ cloning methodology utilises the bacteriophage λ recombination system to facilitate the exchange of heterologous DNA sequences between different vectors. A number of modifications from the native recombination system ensure fidelity of recombination and in the case of a modified *attR* site, that the LR reaction is irreversible. All Gateway vectors also contain the *ccdB* gene, which interferes with the *E.coli* DNA gyrase [69] and acts to allow negative selection within cloning. Two recombination reactions form the basis of the system: the BP reaction and the LR reaction.

BP reaction

In the BP reaction, an *attB* flanked PCR product is recombined with an *attP* substrate to create an *attL* containing entry clone, Figure 1.2.5.

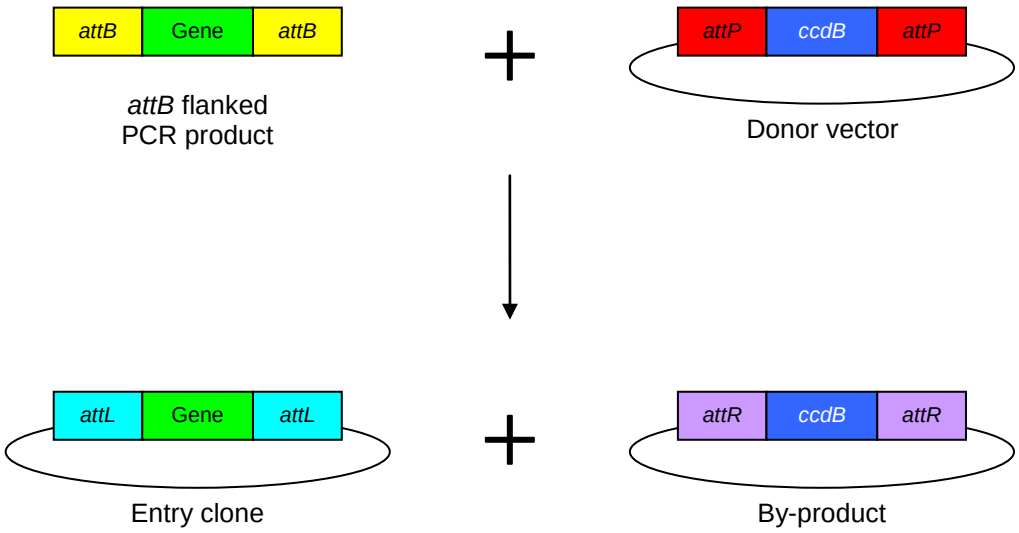


Figure 1.2.5 The BP reaction
Recombination between an *attB* containing PCR product and an *attP* containing donor vector using the BP clonase mixture produces an *attL* flanked entry clone and an *attR* flanked by product containing the *E.coli* lethal *ccdB* gene.

For the OPPF targets, the gene of interest is amplified using *attB* tagged primers designed for each clone and inserted into an entry vector with *attP* sites to form a vector with *attL* sites. Correctly integrated DNA produces viable *E.coli* colonies on ampicillin spiked agar. Although untransformed vector contains the ampicillin resistance gene, the lethal *ccdB* gene is active in this, such that only correctly integrated clones produce colonies.

LR reaction

The LR reaction facilitates the recombination of an *attL* entry clone with an *attR* destination vector to create an *attB* expression clone Figure 1.2.6. The LR Clonase recombination mix facilitates the transfer of DNA segments between the two in the same manner as the BP reaction. Again, selection for correctly ligated products is carried out by testing for ampicillin resistant colonies in *E.coli*.

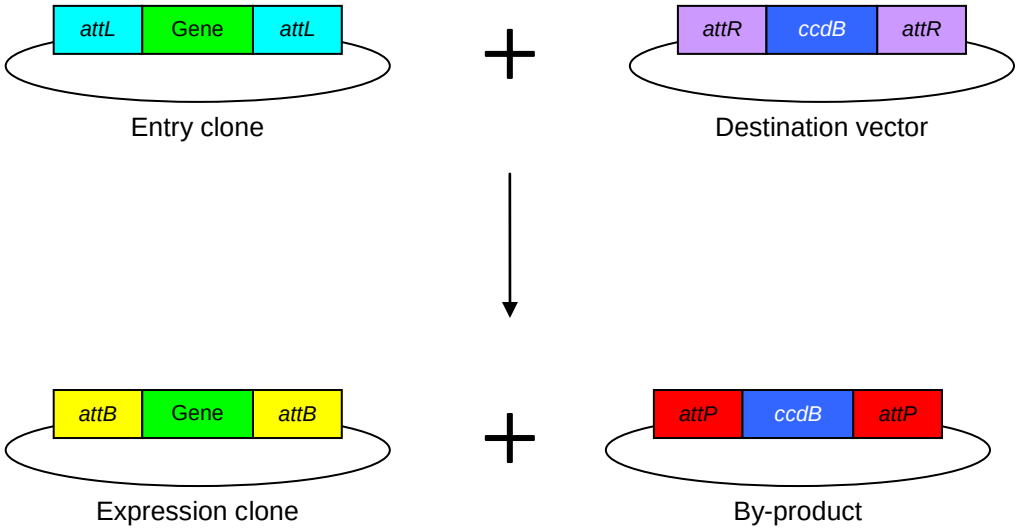


Figure 1.2.6 The LR reaction
Recombination between an *attL* containing entry clone and an *attR* containing donor vector using the LR clonase mixture produces an *attB* flanked entry clone and an *attP* flanked by-product containing the E.*coli* lethal *ccB* gene.

All recombination reactions were carried out using the standard Invitrogen protocol for Gateway™.

In the event of a cloning failure, the step at which failure occurred is identified and the process is restarted from the last successful step. In the case of subsequent failures, the process is restarted from the PCR stage, or new template DNA is brought into the pipeline.

1.2.6 Protein expression in a prokaryotic system

Constructs that were successfully cloned were first tested for protein expression and then scaled up for protein crystallography.

a. Protein expression trials

Expression competent Rosetta pLysS *E.coli* cells (Novagen) were dispensed in 50 µl aliquots in a 96-tube matrix format and 1-2 µl of prepared pDest14 plasmids were added to each tube. These were incubated on ice for thirty minutes, and subjected to heat shock for one minute at 42 ºC on a heating block, before being returned to ice for two minutes. 450 µl of Luria-Bertani (LB) Media was added to each tube and then these were incubated at 37 ºC for one hour to allow cells to recover.

After this incubation, the samples were plated on LB-Agar supplemented with 1% glucose plus 50 µg ml^{-1} Carbenicillin (using 30 µl of culture medium to one well of a 24 well plate) and incubated overnight.

b. Large Scale Expression

A single colony of transformed cells was inoculated into 20 ml of LB-Media containing appropriate antibiotic at 37 ºC overnight. One ml of this stock culture was used to seed 1 litre of LB-media split between two 500 ml cultures in 2 litre shake flasks and grown to mid-log phase (OD_{600} = 0.4 - 0.6). At this stage, IPTG was added to final concentration 0.5 mM to induce protein expression and the incubation was continued for 3 hours at 37 ºC. Cells were harvested by centrifugation at 5000 g, 4 ºC for ten minutes and re-suspended in 40ml of PBS.

c. Protein expression tests

To test for protein over expression and the presence of soluble or insoluble protein, from both initial small-scale tests and large-scale expression, samples of 1 millilitre of cell culture media were taken pre-induction and upon harvesting of final cultures. These samples were centrifuged at 6,000 g for five minutes to pellet cells. They were then re-suspended in 1 ml PBS and sonicated at 4 ºC to disrupt cells and then centrifuged at 25,000 g, 4 ºC, for ten minutes. The supernatant was transferred to a clean eppendorf tube with 1 ml of 4X SDS loading buffer, while the insoluble pellet was re-suspended in the same volume of loading buffer. Samples were boiled for at least ten minutes and subsequently loaded on a 15 % Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) gel for analysis.

1.2.B SDS-PAGE analysis

Sodium Dodecyl Sulphate (SDS) denatures proteins through interaction with the peptide backbone and binds with the specific mass ratio of 1.4:1. In the process of this the SDS confers a negative charge upon the protein in proportion to its length. Disulphide bonds must be broken using a reducing agent such as dithiothreitol (DTT) to ensure that the protein is fully denatured. Because the protein is fully denatured and the binding of SDS is stoichiometric, the migration of proteins in a polyacrylamide gel is determined by their weight and thus with the aid of known standards one can estimate the molecular weight of the proteins. It is also a convenient method to assess the purity of a protein.

SDS-PAGE is carried out in a discontinuous buffer system where the gels and running buffer have different pH and ionic strengths. The gel itself consists of a large-pore

stacking gel layered on a resolving gel. Samples are focused through the stacking gel by a steep voltage gradient and the resolving gel separates the proteins by size in a uniform voltage.

All gels were cast as 15 % resolving gels with a 5 % stacking gel in 1 mm x 10 cm x 10 cm Xcell cartridges and run using the Invitrogen Xcell system at 150 Volts until the dye front had run off the bottom of the gel.

Gels were stained using Invitrogen Simply Blue™ SafeStain. The fast microwave based staining protocol was used. Gels were removed from the cartridge with a spatula and placed in a microwave-safe container with 100 ml deionised water and microwaved on high power for one minute and left to stand for a further minute. The water was poured off and replaced and the procedure was repeated three times. After the fixing step, 50 ml of SafeStain was added to the container and the gel was microwaved for one minute, in thirty-second bursts to prevent the stain from boiling. The gel was left to stand for ten minutes, or until bands were visibly stained, and then the stain was poured off and replaced with 100 ml deionised water and microwaved for one minute to de-stain the gel. Gels were photographed using an AlphaImager system and saved as TIFF files.

1.2.7 Refolding of insoluble prokaryotic expressed protein products

Proteins expressing in insoluble inclusion bodies in prokaryotic expression were washed and refolded using a rapid dilution technique developed in house.

a. Preparation of protein inclusion bodies

Bacterial pellets from large scale expression were resuspended in cold PBS and then disrupted by sonication at 4 ºC before being centrifuged at 25,000 g, 4 ºC for thirty minutes. The pellet containing inclusion bodies was washed with ice-cold Triton wash buffer (50mM Tris pH 8.0, 100 mM NaCl, 0.5 % Triton X-100), re-suspended using a homogeniser and then centrifuged at 25,000 g, 4 ºC for ten minutes. This was repeated until the supernatant became clear, before a final wash with cold, detergent free, wash buffer. The inclusion bodies were solubilised in denaturing buffer (6 M guanidine, 50 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA) at 4 ºC overnight on a bench top shaker.

b. Protein refolding

Two millilitres of solubilised, denatured protein (20 mg ml^{-1} concentration) were incubated with DTT to a final concentration of 10 mM for one hour at room temperature to reduce any cysteine residues present. This reduced protein solution was then added drop-wise to 250 ml of cold refolding buffer (200 mM Tris pH 8.0, 1 M L-arginine, 10 mM EDTA, 6.5 mM cysteamine, 3.7 mM cystamine), stirring at high speed to aid rapid dilution and to prevent aggregation; stirring continued overnight at 4 ºC. The solution was centrifuged at 3,500 g, 4 ºC, for thirty minutes to remove large aggregates of protein and then concentrated to 5 ml, using a Vivacel 250 ml concentrator (Vivascience), with either a 5 kDa Molecular Weight Cut Off (MWCO) filter insert for TNF receptors, or a 10 kDa MWCO insert for TNF ligands.

1.2.8 Purification of prokaryotic expressed proteins

Both refolded and soluble *E.coli* produced proteins were purified by standard chromatography methods. Soluble proteins were purified using a two-step method. First His-tagged protein was separated from cell lysate by Immobilised Metal-ion Affinity Chromatography (IMAC) using nickel charged chelating sepharose beads (GE Healthcare), and then they were purified by size exclusion gel filtration chromatography.

a. Initial purification of soluble protein

Protein that expressed in the *E.coli* cytosol was initially purified by nickel-histidine affinity. Cell lysate was first clarified with a 0.45 μM minisart filter (Vicascience) and then passed through a 5 ml nickel chelating sepharose column (GE Healthcare) that had been equilibrated with His-tag binding buffer (20 mM $NaPO_4$ pH 7.4, 500mM NaCl, 10 mM imidazole). The column was then washed with seven column volumes of binding buffer to remove all non-histidine-tagged protein in the lysate. A His-tag elution buffer (20 mM $NaPO_4$ pH 7.4, 500mM NaCl, 500 mM imidazole) was then run through the column over a three column volume gradient and elution peak fractions were pooled.

b. Size exclusion chromatography

All proteins were subjected to size exclusion gel-filtration purification. Both refolded protein and soluble protein were first filtered through a 0.22 μM minisart filter and loaded onto a gel filtration column for purification. TNF receptor targets were purified on a HiLoad Superdex 75 16/60 column, while TNF ligand targets were purified using a Superdex 200 16/60 column. These were equilibrated with 1.5

column volumes of gel filtration buffer (50 mM Tris-HCl, 150 mM NaCl). The peak fractions were pooled.

1.2.9 Eukaryotic protein expression

Protein constructs for TNF receptors that were to be expressed in mammalian HEK-293 T-cells were transfected using polyethylenimine (PEI) transfection (Sigma Aldrich) and grown for protein expression over four days. All DNA was prepared using the endotoxin-free mega-kit from Qiagen.

HEK-293 T-cells were grown to 90 % confluence in 15 centimetre dishes using DMEM/L-Gln/Non-essential amino acids, plus 10 % foetal calf serum (Gibco). For PEI transfection, 50 μg of DNA was added to 5 ml of serum-free media and 100 μl of 1 mg ml^{-1} PEI. This was left to incubate for ten minutes at room temperature to allow the DNA/PEI complex to form. Old media was removed from the cells and 10 ml of media with 2 % serum was added to the PEI/DNA complex. The mixture was then added to the cells. After three hours of incubation, the media in the dish was topped-up to 25 ml.

Cells were left to express protein for 3 – 4 days and the media was collected, centrifuged and filtered using a 0.22 μm filter device and purified using methods developed within the OPPF.

1.2.10 Purification of Eukaryotic expressed proteins

a. Purification of Fc tagged secreted protein

Protein was loaded onto a Protein A sepharose column (GE Healthcare), and washed with wash buffer (100mM Tris-HCl pH8, 500mM NaCl), and one column volume of His-3C protease added (Novagen) (100mM Tris-HCl pH8, 500mM NaCl, 1mM EDTA, 2mM DTT). The flow was stopped and the column incubated for eight hours to allow cleavage to occur. After this, the cleaved protein was washed from the Protein-A column with wash buffer, through a histidine column (50mM Tris-HCl pH7.5, 500mM NaCl, 20mM imidazole) to the fraction collector. The Histidine column was eluted (50mM Tris-HCl pH7.5, 500mM NaCl, 500mM imidazole) to remove the His-3C protease and an acid elution was performed on the Protein A column to remove un-cleaved protein and the free Fc tag (100mM Citric Acid pH3, 500mM NaCl).

All fractions were analysed by SDS-PAGE and those containing the protein of interest were pooled. Pooled fractions were run on a Superdex 200 16/60 gel filtration column (20mM Tris-HCl pH7.5, 200mM NaCl) as the final purification step. All steps were carried out at 4 ºC.

b. Purification of Histidine tagged secreted protein

Protein was loaded onto a nickel-chelating sepharose column and non-bound protein was washed from the column (50mM Tris-HCl pH7.5, 500mM NaCl, 20mM imidazole). Protein of interest was then eluted from the column with elution buffer (50mM Tris-HCl pH7.5, 500mM NaCl, 500mM imidazole).

All fractions were analysed by SDS-PAGE and those containing the protein of interest were pooled. These pooled fractions were run on a Superdex 200 gel filtration column as the final purification step (20mM Tris pH7.5, 200mM NaCl). All steps were carried out at 4 ºC

1.2.C Light Scattering by proteins

When light passes through a suspension of molecules of a much lower size than the wavelength of the light the intensity of light scattered is uniform in all directions, this is known as Raleigh scattering. If the light is from a coherent source, such as LASER light, the fluctuations in scattering can be measured using a suitable photomultiplier. These fluctuations arise from interference caused by the Brownian motion of particles in solution affecting the distance between particles. Analysis of this motion can yield the diffusion coefficient of the particles and using the Stokes-Einstein equation, the hydrodynamic radius of the particle can be calculated.

Using a digital correlator, the relationship between the intensity of light fluctuations and a delayed set of intensity fluctuations is followed. The decay of this autocorrelation function can be used to model the diffusion coefficient of the particles. A standard molecular weight approximation method can then be applied to the calculated hydrodynamic radius.

The most effective use of DLS in protein science is to analyse a solution for aggregates using the distribution of hydrodynamic radii to identify potential aggregates and to check the monodispersity of the sample. For protein crystallization the sample quality is of great importance and the presence of aggregates and potential foci of aggregation can have a dramatic affect on the propensity of a particular protein solution for crystallization.

1.2.11 Quality control

To ensure that protein products are correct, and that all material to be used for crystallization trials is of high quality, a number of quality control steps were performed on each protein sample.

a. Protein purification. Purity of protein fractions from gel filtration was determined by running 10 µl of each peak fraction on an SDS-PAGE gel and staining as described in Chapter 1.2.B. The oligomeric state of each protein was approximated from calibration data for each column. In the case of multiple oligomeric states existing on gel-filtration, both peaks are treated separately.

b. Dynamic Light Scattering. Dynamic Light Scattering (DLS) was carried out on a Dynapro Light Scattering instrument (Protein Solutions) to ensure mono-dispersity and to assess the potential oligomeric state of the protein in solution. Each protein sample was tested at a concentration of 1 mg ml^{-1}.

c. Mass spectroscopy. All proteins were submitted for in-house Mass Spectroscopy (MS) analysis to ensure the product is of the correct calculated mass. All MS was carried out by J.Nettleship or R.Aplin using a Dionex Ultima HPLC and a Waters Q-TOF micro instrument.

d. Protein concentration. Pooled fractions from gel-filtration were concentrated in Vivaspin 15 ml and 2 ml concentrators with an appropriate MWCO at 3,500 g, 8 ºC. Protein concentration was estimated by UV spectroscopy at 280 nm using a calculated molecular extinction coefficient. For initial crystallization screening, proteins were concentrated to between 10 and 20 mg ml^{-1}.

1.2.12 Crystallization

TNF ligands and receptors were screened for crystallization using the OPPF Cartesian micro-drop dispensing system and the full range of available crystallization screens [70]. The following commercially available screens (Hampton Research, Emerald Biostructures) [71-73] have been adapted to the 96 well format:

1. Hampton Crystal Screen 1 and 2
2. Emerald Wizard I and II
3. Hampton PEG/ION screen, Hampton Grid Screen PEG 6000, Hampton Grid Screen Ammonium Sulphate
4. Hampton Natrix, Hampton Crystal Screen Cryo
5. Hampton Grid Screen: PEG/LiCl, NaCl, MPD; Hampton Quik Screen
6. Hampton Index Screen
7. Hampton Salt RX
8. Limited protein screen. Selected conditions from Hampton Crystal screen 1, 2 and PEG/Ion.

A 96-channel Hydra pipetting robot (Robbins Scientific) is used to dispense 95 µl of each crystallization screen from a master block of the screen into a 96-well Greiner Crystal Quik crystallization plates (Greiner bio-one). The Cartesian Technologies Microsys MIC4000 dispensing robot is then used to dispense 100 nl drops of protein into the crystallization wells of the plate and 100 nl drops of reservoir solution are then added to this. These sitting-drop vapour-diffusion experiments are then sealed with a clear, self-adhesive, sealant and inserted into an automated storage vault (The Automation Partnership, Royston) which keeps the experiments at 21 ºC or placed in a cold room at 4 ºC. Crystallization plates are automatically imaged at pre-set time intervals using an Oasis 1700 imaging system (Veeco, Cambridge); these images are stored on a central database with a web-interface, which allows analysis of crystallization experiments remotely [70].

1.3 OPPF TNF project

The number of proteins ascribed to the TNF receptor and ligand families increased dramatically in the late 1990s due to genomics initiatives [11], but only a small, yet interesting, sub-set of these have had their structures solved by X-ray crystallography. TNF ligands have been more amenable to structural studies (ten structures), than the TNF receptors where only two structures have been solved in isolation. Five different receptors have been solved in complex with their ligands (see Appendix Xtal for details and references).

From the crystallographic perspective, globular TNF ligands present a more attractive target than the cysteine knot receptors, as phasing X-ray data by molecular replacement is more easily achieved with globular proteins, than with long, thin mainly beta sheet proteins like the TNF receptors. From a biological perspective, the receptor/ligand complexes are of most interest since they inform on the nature of complex formation, protein recognition, and the signalling mechanisms initiated by the trimeric ligand.

When the structure of the TNFR1 was solved, questions were raised about the potential for pre-ligand association that appears to back up biochemical data for FAS [14], and these have still to be satisfactorily answered from a structural perspective.

Since relatively few different TNF family members have been solved, and given the high-level of interest in their biology and signalling mechanisms, the receptor and ligand families were chosen to be part of the Oxford Protein Production Facility (OPPF) high-throughput structural genomics pipeline. This pipeline aims to apply

parallelisation and miniaturisation to the production of large numbers of protein targets of biomedical relevance for high throughput structural genomics. Steps in the pipeline include cloning of the protein-encoding gene into a suitable expression vector, protein production in an appropriate host organism, purification of the resulting protein and then the set up and automatic imaging of crystallization trials for the purified, concentrated protein.

The progress of the TNF proteins selected as targets for the OPPF pipeline including analysis of purification and crystallization trials are discussed in this chapter. Crystallographic studies resulting from those protein targets that produced diffraction quality crystals are discussed in Chapter 1.4.

1.3.1 Overview of OPPF TNF project results

As described in Chapter 1.2.1, TNF ligands and receptors were cloned and expressed as OPPF target proteins. The success of the various stages of this pipeline are shown in overview in Table 1.3.1 and subsequently discussed in further detail in this chapter.

	Cloning	Expression	Refolding/ Purification	Crystallization
TNF Receptors				
LTβR				
CD40				
FAS				
DcR3				
CD27				
CD137				
DR4				
DR5				
DcR2				
OPG				
TROY				
DR6				
DcR1				
GITR				
CD30				
HVEM				
TNF Ligands				
LTα				
CD27L				
TRAIL				
APRIL				
LIGHT				
CD30L				
GITRL				
VEGI				
4-1BBL				
CD40L				
RANKL				
LTβ				
FASL				
Success	15/29	11/29	6/29	4/29

Table 1.3.1 Overview of OPPF TNF project results
For each stage in the process, success or failure is indicated by green or red shading. Successful cloning was determined by sequencing of construct vectors. Expression of protein at levels detectable at the correct apparent molecular weight on a standard SDS-PAGE gel was used as the benchmark for protein expression. Successful purification of refolded, or soluble protein, was marked by a well-defined gel-filtration peak and pure protein on SDS-PAGE analysis of fractions. All proteins that were successfully purified were screened using the OPPF crystallization screens; the formation of crystals in 96-well crystallization trials is noted in yellow, with those proteins producing diffracting crystals in green.

1.3.2 Cloning Results

The OPPF semi-automated Gateway™-based cloning methodology using IMAGE clones proved to be successful for just over fifty percent of the TNF targets. This level of cloning success compares favourably with standard cloning procedures used, but is low for the overall results seen within the OPPF. The potential to rescue cloning failures is not as simple as in standard cloning procedures, because these must be reformatted into the 96-well format and re-inserted into the pipeline. However, a high-proportion of the targets were successfully cloned in initial trials or after one rescue attempt. Outright failures of cloning could be ascribed to the initial stages of the cloning process failing at the single PCR condition used for the whole 96-well block, or to the template material from the IMAGE clones being incorrectly specified in some way.

1.3.3 Protein Expression Results

Expression of protein from small-scale expression tests of Gateway™ vector based protein constructs was relatively successful in that a high proportion of those proteins that were successfully cloned produced detectable levels of protein expression. On the transfer to large-scale expression, nearly half of those protein constructs that produced detectable levels of protein expression in initial testing failed to produce any protein. One problem that was apparent was that while the majority of expressed proteins were found in the form of insoluble Inclusion Bodies (IB) that needed further processing and refolding, the actual protein content of these IBs was highly variable.

The yield of protein from IBs was variable, with the physical quality of the actual IBs themselves varying from a light putty-like paste to being brown and grit-like. The

best quality inclusion bodies, i.e. those with the highest levels of the desired protein in them, came from TRAIL, DR6 and OPG. That the two TNF receptor targets, DR6 and OPG, should form inclusion bodies is unsurprising; the high-cysteine content of these proteins requires an oxidising environment for correct folding, such an environment is not present within the cytosol of the *E.coli* strains used for expression. TRAIL has previously been reported to form IBs in *E.coli* [13].

The TRAIL construct used in protein expression was not the OPPF construct, as this failed initial cloning screening. A pET9C based expression vector from the laboratory of Professor Gavin Screaton was used as a control for the production of large quantities of protein for protein refolding. VEGI was the only TNF target that expressed as soluble product, which correlates with the observation that this protein can exist in a soluble, non-membrane anchored form in certain human tissues [74]. Other TNF ligands are also found in soluble form, but those that were included in this project, such as FAS ligand and APRIL, either failed the initial cloning steps or produced no detectable levels of protein. Lymphotoxin β receptor was expressed both as an Fc-fusion protein and a histidine tagged construct using HEK-293 T-cells.

1.3.4 Purification of recombinant TRAIL

As a control for the protein refolding experiment, an N-terminally truncated TRAIL construct (amino acids 91-281, predicted molecular weight 22 kDa) in a pET9c vector was obtained from Professor Gavin Screaton and protein was expressed in *E.coli* as inclusion bodies (As described in Chapter 1.3.3). After refolding, the protein was purified on a Superdex 200 26/60 column, showing a sharp product peak at 55 kDa apparent molecular weight (Figure 1.3.1). This result is consistent with previously published results, where TRAIL migrates at a slightly lower apparent molecular weight than the 66 kDa mass of the active trimer. The cysteine bridge that occurs between two monomers in the functional trimer is also present; under non-reducing conditions in SDS-PAGE two bands are visible, one at 25 kDa and one at 50 kDa, corresponding to the apparent molecular weights of slightly more than the monomeric TRAIL and the TRAIL dimer (Figure 1.3.2).

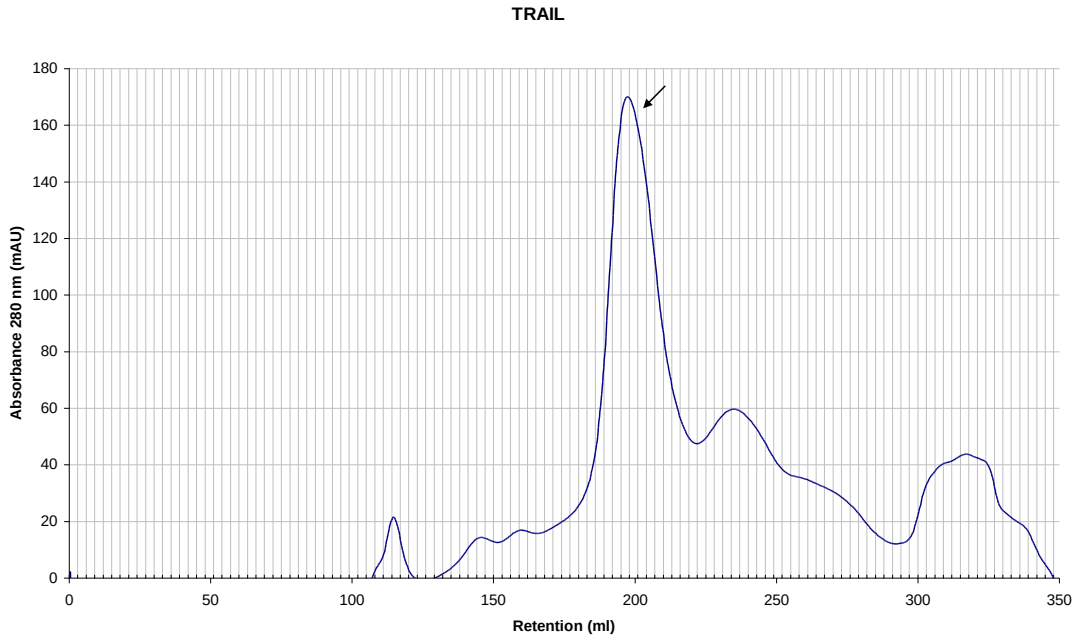


Figure 1.3.1 purification of recombinant TRAIL
Refolded TRAIL was purified on a Superdex 200 26/60 column. The single major peak, labelled with an arrow, at 200 ml corresponds to pure trimeric TRAIL.

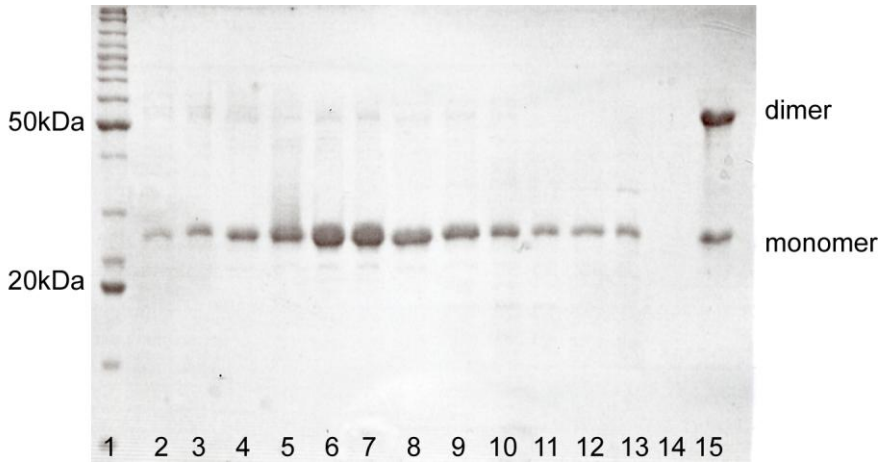


Figure 1.3.2 Characterisation of purified TRAIL by SDS-PAGE
Samples of 3 ml FPLC fractions between 189 and 225 ml were subjected to SDS-PAGE analysis. Pure TRAIL is visible in all lanes. Molecular weight marker is shown in lane 1, TRAIL fractions in lanes 2-12 and non-reduced TRAIL is shown in lane 15. Non-reduced TRAIL migrates as two bands, one corresponding to monomeric trail at around 22 kDa and the other at 44 kDa corresponding to the disulphide linked trail dimer.

Characterisation of the TRAIL trimer using dynamic light scattering at 1 mg ml $^{-1}$ protein concentration, gives an apparent molecular weight of 71.7 kDa. This is consistent with the protein existing as a trimer in solution. Mass spectroscopy analysis of the protein shows the mass to consist of two major peaks at 24,755 and 49,505 kDa, which is consistent with the protein existing as the single TRAIL monomer and the disulphide linked dimer, although the mass is somewhat higher than expected.

The yield of protein is lower than has been previously reported (J. Mongkolsapaya, Thesis). Each litre of bacterial culture would give around 15 mg of pure protein after refolding, representing a yield of ~ 10% from the purified inclusion bodies; the majority of protein is probably lost as aggregate or incorrectly folded protein during the rapid-dilution refolding step, or as aggregate in the flowthrough of the gel filtration step.

1.3.5 Purification of Recombinant DR6

An N and C-terminally truncated DR6 construct consisting of the extra-cellular CRD regions of the protein (amino acids 42 – 214, predicted molecular weight 19.8 kDa) was expressed in *E.coli* as inclusion bodies. After refolding, the protein was purified by size exclusion chromatography on a Superdex 75 column showing three peaks, corresponding to high-molecular weight aggregates, high-molecular weight contaminants and soluble DR6 (Figure 1.3.3). The peak containing DR6, elutes at a molecular weight of ~ 20 kDa based on standards run on the same column, corresponding to a monomeric state.

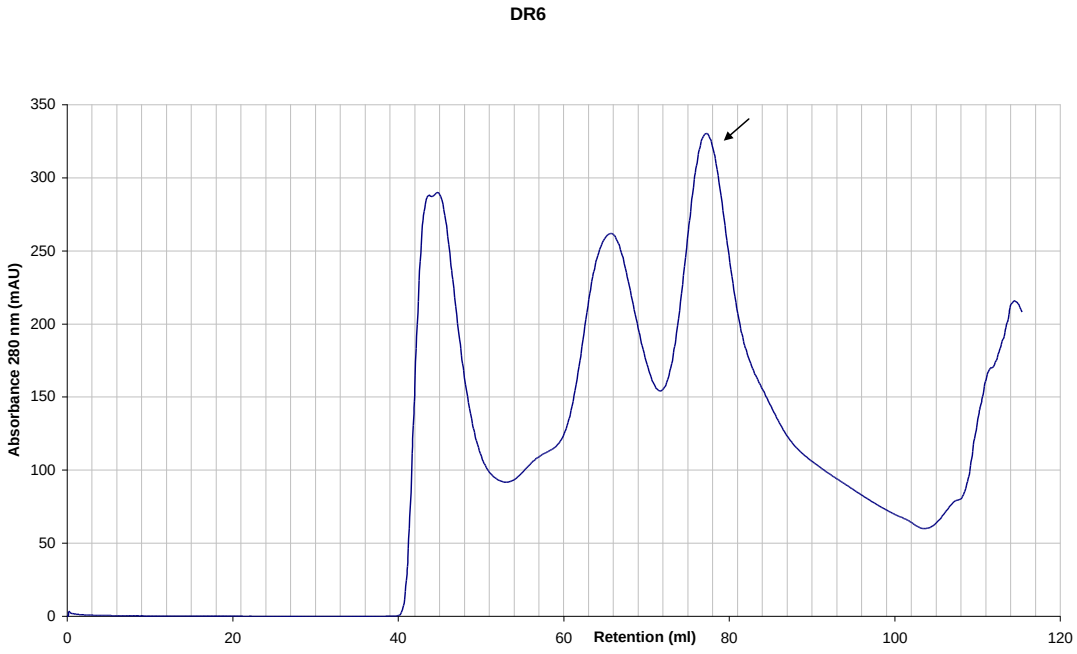


Figure 1.3.3 Purification of refolded DR6
Refolded DR6 was purified on a Superdex 75 16/60 column.
The peaks correspond to 45ml - high molecular-weight aggregate; 65 ml – high molecular-weight contaminants; 79 ml – soluble DR6, labelled with an arrow. The peak at the end of the purification corresponds to arginine from the refolding buffer.

Purified recombinant DR6 migrates as a single band under reducing conditions on SDS-PAGE (apparent MW of 20 kDa), but runs with a slightly lower apparent mass under non-reducing conditions (Figure 1.3.4). This is because the protein is not fully unfolded, retaining the characteristic compact form of a TNF Receptor cysteine-knot and thus being more mobile on an acrylamide gel than the denatured protein.

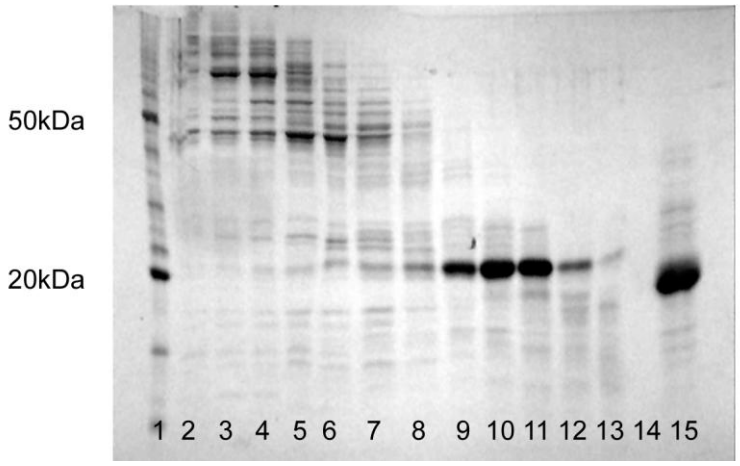


Figure 1.3.4 Characterisation of DR6 by SDS-PAGE
Samples of FPLC fractions between 51 and 87 ml were subjected to SDS-PAGE analysis. Pure DR6 is visible in lanes 8 to 11. Non-reduced DR6 is shown in lane 15; this migrates at a slightly lower apparent molecular weight to the fully denatured and reduced protein.

Dynamic light scattering experiments carried out at 1 mg ml^{-1} protein concentration show that DR6 has an apparent molecular weight of 38 kDa. Although this might be a consequence of the protein dimerising in solution, this is not consistent with the gel-filtration results. A more likely explanation is that the protein hydration model used by the Dynamics analysis software assumes that the protein is globular and has a roughly spherical shape, whereas TNF receptors are long and thin, hence any estimation of their size based on a globular estimation would be somewhat higher than expected. The polydispersity of the protein is relatively low at 7% (Appendix DLS).

Mass spectroscopy analysis shows that the protein has a molecular weight of 19,688 Da (Appendix MS). This is eighteen Daltons short of the calculated mass of the protein, which corresponds to the number of hydrogen atoms that would be lost on the formation of the nine disulphide bridges found in the protein. Taking the MS data into consideration the DR6 product is in all likelihood correctly folded.

Crystallization of DR6

Potential crystallization conditions for DR6 were screened using the OPPF Cartesian microdrop dispensing robots and the eight standard 96 condition screens available. - DR6 was set up in crystallization experiments at a concentration of 15 mg ml^{-1} in 50 mM TRIS pH8, 150 mM NaCl. Crystals grew in Hampton PEG/Ion Screen (38) (200 mM di-ammonium tartrate, polyethylene glycol 3350 20 % w/v) after three days and had a rugby ball like shape (Figure 1.3.5)

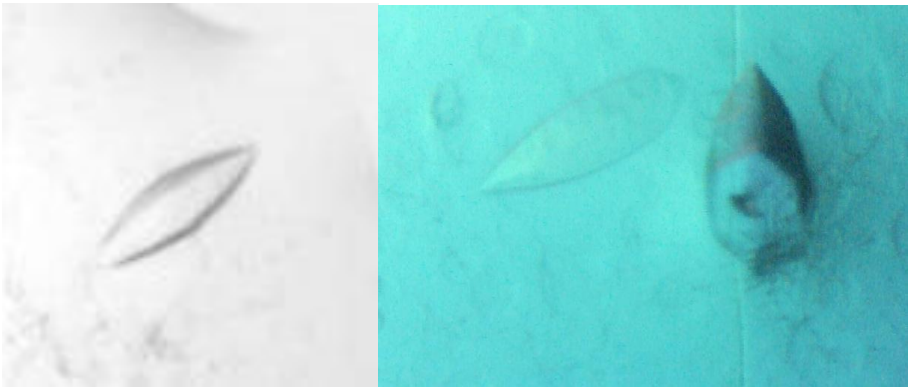


Figure 1.3.5 Example DR6 crystals
These crystals of DR6 were grown in 200 mM di-ammonium tartrate, 20 % PEG 3350.
The crystals were taken to the ESRF (Grenoble, France) and were mounted on beamline ID23-2 using a cryoprotectant of 50 % reservoir solution and glycerol. A dataset was collected to 3.4 Å and subsequent indexing gave a spacegroup of $P6_{1/5}22$ with cell dimensions a = b = 77.6 Å, c = 182.9, α = β = 90 º, γ = 120 º. Further analysis of this data is discussed in Chapter 1.4.

1.3.6 Purification of Osteoprotegerin

An N-terminally histidine tagged N and C-terminally truncated Osteoprotegerin corresponding to amino acids 26 – 188 of 19.6 kDa was expressed in *E.coli* as inclusion bodies. After refolding, the protein was purified by size exclusion chromatography on a Superdex 75 column, showing a large peak at 80 ml. This corresponds to osteoprotegerin existing in a monomeric state (Figure 1.3.6).

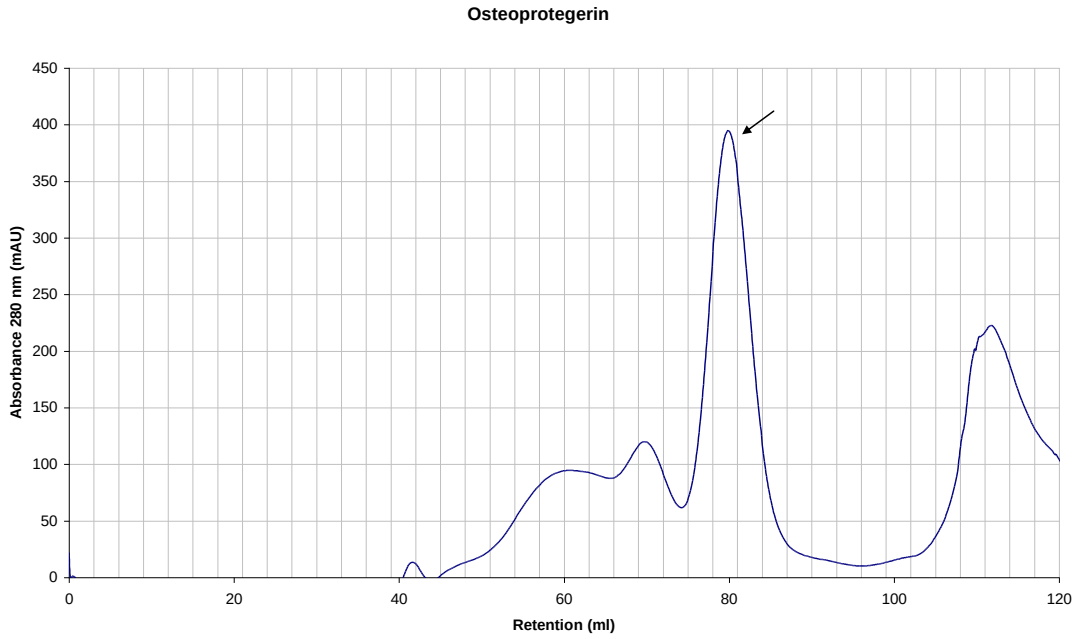


Figure 1.3.6 purification of refolded Osteoprotegerin
Refolded Osteoprotegerin was purified on a Superdex 75 16/60 column. Soluble protein is found in the major peak at 80 ml retention, labelled with an arrow.

The purified recombinant OPG was somewhat unstable in solution, with the majority of the protein precipitating within minutes of purification. The addition of 100 mM KCl to the FPLC buffer prevented the rapid precipitation of the protein, but if left at 4 ^{o}C for 24 hours the majority of the protein would leave solution. It was initially thought that this was due to the isoelectric point of the construct being very close to the pH of the FPLC buffer, but attempts to refold and purify the protein using pH 6

MES buffers did not to stop this precipitation. SDS-PAGE analysis of the peak fractions from gel-filtration show pure protein exists in the peak fractions (Figure 1.3.7). Under non-reducing conditions, the protein exhibits similar behaviour to that seen for DR6, with the protein migrating at a slightly lower apparent molecular weight, due to it being incompletely denatured.

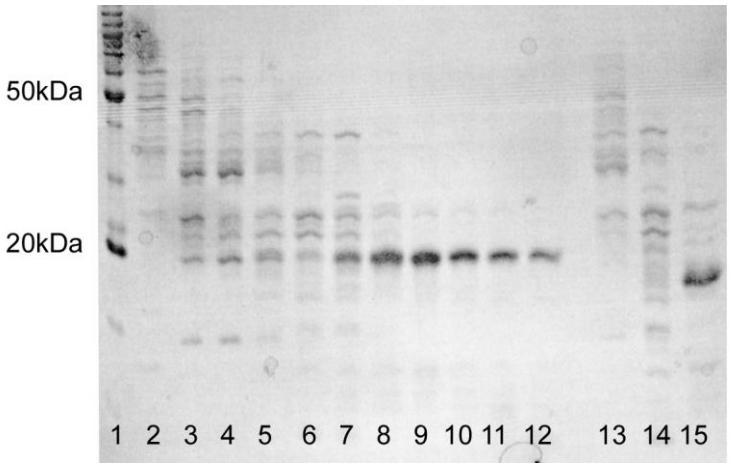


Figure 1.3.7 SDS-PAGE analysis of Osteoprotegerin
Samples of FPLC fractions between 63 and 96 ml were subjected to SDS-PAGE analysis. Pure OPG is visible in lanes 8 to 11. Non-reduced OPG is shown in lane 15; this migrates at a slightly lower apparent molecular weight than the fully denatured and reduced protein.

DLS analysis of the protein solution gives an estimated major mass peak at 40.6 kDa with no polydispersity. Mass spectroscopy analysis of OPG showed that the protein had a mass of 19,110 Da, which is 359 Da short of the predicted mass that the protein should show under MS, even with the N-terminal methionine absent. It is possible that the loss of two or three residues from the C-terminus of the protein, including a CRD cysteine, could lead to the destabilisation of the final CRD and thus cause the protein to be insoluble. If this is the case, it is unclear why the protein should lose these residues given the sequencing of the vector was correct.

Protein from gel-filtration purification was used to set up crystallization trials as quickly as possible, to mitigate the potential for precipitation prior to crystallization. Thirteen crystallization plates, giving over 1200 individual crystallization drops, were set up at 5 and 10 mg ml^{-1} protein concentration. No crystals formed over the time course of the project.

1.3.7 Purification of soluble VEGI

An N-terminally truncated VEGI construct (amino acids 90-252, calculated molecular weight 19,447 Da) was expressed as a soluble product in the cytosol of *E.coli* and purified by IMAC using a 5 ml Hi-trap chelating sepharose column with nickel and size exclusion gel filtration using a Superdex 200 16/60 column. A single major product peak is visible at 85 ml, corresponding to an apparent molecular weight of 46 kDa, which is close to the 58 kDa of the active trimeric complex (Figure 1.3.8). This result is consistent with the fact that TRAIL migrates as an apparently smaller complex on gel filtration.

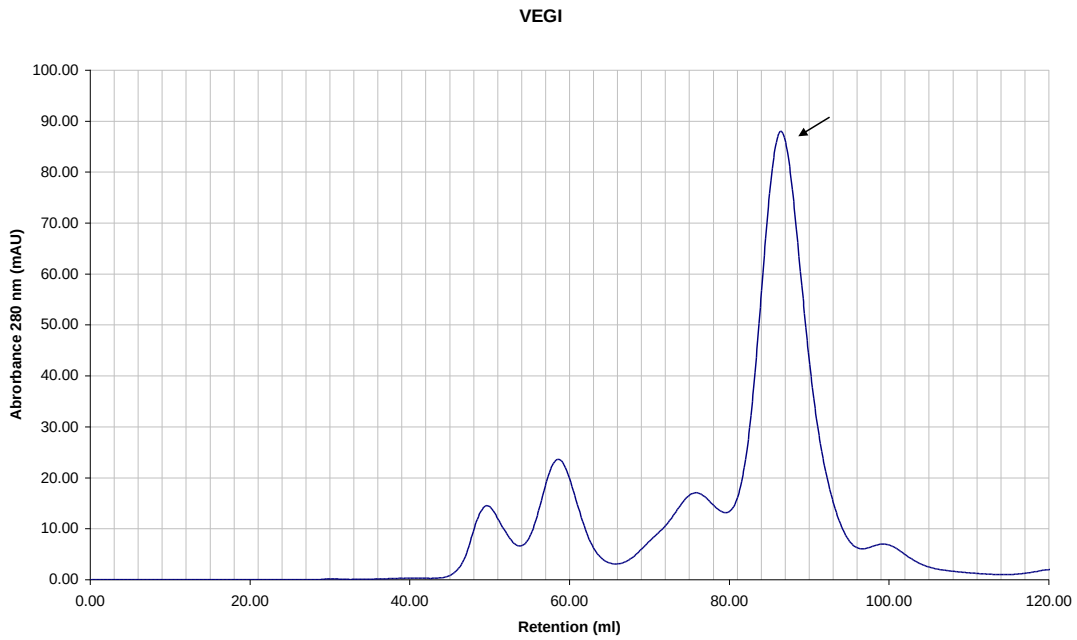


Figure 1.3.8 Purification of VEGI
Soluble VEGI was purified on a Superdex 200 16/60 column. A sharp peak is seen at 85ml, corresponding to the soluble trimeric ligand.

SDS-PAGE analysis of the peak fractions shows that peak fractions between 81 and 93 ml contain pure VEGI (Figure 1.3.9).

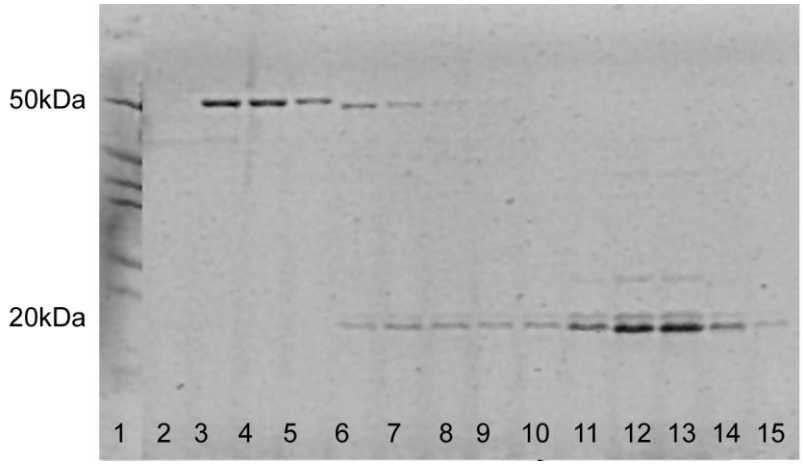


Figure 1.3.9 SDS-PAGE analysis of VEGI
Analysis of FPLC peak fractions of soluble VEGI by SDS-PAGE, 3 ml fractions between 57 and 96 ml are shown in lanes 3 to 15. Soluble VEGI is in fractions between 81 and 93 ml.

DLS analysis of the protein shows a major peak with an estimated mass of 71 kDa, and very good monodispersion. Mass spectroscopy analysis shows that the protein has a mass as expected at 19,315 kDa.

Crystallization conditions for soluble VEGI were screened at 5, 10, 15 and 20 mg ml^{-1} using the whole range of available screens (almost 3,000 crystallization experiments). Crystals grew in the Hampton Grid Screen MPD, condition containing 65 % 2-Methyl-2,4-Pentanediol (v/v), 0.1 M MES pH 6.0 (Figure 1.3.10). The crystals from this condition were mounted on ID14.4 at the ESRF (Grenoble, France) and a dataset was collected at 0.931 Å wavelength. The data was highly anisotropic giving diffraction to 3 Å in one crystal orientation and 5 Å in a different orientation, images and further discussion is in Chapter 1.4.

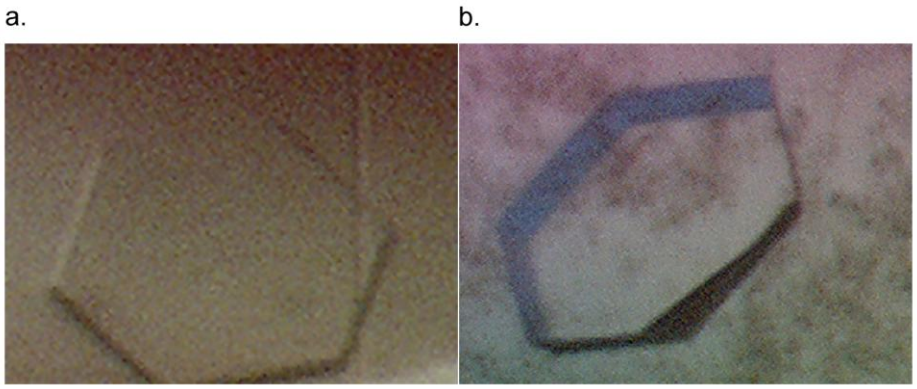


Figure 1.3.10 VEGI crystals.
Panels a. and b. show two crystals of VEGI from the same crystallization drop with a precipitant solution of 65 % 2-Methyl-2,4-Pentanediol (v/v), 0.1 M MES pH 6.0 .

1.3.8 Prokaryotic Protein refolding

One of the major bottlenecks in the protein production pipeline is the refolding of protein from prokaryotic inclusion bodies [75]. The reasons for the formation of inclusion bodies are not fully understood, but the differing environment of the eukaryotic and prokaryotic cytosol and the lack of common post-translational modification steps could contribute to the misfolding of eukaryotic proteins expressed in bacterial culture.

A number of different protein refolding methodologies exist [76-79], all relying on the basic premise that insoluble protein is initially denatured and then subjected to such conditions that proper folding of the protein of interest would occur. Within the mammalian cell, chaperones are involved in protein folding and research carried out into the use of *ex vivo* chaperones has had mixed success [80].

Aggregation of folding intermediates, or incorrectly folded protein, is a major problem with protein refolding, so all protocols go some way to ensure that the formation of aggregates is minimized as protein is transferred from denaturing conditions to refolding conditions. In the dialysis and diafiltration methods, the denaturing solution containing the protein is slowly dialysed against the refold buffer, or the protein solution is rapidly added to an intermediate refolding buffer and then dialysed against the final buffer. For both cases, the key premise is that the gradient from denaturing conditions to those promoting correct folding is achieved slowly enough to prevent aggregates forming in intermediate concentrations of denaturant.

In the rapid dilution method, denatured protein is quickly pulsed into the refolding buffer, which usually has a high-concentration of a chaotropic agent (such as polar amino acids, or carbohydrates) to encourage solubilisation of folding intermediates and to prevent aggregation upon removal of denaturant molecules by dilution. The on-column methodology relies on the fact that while immobilized upon a column, denatured and subsequently correctly folded protein, has limited scope for contact with other protein chains and thus aggregation is minimized. Those methods that involve any kind of dilution are not particularly well suited to very large scale refolding due to the need to concentrate the product to a useable volume. This concentration step is highly rate limiting.

A number of esoteric means to further prevent aggregation have been reported [81]. These usually take the premise that at particularly high or low temperatures, protein does not form aggregates so readily as at room temperature.

One of the problems with *in vitro* protein refolding is that the specific conditions for efficient refolding of any given protein cannot be readily predicted from all of the variables involved. Screening of multiple refolding conditions in small-scale trials has been used as the main empirical determinant of refolding, and a number of kits for this are commercially available [82]. However, the majority of investigators involved in protein refolding use a favoured method and a narrow range of conditions. The empirical approach to protein refolding is not widely used, although people are becoming more interested in factorial screening of refolding conditions [83]; there is also a useful website database of protein refolding conditions that have been

successfully applied to protein renaturation curated at Monash University in Australia (http://refold.med.monash.edu.au/) [84].

Since a number of the TNF targets were expressed in an insoluble form, a robust and highly reproduceable method of protein refolding had to be adopted. Within the Strubi laboratory a number of different protein families have been successfully refolded and crystallized using the rapid dilution method [85]. This method is more conceptually simple than any dialysis method and because the TNF ligands cannot be efficiently refolded on-column, due to the requirement for the formation of protein trimers in the stabilisation of the hydrophobic core of the protein.

Guanine hydrochloride was used as the initial denaturant, denatured protein was then reduced with 10 mM DTT. The refolding solution contained TRIS-HCl buffer, L-arginine as the chaotrope, and a cystamine/cysteamine redox couple to catalyze the formation of disulphide bonds. This particular set of conditions has previously proven successful for the refolding of the TNF ligand TRAIL [13] amongst other things and thus was adopted after initial unsuccessful commercial refolding trials (Avidis, personal communication). A commercial refolding screen (Novexin, Cambridge) utilising a number of proprietary protein protection agents has been tested against a number of OPPF target proteins, but the yield from this was of the order of two times lower than found using the rapid dilution method described (J. Brown, personal communication).

The concentration step, which can be highly throughput limiting, was made easier with the use of Vivacel concentrators instead of Amicon stir cells. The Amicon stir

cell uses constant gas pressure to achieve flow of liquid through the cell and must be closely monitored to ensure that the solution is not over-concentrated, or possibly left to concentrate to dryness. The Vivacel concentrator has a sealed pressure lid and a void volume of 500 µl. It works in such a way that the cell can be pressurised to up to four atmospheres, the gas disconnected and then left for the pressure in the cell to equilibrate by forcing liquid through the membrane into the flow-through chamber. With 250 ml of refold buffer in the concentrator and a new membrane in place, a single charge of 4 Atmospheres pressure is enough to concentrate the solution to 500 µl, overnight with no risk of loss of protein due to over-concentration. Additionally the entire cell can be placed on a rocking platform to reduce gradient formation within the concentration chamber.

TRAIL was used as a control within the refolding experiment to ensure that the particular conditions used were replicable and indeed, multiple attempts at the refolding of this ligand were successful and gave a good yield of 10 % from the initial inclusion bodies to the purified protein.

Refolding of DR6 was also successful and easily replicated. That the protein should refold in the conditions used was fortuitous given initial concerns that this protein, with its high cysteine content and an absolute requirement for the formation of correctly ordered disulphide bridges, may be unsuitable for the initial refolding conditions. The well-defined peak on gel-filtration and mono-dispersity of the sample under dynamic light scattering implies the existence of a single species. Mass spectroscopy analysis of this, giving a mass of 19,668 kDa, corresponds to the protein with all potential disulphide links being formed within a single chain.

Osteoprotegerin proved somewhat more difficult to successfully refold and purify. Initial results showed a high yield of protein from the inclusion bodies, but post-purification the majority of the pure protein would fall out of solution. The solubility of the protein in the pH 8 refold and gel filtration buffer was questionable, given the fact that the construct had a calculated pI of 7.8, attempts were made to refold and purify the protein at pH6, but again, if left for any length of time, the protein would precipitate. Any protein that was left in solution would migrate on gel-filtration at the same apparent mass as in the initial purification steps. Mass spectroscopy analysis of the protein showed that the pure protein had a lower mass than expected, and may have lost a number of residues. Should these have come from the C-terminus, a structural cysteine would be missing and thus it would be likely that the protein would be unstable in solution due to the lack of the final stabilising cysteine bridge.

1.3.9 Purification of Eukaryotic expressed Fc-tagged LTβR

LTβR was initially produced as an Fc-tagged N- and C-terminally truncated protein (amino acids 28-215, calculated molecular weight 21 kDa) in HEK 293 T-Cells. This was purified by Protein A-Fc affinity chromatography initially, followed by cleavage of the Fc and then a final size-exclusion gel filtration was performed on a Superdex 200 16/60 column (Figure 1.3.11).

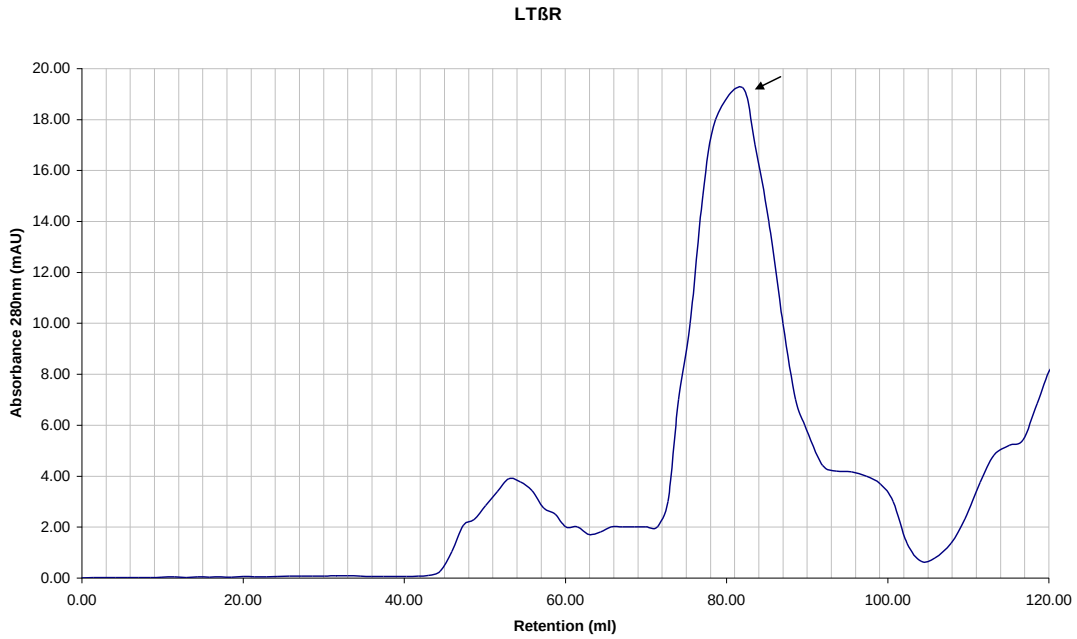


Figure 1.3.11 purification of Fc-tagged LTβR
LTβR was purified on a Superdex 200 16/60 column after Fc-affinity chromatography using protein A column. The tag-free protein migrated at an apparent mass of 20 kDa.

SDS-PAGE analysis of LTβR fractions from gel filtration reveals two bands of equal intensity and closely related molecular weights, which are not separated by gel-filtration. This is likely to be the result of two differently glycosylated species with the sharp nature of the bands implying a simple glycosylation pattern (Figure 1.3.12).

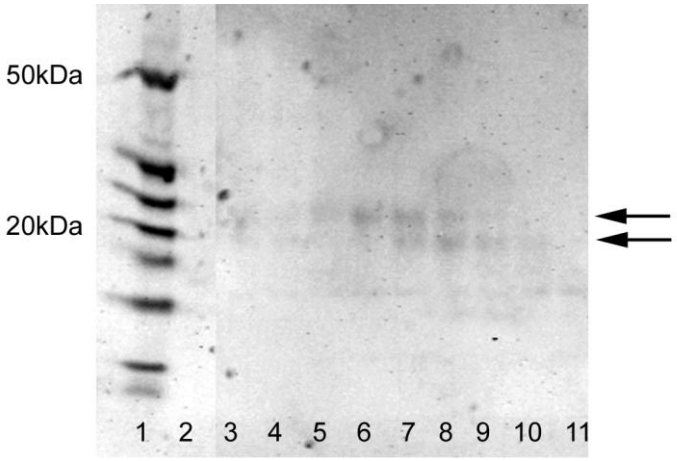


Figure 1.3.12 SDS-PAGE analysis of LTβR
Fractions from gel-filtration of LTβR were subjected to SDS-PAGE analysis, 3ml fractions between 69 and 102 ml are shown. Two bands of LTβ-R are visible in lanes 3-9 corresponding to glycosylated products that were not separated by gel filtration (indicated with arrows).

Analysis of the protein sequence gives two potential N-linked glycosylation sites at N40 and N177 and a single O-linked site at T199. After treatment with PNGase to remove potential N-linked sugars, mass spectroscopy analysis yielded two distinct mass peaks at 22,516 Da and 21,718 Da.

The 22,516 Da mass peak corresponds to the following sequence, which is expected from the construct:

RLRVPGKLSQPQAVPPYASENQTCRDQEKEYYEPQHRICCSRCPPGTYVSAKCSRIRDTVCATCAENSY
NEHWNYLTICQLCRPCDPVMGLEEIAPCTSKRKTQCRCQPGMFCAAWALECTHCELLSDCPPGTEAELK
DEVGKGNNHCVPCKAGHFQNTSSPSARCQPHTRCENQGLVEAAPGTAQSDTTCKNPLEDPEVLFQ

While the 21,718 Da peak has the following sequence with the seven amino acid sequence shown above in red missing:

LSQPQAVPPYASENQTCRDQEKEYYEPQHRICCSRCPPGTYVSAKCSRIRDTVCATCAENSYNEHWNYL
TICQLCRPCDPVMGLEEIAPCTSKRKTQCRCQPGMFCAAWALECTHCELLSDCPPGTEAELKDEVGKGN
NHCVPCKAGHFQNTSSPSARCQPHTRCENQGLVEAAPGTAQSDTTCKNPLEDPEVLFQ

Analysis of the sequences using the SignalP server [86] gives a very weak consensus sequence for secondary signal cleavage that corresponds to the seven amino-acid sequence that is seen to be lost between the two species.

Crystallization of LTβR

LTβR was set up in crystallization trials at 10 mg ml^{-1} and produced spherulite-like crystals (Figure 1.3.13) in Hampton Quick Screen D1 (di-potassium hydrogen phosphate 36 mM, sodium di-hydrogen phosphate 1.764 M, final pH 5) and Hampton Crystal Screen 35 (sodium di-hydrogen phosphate 800 mM, HEPES-Na pH 7.5 100 mM, potassium di-hydrogen phosphate 800 mM).

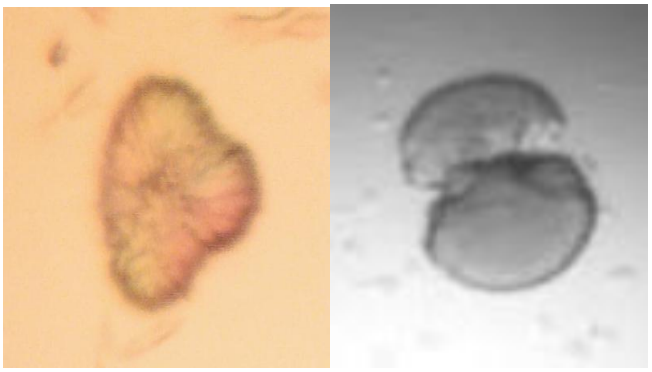


Figure 1.3.13 Example crystals of LTβR
To the left are crystals that formed in Hampton Quick Screen D1 and to the right crystals that formed in Hampton Crystal Screen 35.

An OPPF optimisation plate was set up in an attempt to grow better crystals using Hampton Crystal Screen 35. This optimization produced very small crystals with a similar morphology to those previously grown.

Mass spectroscopy analysis of the protein content of the crystals showed a mass of 22,304 Da, which corresponds to a loss of 212 Da over the 22,516 Da full-length construct. This loss of 212 Da would appear to correspond to protein with a truncation of two amino acids.

1.3.10 Purification of Histidine tagged LTβR

In an attempt to produce an homogenous LTβR product, a second construct was produced as a non-cleaveable N-terminal histidine tagged fusion using a pHLsec vector. The construct was N- and C-terminally truncated to include amino acids 31 to 215 (calculated molecular weight 21,866 Da) and was produced in 293 T-cells.

The protein was purified by nickel histidine affinity chromatography and size exclusion gel filtration using a Superdex 200 16/60 column with a major product peak at 78 ml retention (Figure 1.3.14).

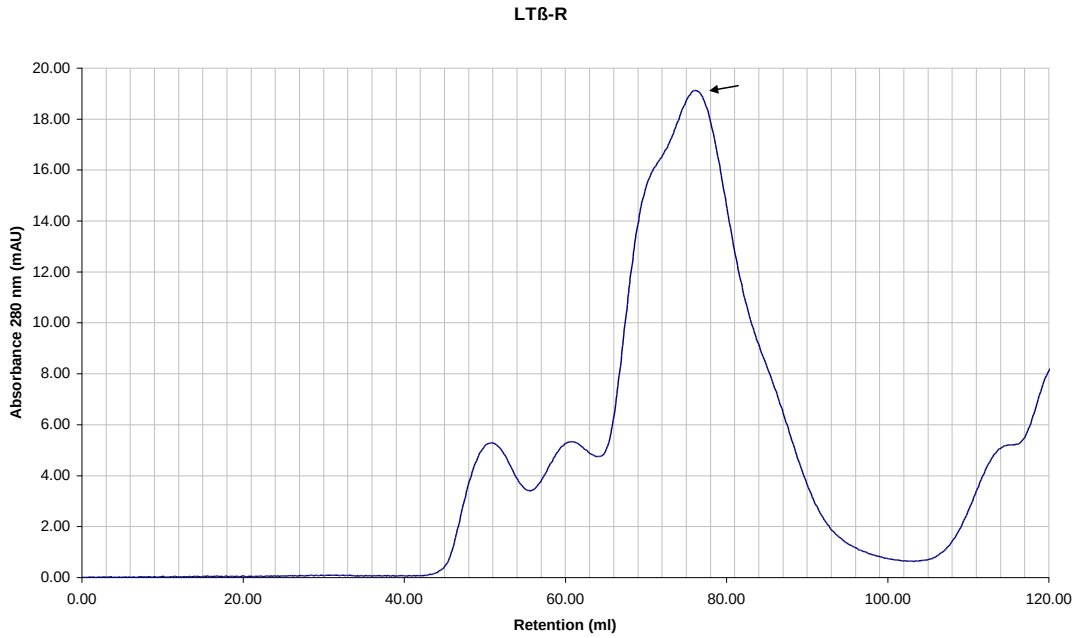


Figure 1.3.14 purification of histidine tagged LTβR
Histidine tagged LTβR was purified on a Superdex 200 16/60 column. Pure, soluble protein is found in the 78 ml peak.

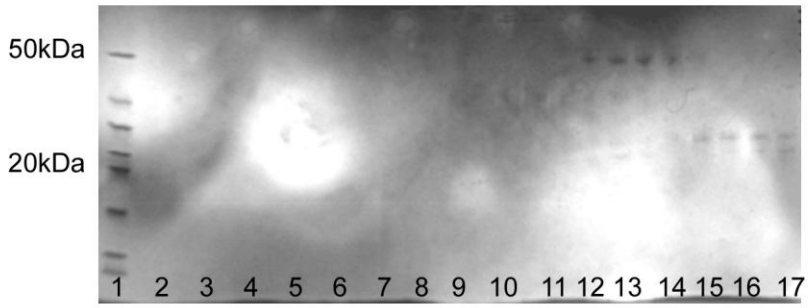


Figure 1.3.15 SDS-PAGE analysis of LTβR
Fractions from gel-filtration of LTβR were subjected to SDS-PAGE analysis. The two glycosylated LTβR products are visible in lanes 14-17.

Again, SDS-PAGE analysis of the protein revealed two bands in the peak fractions (Figure 1.3.15). Mass spectroscopy analysis of this protein gave a major mass peak of 21,848 Da, corresponding to the non-glycosylated product, and another major peak at 26,738 Da, corresponding to the glycosylated product. PNGase treatment of the protein produces a single peak at 21,846 Da, which corresponds to the mass of the protein with all CRD cysteine residues oxidised and the loss of sugars from the two N-glycosylation sites.

Crystallization of histidine tagged LTβR

Yields of this protein were low (0.13 mg per litre of culture) so this protein was put into crystallization trials using the OPPF limited protein screen. To date no crystals have grown.

1.3.11 Eukaryotic production of TNF receptor targets

The production of soluble TNF receptor proteins in a eukaryotic system should produce correctly folded protein, which can be purified from the cell growth media. This mitigates the need for protein refolding; which, in the case of TNF receptors produced in *E.coli*, is required due to the formation of inclusion bodies when the proteins are expressed in a prokaryotic system.

One of consequences of producing protein in eukaryotic cells is that proteins can be post-translationally modified with glycosylation and signal peptide cleavage. Glycosylation of proteins can interfere with efficient crystallization because the bulky, heterogeneous and flexible additions can interfere with the packing of the protein into a crystal lattice. This is not always the case, as a number of proteins do produce diffraction quality crystals even with complex glycosylation [87].

In the case of LTβR, a number of glycosylation products were produced and the mixed population was further added to by differential signal-peptide cleavage events. A second construct was created to work around the potential peptide cleavage event, but the glycosylation problem remained.

Initial attempts to crystallize the protein showed that the non-glycosylated product preferentially formed crystals. Purification of the protein using a lectin column to separate the glycosylated product from the non-glycosylated product proved unsuccessful.

What has proved most problematic is the limited amounts of protein that are produced by the eukaryotic system and the limited number of crystallization screens that can be set up because of this.

1.3.12 Crystallization of TNF targets

Published crystal structures of TNF ligands and receptors show that a wide range of protein concentrations and crystallization conditions have yielded usefully diffracting crystals and the lack of a firm trend for crystallization from these data prevents the formulation of a simple 'TNF' crystallization screen. The utility of the Cartesian microdrop system and the automated imaging system in the OPPF cannot be understated when screening a wide range of potential crystallization conditions. For the TNF project, 178 96-well crystallization plates have been set up, totalling over 17,000 crystallization conditions screened for the various proteins.

Although VEGI, DR6 and LTβR formed crystals, DR6 proved to be the only target to yield any well diffracting crystals and the data obtained from these have proven somewhat difficult to solve due to the nature of the protein and the weakness of the data. The high cysteine content of the protein makes it an ideal target for sulphur SAD phasing, although in the case of the crystals that have been tested for this at a synchrotron, only a very weak anomalous signal is seen, preventing phasing by this method. Attempts to crystallize the protein using selenium labelled methionine residues have yet to yield any crystals from a targeted optimization of the conditions giving crystals with the native DR6 protein. The preliminary analysis of X-ray data produced from DR6 crystals is discussed in Chapter 1.4.

1.4 Analysis of X-ray data from TNF protein crystals

1.4.1 Preliminary X-ray analysis of DR6 crystals

A total of over 2,000 crystallization trials for DR6 were set up at concentrations of 10, 15 and 20 mg ml^{-1} using a Cartesian Technologies Microsys MIC4000 (Genomic Technologies) [88]. This system produces 96-well sitting drop vapour diffusion experiments, with a reservoir volume of 95 µl, and drops of 100 nl reservoir solution plus 100 nl protein. Crystals were found in 23 drops in 19 different conditions, five of which had crystals that were suitable for mounting on an x-ray source.

a. Initial screening of DR6 crystals

A number of crystals of DR6 were found to form in a drop with 200 mM Di-ammonium Tartrate, 20 % PEG 3350 as the precipitant (Figure 13.5). The crystals were taken to the ESRF (Grenoble, France) and were mounted on beamline ID23-2 using a cryoprotectant of 30 % glycerol and flash-cooled to 100 K with a cryostream (Oxford Cryosystems). The crystals gave well ordered, but very weak Bragg Diffraction (Figure 1.4.1)

In total, 540 images were taken at an oscillation range of 0.5 º at a wavelength of 0.98 Å using an ADSC Q4 CCD detector (Area Detector Systems Corporation). Processing, merging and scaling of the data were carried out using the programs DENZO and SCALEPACK (http://www.hkl-xray.com). Processing statistics are shown in Table 1.4.1.

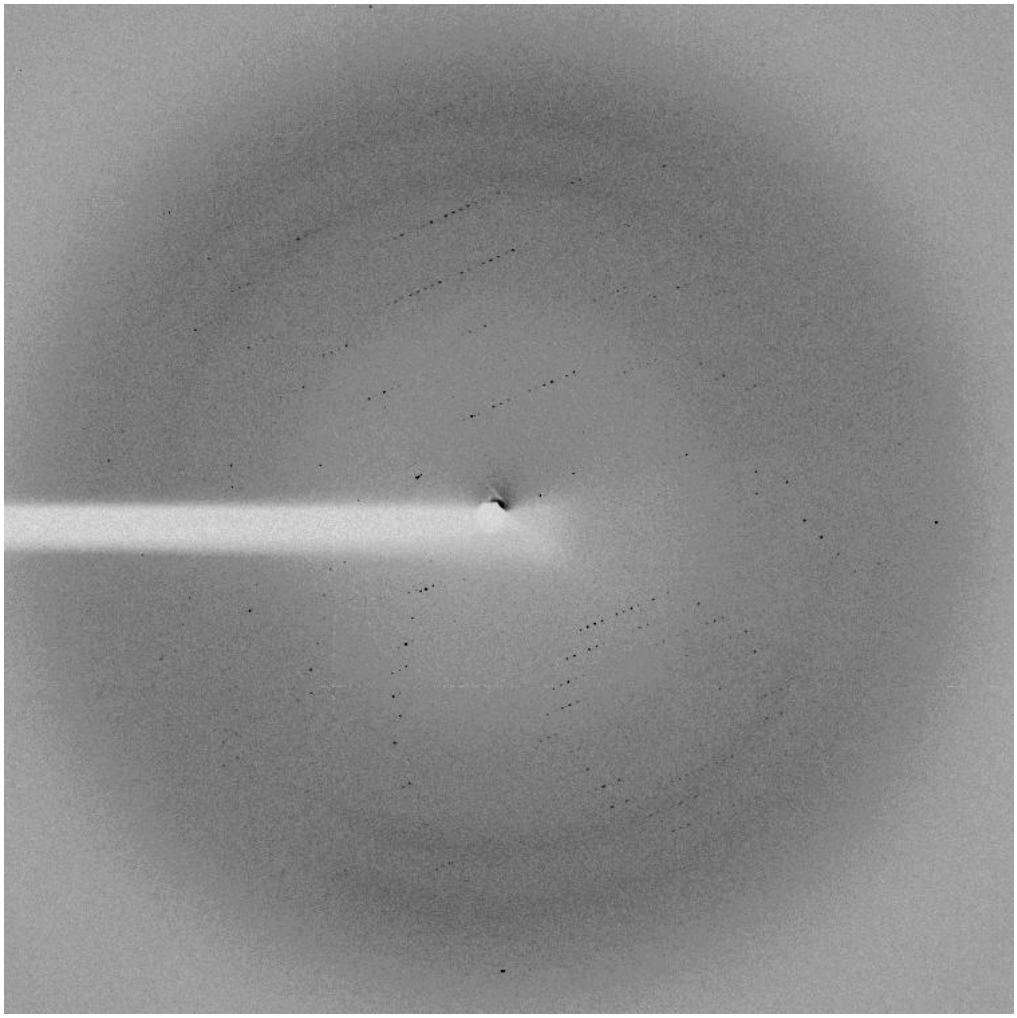


Figure 1.4.1 Diffraction pattern from DR6 crystals.
The diffraction from the crystals shown in Figure 3.5 was well ordered, but very weak.

	DR6 1
Data collection	
Resolution range (Å)	30 – 3.4
Number of collected reflections	149,723
Unique reflections	9,311
Completeness (%) (highest resolution shell)	100 (100)
R_{merge} (%)	34.1 (96)
I/σI	9.1 (2.9)
Space group	$P3_{2/1}21$
Unit cell	
Dimensions (Å) (a, b, c)	77.6, 77.6, 182.9
Angles (°) (α, β, γ)	90, 90, 120
Source	ESRF ID23-EH2

Table 1.4.1 processing statistics for DR6

The high R_{merge} indicates that the space-group of this data could be incorrectly specified, or it could be a consequence of the highly redundant data and the weak nature of observed reflections. Using the Matthews [89,90] method of analysing the content of the unit cell, the solvent content of the cell was estimated to be 54 %, with three molecules per asymmetric unit. Subsequent datasets that were collected for DR6 confirmed that the spacegroup was incorrectly specified and this is discussed futher in later parts of this section.

The structures of DR5, TNFR1 and HVEM were used as potential molecular replacement solutions for the CaspR molecular replacement web-interface [91]. The CaspR web-interface goes some way to automating the steps involved in finding a molecular replacement solution from a data set; taking as its input the structure factors, sequence of the protein in the crystal and structures of homologous proteins. Initially a multiple sequence alignment is calculated for the protein sequences provided, using T-coffee [92,93], which are then used by Modeller [94] to generate an ensemble of homology models for the protein. Each homology model is then screened in search of a molecular replacement solution using AMoRe [95], then these solutions are pre-refined using CNS [96]. In the case of this particular DR6 data set this strategy failed to find any suitable molecular replacement solution.

Two maximum likelihood molecular replacement programmes, BEAST [97] and Phaser [98], were used with an ensemble of the known TNF receptor structures and homology models produced by CaspR. Again, these were both unsuccessful.

The probable shape of the protein coupled with the low-resolution data collected conspired, in this case, to prevent the easy determination of a molecular replacement solution for the phasing of the X-ray data. More crystals were grown, with the intention of attempting phasing using the anomalous scattering of the sulphur in the protein and a seleno-methionine derivative of the protein was expressed to use the anomalous scattering potential of this heavy atom.

1.4.A Molecular Replacement

In the collection of X-ray diffraction data from protein crystals, only the amplitudes of the diffracted beams are measured, presenting a problem in the calculation of electron density maps and the molecular structure solution because the phase angles of the diffracted beams are necessary for these calculations. One way that this phase problem can be solved is by using the structure of a known, related protein, which would be expected to have a similar X-ray diffraction pattern. This model is used to calculate initial phase angles from which electron density maps can be calculated and the model structure refined. This is known as molecular replacement.

The method of molecular replacement to solve the phase problem for protein crystallography was first suggested by Rossmann and Blow [99]. The key to the problem is the correct placement of the known protein structure within the crystal lattice of the unknown protein; this involves the calculation of correct rotational and translational parameters. This can be understood with reference to the Patterson map, that is, the map of the vectors between all pairs of atoms within the molecule. If the pairs of atoms belong to the same molecule, then the Patterson vectors are relatively short and close to the origin of the Patterson map. If there are no inter-molecular

vectors in the Patterson map then the map calculated for the unknown structure and the known structure would be the same, apart from a rotation around the origin. In the case of homologous molecules, the map is not exactly the same, but similar. The longer vectors in the Patterson map represent inter-molecular distances and can be used, after the rotation of the Patterson map, to determine the translation of the molecule within the lattice. For a molecule with only a few atoms, the rotation and translation is relatively simple, but as the number of atoms in the molecule increases, the number of vectors in the Patterson map increases proportionally to the square of the number of atoms (minus the contribution from the origin peak and self-vectors from an atom to itself), i.e. peaks = N^2 - N non origin peaks. Obviously, for a protein molecule with over 1000 atoms this becomes computationally intensive very quickly.

Molecular replacement utilises the difference between self- and cross-vectors in the Patterson map to locate the molecules within the crystal lattice with reference to the model structure. A number of methods using the basis of the Patterson function have been developed, which simplify the calculation of the rotation and translation functions in the search for a solution to the phase problem. Two of the most widely used programmes are AMoRe [100] and CNS [96], the method of molecular replacement as implemented in AMoRe will be discussed here.

AMoRe has a number of separate modules, which carry out the rotation and translation functions and the rigid body optimisation of solutions from the molecular replacement. Firstly, the fast rotation function [101] is performed to maximise overlap between the model and the observed Patterson functions. The resulting peaks in the

Patterson maps are then used to calculate the correlation between the calculated and observed structure factor amplitudes.

The output from this is then used in the translation function, which incorporates a test of the crystal packing for each translation function. These are then tested for the correlation between calculated and observed structure factor amplitudes and this is the basis upon which acceptable solutions are selected.

b. Sulphur SAD experiment

DR6 crystals grew in four separate conditions in a Hampton Research Salt RX screen (see Chapter 2.15 for details) as shown in Figure 1.4.2. The best quality crystal gave well-ordered Bragg Diffraction to 2.95 Å when mounted on beam line BM14 at the ESRF, data processing statistics are shown in Table 1.4.2. A dataset was collected by J-B. Reiser at 0.97 Å wavelength over 0.5 º oscillations and 60 frames (Figure 1.4.2e). Unfortunately, a beam dump meant that a full data set could not be collected. A number of data sets (1,300 frames) were collected at 1.77 Å wavelength for the potential sulphur anomalous signal. The crystal from which data was collected (Figure 1.4.2d) has a $P6_{1/5}$ 2 2 space group with unit cell dimensions of a=b=77.3 Å, c=187.8 Å, α=β=90 º, γ=120 º. This gives one or two molecules in the asymmetric unit, with a solvent content of between 40 and 60 per cent.

Processing, merging and scaling of the data were carried out using the programs DENZO and SCALEPACK (http://www.hkl-xray.com). Processing statistics are shown in Table 1.4.2. SCALEPACK was used to estimate the extent of the anomalous signal using the method of initially scaling the data with the anomalous

flag set and then re-scaling this dataset with the anomalous flag removed. This allows a comparison of the difference between I^+ and I^-, which is reflected in the χ^2 values for the data, with an anomalous signal the χ^2 would be expected to be greater than one if the error model of the data is good. For the DR6 dataset, this method indicated that the anomalous signal extended only to 4.5 Å. Further analysis with SHELXC (http://shelx.uni-ac.gwdg.de/SHELX/index.html) confirmed this and further attempts to use this weak signal were unsuccessful, this was compounded by systematic errors from radiation damage to the crystal (Table 1.4.3).

As in Chapter 1.4.1a attempts were made to find a molecular replacement solution for these data, these were also unsuccessful.

	DR6 Native	DR6 Sulphur
Data collection		
Resolution range (Å)	30 - 2.95	45-3.8
Number of collected reflections	24,385	263,159
Unique reflections	5,106	6,184
Completeness (%) (highest resolution shell)	67.5 (44.1)	98.9 (98.9)
R_{merge} (%)	8.4 (38.4)	16.2 (30.3)
I/σI	15.7 (2.7)	51.2 (26.5)
Wavelength (Å)	0.9762	1.77
Space group	$P6_1$ 2 2	
Unit cell		
Dimensions (Å) (a, b, c)	77.3, 77.3, 187.8	
Angles (°) (α, β, γ)	90, 90, 120	
Source	ESRF-BM14	

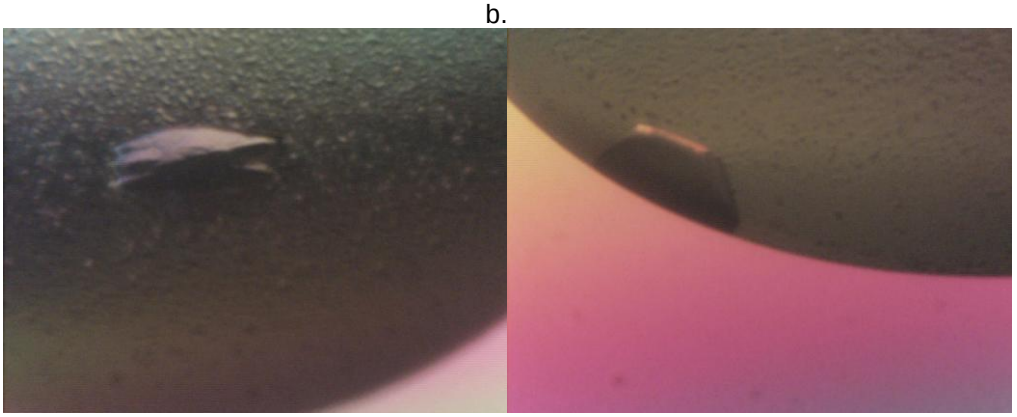
Table 1.4.2 processing statistics for DR6

Resolution	**∞ - 8.0**	**- 6.0**	**- 5.0**	**- 4.4**	**- 4.2**	**- 4.0**	**- 3.8**	**- 3.6**	**- 3.4**	**- 3.2**	**- 3.0**
N(data)	444	556	681	735	333	418	503	611	760	952	1112
<I/sig>	46.3	37.9	33.3	34.6	33.3	30.9	25.8	18.9	13.6	7.3	4.2
%Complete	92.9	98.9	99.3	99.5	99.4	99.5	99.6	99.5	99.7	99.7	95.0
<d"/sig>	1.93	1.36	0.97	0.88	0.73	0.74	0.71	0.59	0.59	0.57	0.53

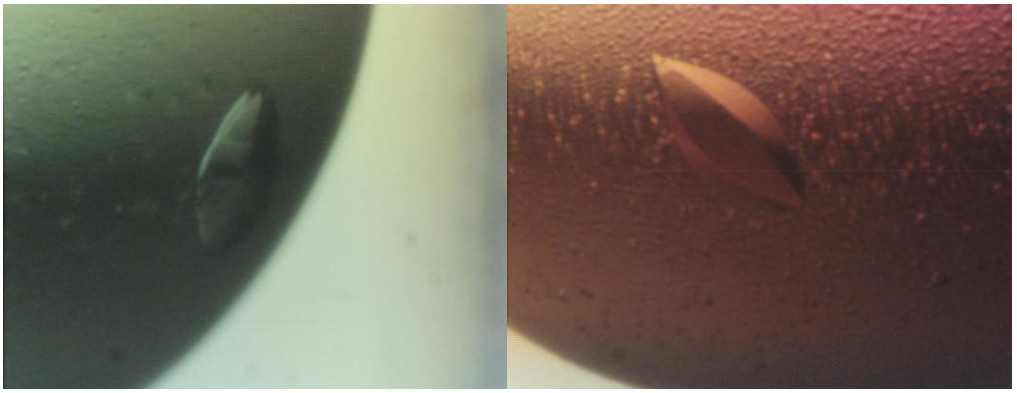
Table 1.4.3 Anomalous signal analysis from SHELXC for DR6

The scaled anomalous DR6 dataset was analysed using SHELXC. A <d"/sig> of 0.8 or less indicates no anomalous signal. For this data set the anomalous signal only extends to the 4.4 Å resolution limit, which is consistent with the results from Scalepack.

a. b.



c. d.



e.

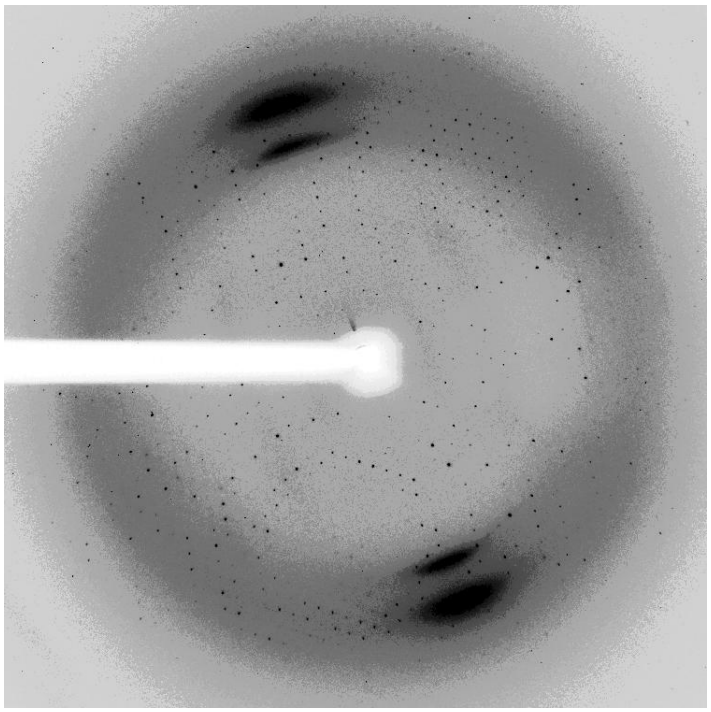


Figure 1.4.2 Crystals of DR6 grown in the Salt RX screen
The crystallization conditions for each crystal are described below. The crystal in panel four gave the best diffraction and data sets were collected from this.

a. 1.5 M Ammonium Sulphate, 100mM Tris, pH 8.5
b. 1.4 M di-Ammonium Tartrate, 100 mM Tris pH 8.5
c. 1 M tri-Ammonium Citrate, pH 7.0, 100 mM bis-Tris Propane, pH 7.0
d. 1.5 M Ammonium Sulphate, 100 mM bis-Tris Propane pH 7.0
e. Diffraction pattern given by the crystal in panel d.

1.4.2 Initial screening of VEGI crystals

Crystals of VEGI from material purified using a 50 mM pH 7.5 Tris-HCl, 50 mM NaCl FPLC buffer grew in the Hampton Grid Screen MPD condition containing 65 % 2-Methyl-2,4-Pentanediol (v/v), 0.1 M MES pH 6.0 (Figure 1.4.3a/b). The crystals displayed hexagonal morphology and were mounted on ID14.3 at the ESRF (Grenoble, France) and a dataset was collected at 0.931 Å wavelength. The data were highly anisotropic giving diffraction to 3 Å in one crystal orientation and only 5 Å in a different orientation (Figure 1.4.3c/d). Across the oscillation range of the data collected, the crystal also displayed a very high mosaicity. An attempt was made to determine unit cell and space group parameters from these data using Denzo [102], this gave a primitive hexagonal space group with a=b=119.3 Å ,c=197.1 Å, α=β=90 º, γ=120 º. Because of the overall poor quality of this data, integration and further analysis could not be performed.

The VEGI crystals formed in a precipitant solution that would allow for dehydration of the crystals in an attempt to improve the quality of diffraction; unfortunately, no crystals have been available for this procedure to date.

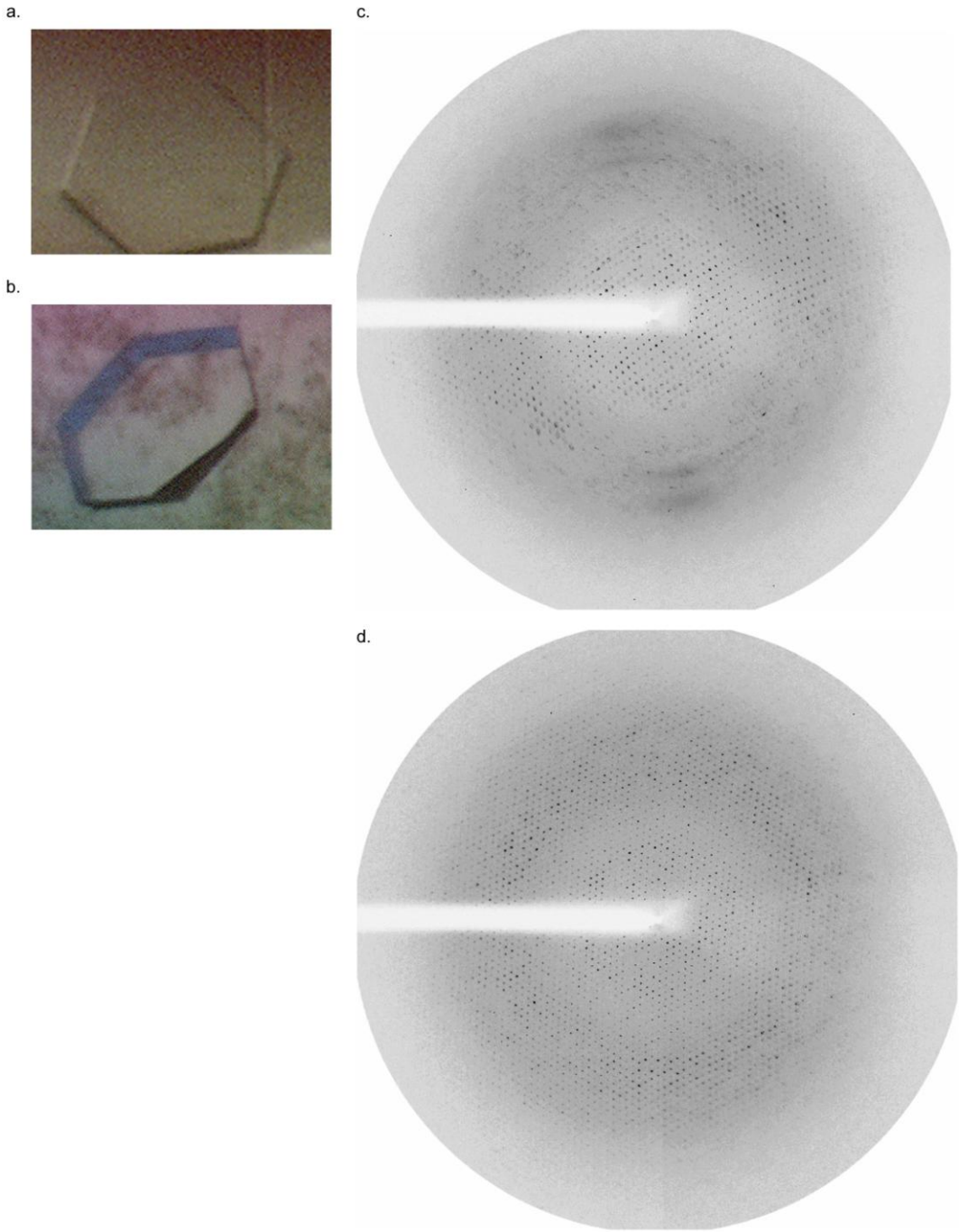


Figure 1.4.3 Crystals of VEGI and x-ray diffraction pattern produced by the crystals
Crystals of VEGI are shown in panels a and b. The diffraction pattern produced by one of these crystals is shown in panels c and d. The diffraction for this crystal is highly anisotropic there is a large degree of crystal mosaicity apparent.

1.5 Conclusions

The TNF receptor and ligand superfamilies were chosen as targets for the pilot of the OPPF protein production pipeline as they are of high biomedical interest and represent a set of proteins, for which a large number of homologous members of each family exist, yet many have not been crystallized, or had their crystal structures solved. In the sixteen years since the solution of the structure of TNFα [18], nine of the nineteen TNF ligand superfamily proteins and only six of the twenty-six TNF receptor-like proteins have been solved.

1.5.1 OPPF pipeline

The implementation of the OPPF pipeline applied to this project was successful in that a large number of protein expression constructs were created and screened for potential protein expression for further crystallization experiments. In total twenty-nine protein expression constructs were created, fifteen of which were potentially suitable for large-scale protein expression. Unfortunately, only four of these constructs produced any quantities of protein that were useful for further study. The failure rate for these constructs is considerably higher than that of bacterial targets within the OPPF. The nature of these human proteins, with the TNF receptor constructs having high cysteine content, may indicate that they are not generally easy protein targets for prokaryotic expression (even as inclusion bodies). The attempts made to produce the TNF receptor like protein LTβR in a eukaryotic system were only nominally successful, with very small quantities of protein produced, from which very little further experimentation could be done.

1.5.2 Refolding of protein material from E.coli inclusion bodies

A great many proteins, when produced in *E.coli* cells form insoluble inclusion bodies and the protein must first be refolded before it can be used in any biochemical or crystallization experiments. A relatively straightforward rapid dilution protein refolding technique (Chapter 1.2.7) was applied to the targets in this project that formed a high yield of inclusion bodies (TRAIL, DR6 and OPG). The production of TRAIL in this manner has been discussed before [13], but it has never before been applied to TNF receptors. In the case of DR6, this method was highly replicable and produced yields of around 10 %, which compares favourably with yields of 5 % for this protein when refolded using commercially available kits (J. Brown, personal communication). This protein produced a number of crystals, which are discussed below.

1.5.3 Crystallization of TNF superfamily targets

There are thirty-two crystal structures of various TNF ligands, receptors and receptor/ligand pairs in the protein databank, although many of these are of the same complex (three TRAIL/DR5 structures [13,38,39], two BAFF/BAFFR [29,30]). In spite of the number of crystal structures available in the family, no pattern has emerged for favoured crystallization conditions and attempts at crystallization remain based on screening of many potential precipitant solutions. The implementation of the microdrop crystallization methodology within the OPPF allows the screening of large numbers of crystallization conditions relatively quickly and easily and the addition of an automated imaging system makes analysis of crystallization plates a simple matter.

This project produced two proteins that formed crystals giving diffraction. In the case of DR6 crystals formed in multiple related conditions, growing over the course of a number of weeks in each case. These crystals gave good Bragg diffraction, although at a relatively poor resolution. Attempts to solve the structure of DR6 using these data and the methods of molecular replacement and the Anomalous Dispersion of the sulphur atoms within the molecule both failed to generate suitable phases from which the structure could be refined. A seleno-methionine derivative of DR6 was created to utilise the potential phasing power of the anomalous signal from selenium, but no crystals of this protein have yet been found in the crystallization trials that were set up.

In the case of VEGI, crystals formed in only one condition. These crystals gave extremely poor diffraction, but showed potential for further optimization of the crystallization condition and manipulation of crystal hydration. Attempts to grow more crystals have been hindered by a lack of protein material with which to set up crystallization optimizations.

1.5.4 Future Perspectives

This project has provided a good test of the capabilities and systems in place at the OPPF. The fact that the rapid-dilution refolding method of protein refolding is effective when used against TNF receptor proteins is very encouraging for further work on this family of proteins. Although no crystal structures have arisen from this work, progress has been made on identifying potentially productive crystallization conditions and strategies for the solution of the structures of DR6 and VEGI.

Chapter 2 TNF/TNF receptor binding and potential small molecule inhibition

This chapter describes the analysis of the extracellular domains of those TNF ligand superfamily members for which a crystallographic model exists and the potential for the design or discovery of potential small molecule inhibitors of receptor/ligand binding. The introduction provides an overview of the current state of the art in TNF targeting therapies and drug design. The aim of this chapter is the analysis of TNF superfamily proteins for determinants of receptor binding and the exploration of methods for the design and discovery of potential small-molecule inhibitors of receptor binding.

2.1 Introduction

The TNF-superfamily ligands share a common scaffold, upon which is built specificity for their receptors. A number of these ligands bind to multiple receptors, yet each displays remarkable specificity and discrimination between receptors [11]. Discrimination of receptors by ligands is in part related to the circumstances of the protein [6]. A number of ligands exist in strict cellular populations and their milieu is specific to interactions with a narrow range of potential receptor partners[4]. If one considers the widespread tissue distribution of some TNF ligands, such as Lymphotoxin and TRAIL, this compartmentalisation is not enough to explain how, and why, this discrimination or specificity takes place [103,104].

2.1.1 Protein Recognition

Protein recognition has been shown to be related to the surface complementarity and the physical properties of the interaction surfaces, the charge distribution of the surface and the potential for favourable electrostatic interactions [105-110]. Understanding of the specificity of TNF receptor/ligand interaction has been greatly aided by the solution of a number of structures, by X-ray crystallography, of ligands and ligand/receptor pairs (Appendix Xtal). With these data, it is possible to dissect some of the key determinants of receptor specificity that ligands show, and analyse these to gain greater knowledge of the subtle processes of protein-protein recognition.

With this knowledge comes the potential for further investigation into how TNF ligand/receptor binding could be interfered with, using small molecules or peptide inhibitors. The involvement of TNF superfamily members in a wide variety of

disease pathologies makes them prime targets for this line of investigation and to date, there are two TNF targeting therapies: Etanercept and Infliximab.

2.1.2 TNF therapies

Etanercept is a recombinant TNFR2-Fc fusion protein, which, in addition to being used as a TNF antagonist to treat patients with tumour necrosis factor associated periodic syndrome (TRAPS) [111], has been used in patients suffering Psoriatic Arthritis [9]. This therapy is not widely used, as it has been shown to cause lupus-like symptoms as a side effect [112]. Infliximab is a monoclonal antibody raised against TNFα that is used as a treatment against poliomyelitis, rheumatoid arthritis and other inflammatory diseases [113]. Both therapies are based on recombinant human proteins and are designed to interfere with the interaction between TNF receptor and ligand to prevent abnormal levels of signalling occurring. One of the main drawbacks with this particular type of immuno-therapy is the likely side effect of immune complex deposition in the kidneys and at joints.

2.1.3 High-throughput and in silico drug discovery

Large recombinant proteins that need complicated storage and administration methods are not ideal drug therapies. The most valuable drugs are usually based on small molecules, hence the interest in the design, or identification, of some kind of small molecule inhibitors of TNF receptor/ligand binding. Although the traditional route of screening large libraries of compounds for their actions against a particular protein has served the pharmaceutical industry well thus far. Recently the tools of crystallography and *in-silico* modelling have become more prominent in drug discovery and screening [105,114-118].

With the growth in small molecule databases such as the one at the Cambridge Crystallographic Data Centre (CCDC) [119], the Available Chemicals Directory (ACD) and the National Cancer Institute Drug Information System Database (NCI-DIS) [120], lead discovery can now be carried out using the techniques of database-mining and molecular docking [121]. Such 'high throughput' drug design, utilising X-ray crystallography structures has been responsible for the development of over 30 drugs including HIV protease and thrombin inhibitors [105,122-125].

2.1.4 Aims

In this chapter, the potential for the design of small molecule antagonists of receptor/ligand binding is assessed, using currently known structures of TNF ligands and their cognate receptors, by the analysis of their receptor binding surfaces. Potential small molecules that may serve as drug leads for the TNF receptors may exist within structural databases, thus two strategies for the generation of search models for screening the NCI small molecule database [126] were pursued and their relative merits discussed. The first strategy used Ligbuilder [127] to generate a complement of *de novo* small molecules which were used as search models for use in a similarity search of the NCI database. The second strategy used molecular docking to screen the NCI diversity set [128] against the target protein; top scoring molecules from this were used as similarity search models in the same way as for those generated using Ligbuilder. Those molecules that were found using the similarity search of the full NCI database were then docked against their protein target to analyse potential binding energies and to gauge the relative efficacy of the search strategies.

2.2 Materials and methods

2.2.1 Analysis of TNF-receptor binding sites on TNF ligands

Crystal structures for the TNF receptor/ligand complexes were split into separate files for receptor and ligand, and crystallographic waters were removed from these structures. TNF ligands for which no co-crystal structure exists were also included in the analysis (Appendix Xtal for references and PDB accession codes). Solvent exposed and buried surface areas in the receptor/ligand complexes were calculated using Naccess [129].

The contours of the surface of a protein often have marked clefts or gaps that correspond to the binding site of a ligand or receptor, therefore surface regions corresponding to surface clefts were calculated for each ligand using Surfnet [130] with a 0.8 Å grid separation, 1.0 Å minimum radius and 4 Å maximum radius for gap spheres. Surface cavities were output from Surfnet as CCP4 format maps and then read into SPDBV [131] with the corresponding ligands for visualisation.

Protein molecular surfaces were calculated for each ligand using SPDBV with electrostatic potential mapped to the surface. These potentials were calculated using the Coulomb method, with a solvent dielectric constant of 80 at pH7 and coloured with +5 kT in blue and -5 kT in red. All figures were prepared using POV-Ray.

2.2.2 Small molecule generation by Ligbuilder

The crystal structures of TRAIL/DR5 (1D4V) and BAFF/BAFFR (1OQE) were edited to give the trimeric ligand and receptor chains in separate files and crystallographic waters were removed. Receptor structures were converted from PDB

files to Tripos MOL2 format using PRODRG [132], with hydrogens added at this stage. The Ligbuilder [127] module Pocket was run on each receptor/ligand pair to generate interaction grids and a pharmacophore for the generation of small molecules, with minimal feature distance set to 3.5 Å and maximal feature number set to eight.

For use in the Link and Grow modules, BAFFR and DR5 were cut down to contain two key functional sites, which acted as the input seeds for these programmes. The DR5 seed molecule contained the carboxylic acid moiety of E151 and the dimethyl sulphide group of M152. The BAFFR seed consisted of the guanidine group from R15 and the isobutane group of L13. All hydrogen atoms were specified as potential growth points for the programmes.

Initially the Link module was run on both receptor/ligand pairs with the following parameters: growing probability 1, linking probability 1, and mutation probability 0.5. The selection of molecules for each population within the genetic algorithm was based on the following chemical viability tests: molecular weight between 300 and 600 Da, LogP between 3 and 6, 3-6 hydrogen bond donors, 3-6 hydrogen bond acceptors, and a binding score between 10 and 5. The top ten molecules conforming to these selection criteria were output as MOL2 files for analysis.

The Grow module was run using the E151 carboxylic acid seed from DR5 and the BAFFR R15 guanidine seed with the following parameters: growing probability 1, linking probability 0.5, mutation probability 0.5 and the same chemical viability tests as for Link. The top ten molecules conforming to these selection criteria were output as MOL2 files for analysis.

2.2.3 NCI database search using generated molecules

The SMILES strings (described in Chapter 2.4.D) for molecules generated by Ligbuilder were calculated using Babel (reference) and these were used as the query strings for searching the NCI structural database. Molecules with a similarity to the query structure indicated by a Tanimoto [133] score of greater than 80% were then docked using the method described in 2.2.4.

2.2.4 NCI Diversity set docking

The workflow of all the steps described below are shown in Figure 2.2.1

a. Target preparation

The crystal structures of TRAIL/DR5 (1D4V) and BAFF/BAFFR (1OSX) were edited to give the trimeric ligand and receptor chains in separate files and crystallographic waters were removed. Polar hydrogens, Kollman charges [134] and solvation parameters were added to each ligand using AutoDock tools [135]. 60 x 60 x 60 $Å^3$, 3D affinity grids were calculated for the following atom types for each ligand: C, A (aromatic C), N, O, S, H, F, Cl, Br, I, P and e (electrostatic) by the use of Autogrid3 [135] (Figure 2.2.1).

b. NCI Diversity Set

The NCI diversity set is a library of some 1990 compounds chosen to represent the structural diversity in the NCI-3D structural database [126]. Each member of the set was set up for use in AutoDock using Autotors [135] to assign rotable bonds and Gasteiger charges were assigned.

c. Screening

AutoDock 3.05 [135] was used for the docking simulation, with the Lamarckian Genetic Algorithm (LGA) used for ligand conformational searching. The translation, rotation and internal torsions were defined as the variables for each ligand and each gene in the LGA represents one of these variables. For each compound the docking parameters were as follows: 100 docking trials, population size of 150, random starting position and conformation, step ranges of 1.5 Å, rotation step ranges of 35º, elitism of 1, mutation rate of 0.02, crossover rate of 0.78, local search rate of 0.06 and 10 million energy evaluations. The jobs were carried out on SGI machines within the Strubi lab computing facility. Final docked conformations were clustered by use of a tolerance of 1.5 Å root mean square deviation (rmsd). The top 1 % of compounds with the best simulated-binding energies that bound within the footprint of the ligand's cognate receptor were selected as potential inhibitors (Figure 2.2.1).

d. Similarity searching

The top 1 % of molecules based on simulated binding energies from AutoDock docking analysis of the NCI diversity set were used as search models for the full NCI-3D database consisting of 213, 628 compounds [136]. Compounds from the full database with greater than 70 % similarity index to the search models were docked using the method described above (Figure 2.2.1).

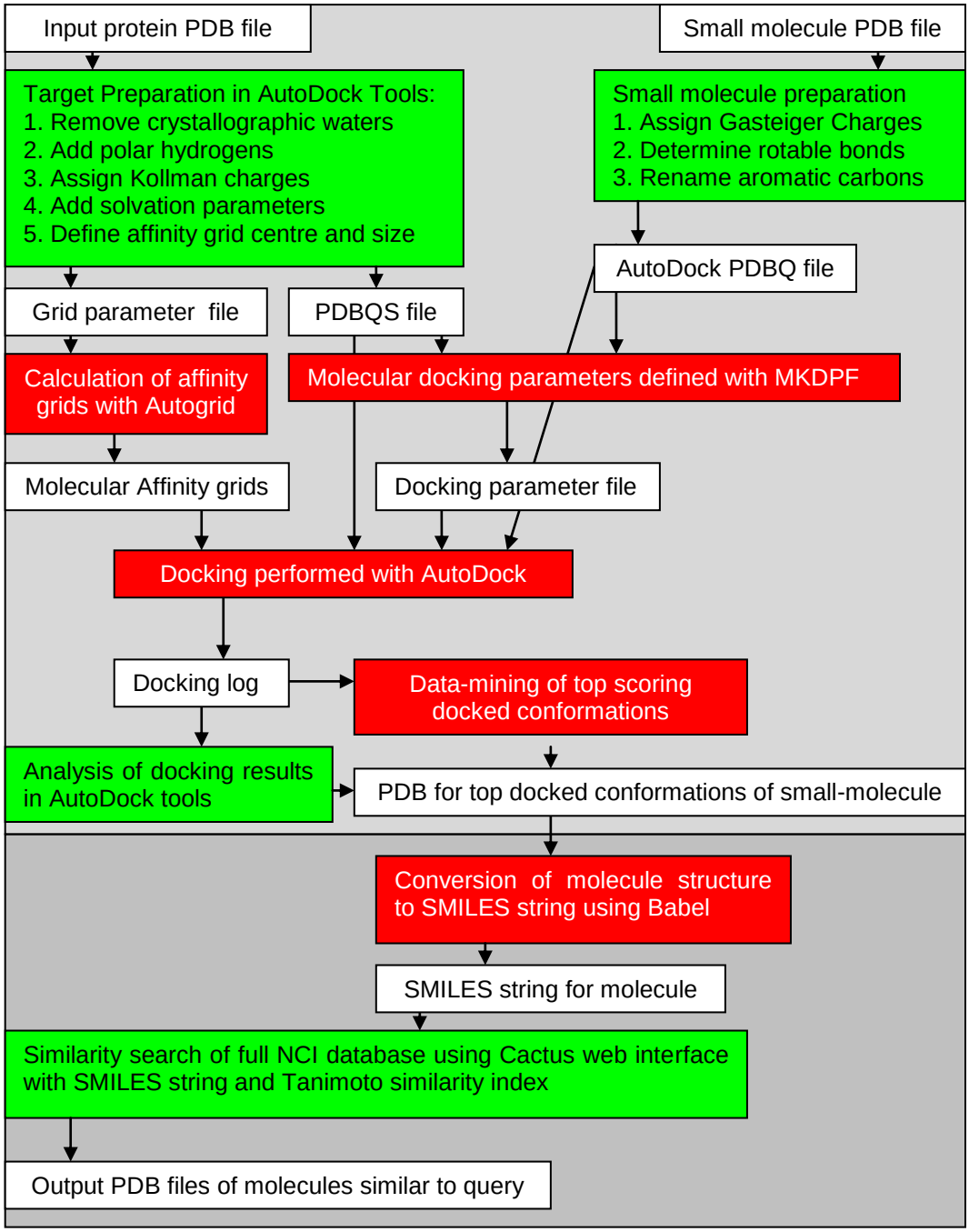


Figure 2.2.1 Conceptual workflow of AutoDock and NCI database search
The steps and programmes involved in the implementation of AutoDock for *in silico* screening and the use of output from this for searching the full NCI database are shown above. Input and output files are shaded in white, automated program driven steps in red and manual analysis in green. The steps involved in docking and then database searching are delineated by different shades of grey boxes.

2.3 Analysis of TNF ligands and their receptor binding sites

2.3.1 Suitability for small molecule binding

The extra-cellular portion of a TNF receptor consists of long and thin cysteine rich domains, which do not present an ideal candidate surface for small molecule binding because they possess no extended surface areas with cavities that would be suitable for such binding. The extra-cellular portion of TNF ligands, however, are globular in nature and represent a bigger target area for any small molecule design and binding considerations. In support of this strategy, existing TNF therapies aim to bind and sequester TNF ligands to prevent interaction with and the activation of TNF receptors [111,113].

2.3.2 How to hit a moving target

Proteins are, by their nature, flexible, with many proteins undergoing distinct domain and side-chain movements upon binding to their target [137]. Any analysis of a binding site must take some consideration of the fact that the crystallographic model is a single snapshot of what is a dynamic molecule [138], and in many instances the crystallographic model may not be the 'definitive' structure of a protein. Fortunately, for the analysis presented here, the extra-cellular domains of TNF superfamily ligands are quite rigid, exhibit no gross structural changes on receptor binding, and minimal side chain movements (the structure of TRAIL superposed onto the TRAIL/DR5 complex with an RMS deviation of 0.73 Å with almost all side-chain residues in the same conformation). For the purposes of this analysis, the TNF ligands are not considered to present a 'moving target'.

2.3.3 Receptor/ligand complexes

The structure of the Lymphotoxin alpha (LTα) complex with TNFR1 was the first structure of a TNF receptor/ligand complex to be solved, this was followed by the TRAIL/DR5 structure some years later; these receptor/ligand complexes stood as the paradigm for analysis of interactions between receptor and ligand in this superfamily [12,13,38,39]. Recently the structure of BAFF in complex with its receptors BCMA and BAFFR and APRIL in complex with BCMA and TACI have been published [28,30,139].

An analysis of the accessible surface areas of these protein complexes and the surface buried on ligand/receptor binding was carried out using Naccess [129]. The surfaces of TNF ligands are between 16,000 and 21,000 $Å^2$ in area with ~ 55 % of this calculated surface area being hydrophobic in nature (Table 2.3.1). The structures of both DR5 and TNFR1 consist of three CRD regions and have ~ 9,000 $Å^2$ of surface area, 55 % of which is hydrophobic. BAFFR, BCMA and TACI have only a third of the total surface area of DR5 and TNFR1, this is because these structures represent only a single CRD. They also show a slightly higher proportion of hydrophobic surface.

On receptor/ligand binding 4,144 $Å^2$ of the surface area of TRAIL and 3,142 $Å^2$ of LTα are buried (representing ~ 20 % of the total surface area of each), while 1,495 $Å^2$ of DR5 and 1,118 $Å^2$ of the TNFR1 surfaces are buried. Between 1,087 $Å^2$ and 1,903 $Å^2$ of the surface of BAFF is buried upon binding to its receptors, with the minimal BR3 loop having the lowest contact area. In the case of APRIL and its receptors around 2,200 $Å^2$ of the ligand surface is buried on receptor binding.

The proportion of surface buried that is hydrophobic is between 50-70% in most cases and surface complementarity of ~ 0.7 (as calculated by SC from the CCP4 suite) is similar for all receptor/ligand pairs, this is somewhat higher than for that seen for some T-cell receptors and their MHC-peptide ligands which have a surface complementarity of around 0.5 [140].

Interestingly, the minimal BR3 loop [28] binds to BAFF with almost the same affinity as the full length BAFFR, yet has only half the contact area of a normal CRD. Analysis of this minimal loop shows that the key contacts between BAFF and BR3 are hydrophobic interactions between V33 and L38 of BR3 packing against Y206 of BAFF. The experimental evidence points to the fact that the majority of the affinity and stabilization of binding between BR3 and BAFF are found in this beta-hairpin region of the protein. The BAFFR represents an outlier in the TNF receptor superfamily, with its minimal CRD-loop that binds as strongly to its ligand as other TNF receptors. This is in stark contrast to the binding surfaces of TNFR1 and DR5 are much more diffuse in their nature and rely on a large number of contacts spread out over a large surface area for their affinity.

Protein Complex	Total surface area (Å 2) (hydrophobic (%))		Accessible surface area (Å 2) when in complex (hydrophobic (%))		Buried surface area (Å2) When in complex (hydrophobic (%))		Surface Complimentarity
	Ligand	Receptor	Ligand	Receptor	Ligand	Receptor	
TRAIL/DR5	20,898 (50)	8,899 (56)	16,754 (52)	7,407 (55)	4,144 (44)	1,492 (60)	0.63
TNF/TNFR1	17,774 (59)	9,701 (55)	14,632 (59)	8,583 (54)	3,142 (59)	1,118 (62)	0.64
BAFF-BAFFR loop	17,590 (56)	1,459 (69)	16,503 (56)	963 (65)	1,087 (50)	496 (76)	0.66
BAFF-BAFFR	17,629 (58)	3,184 (68)	15,726 (58)	2,395 (72)	1,903 (54)	825 (56)	0.68
BAFF-BCMA	17,536 (59)	3,261 (58)	15,747 (59)	2,496 (57)	1,792 (59)	765 (62)	0.65
APRIL-BCMA	16,757 (53)	2,955 (57)	14,586 (52)	2,181 (50)	2,171 (60)	774 (74)	0.70
APRIL-TACI	17,060 (53)	3,454 (55)	14,783 (52)	2,362 (54)	2,277 (56)	1,092 (56)	0.76

Table 2.3.1 Solvent accessible surface area and contact areas of TNF receptor/ligand complexes
Accessible surface area calculated using Naccess, surface complementarity calculated using SC from the ccp4 suite. Surface areas are shown as square angstroms, and are calculated with a 1.4 Å probe. Accessible and buried surface areas are calculated for trimeric ligand binding to three receptors and for a single receptor. Hydrophobic area buried is shown in parentheses as the percentage of the calculated area that they represent. Surface complementarity scores are described as a score between one and zero, with a score of one being a perfect fit and zero representing no surface complementarity.

2.3.4 Receptor binding surface analysis of TNF ligands

TRAIL, LTα and TNFα

The receptor-binding surface of TRAIL shows significant cavities, with a large contiguous cavity stretching most of the length of the TRAIL monomer interface. This corresponds in part with the two major sites of DR5 binding. Surface charge distribution in this cavity is mainly positive with some hydrophobic patches. In contrast to this, LTα and TNFα have no large cavities in their receptor binding sites and show a more even charge distribution. TNF receptor discrimination between LTα and TNFα may rest on obvious charge differences in the cell-proximal receptor-binding site (Figure 2.3.1a).

BAFF and APRIL

BAFF and APRIL are somewhat more compact than LTα, TNFα and TRAIL. Each has only a single receptor-binding patch on their surfaces. The two proteins show divergent surface charge distribution, yet share common receptors in BCMA and TACI. Discrimination in BAFFR has been suggested to be due to steric clashes caused by different side-chain conformations in the receptor binding site between BAFF and APRIL [30]. Both proteins show a deep surface cavity that corresponds to the receptor-binding site and continues down to the D-E loop on both (Figure 2.3.1a).

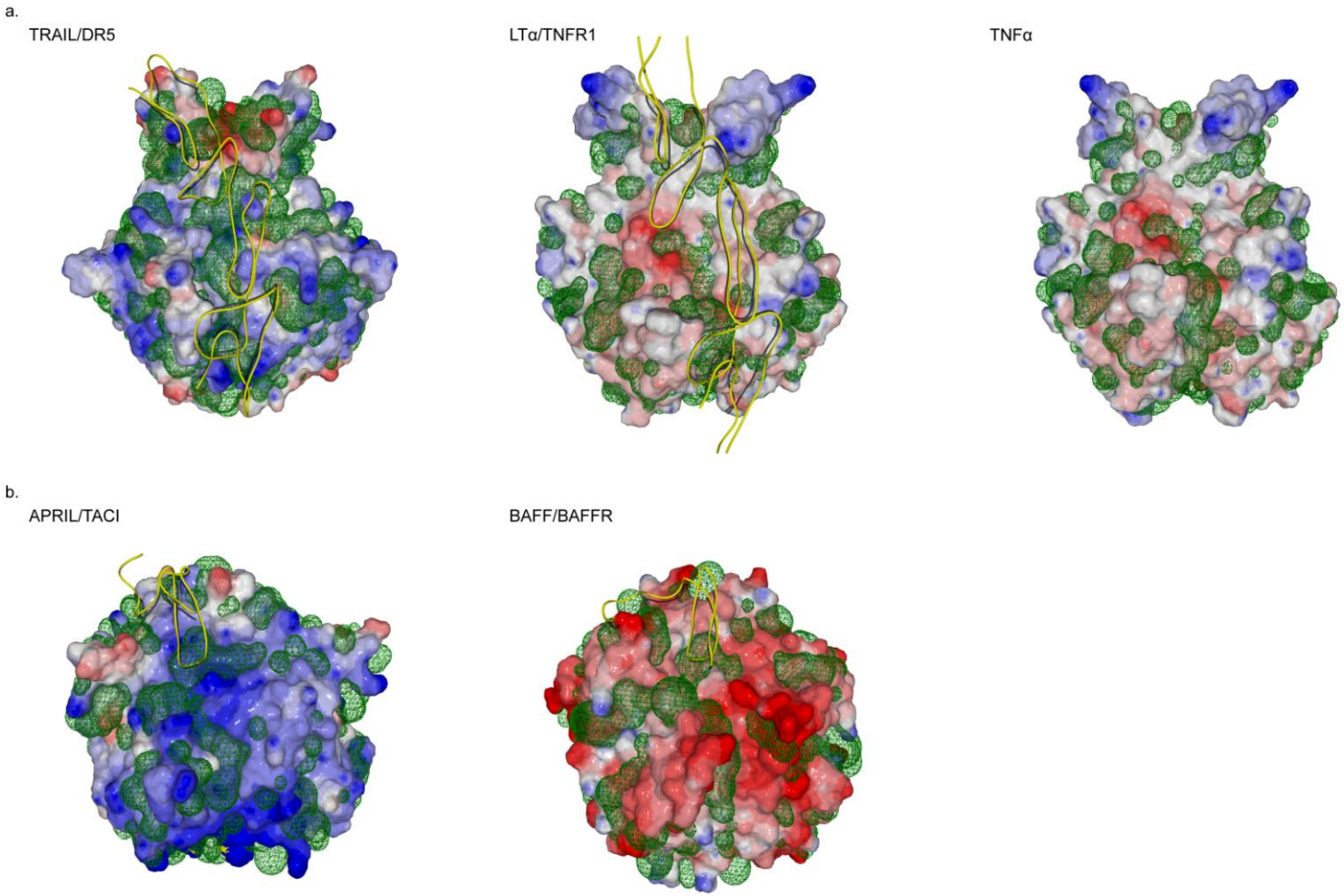


Figure 2.3.1a Electrostatic interaction surface of TNF Ligands
The electrostatic surface of each protein is coloured according to the calculated electrostatic potential with +5 kT in blue and -5 kT in red. Surface cavities calculated using Surfnet are shown in green wire frame and receptors, where present, are in yellow cartoon representation. Putative receptor binding sites are shown as yellow dashed boxes. Proteins are shown with the membrane proximal portion of the TNF ligands at the bottom. (Figure continues overleaf)

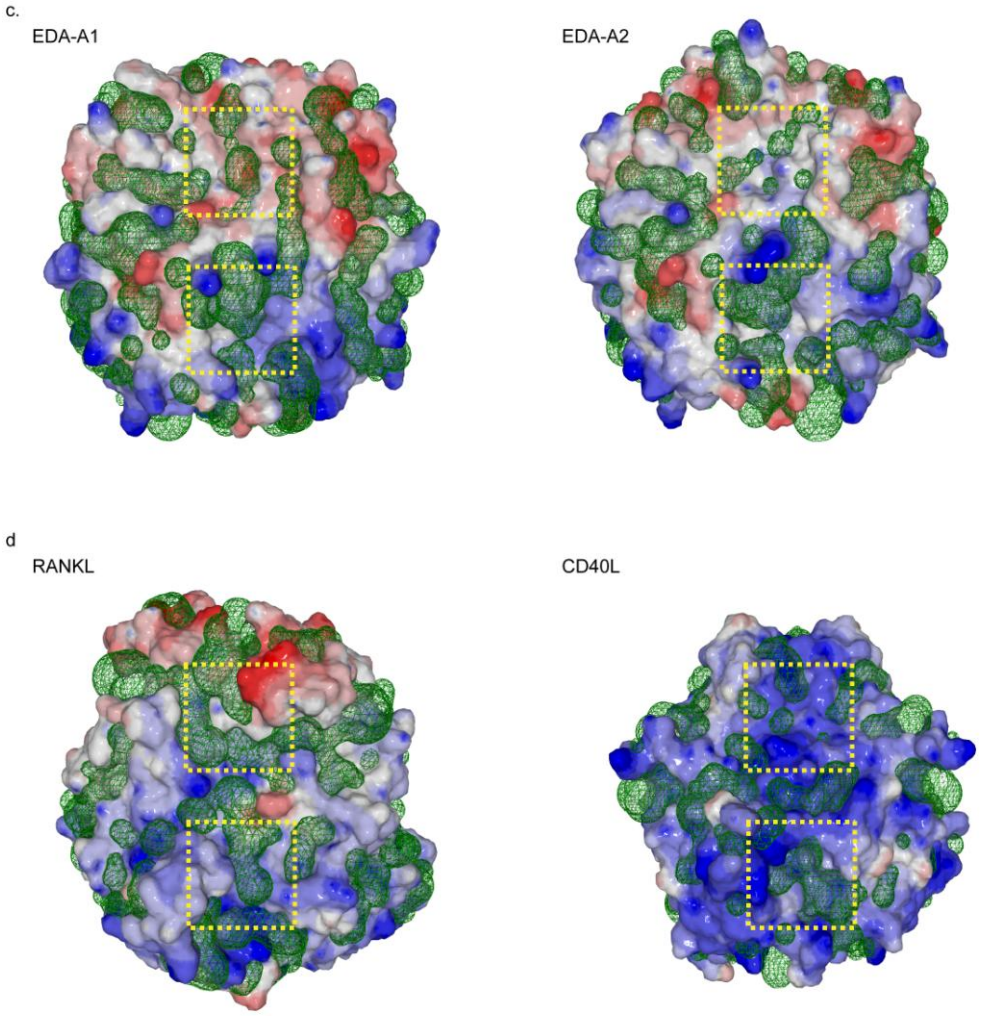


Figure 2.3.1b Electrostatic interaction surface of TNF Ligands
The electrostatic surface of each protein is coloured according to the calculated electrostatic potential with +5 kT in blue and -5 kT in red. Surface cavities calculated using Surfnet are shown in green wire frame and receptors, where present, are in yellow cartoon representation. Putative receptor binding sites are shown as yellow dashed boxes. Proteins are shown with the membrane proximal portion of the TNF ligands at the bottom.

EDA-A1 and EDA-A2

The two splice variants of the TNF-like molecule Ectodysplasin A (EDA) differ by a two residue motif (Glu 308 and Val 309) present in EDA-A1, but not EDA-A2, which affects the discrimination between the receptors EDAR and XEDAR [141]. The EDA-A2 surface is flatter and less charged than in EDA-A1, with the main difference being caused by the presence of the charged glutamic acid in the cell-surface distal receptor-binding region of EDA-A1 (Figure 2.3.1b).

RANKL and CD40L

RANKL is an important mediator of bone formation and loss, while CD40L is important in the humoral immune response. Both present ideal targets for small-molecule inhibition, although no receptor/ligand complex structures have yet been solved. The proposed receptor-binding site of RANKL has two large contiguous cavities that correspond closely to the CRD binding regions seen in TRAIL and LTα. The top site has mixed hydrophobic and negatively charged character, while the bottom site is more positively charged. CD40 ligand shows a large surface cleft in the membrane proximal receptor-binding site and has a predominantly basic surface charge pattern (Figure 2.3.1b).

2.3.5 Determinants of receptor binding

The analysis of the receptor/ligand complexes and receptor binding surfaces of the TNF ligands shows that the common scaffold of the TNF protein families can accommodate a large range of different interactions. Stark differences in surface charge patterns and shape allow discrimination between binding partners; yet BAFF and APRIL share receptors in spite of these differences. Surface complementarity

between ligands and receptors is relatively high, this finding corresponds with the high binding constants that have been reported for receptor interactions [30]. Large surface clefts within the receptor-binding site of ligands show areas where small molecules would be easily accommodated to allow high binding energies through surface complementarity.

2.4 Design of small molecule ligands for TNF family members

2.4.1 3D Pharmacophores and Database mining

The initial stage in small-molecule ligand design is modelling the active site of a protein to create a pharmacophore model. The creation of a pharmacophore model traditionally involved taking a great deal of Structure/Activity Relationships (SAR) from binding studies of proteins and their ligands and using chemical knowledge to construct a query. A number of automated procedures have been developed to determine SARs [142,143]. One of the primary uses for automated SAR programmes is in determining average binding energies for active interactions between multiple ligands and the target and using this averaged ensemble as a model for database searching.

2.4.A Pharmacophores

A pharmacophore is defined as the group of atoms within a drug molecule that are responsible for their pharmacological action. To model the interaction between protein and drugs a pharmacophore model can be generated from biochemical and structural data [144,145]. These are useful for analysing the efficacy and binding scheme of families of drugs within the active site of a protein and in the optimisation of lead compounds.

The active site of a protein can be modelled in terms of a pharmacophore for which binding partners, be they protein or small molecule, are seen to be active. This model, once created, can then be used as the conceptual framework around which the design of other molecules that bind to the target can be accomplished.

With a suitable pharmacophore model of a protein and potential docked molecular fragments from a receptor or a ligand one can search a similarly arranged database of small molecules and potential drug targets. This is a computationally intensive task which customarily involves the use of proprietary tools that are expensive to license or are closely guarded by their authors. The Accelrys Catalyst [146] suite of programmes allows automatically generated 3D pharmacophore models to be screened *in silico* against libraries of small molecules for potential drug leads. A number of other 3D database searching programmes exist, including LIDAEUS [125] and the father of such programmes CONCORD. These programs utilise 2D representations of interactions between the constituent atoms of small molecules within the database and the fitting of a calculated 2D profile from the active site of the protein to search within their databases.

2.4.2 Docking

One of the most widely used methods of screening for potential drug leads is to dock a library of compounds against the active site of the protein of interest *in silico*. The results of this docking are then used to inform *in vitro* binding studies, such that one can narrow a large library of molecules down to a few candidates that are likely to have some interaction *in vitro* [147]. The original docking programme, still widely used, is DOCK [148]. This uses an algorithm, which represents surface clefts on a molecule as spheres, which are then culled to include only those of interest (i.e. those within the potential ligand-binding site). These sphere centres are then matched to flexible ligands, and a shape, or a simple energy algorithm, is applied to score the binding of the molecule within the active site of the protein.

AutoDock [135] uses calculated surface energy grids and an empirical scoring function to dock flexible ligands to a protein using a modified Lamarckian Genetic Algorithm (LGA). AutoDock works such that populations of ligands are varied in position and conformation (if flexible) against the active site of the protein and then scored according to an empirically derived scoring function. If the new conformation is energetically more favourable than the old one it is automatically accepted, if not it is accepted probabilistically. This implementation allows a wide search of possible conformations by seeding a population of ligands in random orientations and conformations and then selecting those elite members, which go on to populate the next generation.

The scoring functions within automated docking programs are major points of contention when considering the usefulness of information from docking. The DOCK scoring function uses simplified Van der Waals and electrostatic terms to score the docked conformations, while AutoDock uses a more complex empirical scoring method based on the paramaterisation of data from ligand receptor complexes, considerations of the desolvation of protein on ligand binding, and torsional stabilisation of the ligand. The simplified algorithm of DOCK may miss conformations of ligands that might be useful and it may in fact have limited utility outside of a simple comparison of docked molecules. In this scoring algorithm there is potential for bias towards Van der Waals energy minimization, which in other cases have proven to be unproductive in selecting for molecules that will bind to the target *in vitro* [149].

An empirically derived scoring function is conceptually more attractive than an arbitrary one, although an obvious flaw is that this may be subject to bias from the test set used to determine the energy terms and could in theory, force certain 'favoured' conformations of ligands within their target site. The scoring functions from different programs may be incompatible to such an extent that molecules with a high simulated binding energy from one programme may have a low simulated binding energy when docked using a different scoring function. It is also difficult to account for the entropic portion of any binding stabilisation in a scoring function. An appreciation of these caveats when interpreting any data from docking can prevent overly grand conclusions being drawn about the real world applications of such data.

2.4.3 Pharmacophore modelling for TNF ligands

TNF-ligands present much better targets for any structure-based ligand design than TNF-receptors because of their larger size and more globular structure. If either could be said to have an 'active site' then it is the ligands, as discussed in Chapter 2.3.1.

As discussed in Chapter 4.1, the pharmacophore model is central to any database search or any attempts at structure-based ligand design. From the analysis of receptor ligand complexes, a number of proteins present themselves as potential targets for small molecule inhibition. TRAIL and BAFF stand out as ideal models to take further and model their ligand-receptor interface: TRAIL has mixed surface charge and relatively large cleft regions with large contiguous receptor binding sites, and BAFF is unique in having a minimal receptor that binds almost as strongly as a full-length receptor. That these structures exist, with the receptor and the ligand, makes the

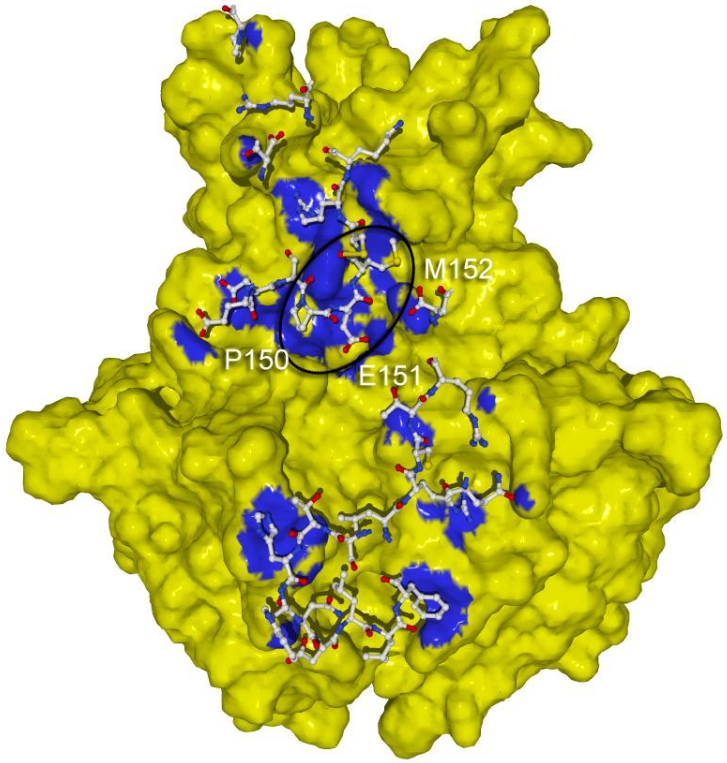
dissection of key binding interactions much easier than if the structures were of isolated ligands.

To model a ligand/receptor pharmacophore, the pdb files for complexes of TRAIL/DR5 and BAFF/BAFFR were cut down to contain just a single receptor chain and the trimeric ligand. Receptor amino acids within 4.5 Å of the ligand were considered to be contact residues, this correlates with the calculated contact surface for both receptor ligand complexes (Figure 2.4.1) and previously reported data for both BAFF and TRAIL [13,29].

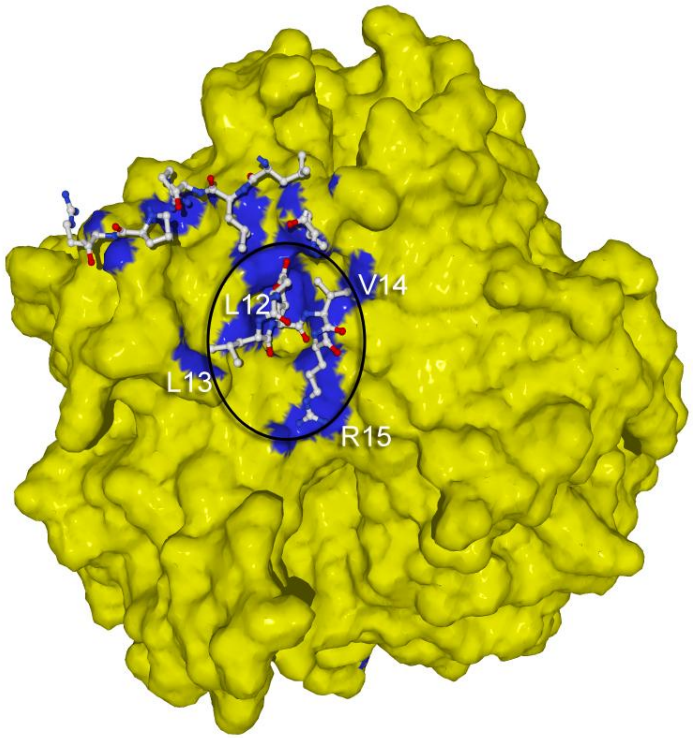
Contact residues for DR5 are split roughly equally between CRD1 and CRD2, with the CRD2 contacts residing in a deeper cleft than those of CRD1. The P150-E151-M152 motif in CRD2 represents a good mix of electrostatic and hydrophobic contacts within a pocket, for use as a potential small molecule pharmacophore.

The BAFFR contacts with BAFF occur within the relatively deep pocket at the top of the BAFF molecule. Although BAFFR possesses only a single CRD, the binding affinity between it and BAFF is comparable to TRAIL/DR5. This is most likely due to the larger contact surface offered by the deeper surface pocket in BAFF. The L27-L28-V29-R30 motif of the minimal BAFFR loop represents a similarly placed potential small linear pharmacophore to that chosen for TRAIL/DR5.

TRAIL/DR5



BAFF/BAFF-R



DR5 69 PQQKRSSPSEGLCPPGHHISEDGRDCISCKYG**Q**DY**STHWNDLLF**C**LRCTR**C**D**SGEVEL
SPCTTTRNTVCQCEEGTFRE**EDSPEMCRK**C**R**TGC**P**RGMVKVG**D**CTPWSDIECVHKESGD

BAFF-R 16 PTPCVPAECF**DLLVR**HC**V**ACG**LLR**T**PR**PKPA

Figure 2.4.1 Receptor ligand contacts for TRAIL/DR5 and BAFF/BAFFR
The contact surfaces between the receptor ligand pairs for the structures of TRAIL/DR5 (1D4V and BAFF/BAFFR (1OQE) were calculated using SPDBV and are shown as blue areas on the yellow protein surfaces. Receptor residues within 4.5 Å of their corresponding ligand are shown in ball and stick representation. The extent of the receptor pharmacophore model is shown with an ellipse and key residues are labelled. Within the receptor sequences, residues involved in ligand contact are highlighted in red.

2.4.4 Small molecule design using Ligbuilder

2.4.B The Ligbuilder methodology

The Ligbuilder programme developed by Wang *et al* [127] is a multi-step approach to structure based drug design. Initially the programme calculates a pharmacophore model from a protein and a docked molecule, taking as its input a protein structure and ligand file and giving an output of a pharmacophore model of the interaction between the two and a selection of important hydrophobic and electrostatic interactions in a 'key site' file.

This pharmacophore model is then use as input to the main programme along with a seed structure derived from the receptor to build a molecule within the protein binding site using a genetic algorithm to evolve the molecules from a pool of molecular fragments. These molecular fragments are functional groups, such as carboxylic acids, alcohols and sulphur containing groups, along with structural elements like benzene rings and carbon chains. Molecules it creates are then scored with consideration of their chemical viability based on Lipinski's rules [150] and their binding affinity for the protein using an empirical scoring function.

Two strategies of molecule creation are available within the program: linking and growing. With the growing strategy, a pre-docked seed is placed within the binding pocket and Ligbuilder adds fragments to specified special hydrogen positions on this to build a molecule. The primary benefit of this strategy is in the case when a lead compound is already available and derivatives are sought, the grow strategy will allow semi-automated lead-optimization. The linking strategy takes multiple seed

fragments and attempts to integrate these within a single molecular framework. The obvious use of this is when initial lead compounds are sought.

2.4.C The Lipinski rules

The Lipinski rules are a set of factors determined from a survey of orally administered drugs that affect the bioavailability and activity of small molecule drugs with the following characteristics of a molecule being detrimental to its activity:

1. Molecular weight of over 500 Da
2. LogP (the partition coefficient between octanol and water) over 5
3. There are more than 5 hydrogen bond donors (the sum of OH's and NH's)
4. There are more than 10 hydrogen bond acceptors (the sum of O's and N's)

Using the Ligbuilder pharmacophore modelling programme, pharmacophore and key-site models for both BAFF and TRAIL were calculated using the whole ligand and receptor structures as the input. Initially, the LINK module was run with two functional groups from each pharmacophore. For BAFF the guanidine group from R15 and the isobutane group of L13 from BAFFR were used as the seed for the link programme. For TRAIL, the carboxylic acid moiety of E151 and the dimethyl sulphide group of M152 from DR5 were used. Both runs of the program were carried out using the standard parameters described in Chapter 2.2.2.

The BAFFR seed produced only two molecules that fulfil the selection criteria for potential leads. These are shown in Figure 2.4.2a and a representative molecule is shown in Table 2.4.1. The seed structure derived from DR5 produced a large number

of potential molecules, which were built around a common scaffold and were culled to twenty molecules (Figure 2.4.3, Table 2.4.2 as above).

It is clear that the input seed-structure orientation is very important in generating a diverse pool of potential ligands. Because of the limited number of molecules generated for BAFF and the single scaffold that was generated for the TRAIL ligands, an attempt was made to generate a more diverse pool of molecules using the GROW module.

Taking the glutamic acid carboxylic acid group from DR5 and the BAFFR guanidine as input seeds for the GROW module, a larger range of molecules is produced (Figure 2.4.2b, Table 2.4.1). Again, these molecules appear to share a core common to each set of ligands, with the major of diversity associated with extended, flexible, portions of the molecules. The restraints placed upon the growth of molecules by the single seed structure in the GROW module allow for a wider variety of these 'conformationally' flexible regions, than those produced by the strict constraints of LINK. In the case of those molecules generated against TRAIL, the binding orientation is perpendicular to that of the receptor. These molecules may not efficient competitors for ligand binding with the receptor because of this unexpected orientation.

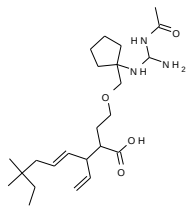
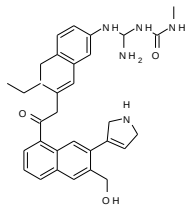
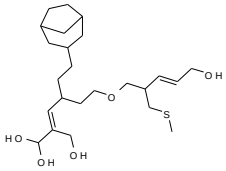
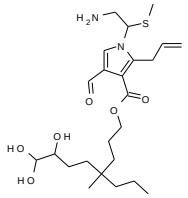
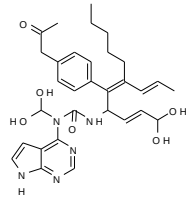
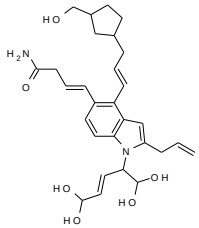
Chemical Structure	Ligbuilder run	MW (Da)	Rotable Bonds	Binding Score (Kcal/mol)
BAFF Link				
	1	437	13	-0.09
BAFF Grow				
	1	555	11	-11.61
TRAIL Link				
	1	442	17	-10.11
	13	484	20	-3.91
TRAIL Grow				
	3	577	13	-14.48
	18	510	14	-8.00

Table 2.4.1 Representative molecules generated by Ligbuilder against BAFF and TRAIL using the growing and linking strategies of the programme.
The binding score for each molecule is calculated using the AutoDock scoring function to ensure consistency throughout the analysis.

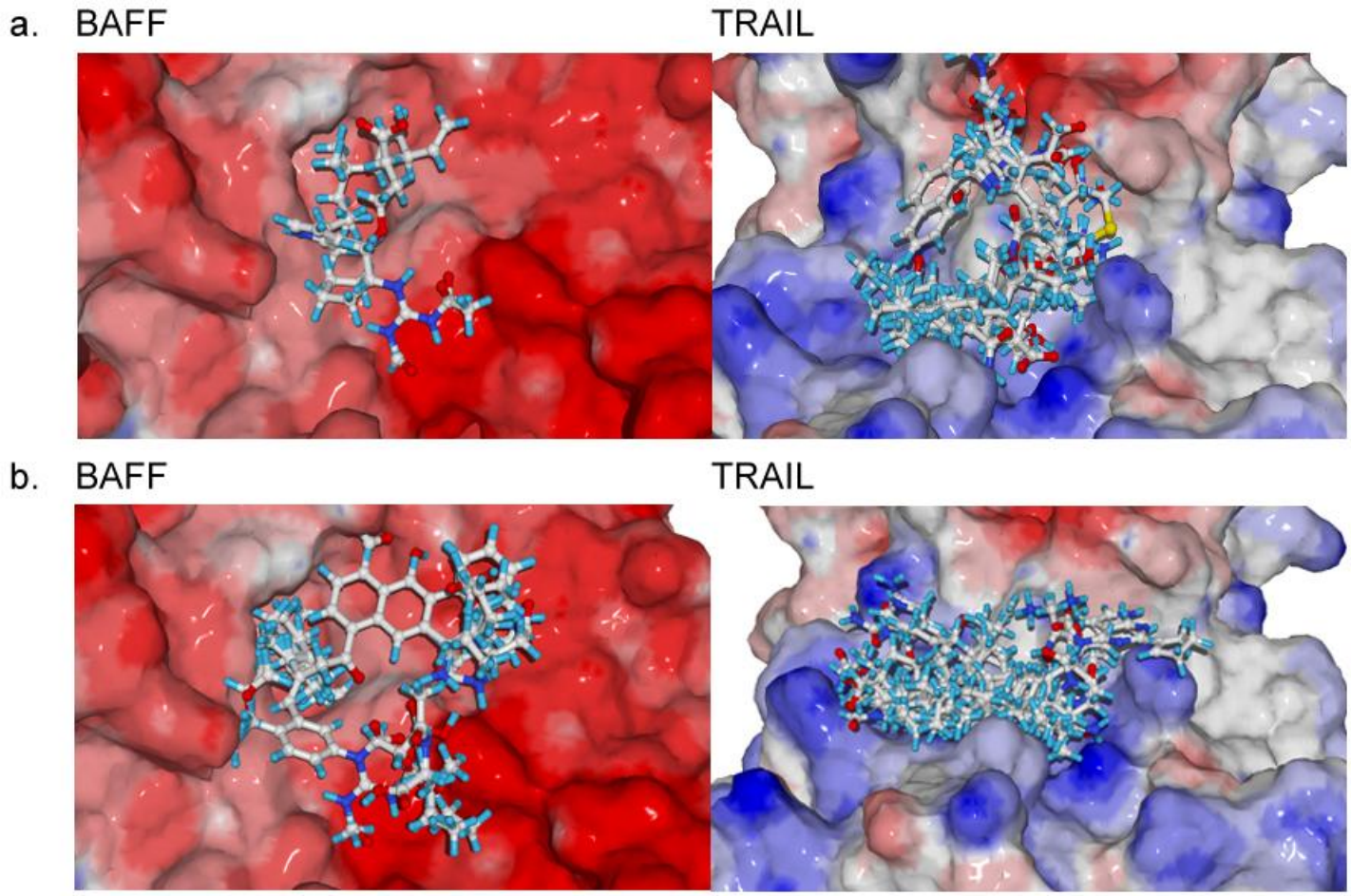


Figure 2.4.2 Small molecule products of Ligbuilder
Molecules generated by the two Ligbuilder modules are shown docked within their protein targets. Protein surfaces are coloured by charge and were calculated using SPDBV, small molecules are shown in ball and stick representation.
a. Molecules generated using the Ligbuilder LINK module, the seed structures are visibly preserved in the product molecules.
b. Those molecules generated with GROW are more structurally diverse than those created with the LINK module and for those against TRAIL show a very different binding orientation on the surface of the molecule.

These trials with two simple seed structures show that the positioning of the seed atoms, and geometrical constraints within the pharmacophore, are key determinants for the generation of diverse populations of small molecules by Ligbuilder. The flexibility that the program has in building molecules is entirely reliant on well-placed initial seeds within the pharmacophore and user defined growth points on the seed.

An awareness of the active site of the protein is critical when choosing seed structures for the program. The small seed used for both ligands for the GROW module produces a diverse, yet potentially ineffective, range of molecules. This is because they are not as constrained within the actual receptor-binding site, as with the greater number of input constraints of the LINK module.

The diversity of molecules produced in the above cases imply that the most suitable usage for this suite of programmes would be in the development of lead compounds, where a previously identified 'scaffold' molecule had been previously docked within the active site. With this scenario, the addition of chemical groups to the lead compound could enhance its activity of binding affinity for the protein and be useful in informing rational design.

2.4.5 Database search with generated ligands

The utility of 'designed' small molecule ligands is somewhat mitigated if said molecules are not synthetically viable. One way that this is taken into consideration within molecular design programmes is the adoption of chemical viability rules of various types. The Lipinski rules have been adopted as the yardstick by which small-molecule drugs are measured. In the case described in 2.4.4, where molecules are generated *a priori* from a suggested protein-ligand pharmacophore, the potential availability of similar molecules within small-molecule databases is of great interest. A browser exists for the NCI cancer drug database [126], which allows open queries to be run using SMILES strings to look for molecules with similarity to the search model.

2.4.D SMILES

Simplified Molecular Input Line Entry System (SMILES) [151] strings are chemical descriptors that describe a valence model of single small molecules. The language is non-platform dependent and uses basic characters to represent molecules with branching represented by enclosing atoms within parentheses and cycles denoted by numerals linking atoms that close a cycle. Thus, ethane would be described by 'CC', isobutyric acid is represented as 'CC(C)C(=O)O' and benzene is shown as 'c1ccccc1' (the lower case c denoting aromatic carbons and 1 the cyclic link). Using these simple one-dimensional descriptors, information exchange and database searching is easily facilitated.

The molecules produced by Ligbuilder from TRAIL and BAFF were converted to SMILES strings using openbabel [152] and used to search the NCI database using the

cactus browser [136]. Molecules with a similarity score of above 80%, as determined using the Tanimoto scale [133], were downloaded from the database and docked within the receptor binding site of BAFF, or TRAIL, using the AutoDock protocol at described in Chapter 2.2.4.

Unfortunately, no molecule was found to match perfectly any of those that were generated by Ligbuilder. Twenty-one molecules were identified with high similarity to those generated against the TRAIL receptor-binding site. They exhibited simulated binding energies of between -9.85 and -7.62 Kcal/mol (Table 2.4.2). Ten molecules were identified as having high similarity to those generated against the BAFF receptor-binding site and they exhibited binding energies between -9.95 and -5.36 kcal/mol (Table 2.4.3). These binding scores are not encouragingly high when compared to those previously reported for applications of the use of AutoDock in lead discovery [124,149,153]. Because of this apparent failure of the Ligbuilder program to identify any promising lead compounds in a small-molecule database, the NCI diversity set of small molecules was used to compare the technique of selective high-throughput database screening with the use of *a priori* generated ligands in lead discovery.

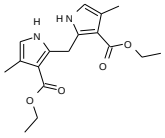
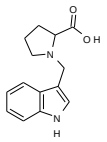
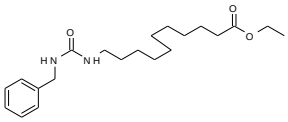
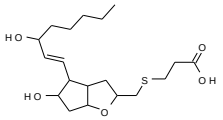
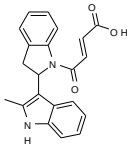
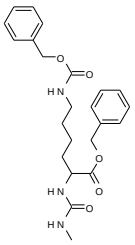
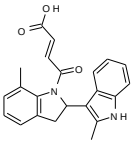
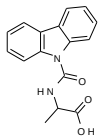
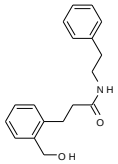
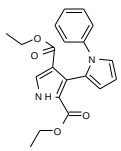
Chemical Structure	NSC Number	MW (Da)	Rotable Bonds	Binding Score (Kcal/Mol)
	22704	318	8	-8.21
	99785	244	3	-7.62
	231412	362	16	-8.85
	603587	372	11	-9.37
	631311	346	3	-9.07
	634764	427	14	-8.74
	636507	360	3	-9.85
	649031	282	3	-8.32
	686928	283	7	-8.3
	693227	352	8	-8.6

Table 2.4.2 Molecules identified by NCI similarity searching using molecules generated against the TRAIL receptor-binding site by LigBuilder.

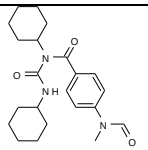
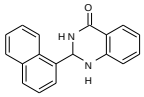
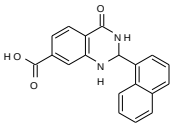
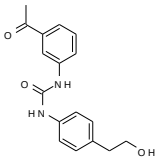
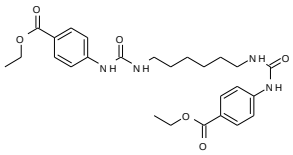
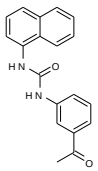
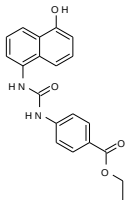
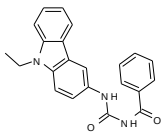
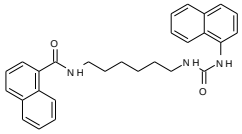
Chemical Structure	NSC Number	MW (Da)	Rotable Bonds	Binding Score (Kcal/Mol)
	114921	385	4	-7.29
	145669	274	1	-7.26
	170414	318	2	-7.38
	203973	298	6	-6.88
	205577	498	17	-8.12
	205727	304	4	-6.4
	215649	350	6	-7.48
	340543	357	4	-5.36
	682235	439	10	-9.95

Table 2.4.3 Molecules identified by NCI similarity searching using molecules generated against the BAFF receptor-binding site by LigBuilder.

2.4.6 In Silico screening with the NCI diversity set

The NCI diversity set of small molecule drugs, a library of some 1992 compounds available from the National Cancer Institute (NCI) [128], have been selected from the 140,000 compounds on file at the NCI to represent a statistically broad cross-section of the pharmacophores that these molecules contain. This library has been made available in a format that is suitable for use within AutoDock and represents one of the few freely accessible small molecule databases.

Using scripts provided as part of the AutoDock package, each member of the NCI-diversity set was docked to the receptor binding regions of both BAFF and TRAIL. This task is computationally intensive, taking roughly 4 minutes per small molecule on a 768 Mb, 1 GHz Silicon Graphics machine. At this rate, the whole set is docked in around five days on a single machine. With parallelisation and more computer power this can be exponentially increased (as it is in drug companies). For the small-scale tests with a few targets, one can set the entire library to dock automatically, and the task can be set up and left running.

Analysis of the data produced however, is somewhat more difficult, with some 20,000 docked conformations produced per protein target for the set. Using simple shell scripts, the docking energies for each ligand can easily be mined from the docking results files and imported into a spreadsheet package. Molecules that exhibit high simulated binding energies against the protein of interest are considered to be potential targets for *in vitro* screening and the docking orientations were analysed further.

An analysis of those docked conformations in the top percentile of binding scores is more selective and therefore computationally less intensive, than manually screening each docking conformation (20,000 per protein target), many of which have low, or even positive binding energies. For both TRAIL and BAFF the top scoring molecules found to bind within the receptor-binding site are shown in Figure 2.4.3 and Table 2.4.4/5

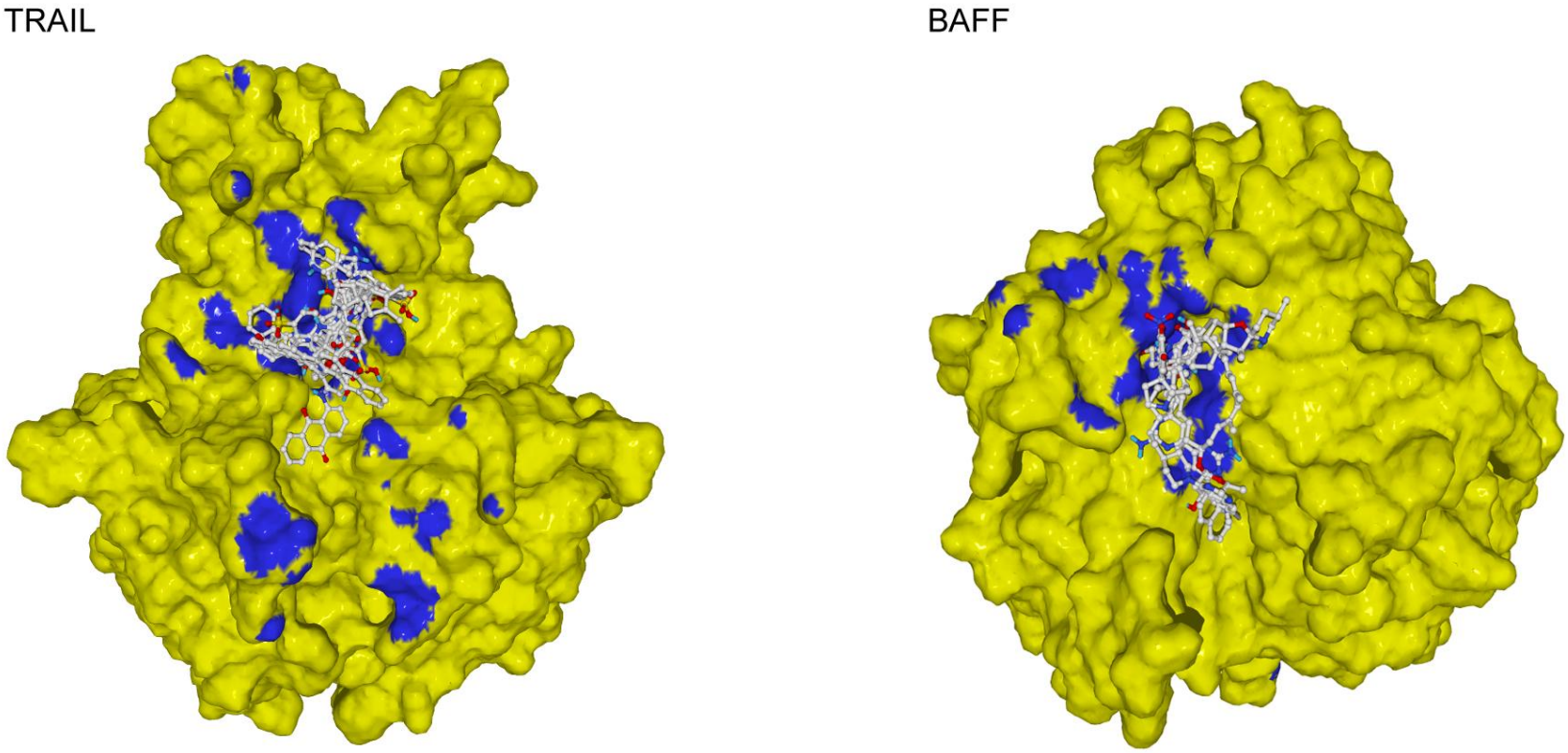


Figure 2.4.3 NCI molecules binding within the TRAIL and BAFF receptor-binding grooves from AutoDock screening.
The top scoring molecules from AutoDock screening are shown in their final docked conformation within the receptor binding grooves of both TRAIL and BAFF. The ligands are shown in surface representation with blue representing the footprint of the cognate receptor for each ligand. Small molecules are shown in ball and stick representation.

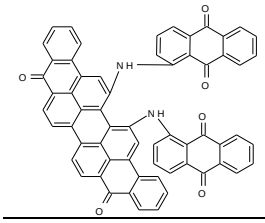
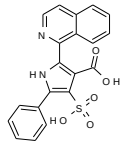
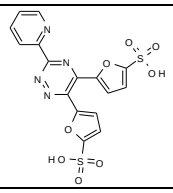
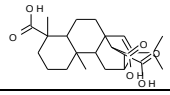
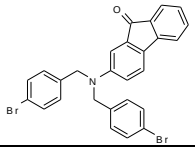
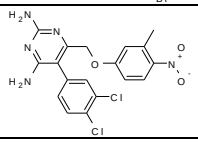
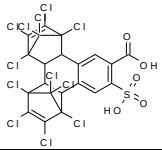
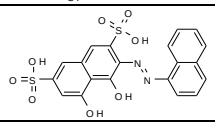
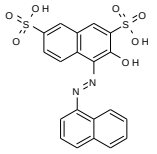
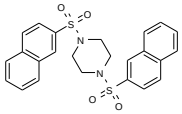
Chemical Structure	NSC Number	MW (Da)	Rotable Bonds	Binding Score (Kcal/Mol)
	23128	899	4	-11.94
	116654	394	4	-11.6
	601364	450	5	-11.31
	15520	419	4	-11.22
	18451	533	5	-11.12
	126710	420	4	-11.06
	270718	798	2	-10.98
	45208	474	4	-10.94
	45583	458	4	-10.73
	37245	467	4	-10.72

Table 2.4.4 Potential TRAIL receptor antagonists from the NCI diversity set identified by virtual screening

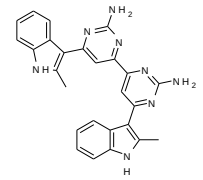
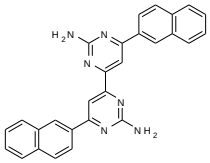
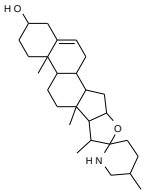
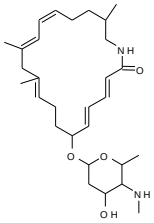
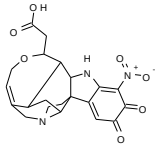
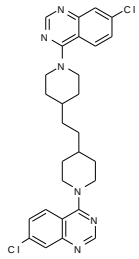
Chemical Structure	NSC Number	MW (Da)	Rotable Bonds	Binding Score (Kcal/Mol)
	371884	446	5	-11.38
	371878	440	4	-11.17
	179187	413	1	-10.91
	641691	486	4	-10.72
	5069	427	4	-10.71
	343227	520	4	-10.55

Table 2.4.5 Potential BAFF receptor antagonists from the NCI diversity set identified by virtual screening

Those small molecules with a high affinity for TRAIL all bind within the DR5 CRD2 binding site, which exhibits the deepest surface pocket region and the highest contiguous area of ligand receptor contact. For those ligands that are found to have high binding energies against BAFF the primary site of binding is the deep pocket within which major contacts between ligand and receptor are found. A number of high-scoring molecules do not bind within this pocket, binding instead to the right of this in the extended cleft on the protein surface. These were disregarded because that they did not bind within the footprint of the receptor, and would thus not be useful as drug targets against ligand/receptor binding.

Because the NCI Diversity set represents only a small range of the potential molecules within the NCI database, related compounds are present in the full database and these molecules may have a higher binding affinity for the protein of interest. Thus, those molecules from the NCI Diversity set that exhibited a high binding affinity for the proteins of interest were used to search to full NCI database and the resulting hits were screened in the manner described above.

For the those molecules that scored high binding affinities for TRAIL none of the molecules that were found to have similarity to the top ten diversity set molecules had a binding affinity for the protein that was above that found for the top 100 from the diversity set. A similarity search using the molecules that bound with high affinity to BAFF yielded three molecules that bound with a similar affinity to those used as the search models, with the top scoring molecule NCI640469 binding within the footprint of the receptor-binding site. This molecule is the drug Dehydrosolasodine, which has

been marked within the NCI database as a potential anti-inflammatory and anti-tumorigenic molecule.

Those small molecules that bind in positions that overlap the receptor-binding site are the most likely potential inhibitors for receptor/ligand binding. One problem with the TNF ligand 'active site' is that it is somewhat diffuse and flat in nature (it does not conform to the traditional stereotype of a small molecule ligand-binding surface) and any results from *in silico* docking studies might be expected to be very different to *in vitro* binding studies. The manner that the grid within which molecules are docked is chosen, can force molecules into certain conformations, although this would produce unfavourable *in silico* binding energies. Consideration should also be made that these molecules were screened against a single protein *in silico*, where they would encounter many thousands of different proteins, including cytochromes, which are responsible for the breakdown of the majority of drug targets within the clinical setting of a live human.

The molecules presented here represent small molecules that have potential for being competitive inhibitors of receptor/ligand binding. Ideally, these molecules would be tested for activity in some form of *in vitro* competition assay against their targets, but this was beyond the scope of this project.

2.5 Conclusions

2.5.1 Pockets, Avidity and active sites

One can see from the structures of the TNF ligands that the idea of an active site for directed small-molecule binding presents some difficulty. The natural binding state for TNF-receptors is that of an avidity mechanism, where multiple contact sites, multiplied by multiple chains coming into contact with the ligand, create the necessary binding stabilisation energy to facilitate downstream activation of signalling pathways. The solution of the structure of BAFF with its receptors was of great interest as the single CRD receptors bind within what is a well-formed pocket on the surface of the molecule and have a relatively continuous binding surface with the ligand.

The design of a soluble LFA-1 antagonist using the ICAM-1 immunoregulatory element as the starting point for small molecule design [154] exemplifies a rather elegant solution to a similar problem of ligand design. In this experiment, the key residues involved in the ICAM-1/LFA-1 interaction were used to generate a cyclic peptide, which was then grafted onto a non-peptide template to overcome the inherent problems of using a cyclic peptide. These small molecules were then combinatorially optimised in the chemistry laboratory to maximise binding efficiency. This approach would seem ideally suited to targets against BAFF, given the receptor can be reduced to a single cysteine knot. It remains to be seen whether anyone follows this particular approach using crystallographic data in concert with combinatorial chemistry.

2.5.2 Ligbuilder

Ligbuilder appears to be most useful in situations when a lead compound has been identified and is known to bind to the protein of interest. In this case, the programme can be used to create a library of potential derivataves of the lead molecule, which can then be synthesised chemically and screened using SAR analysis. For the purposes of creating a library of molecules for use in screening available small molecule databases this programme has proven to be less effective than a simple high-throughput screen of a representative 'diversity' set of small molecules taken from a large structural database.

2.5.3 Automated docking in drug screening

The value of screening large databases of molecules against a protein is difficult to gauge given the majority of such efforts take place behind closed doors and use proprietary software and databases. Results from various groups show that AutoDock can be successfully used to screen for potential lead compounds against protein targets [153,155-157]. Those molecules identified by *In vitro* screening of the small molecule database against BAFF and TRAIL may prove to be effective inhibitors of receptor/ligand binding.

Chapter 3 T cell receptor recognition of CD1d

This chapter describes the recognition of CD1 proteins by T cell receptors. The introduction provides an overview of T cell receptor biology and the recognition of MHC and CD1 presented antigens. The aim of this chapter is the determination of a number of NK T cell receptors and an analysis of their structures with reference to the engagement of CD1d presented lipid.

3.1 Introduction

3.1.1 The Cellular Immune Response

The discrimination between self and non-self is a central concern within the immune system for the surveillance of infection by various pathogens. T cells constitute the cellular arm of specific immunity and their response to antigens displayed by cells is mediated by the T cell Receptor (TCR) family of proteins [158]. The role of T lymphocytes in controlling the acquired immune response is mediated by the production of lymphokines (including TNF superfamily molecules) on the engagement of the TCR complex by antigen presented by the appropriate Major Histocompatibility Complex (MHC) or Cluster of Differentiation 1 (CD1) protein [159-161].

Traditionally T cells have been split into groups, classified based on differing effector function. Firstly, by which MHC protein they are restricted to and, secondly, their accessory molecule usage. MHC class I proteins are expressed on the surface of nearly all nucleated cells and present endogenously synthesized peptides for recognition by cytotoxic T cells, the co-receptor CD8 is always found associated with these T cells. MHC class II proteins are more restricted in their expression, being found on specialised antigen presenting cells, such as B cells, macrophages and monocytes, but can be induced on other cell types during immune response. This class of MHC is recognised by Helper T cells [162,163] which use CD4 as their co-receptor. Populations of CD4/CD8 negative T cell also exist that respond to CD1 restricted antigens [164,165].

3.1.2 T cell development

The expression of T cell receptors is limited to T cells, which start their life as emigrants from the bone marrow [166]. T cells begin their developmental pathway to becoming fully mature CD4 or CD8 positive cells within the thymus. The first stage of T cell development involves the rearrangement of the variable portion of the β TCR genes [167], which is similar to the rearrangement of the Immunoglobulin (Ig) heavy chain locus. The T cell receptor gene clusters are found on chromosome 7 for genes encoding the β and γ chains and chromosome 14 for the α and δ chains. Each locus consists of variable (V), joining (J) and constant (C) region genes and the β region contains additional diversity (D) segments. Humans have approximately 50-100 Vα, 50-100 Jα and a single Cα gene segment. For the beta chain, the genome contains 75-100 Vβ, and two clusters containing one Dβ, 6-7 Jβ, and one Cβ segment. It is estimated that the TCR repertoire is as wide as 10^{15} different receptors [168]

The rearrangement of the T cell receptor genes is mediated by specific recombinases. The β locus rearranges prior to the α locus. Functional rearrangement of the β chain locus stimulates the rearrangement of the α locus and the process of differentiation begins. At this point T cells express both CD4 and CD8 and the expression of a functioning α chain leads to the T cells undergoing positive selection and then negative selection (Figure 3.1.1) [169-171].

In negative selection, those T cells possessing a TCR that has a high affinity for self-antigen are deleted by apoptosis. This ensures that T cells do not 'over-react' to self-proteins, thus ensuring self-tolerance. The role and process of positive selection is not

fully understood, but it is thought that this stage selects for T cell receptors that have a moderate affinity for self-peptides [171-173]. The role of negative selection is in the prevention of autoimmunity and that of positive selection in preventing T cell clones that fail to recognise MHC molecules entering the peripheral circulation.

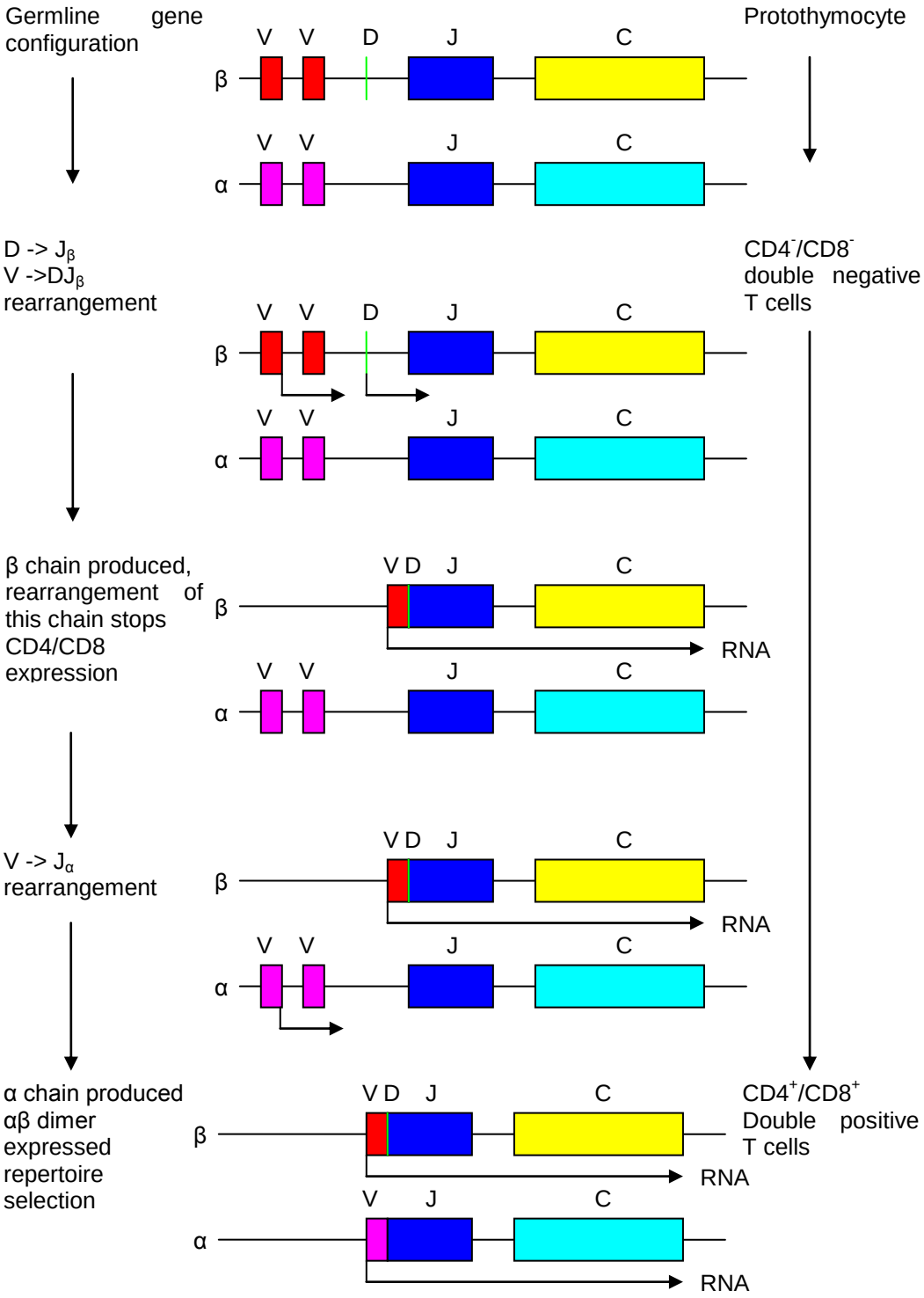


Figure 3.1.1 T cell Receptor gene rearrangements
Rearrangement of the V-D-J and C gene segments for the β chain and then V-J and C segments for the α chain takes place to increase diversity of T cell Receptor proteins in a manner similar to that seen for immunoglobulin proteins.

3.1.3 TCR recombination and specificity for MHC

As detailed in the previous section T cell receptors are assembled from genes that undergo recombination, in the manner of antibodies, from V, D, J and C segments [174], although they do not undergo somatic mutation to increase affinity for their ligand as antibodies do. TCR/MHC affinities are substantially weaker than those of antibodies for their antigens [175,176] (of the order of KD = 4 x 10^{-5} weaker) and thus they rely on the co-receptors CD4 or CD8 for signalling. From the diversity seen in the genetic loci of T cell receptor genes, the determinants of MHC binding were suggested to reside in the loop regions in the variable domain in an analogous fashion to that seen in immunoglobulins, with three surface loops per chain that were termed the Complementarity Determinant Regions (CDR) [177]. The major region of diversity in the TCR is found in the third CDR, as this region of the protein is encoded at the junctions of gene segments and the recombination of these is often imprecise, giving enhanced diversity. This region interacts directly with the MHC protein/antigen complex [178].

3.1.4 The T cell receptor complex

The T cell receptor consists of two protein chains, each with two Ig like domains [140]; the structure of one domain remains constant between T cell clones, while the other variable domain is where antigen and MHC recognition takes place and differs between clones. TCRs are membrane anchored, but do not possess any functional intracellular domain, the receptor forms a signalling complex with the various chains of CD3 and either the CD4, or CD8 co-receptor (Figure 3.1.2). When antigen presented by MHC or CD1 binds to the TCR, the associated CD3 complex transduces the activating signals to the cytoplasm of the T cell.

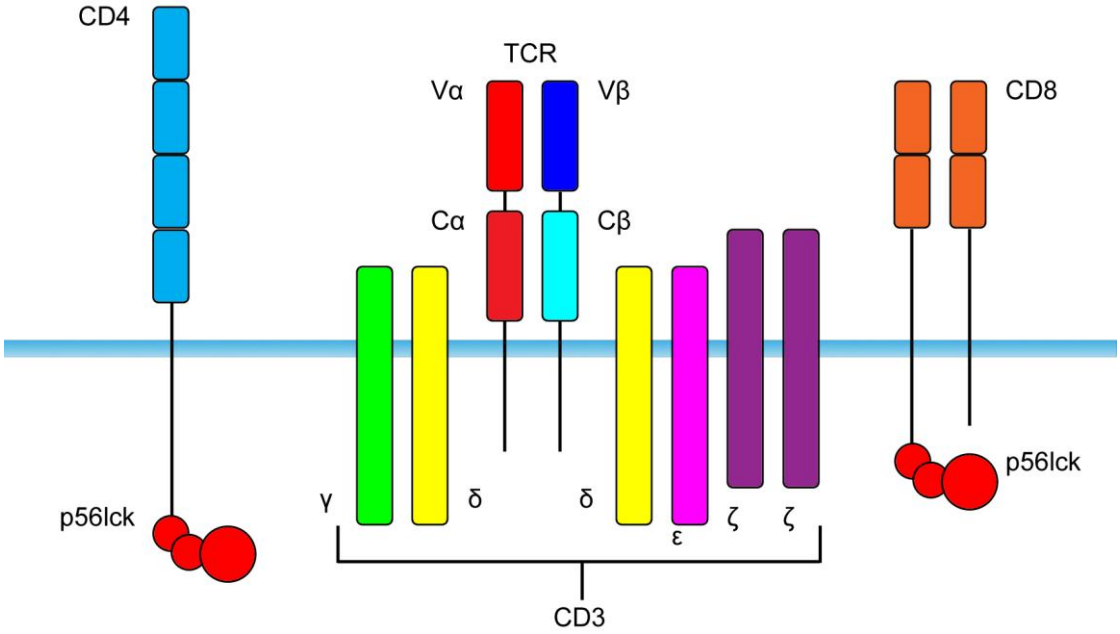


Figure 3.1.2 The T cell receptor signalling complex
Engagement of the T cell receptor complex, consisting of the six chain CD3 complex and either CD4 or CD8, by MHC presented antigen leads to the activation of intracellular tyrosine kinases (p56lck) which in turn activates further kinases and leads to the activation of transcription factors for the T cell cytokines.

3.1.5 Structural biology of the T cell Receptor

The T cell receptor consists of two Ig-like chains, each with two beta-sandwich domains, a number of these proteins have been solved in isolation and in complex with MHC-presented antigens (see Appendix TCR for details and references). The variable domain is involved in the recognition of MHC or CD1 presented antigen. While the constant domain of the TCR acts in the recruitment of the CD3 signalling complex, and allows a certain degree of flexibility in the hinge-region joining the variable and constant chains that may be utilised in ligand binding (Figure 3.1.3). Structures of TCRs determined to date indicate that this flexibility is of the order of 5º, and is considerably less than that seen in antibodies.

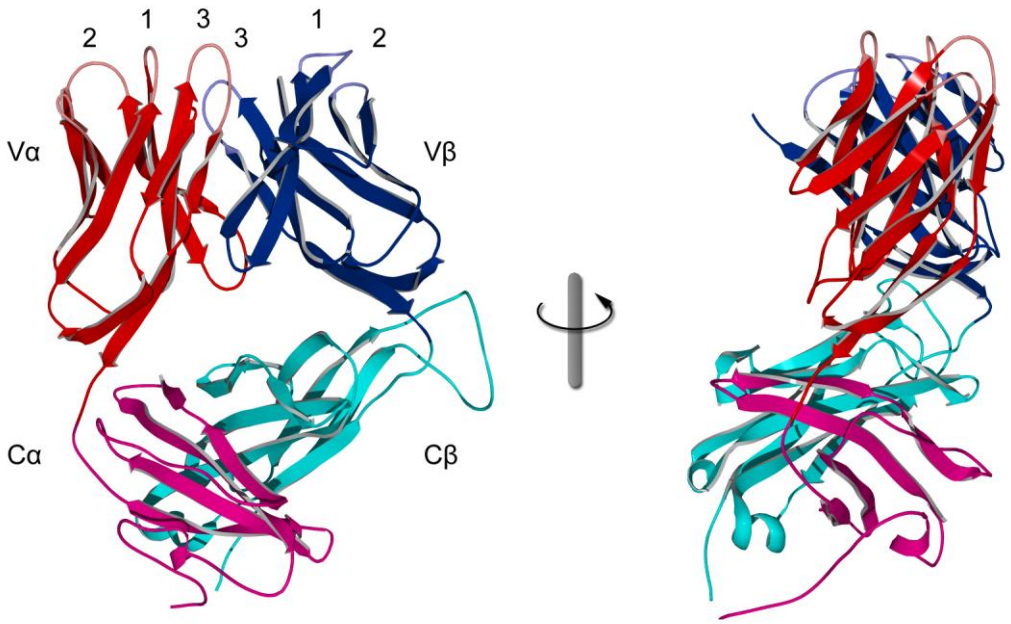


Figure 3.1.3 T cell Receptor structure
The structure of the T cell receptor showing both chains, with the variable and constant domains for the alpha and beta chains indicated and the CDR loops numbered. The view on the right shows a rotation by 90 degrees into the plane of the paper from the view on the left

3.1.6 TCR/MHC binding

The binding of MHC class I or II presented peptide (pMHC) by the TCR is key to the action of T cells and a number of structures have been reported with the TCR in complex with pMHC [179-182]. The MHC gene cluster was initially identified as the major source of tissue transplant incompatibility (4133807), within this gene cluster in humans are genes for the class I and II MHC proteins, a number of cytokine genes (LT and TNF) and also for the complement proteins. Within the class I and II MHC genetic regions a number of polymorphic loci exist for these and in the wider population, a large number of alleles exist for these.

Class I MHC molecules consist of two peptide chains, an MHC-encoded α-chain of around 44 kDa and a non-MHC-encoded β-chain of 12 kDa known as β_2Microglobulin (β_2M). The extra-cellular portion of the α chain consists of three domains, an immunoglobulin-like domain, α3, which interacts with β_2M and the peptide binding region which consists of around 180 amino acids at the N-terminus of the protein chain. This peptide binding region forms an eight stranded β-sheet which supports two α-helices, half of which is contributed by each of two α-domains, α1 and α2 (i.e. four strands and one helix from each).

The peptide-binding groove formed by class I MHC molecules has closed ends and can accommodate peptides between eight and ten amino acids in an extended conformation. The surface of the peptide-binding groove can form a number of pockets in which to accommodate amino acid side-chains, designated A-F. Although the A and F pockets are relatively conserved, the B to E pockets show a high degree of polymorphism across the allelic variants of MHC I.

Class II MHC molecules differ from class I MHC in that they consist of two similar protein chains of around 30 kDa, these each have two extra-cellular domains, one an immunoglobulin-like domain (designated α2 and β2) and the other, N-terminal domain, which forms the peptide binding groove (α1 and β1 domains). The peptide-binding groove of class II MHC molecules is different in character to that of class I MHC in that it can accommodate longer peptides as it has an open-ended groove, due to it being formed by two protein chains. Because of this, a single class II MHC protein can bind a greater number of potential peptides than a class I MHC molecule.

The allelic diversity found for MHC genes and the fact that each individual possesses two copies of the MHC gene cluster (one from each parent) allows for the recognition of a wide range of antigens. In terms of pathogenic antigens, this diversity would prevent a single pathogen from evading all responses from a random sample of the population. The majority of diversity found in the MHC proteins is found in the peptide-binding groove.

In TCR/pMHC complexes solved to date, the TCR docks on to pMHC with the Vα domain positioned over the N-terminal half of the peptide and the Vβ domain over the C-terminus (Figure 3.1.4). Interactions between the the TCR and binding-groove α-helices and presented peptide account for the specificity of interaction. There is a degree of variation in the twist of the TCR over the long axis of the pMHC [183] with a 70 º difference in orientation seen between some structures [184]. There are no strictly conserved interactions between different TCR/pMHC complexes, with a high degree of variability in the shape of the binding surfaces presented by pMHC to TCR

[140,185,186]. The solution of structures of TCR molecules in isolation and bound to pMHC have shown that the CDR3 loops of the TCR exhibit the highest degree of conformational flexibility in the pMHC binding surface [140,181]. This flexibility has been suggested as an explanation for the apparent promiscuity of TCRs for different pMHC complexes [187]

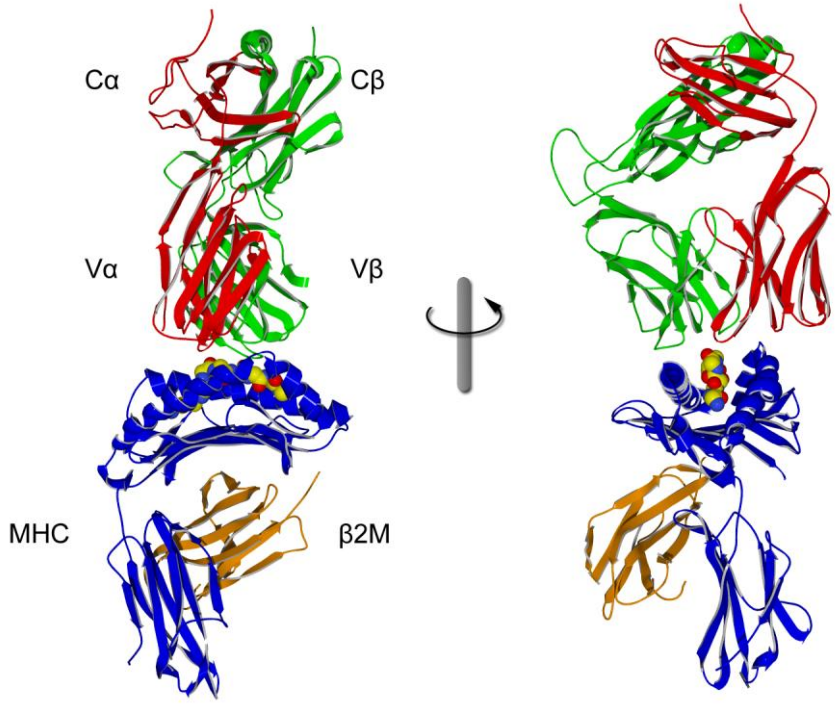


Figure 3.1.4 TCR/MHC I binding
T cell receptor binds peptide/MHC complex in a roughly diagonal arrangement, with the CDR loops of the variable domains forming the major contacts to the protein. MHC presented ligand is held in the vice-like beta-sheet and helix peptide binding groove. The MHC heavy chain is in blue, β_2M in orange, MHC presented antigen is shown in CPK representation with the C-terminus of the peptide to the left and the N-terminus to the right. The TCR α and β chains are shown in red and green, respectively.

3.1.7 TCR/CD1 binding

Recent studies have shown that the glycolipid content of cells is surveyed by specific T cell receptors that bind to CD1/glycolipid complexes. CD1 molecules, found in all mammalian genomes studied to date, are structurally related to class I MHC and are thought to have arisen by divergent evolution from a common ancestor. CD1 heavy chains are expressed as a heterodimer with β_2M on the surface of professional antigen presenting cells, such as myeloid dendritic cells, Langerhans cells, and B-cells [159].

The CD1 genes are found in two clusters on chromosome one in humans [188] with CD1a, b, c and e forming a separate group to CD1d. The group I CD1 proteins interact mainly with foreign lipid antigens, while CD1d (the only group II CD1) is involved in the presentation of self-lipid antigens and those produced by the α-proteobacteria, such as *Sphingomonas capsulate* and *Erlichia muris* [189,190]. The lack of polymorphism in CD1 molecules when compared to the MHC proteins is not mirrored in the diversity of antigens that CD1 can bind, with lipopeptides, sphingolipids, polyketides, small molecules and polyacylated carbohydrates all presented by the molecule.

Recent structural studies of CD1 molecules [191,192] have shown that the diverse ligands are accommodated within the glycolipid binding groove that has individual binding pockets, denoted by A', C', F' and T' in the case of human CD1b[193]. These deep pockets are highly hydrophobic and accommodate the alkyl chains of the glycolipid antigen, with the hydrophilic head-group protruding from the groove (discussed in Chapter 3.1.8). Due to the similarities observed between CD1/glycolipid and pMHC structures, the formation of the CD1/glycolipid/TCR complex is proposed to be highly analogous to that seen of pMHC/TCR.

Controversy over CD1 isoform specificity for ligands exists, but there is no definite pattern of expression or lipid binding, other than the fact that the long groove of CD1b allows longer lipids to be accommodated than that of any of the other isoforms [194]. It remains to be seen whether, with more structural and biochemical information, distinct patterns will emerge. CD1 molecules that bind to myelin lipids have been identified in multiple sclerosis and experimental encephalomyelitis and thus a role for self-reactive TCRs has been identified in autoimmune disease [195].

3.1.8 CD1 ligand binding

Structures for the mouse and human CD1d [196-198], human CD1b [191,194] and human CD1a [199,200] have been solved (Figure 3.1.5). For all CD1 molecules, the interaction with lipid occurs in a defined ligand-binding pocket, analogous to the class I MHC binding groove. The CD1 ligand binding pockets are somewhat deeper and narrower in form than the class I MHC binding pockets. The pockets are lined with hydrophobic residues, which interact with the acyl chains of the glycolipid, while the polar head-group protrudes from the centre of the groove and is the main portion of the glycolipid antigen that is exposed for recognition by TCRs. The differences found in the binding grooves of CD1 isoforms fine-tunes their preference for antigen. CD1b has a large antigen binding region formed by long A' and F' pockets with a joining T' tunnel and short C' pocket, allowing it to bind acyl chains up to 80 carbons in length [191]. The CD1d pocket is much smaller in comparison, with the T' tunnel partially blocked by large hydrophobic residues, and it can accommodate acyl chains of up to 26 carbons in length [201].

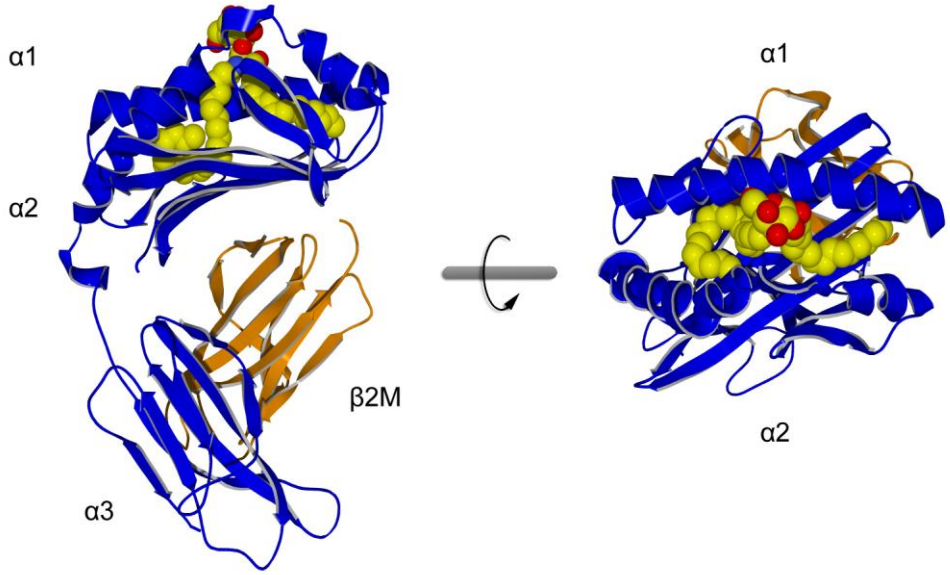


Figure 3.1.5 The structure of human CD1d with α-GalCer
CD1d structure with α1-α3 domains (blue), β_2M (orange) with bound α-GalCer shown in CPK representation with carbon atoms shown in yellow and oxygen atoms in red. Viewed from the front (left) and orthogonal to this (right). (After Koch *et al* [197])

3.1.9 Recognition of CD1d bound α-GalCer by NKT cell receptors

CD1d is recognised by the Natural Killer T cell receptors (NKTCRs) on the surface of NKT cells [202-204], the canonical NKTCR has Vα24-JαQ chain paired with a Vβ11 chain and is referred to as the invariant NKTCR (*i*NKTCR). The glycolipid α-galactosylceramide (α-GalCer), originally isolated from a marine sponge, binds to both human and murine CD1d with high affinity and specifically activates invariant NKT cells. This leads to the release of T_H1 and T_H2 type cytokines, that lead to the suppression or enhancement of immune responses [205,206]. The binding of α-GalCer to CD1d is proposed to be similar to that of bacterial α-glycuronosylceramide ligands to CD1d and provides insight into potential design of agonists for specific T_H1 or T_H2 activation [198]. The conserved Vα CDR3 region of the *i*NKTCR has been suggested to be essential for recognition of CD1d bound ligand [197,198].

These *i*NKTCRs were found to co-express with the NK specific C-type lectin NKR-P1 [207]. All Cd1d-αGalCer specific *i*NKT cells so far described are $CD4^+$, $CD8\alpha\alpha^+$ or $CD4^-CD8^-$ double negative (DN). $CD8\alpha\beta^+$ *i*NKT cells have never been described and are thought to be deleted during development due to their high binding avidity to CD1d [208].

The *i*NKT cells are capable of secreting large quantities of regulatory cytokines, such as IL-4 and IFN-γ and the loss of these cell lineages is responsible *in vivo* for the development of various autoimmune diseases [209-211]. Recently, the existence of populations of $V\alpha24^-$ T cells that bind to CD1d-α-GalCer was described [212]. These T cell receptors were derived from the standard T cell lineage and were found to be either $CD4^+$, or $CD8\alpha\beta^+$, showing that this can be used as a co-receptor for CD1d

specific T cells, and that the recognition of CD1d-α-GalCer is not restricted to the *i*NKTCR and may also be mediated by other randomly generated TCRs.

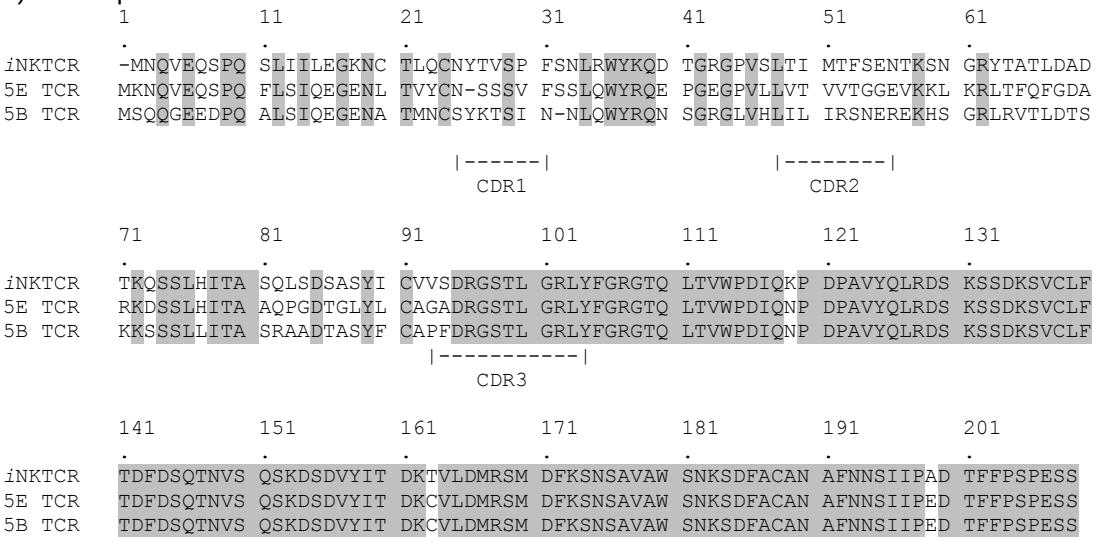
3.1.10 TCR/CD1 binding scheme

The similar structures of both class I MHC and CD1 would imply a similar engagement of TCR by both molecules. To date, no CD1/TCR structure has been solved and the details of TCR engagement of the widely varying CD1 presented antigens remain to be seen. Questions as to why particular TCRs are specific for either CD1, or class I MHC, and how antigen recognition is achieved by the CDR loops of the protein have yet to be answered.

3.1.11 Aims

Although a large number of structures of T cell receptors that bind to MHC have been solved, many in complex with MHC (Appendix TCR), no structure of a CD1 binding T cell receptor exists in the structural databases. Three NKT cell restricted TCRs, the *i*NKTCR and two non-invariant NKTCRs, 5E and 5B TCR, were cloned, expressed and purified by Stefan Gadola (Insel, Switzerland) (A sequence alignment of these is shown in Figure 3.1.6). The crystallographic data-collection and analysis of two of these, 5E and 5B TCR, was performed, while the structure of the *i*NKTCR, was solved by M. Koch. The crystallographic analysis of these and implications for CD1d binding will be described in this Chapter.

a) TCR alpha chains



b) TCR beta chains

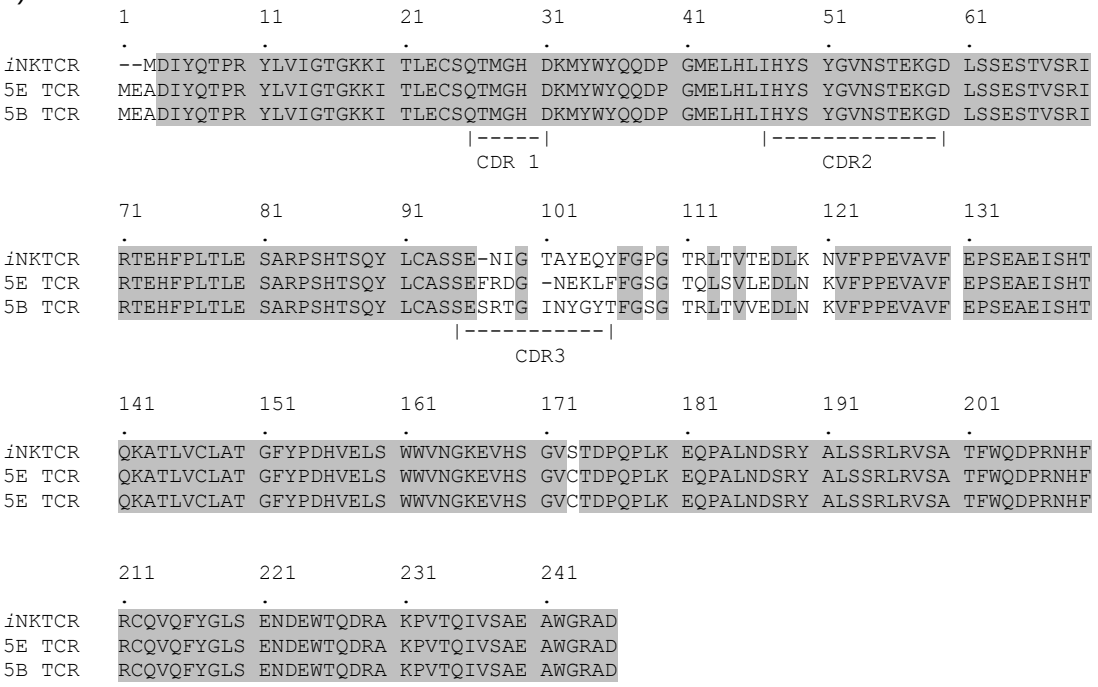


Figure 3.1.6 Alignment of three NKT cell receptor sequences
The *i*NKTCR is shown aligned with two non-invariant CD1-αGalCer recognising TCRs. Sequence identities between all three proteins are shaded in grey. The position of the CDR loops, as defined by Chothia *et al* [177], are shown under the sequences. The *iNKTCR* has Vα24, Vβ11 while the other two receptors each have different Vα domains (5B TCR Vα10, 5E TCR Vα3) and share a Vβ11 domain.

3.2 Determination of the structure of three NKT cell receptors

3.2.1 Crystallization and data collection

All crystallizations were carried out by the sitting drop vapour diffusion method and were set up as described in Chapter 1.2.15, at a concentration of 10 mg ml^{-1} protein in 100 nl drops of protein with 100 nl reservoir solution [88].

Crystals of 5E TCR grew at room temperature in 200 mM magnesium sulphate and 20 % polyethylene glycol 3350. Crystals were soaked briefly in per-fluoroether oil and flash cooled and maintained at 100 K in a cryostream (Oxford Cryosystems). 5B TCR crystals grew at room temperature at a final concentration of 10 mg ml^{-1} in 200 mM di-ammonium tartrate, 20 % polyethylene glycol 3350. *i*NKTCR crystals grew at room temperature at a final concentration of 10 mg ml^{-1} in 0.5 M NaCl, 11 % polyethylene glycol 8000, 50 mM HEPES, pH 7.0.

Diffraction data for the 5E TCR were recorded by K. Harlos at station ID14-EH2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) with an ADSC Q4 CCD detector (Figure 3.2.1). Because of detector overloads at low crystal to detector distance both a high-resolution (175 mm detector distance) and a low-resolution (300 mm detector distance) dataset were collected from the same crystal. The crystal belonged to the spacegroup $P3_221$ (a = b = 64.5 Å, c= 184.9 Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$) and both datasets were autoindexed with Denzo and scaled together using Scalepack (http://www.hkl-xray.com) (Table 3.2.1). There was a single molecule in the asymmetric unit and 44 % solvent.

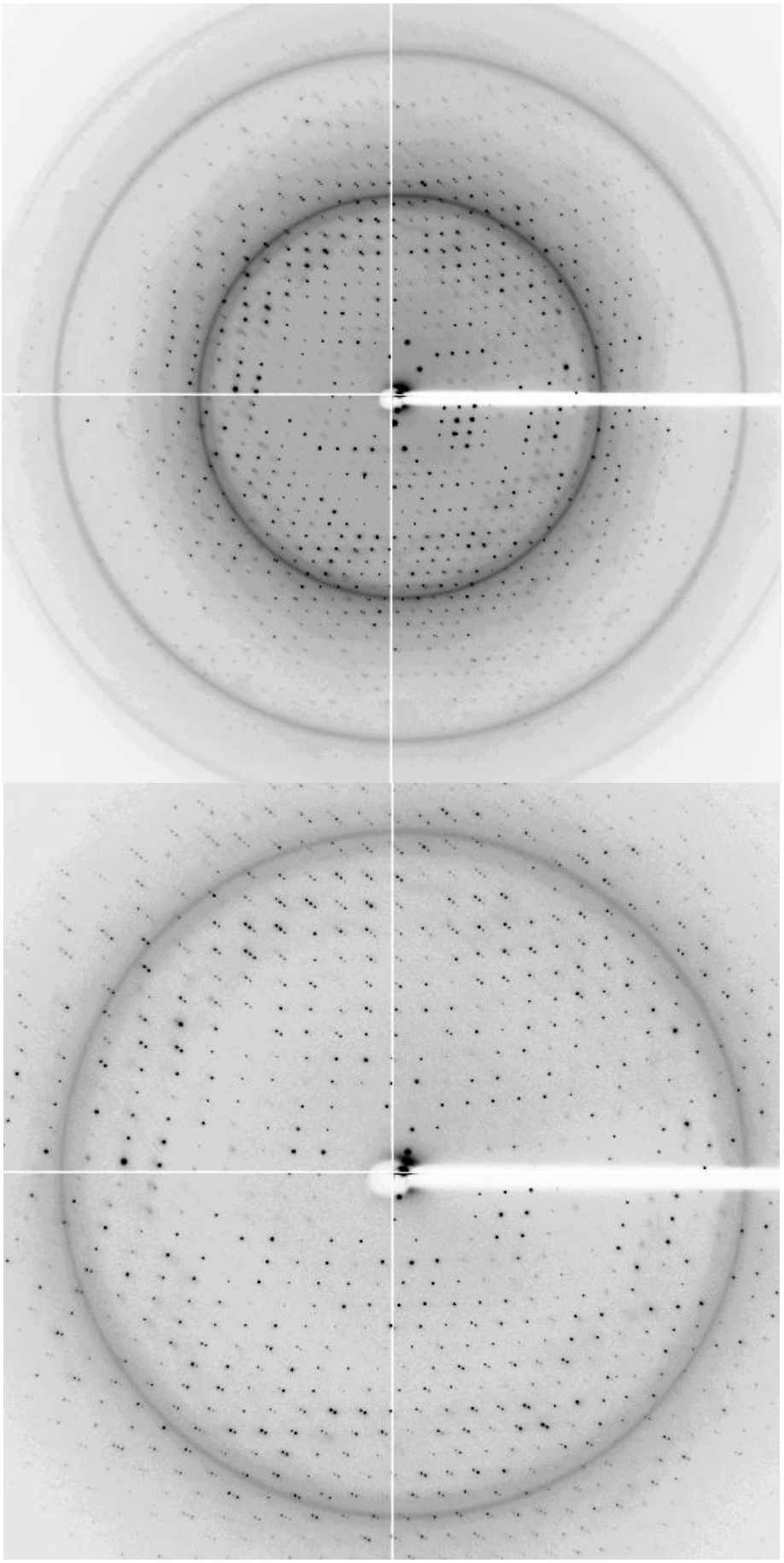


Figure 3.2.1 X-ray diffraction images for 5E TCR

Data for the 5B TCR were recorded at station 14.2 at the SRS Radiation Facility at the Daresbury Laboratory (SRS Daresbury, Warrington) with an ADSC Q4 CCD detector. The crystal belonged to the spacegroup $P3_121$ (a = b = 64.0 Å, c = 185.0 Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$), with a single molecule in the asymmetric unit and 43 % solvent. The diffraction from this crystal gave somewhat smeared spots (Figure 3.2.2), which presented some difficulties for processing of the diffraction images, a large spot size was set in DENZO to ensure that all collected data were indexed and these were integrated using SCALEPACK (Table 3.2.1).

That 5E TCR and 5B TCR should have almost identical unit cell parameters, yet have different space groups was surprising. An attempt was made to scale the two data sets together to confirm that they were different; this gave very high chi squared and R_{merge} values, which indicated that they did indeed belong to different space groups.

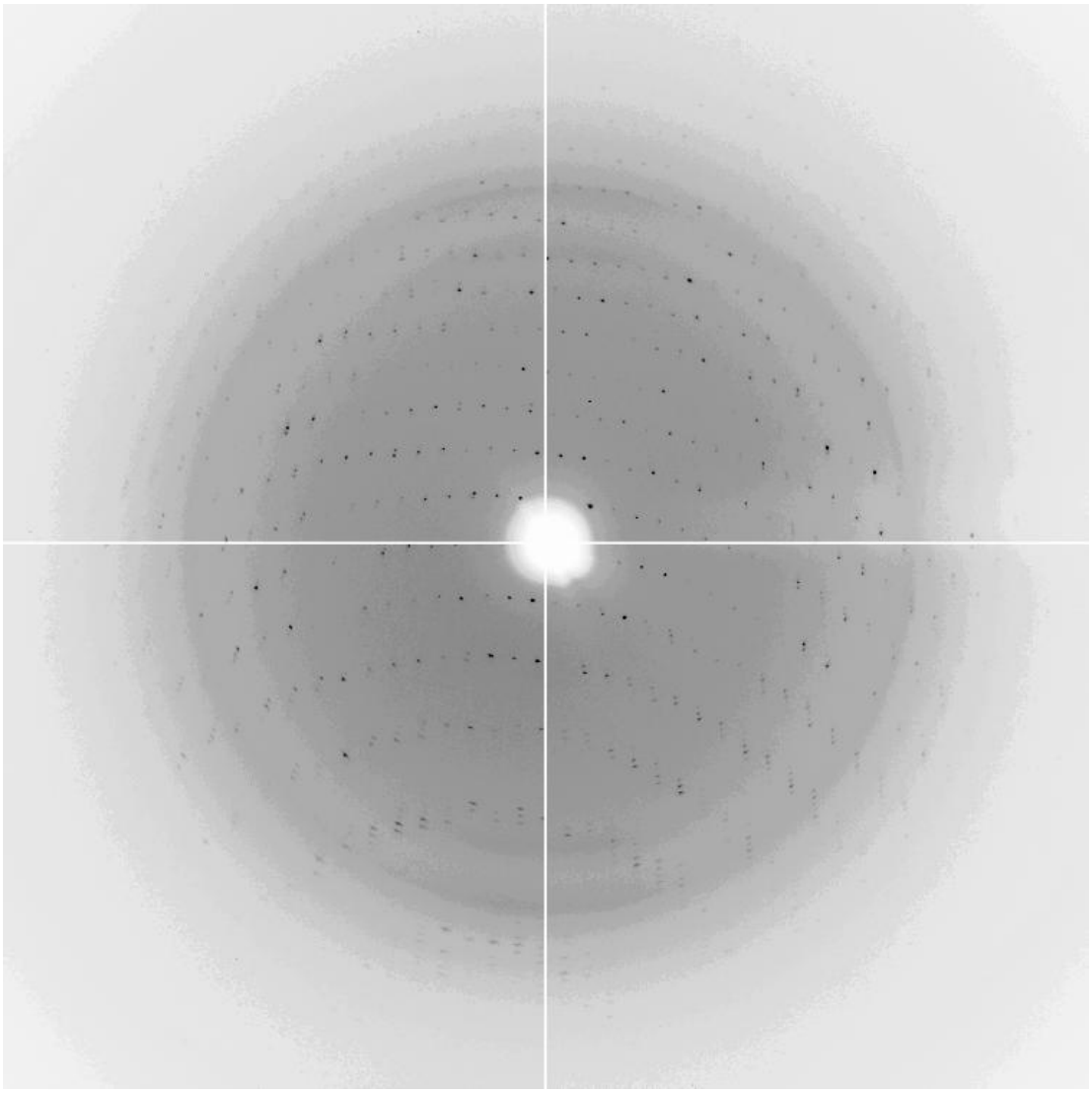


Figure 3.2.2 X-ray diffraction image for 5B TCR

Data was collected for *i*NKTCR by K. Harlos at station ID14-EH1 at the ESRF using an ADSC Q4R CCD detector. The crystal belonged to the spacegroup C2 (a = 289.4 Å, b = 85.0 Å, c = 78.9 Å, $\alpha = \gamma = 90^{o}$, $\beta = 103.9^{o}$), with three molecules in the asymmetric unit and 60 % solvent. Due to variance in the x-ray beam intensity, data processing statistics were somewhat high when the data was indexed and integrated with DENZO and SCALEPACK (Table 3.2.1).

3.2.2 Structure determination and refinement

Both the Crystallography and NMR system [96] (http://cns.csb.yale.edu/) and REFMAC from the CCP4 [213] (http://www.ccp4.ac.uk) suite of programmes were used for refinement. Approximately 5% of reflections were set aside for the R_{free} calculations. See Table 3.2.1 for refinement statistics.

5E TCR

The structure of 5E TCR was determined by molecular replacement using the JM22 T cell receptor structure [184] as the search model in the molecular replacement module of CNS [96]. A single strong rotation function peak was found and was used in a translation search; two symmetry related peaks in the space group $P3_221$ were found with a high correlation co-efficient and good packing scores. The top scoring solution from this stage was used in subsequent refinement.

After initial rigid-body refinement of the Vα, Vβ, Cα and Cβ domains using CNS, the sequence of JM22 was replaced by the 5E TCR sequence rebuilding into F_o-F_c and $2F_o$-F_c electron density. This modelling was followed initially by cycles of simulated annealing using CNS and, as the quality of the model improved, subsequent cycles of

refinement included positional refinement, individual B-factor refinement with bulk solvent scaling and overall anisotropic B-factor scaling. This refinement was cycled with manual rebuilding using COOT [214] and O (http://www.bioxray.dk/~mok/o-files.html), in the final stages of refinement water molecules were added using ARP-wARP [215] on the basis of peaks of at least 3 σ in the F_o-F_c electron density maps. The CDR 3 loops had weak electron density in the $2F_o$-F_c maps and were rebuilt from simulated annealing omit maps calculated with CNS with the CDR loops omitted from the calculation of the maps. This gave clear density for the path of both the CDR 3 alpha and beta. In the final stage of refinement the model was subjected to TLS and restrained refinement using REFMAC [216,217] with the Vα, Vβ, Cα and Cβ domains defining the TLS groups.

The final refined structure had good stereochemistry, as assessed by the programme PROCHECK [217] (Table 3.2.1 and Ramachandran plot in Figure 3.2.3). The final model had an R_{work} of 18.8 % (R_{free} 26.8 %) and comprised residues 3 - 193 of the alpha chain, 2 - 245 of the beta chain and 286 water molecules.

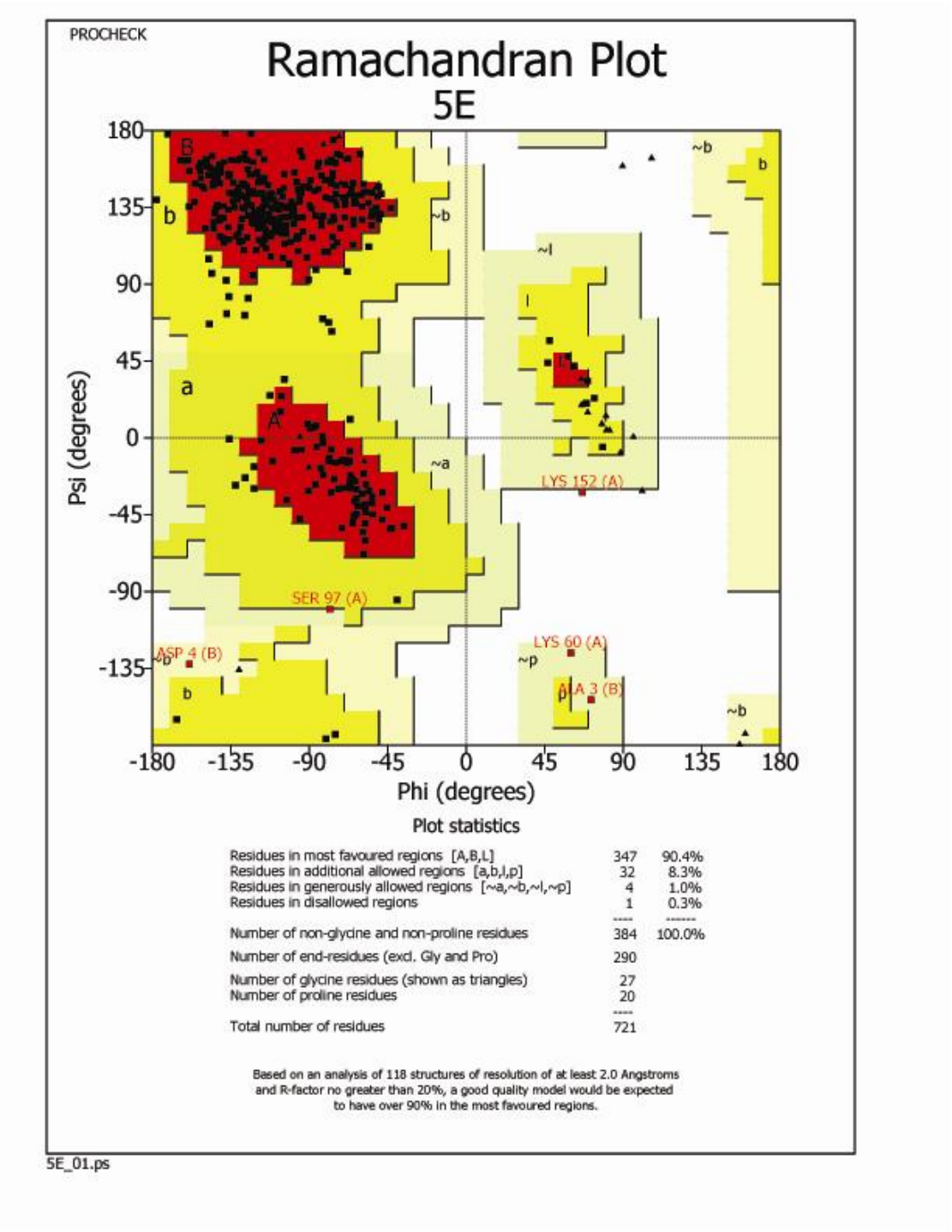


Figure 3.2.3 Ramachandran plot for stereochemistry observed in the 5E TCR model
This figure was produced using Procheck.

5B TCR

The structure of 5B TCR was determined by molecular replacement using the 5E TCR structure as the search model in CNS. A single strong rotation function peak was found and was used in a translation search. Two symmetry related peaks in the space group $P3_121$ were found with high correlation coefficients and good packing scores.

After initial rigid-body refinement of the model Vα, Vβ, Cα and Cβ domains using CNS, the sequence of 5E TCR was replaced with that of 5B TCR guided by F_o-F_c and $2F_o$-F_c electron density maps calculated with CNS. This modelling was followed by refinement as described for 5E TCR.

The final refined structure had good stereochemistry, as assessed by the programme PROCHECK [217] (Table 3.2.1 and Ramachandran plot in Figure 3.2.4). The final model had an R_{work} of 21.7 % (R_{free} 31.8 %) and comprised residues 10 - 193 of the alpha chain, 3 - 246 of the beta chain and 61 water molecules. The high R_{free} value and large difference from the R_{work} seen for this structure is not uncommon for TCR structures (Appendix TCR) and may in the case of this dataset be a consequence of the low completeness of the data in the highest resolution shell used in refinement (87 % between 2.6 and 2.67 Å).

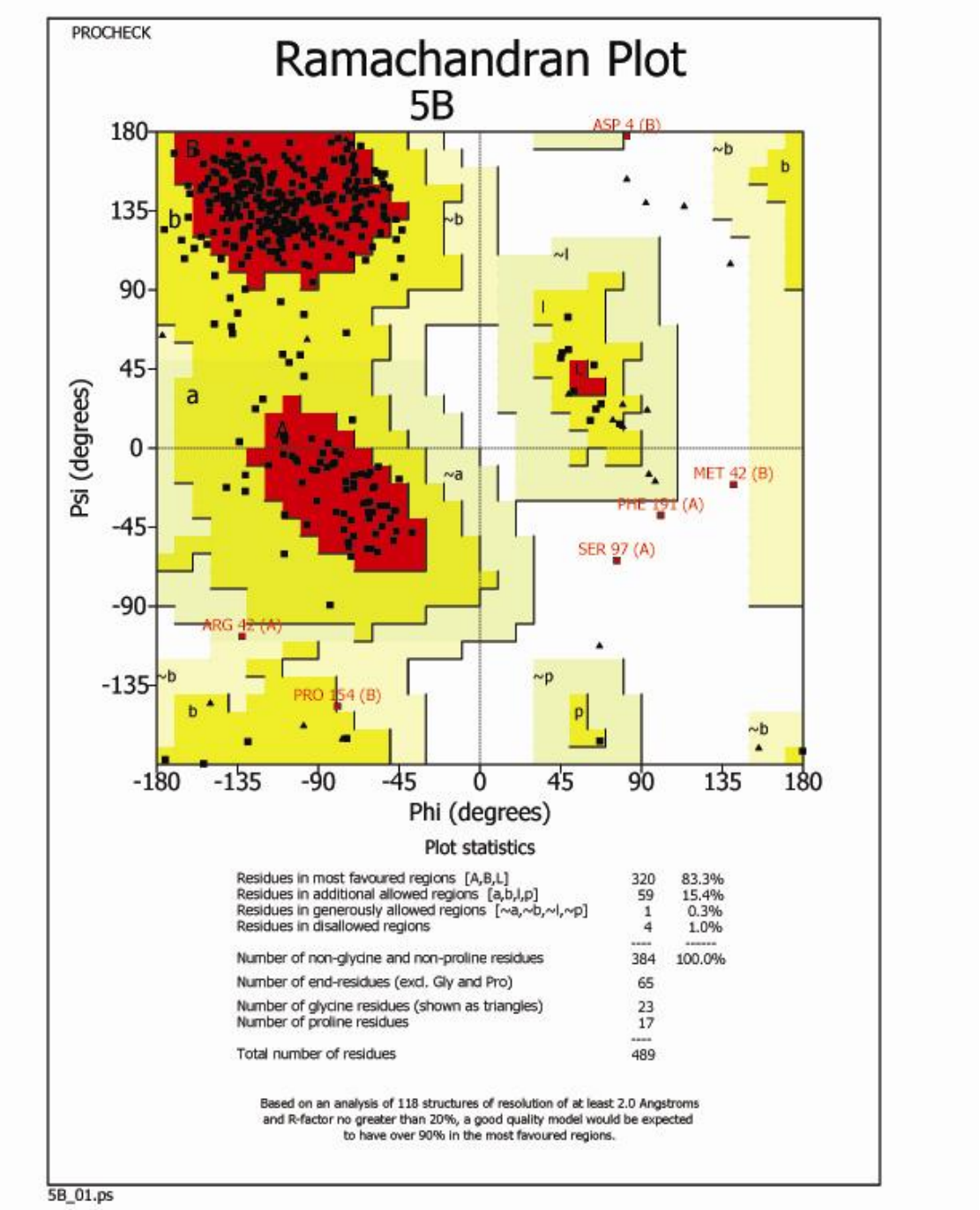


Figure 3.2.4 Ramachandran plot for stereochemistry observed in the 5B TCR model
This figure was produced using Procheck

iNKTCR

The structure of the *i*NKTCR was solved by molecular replacement using the structure of 1MI5 by M. Koch. The chosen molecular replacement model was used as the input model for the CaspR web interface (discussed in Chapter 1.4.1). This gave three clearly defined solutions for the three molecules in the asymmetric unit. NCS restraints were set up for the three molecules within the asymmetric unit; SHP (D. Stuart unpublished programme) was then used to superimpose the domains from 5E TCR onto the three *i*NKTCR molecules in the asymmetric unit. CNS was then used to carry out rigid body refinement on the twelve domains (Vα, Cα, Vβ and Cβ for each of the three NCS related molecules), with low restraints placed on the CRD loops. Omitting the CDR loop regions of the molecule, $2F_o$-F_c and F_o-F_c electron density maps were calculated with CNS and the sequence of 5E TCR was replaced with sequence of the *i*NKTCR in O. Simulated annealing refinement was cycled with manual rebuilding of the molecule guided by $2F_o$-F_c and F_o-F_c electron density maps in O. The CDR loop regions were built into the resulting electron density from simulated annealing omit maps, initially as poly-alanine chains and as the side-chains became apparent in subsequent cycles of refinement these were replaced with the correct sequence. The stereochemistry of the model was corrected using Calpha (R. Esnouf, unpublished programme). Positional and individual B-factor refinement was carried out with CNS using NCS restraints (again with low restraints placed on the CDR loops) in later stages of refinement.

The final refined structure had good stereochemistry, as assessed by the programme PROCHECK [217] (Table 3.2.1 and Ramachandran plot in Figure 3.2.5). The final model had an R_{work} of 32.1 % (R_{free} 36.4 %) and comprised three receptors in the

asymmetric unit comprising residues 2 - 193 of the alpha chain, 2 - 245 of the beta chain and no water molecules.

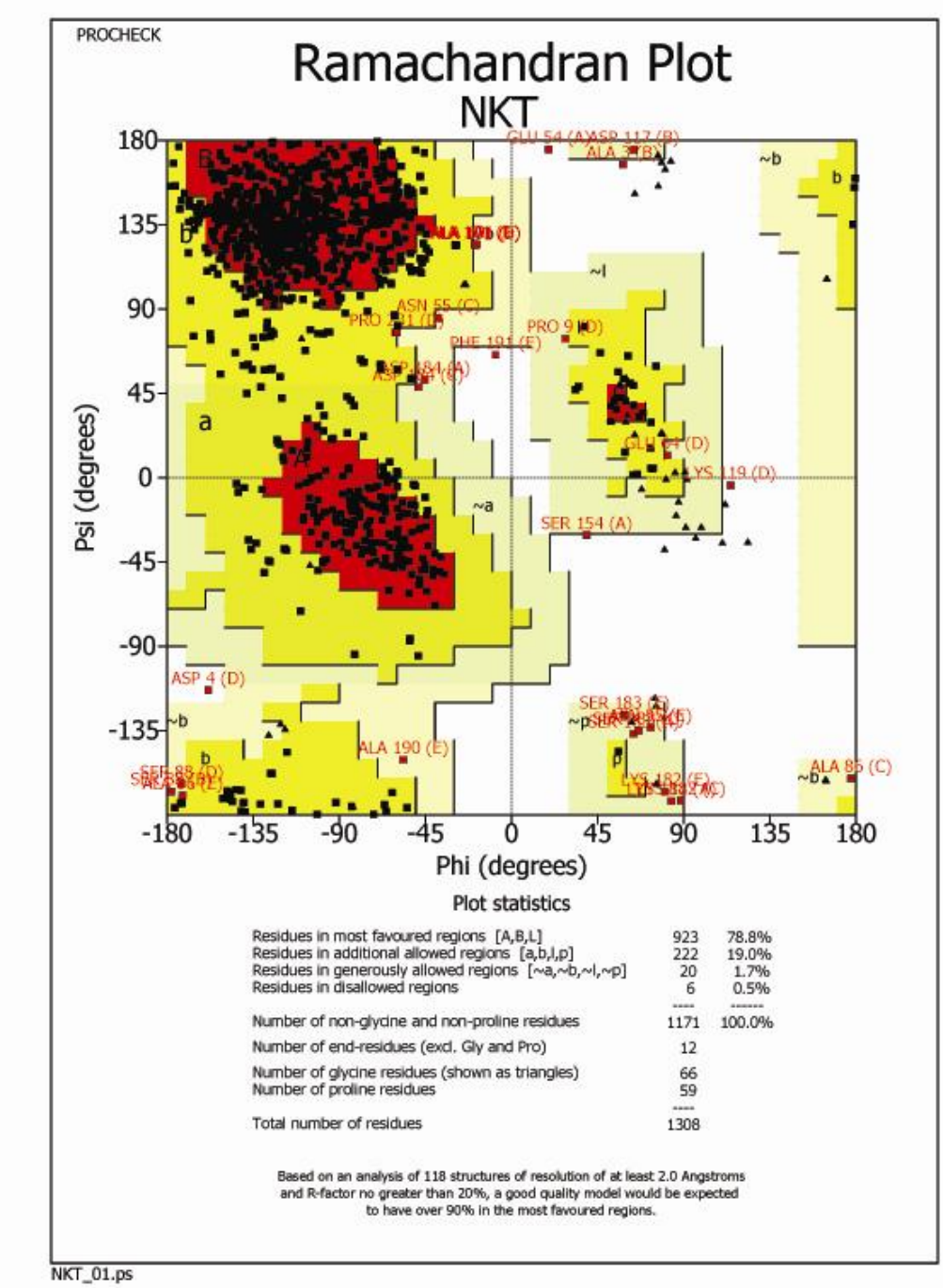


Figure 3.2.5 Ramachandran plot for stereochemistry observed in the *i*NKTCR model. This figure was produced using Procheck.

Crystallographic Statistics			
	5E TCR	**5B TCR**	***i*NKTCR**
Data collection			
Resolution range (Å) [a]	30–2.25	18–2.6	30–3.5
(highest resolution shell)	(2.3–2.25)	(2.67–2.6)	(3.62–3.5)
Number of collected reflections	271,030	73, 961	75,995
Unique reflections	21,795	13,750	23,622
Completeness (%)	98.8 (89.5)	96.6 (87.4)	98.1 (90.9)
R_{merge} (%)[b]	7.1 (23.8)	12 (55)	18.3 (88.0)
I/σI	31 (7.9)	10.4 (2.2)	6.4 (1.4)
Space group	P 3_221	$P3_121$	C2
Unit cell			
Dimensions (Å) (a, b, c)	64.5, 64.5, 184.9	64.0, 64.0, 185	289.4, 85.0, 78.9
Angles (°) (α, β, γ)	90, 90, 120	90, 90, 120	90, 103, 90
Source	ESRF ID14-EH2	SRS Daresbury 14.2	ESRF ID14-EH1
Model Refinement			
Resolution range (Å)	30–2.25	18–2.6	30–3.5
Number of reflections (test set)[c]	20,555 (1,098)	13,033 (656)	23,453 (1,193)
R_{cryst} (%)[d]	18.8	21.7	32.1
R_{free} (%)[e]	26.8	31.8	36.4
Number of non-hydrogen protein atoms	3436	3392	10,330
Number of water molecules	286	61	0
Average B-factors			
Protein (Å^2)	30.6	22.8	46.1
Water (Å^2)	30.2	23.5	n/a
r.m.s deviation from ideality			
Bond lengths (Å)	0.012	0.009	0.006
Bond Angles (°)	1.41	1.32	1.21
r.m.s deviation B factors (bonded atoms)			
Main Chain (Å^2)	2.3	1.8	2.4
Side Chain (Å^2)	3.1	1.9	4.6
Ramachandran Plot			
Favoured (%)	90.4	83.3	78.8
Allowed (%)	8.3	15.4	19.0
Generous (%)	1.0	0.3	1.7
Unfavoured (%)	0.3	1.0	0.5

Table 3.2.1 Statistics for crystallographic data collection and refinement

[a]Values in parentheses refer to the highest resolution shell of data. [b]R_{merge} = Σh Σi | I*i*(hkl) - <I(hkl)> I / Σh Σi Ii(hkl), where I_i(hkl) is the 'ith' measurement of reflection hkl and <I(hkl)> is the weighted mean of all measurements of reflection hkl. [c]Test set is a randomly chosen set of reflections omitted from the refinement process. [d]R_{cryst} = Σh || F_{obs}(hkl)| - | F_{calc}(hkl)|| / Σh|F_{obs}(hkl)|, where F_{obs} and F_{calc} are the observed and calculated structure factor alplitudes, respectively. [e]R_{free} is equivalent to R_{cryst} but calculated for the test set of reflections.

3.2.3 Structure of the *i*NKTCR, 5E and 5B TCRs

Soluble 5E, 5B and *i*NKT cell receptors were produced using published refolding methods [184] and crystallization trials were set up for these [88]. 5E TCR crystallized in the space group $P3_221$ with a single molecule in the asymmetric unit; 5B TCR in space group $P3_121$, again with a single molecule in the asymmetric unit; and *i*NKTCR in C2 with three molecules in the asymmetric unit. These were solved by molecular replacement and refined to 2.25, 2.6 and 3.5 Å respectively (Figure 3.2.6 and Table 3.2.1).

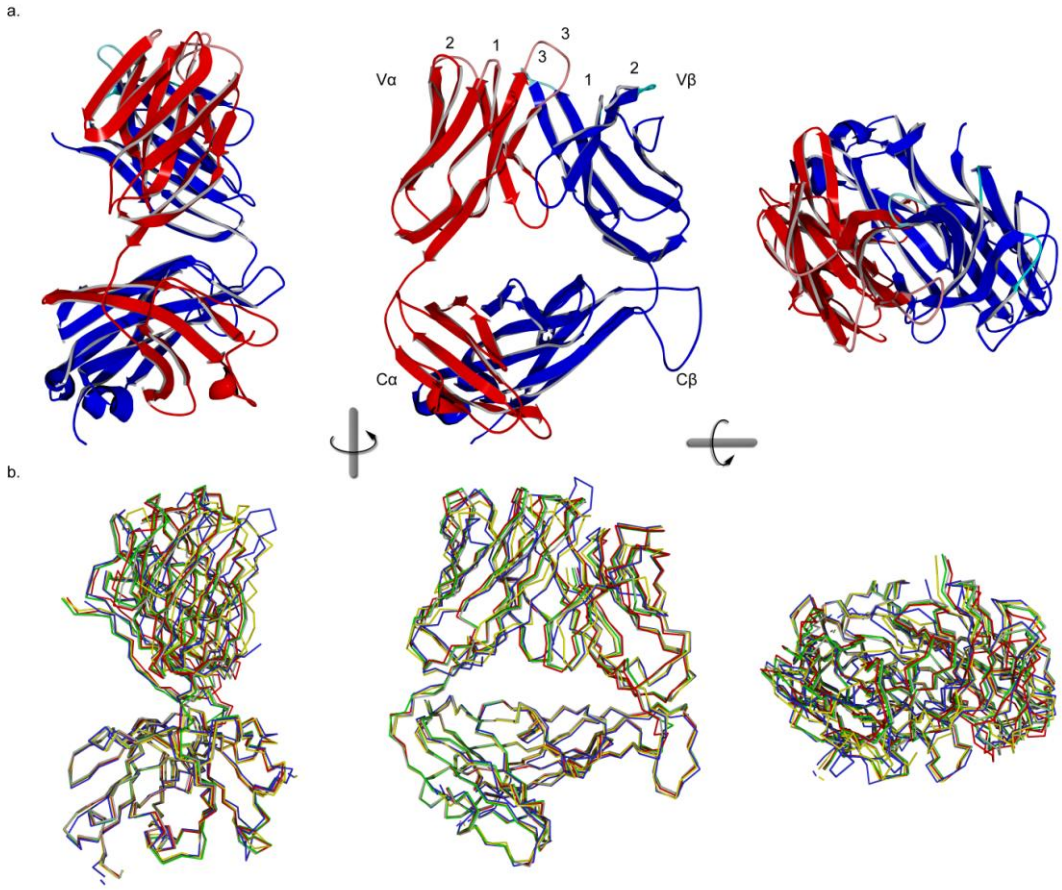


Figure 3.2.6 Structure of The 5E, 5B and NKT T cell receptors
a. The structure of 5E TCR presented as a secondary structure cartoon. The four protein domains are labelled and the CDR loops are shown in light red and light blue and numbered.
b. Alignment of the 5E TCR, 5B and three copies of the *i*NKT receptor, shown as a c-alpha trace. 5E - yellow, 5B - blue, *i*NKT1 - green, *i*NKT2 - red, and *i*NKT3 - grey. The three proteins all share a beta chain and thus this region has the highest level of structural similarity, with the variable portion of the alpha chain showing most difference.

The overall architecture of all three proteins is similar and conforms to the expected TCR paradigm. In the case of the *i*NKTCR structures, the three molecules in the crystallographic asymmetric unit align with an rmsd of 0.5 Å over all atoms and for the purposes of further analysis the first copy of the NKT receptor will be taken as the representative structure. The CDR loop regions of both the 5E TCR (Figure 3.2.7) and 5B TCR (Figure 3.2.8) regions presented weak electron density initially, but with the calculation of OMIT maps centred on these, each loop could be unambiguously traced (Figure 3.2.10). The B-factors of these loops are around of 50 $Å^3$ and no negative density was seen in the F_o-F_c electron density maps to imply that the loops were incorrectly traced. With the NCS averaging used in the calculation for *i*NKTCR electron density maps the loops were seen quite clearly (Figure 3.2.9).

3.2.4 CDR loop architecture

In an analysis of the hypervariable CDR1, 2 and 3 loop regions of a number of crystal structures and sequences of both human and murine T cell receptors Chothia *et al* [218] assigned the architecture of the CDR loops to a number of canonical forms. These, they suggested, would be found in subsequent T cell receptor structures (see Chapter 3.A for a description of these).

The CDR1α loop of 5E TCR conforms most closely to the α1-1 canonical group, with the S26-Oγ forming a hydrogen bond to the backbone nitrogen of S28 and packing against the edge of F30 (Figure 3.2.7). In 5B TCR the loop has a similar path to 5E TCR, with the edge of Y26 packing against the side-chains of I30 and L33 (Figure 3.2.8). In the three NCS related *i*NKTCR alpha chains the α1 loops are all nearly identical with Y25 packing against the F30 and L33 side-chains, the path of this loop

conforms to the α1-2 designation, although Y25 does not interact with P29 (Figure 3.2.9)

The CDR2α loop of 5E TCR adopts a type II turn conformation, characteristic of α2-4 with a main chain hydrogen bond between the carbonyl oxygen of T52 and the nitrogen of E54 (Figure 3.2.4). 5B TCR has a α2-2 class loop, with the loop packing around the I50 side-chain (Figure 3.2.8). The *i*NKTCR CDR α2 loops align perfectly up to S53 with residues following this deviating slightly for each related molecule from this point, (Figure 3.2.9) like 5B TCR the *i*NKTCR loop has a canonical α2-2 structure with the side-chain of M50 making extensive Van der Waals contacts with the K57 side-chain.

Because of the diversity of CDR3 loops, no canonical structures have been described for these. The sequences of the CDR3 loops of all three receptors are almost identical and the differences seen in their structures represent some of the flexibility that is inherent in this region. In the 5E TCR structure a main chain hydrogen bond between D94 and S97 stabilises the observed conformation of the loop (Figure 3.2.10). The *i*NKTCR and 5B TCR CDR3α loops are somewhat more extended than the 5E TCR loop, with the three *i*NKTCRs having almost identical loop structures.

All three proteins share a Vβ chain, with the J region between the variable and constant domains different in each. Consequently, the structures of the CDR β1 and β2 loops are identical in all three structures and the CDR β3 loops are different for each.

The CDR1β loop (Figures 3.2.7/8/9) is a β1-1 canonical loop, with Oε1 and Nε2 of Q26 forming hydrogen bonds with the main chain Nitrogen M28 and carbonyl oxygen of H30 respectively. Histidine 30 is also key in this loop, as the Nδ1 forms a hydrogen bond to the S94 carbonyl oxygen at the beginning of the CDR β3 loop [219]. The β2 loop is of the form β2-1, with the carbonyl oxygen of S50 forming hydrogen bonds to the main chain nitrogen of G52. Arginine 69 packs across the loop and can form a hydrogen bond with the main chain of G52. The CDR2β loops from all structures align with no main-chain differences and only minor side-chain movements (Figures 3.2.7/8/9)

In each structure, the CDR3β loop has a different sequence and consequently their structures vary widely. The 5E TCR has an extended CDR3β loop, with the side-chains from all residues pointing out of the loop, with a number of main-chain hydrogen bonds stabilising this conformation (Figure 3.2.10). In the structure of TCR 5B, the CDR3β loop is less extended with more interactions between side-chains than in TCR5E. Finally, the CDR3β loop of the NKT structure adopts an extended conformation, with main chain hydrogen bonds between the nitrogen of G99 and the carbonyl oxygen of A101; and the nitrogen of E96 and the carbonyl oxygen of Y102. There are no differences in the loops between the three molecules within the asymmetric unit.

3.A Canonical structures in the TCR Complementarity Determining Regions

Chothia *et al* [218] described canonical structures for the CDR loops of TCRs based on a detailed analysis of published TCR structures and described the structural features that give rise to particular loop conformations. These canonical structures are briefly described here and used in an analysis of the human T cell receptors that currently exist in the protein databank and those that are reported in this chapter.

The 1α hypervariable regions exist in one of three conformations:

α1-1. This region is nine residues long, with Ser/Phe29 pointing towards the centre of the loop and packing against the framework residues Tyr24 and Phe/Leu32. Residues from the α3 region are also found to pack against these residues.

α1-2. Again, this loop is nine residues long with Pro30 pointing into the centre of the loop and packing against Tyr24 and Phe/Leu32.

α1-3. Ten residues long with Thr26 and Thr30 pointing into the loop and forming close contacts with potential hydrogen bonding between the two. There are extensive side chain contacts with Tyr24 and Phe32.

The 2α hypervariable regions exist in one of four conformations:

α2-1. This region is five or six residues long with an extended main-chain conformation and side chains pointing towards the solvent.

α2-2. Four loop residues pack around a medium sized hydrophobic residue, Ile49, and pack against Phe32 and Phe66.

α2-3. Tyrosine 49 acts as the centre for the packing of four loop residues with interactions with Leu32 and Phe66.

α2-4. Four residues fold around the small side chain of Ala49 and pack against Leu32 and Tyr66. The fold of this loop gives a main-chain hydrogen bond between residues 51 and 54.

The 1β hypervariable regions exist in one of three conformations:

β1-1. In this eight residue loop, His29 and Gln25 face toward the centre of the loop, His29 forms a hydrogen bond to Ser94 while Gnl25 forms hydrogen bonds with the main-chain atoms of residues 27 and 29.

β1-2. This loop has four key residues and is eight residues long. Pro25 and Val32 pack against His29, which, like in β1-1 forms a hydrogen bond with ser94.

β1-3. At nine residues long, this loop has four key residues. Pro30 forms a hydrogen bond to Arg69 with the loop structure stabilised by Leu25 packing against Met32.

The 2β hypervariable regions exist in one of three conformations:

β2-1. In this loop, Ser49 forms hydrogen bonds to the main chain atoms of residues 51 and 51. Arg69 packs across the loop and hydrogen bonds to residue 53.

β2-2. Main chain hydrogen bonds between residues 50 and 54 stabilise this loop, with His49 packing inside the loop next to Gnl69.

β2-3. Four residues, 50 to 53, form the core of the loop with hydrogen bonding between residues 51 and 54.

The α3 and β3 regions are much more variable then the other CDR loops and thus no canonical structures for these have been described.

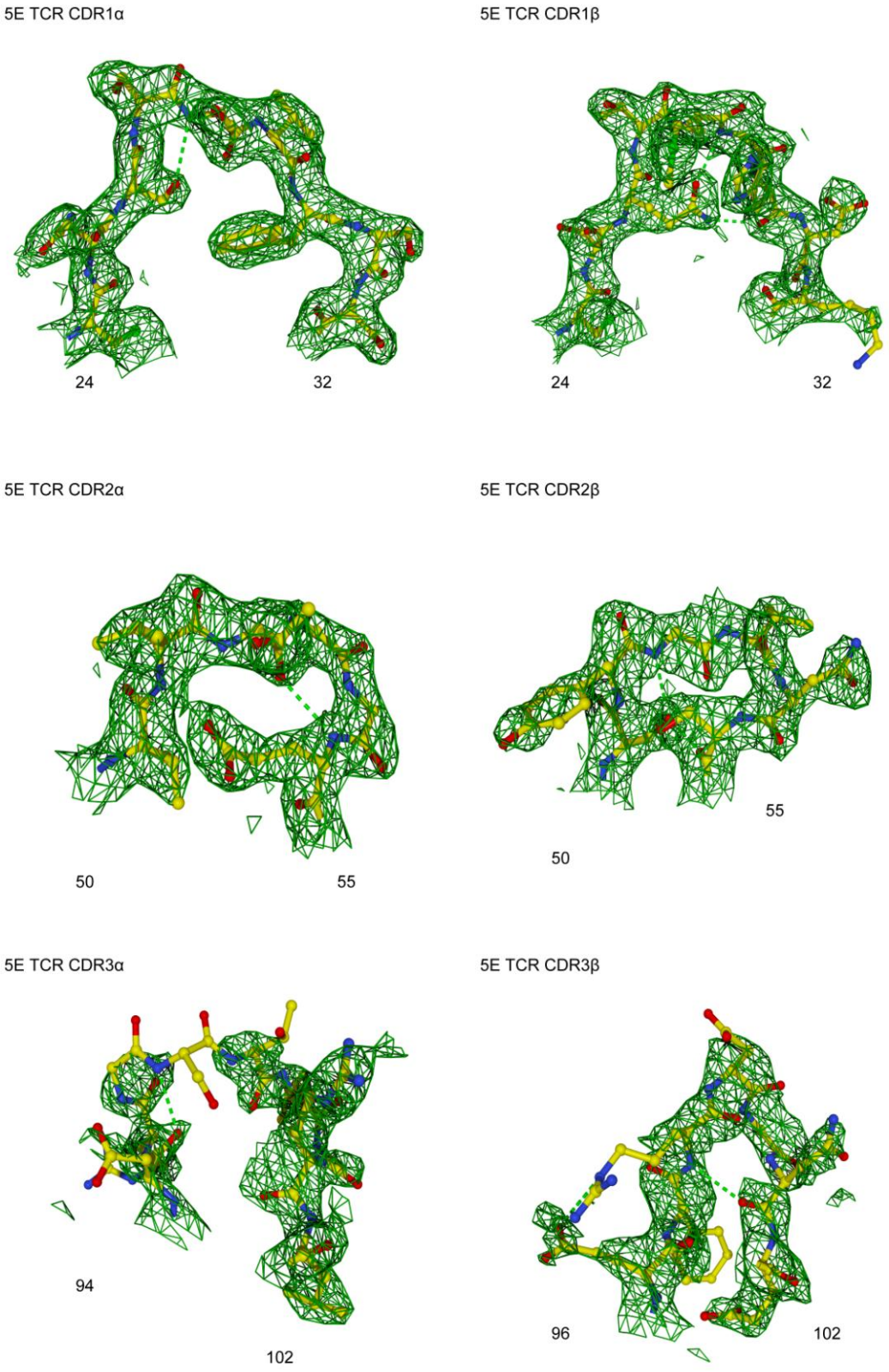


Figure 3.2.7 $2F_o$-F_c Electron density of 5E TCR CDR loops
The structure of the 5E TCR CDR loops and the $2F_o$-F_c electron density for these are shown to illustrate the quality of electron density seen and the canonical structure of these.
The protein is shown as ball and stick representation with carbon atoms in yellow, nitrogen (blue), and Oxygen (red). The $2F_o$-F_c electron density is contoured at 1σ and shown as green wire-frame.

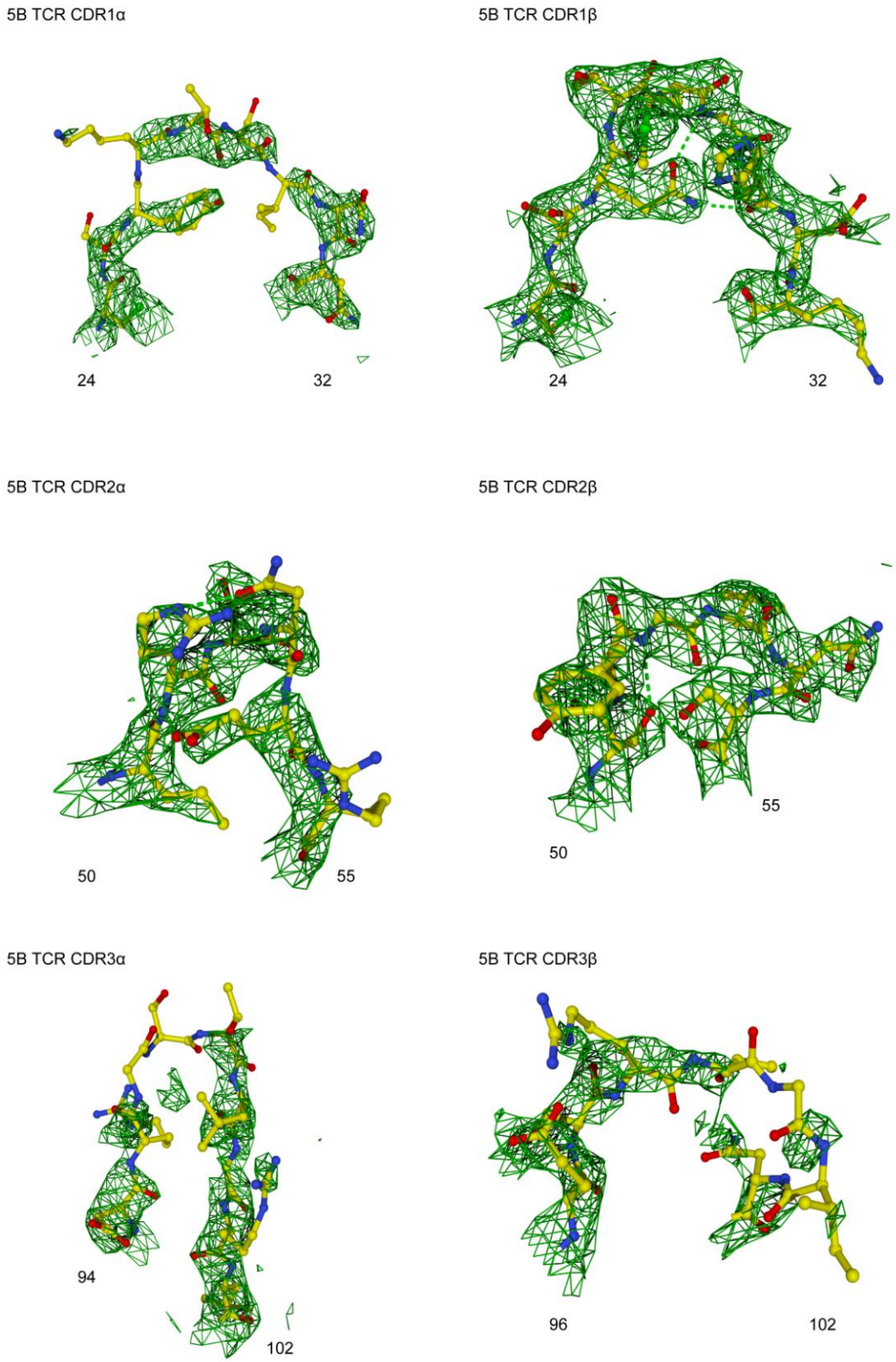


Figure 3.2.8 $2F_o$-F_c Electron density of 5B TCR CDR loops
The structure of the 5B TCR CDR loops and the $2F_o$-F_c electron density for these are shown to illustrate the quality of electron density seen and the canonical structure of these.
The protein is shown as ball and stick representation with carbon atoms in yellow, nitrogen (blue), and Oxygen (red). The $2F_o$-F_c electron density is contoured at 1σ and shown as green wire-frame.

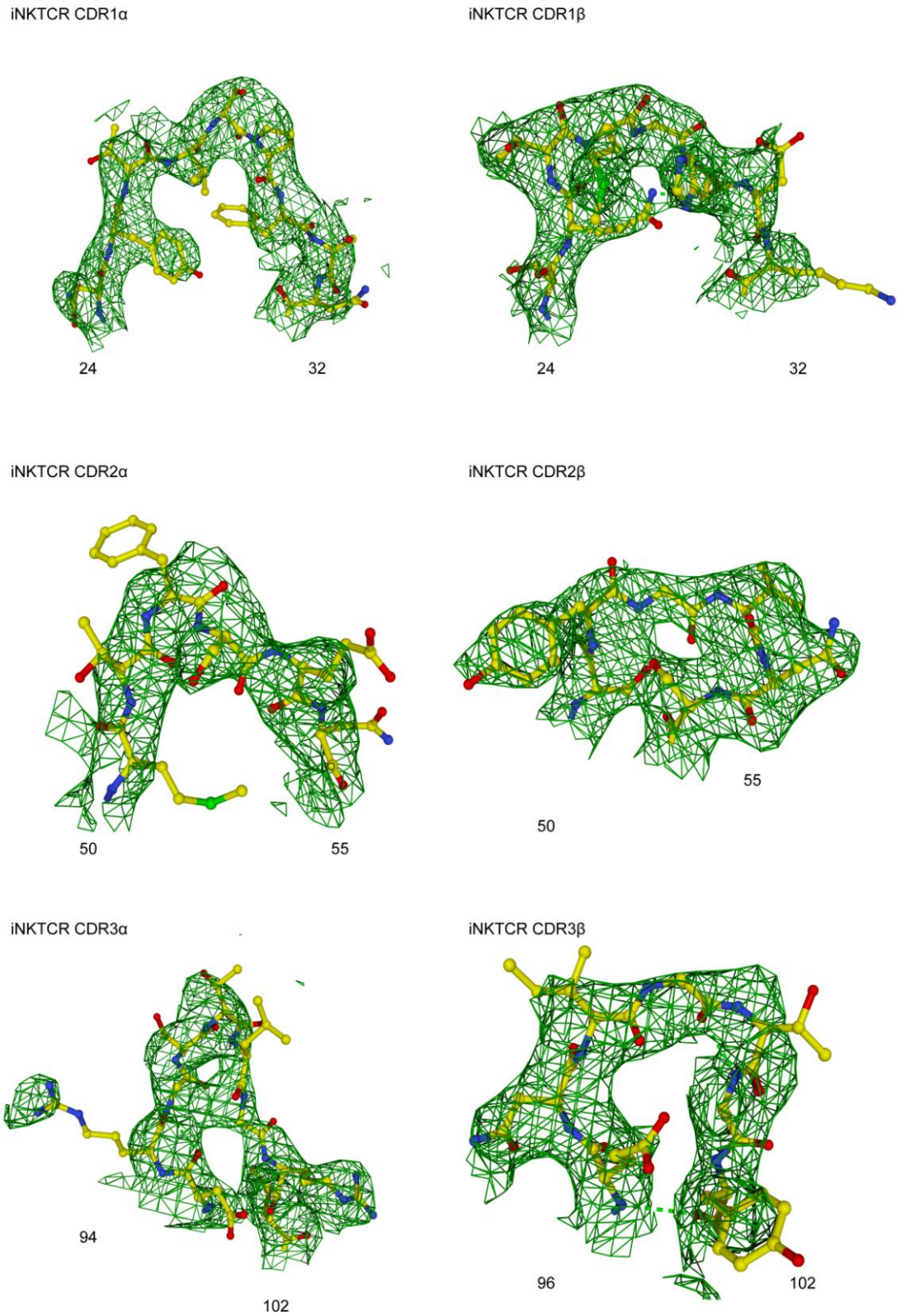


Figure 3.2.9 $2F_o$-F_c Electron density of TCR NKT CDR loops
The structure of the TCR NKT CDR loops and the $2F_o$-F_c electron density for these are shown to illustrate the quality of electron density seen and the canonical structure of these. The protein is shown as ball and stick representation with carbon atoms in yellow, nitrogen (blue), and Oxygen (red). The $2F_o$-F_c electron density is contoured at 1σ and shown as green wire-frame.

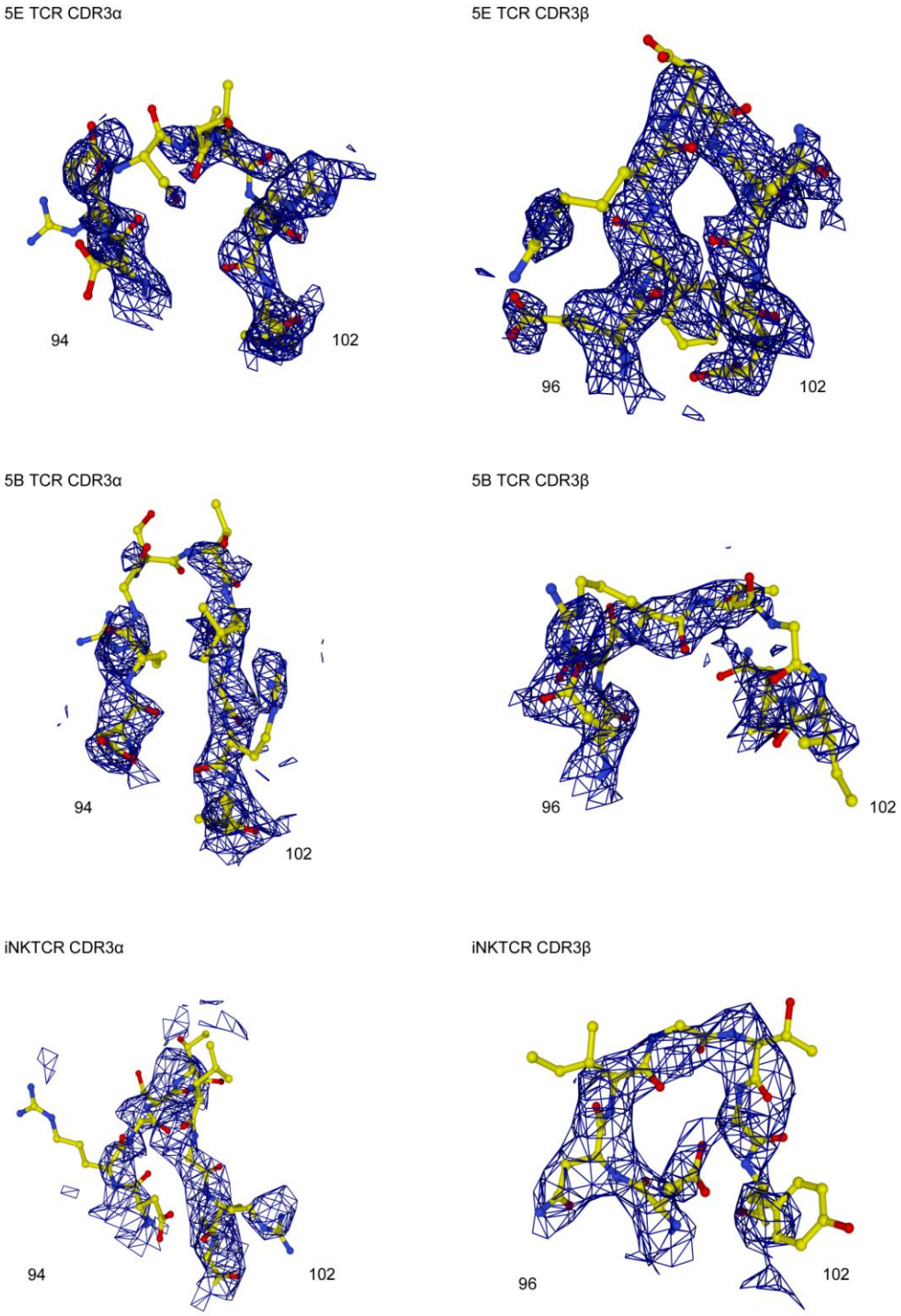


Figure 3.2.10 Omit maps for CDR3 regions of The NKT cell receptors
The structure of the CDR3 loops from the three NKTCR structures are shown with electron density from simulated annealing OMIT maps calculated with these loop regions excluded. The protein is shown as ball and stick representation with carbon atoms in yellow, nitrogen (blue), and Oxygen (red). The OMIT map is contoured at 1σ and shown as blue wire-frame.

3.2.5 Structural comparison of human T cell receptors

The structures of nine human αβ TCRs have been solved prior to the three NKTCRs reported, seven of which were solved as complexes with MHC/peptide (Appendix TCR and references therein). A structural alignment of these molecules gives an impression of the diversity seen in these structures (Figure 3.2.11).

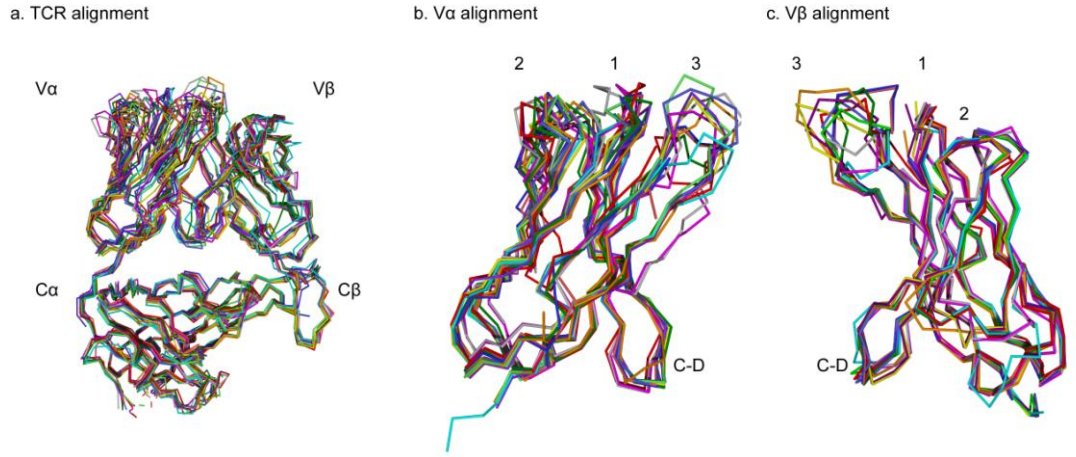


Figure 3.2.11 Structural alignment of human T cell receptors
Human T cell receptors were aligned using SHP to assess structural variation in the peptide chains. 5E TCR is shown in yellow, 5B TCR in orange, TCR NKT in lilac and the following other T cell receptor chains are shown in their respective colours: 1A07, blue ; 1BD2, green ;1J8H, red; 1KGC, grey; 1MI5, pink ; 1OGA, cyan; 1QRN, flesh; 2BNQ, indigo; 2BNU, light green (references for the TCR structures are in Appendix TCR).

a. Structural alignment of the full αβ TCR models.
b. Alignment of the Vα domains, the major diversity is found in the CDR and C–D loops, labelled.
c. Vβ alignment showing more structural conservation than in the Vα regions.

The structures for all of the solved T cell receptor constant domains are remarkably similar, with an average RMSD over the whole protein of 2 Å (Figure 3.2.11a). In all structures reported so far and for the three NKTCRs, the Cα domain has the weakest electron density and highest B-factors for the protein outside of the CDR loops [140].

An alignment based on the variable domains of the proteins shows clearly that the main source of variation between these structures is located in the CDR loops, and the

loop between the C and D strands (C–D loop) points to the gap where the variable and constant domains of both chains meet (Figure 3.2.11b/c). There appear to be no major domain movements observed between TCR structures on their own and in complex with MHC, so it has been argued that the mobility seen in the C–D loop does not represent some rearrangement of chains on ligand binding [140]. The likely explanation for this variation seen is that the quaternary packing of the Vα against the Vβ domain, which show rotational differences of between 3 and 14° [140], affects this flexible region of the protein. The Vα domains align with an average rmsd of 1.8 Å, while the Vβ regions have a slightly lower average rmsd of 1.7 Å.

The TCR combining region, made up of the CDR loops, shows the most variation between published structures, with the Vα loops showing the most diversity between the structures (Figure 3.2.12). In the case of the three NK TCRs their shared CDRα3 regions adopt three distinct conformations, dependent on their crystallographic environment.

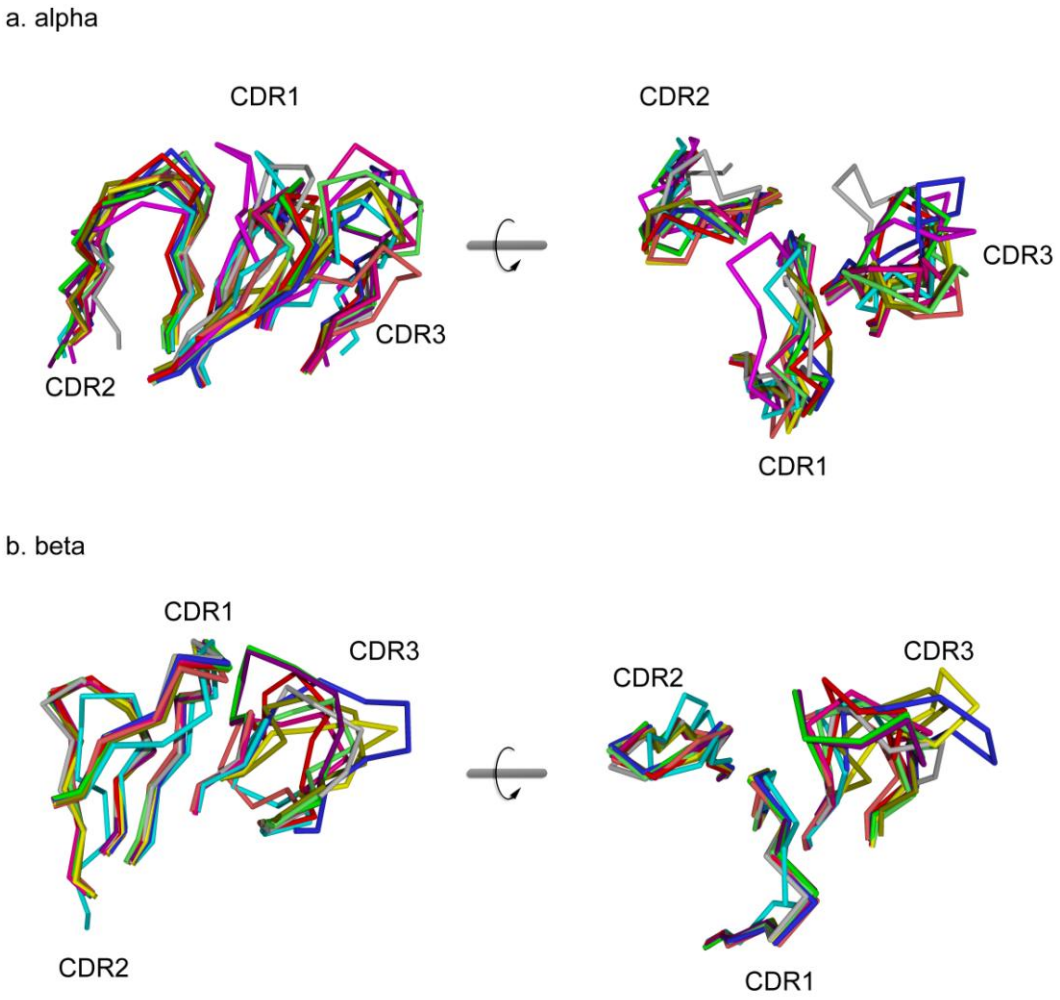


Figure 3.2.12 Alignment of the structures of human T cell receptor CDR loops
The CDR loops of the TCR structures from Figure 3.2.8 are shown with the other protein areas removed for clarity. Protein chains are coloured as before.

All four canonical CDR1α loop architectures are represented in the human T cell receptor structures, with α1-3 (Y24-F32 interaction) the most common (Figure 3.2.13). For the CDR2α loops all four types are present, with the α2-2 (I49 packing against F32/66) variation the most common (Figure 3.2.13). Only two Vβ chain architectures are presented by the structures, with the β1-1/β2-1 paring found in all but two structures, which display β1-2/β2-1 pairing (Figure 3.2.14). The CDR3 loops display high variation between all of the structures. It can be seen that in the TCR/MHC structures the CDR1 and 2 loops are predominantly in contact with the MHC molecule and the CDR3 loops typically make direct contact with the presented antigen [184,220,221]. It follows from this that the majority of variation seen would naturally be in the CDR3 loops to allow for the recognition of a wide variety of presented ligand and this is clearly seen in the diversity seen in these loop regions (Figure 3.2.12/13/14).

TCR	CDR α1	CDR α2	CDR β1	CDR β2
5E	1	4	1	1
5B	1	2	1	1
NKT	2	2	1	1
1A07	3	1	1	1
1BD2	3	2	1	1
1J8H	2	3	1	1
1KGC	2	1	2	2
1MI5	2	1	2	2
1OGA	1	4	1	1
1QRN	3	1	1	1
2BNQ	3	2	1	1
2BNU	3	2	1	1

Table 3.2.2 CDR loop architecture of human T cell receptors
The canonical forms, as described by Chothia *et al,* adopted by the CDR loops of human T cell receptors for which a structure is available. These are described in Section 3.A and illustrated in figures 3.2.12-15.

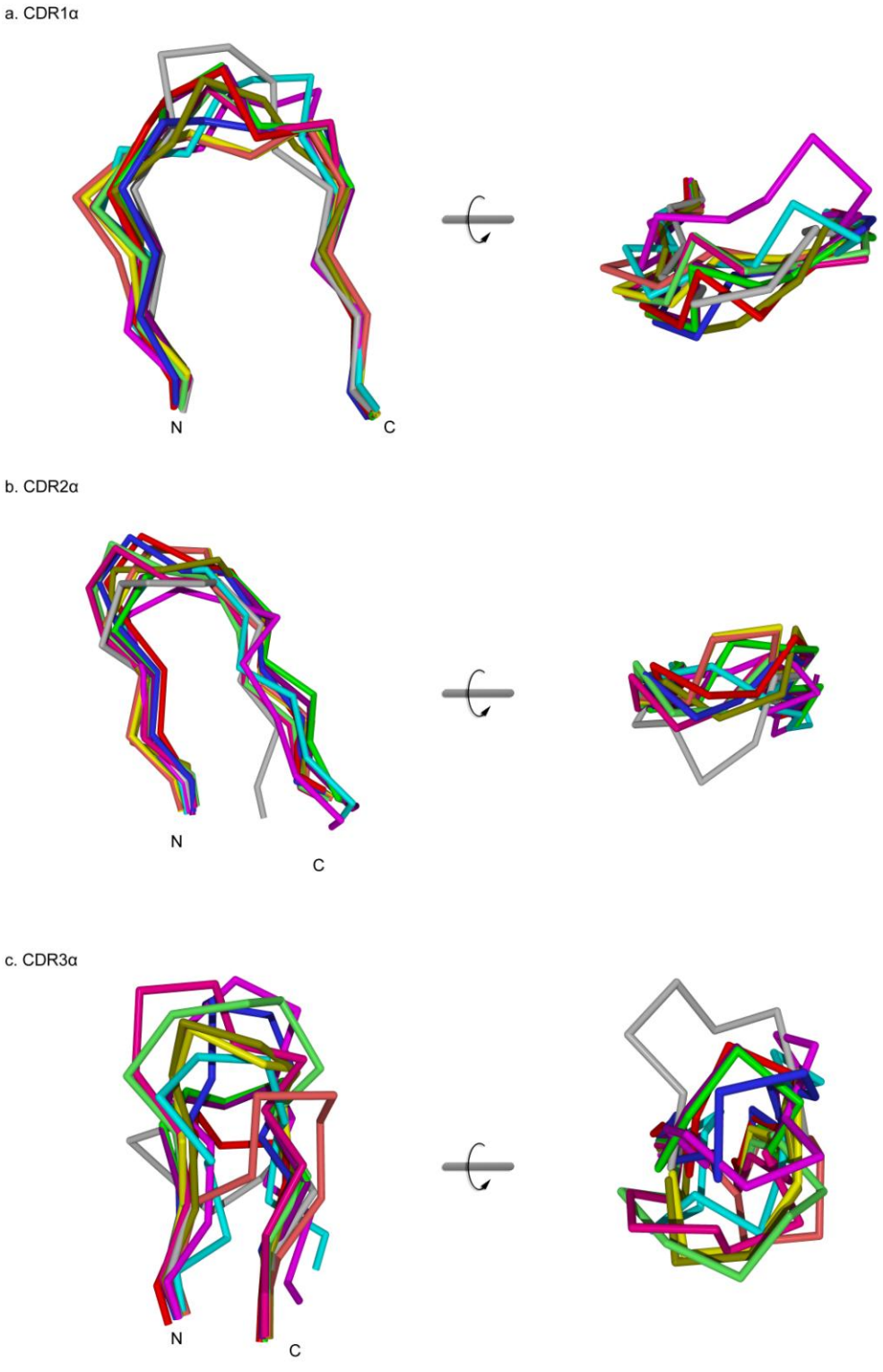


Figure 3.2.13 Alignment of T cell receptor CDR alpha regions
The CDR loops of the TCR structures from Figure 3.2.8 are shown in isolation. Protein chains are coloured as before. The CDR3α loops show the most structural diversity between TCRs.

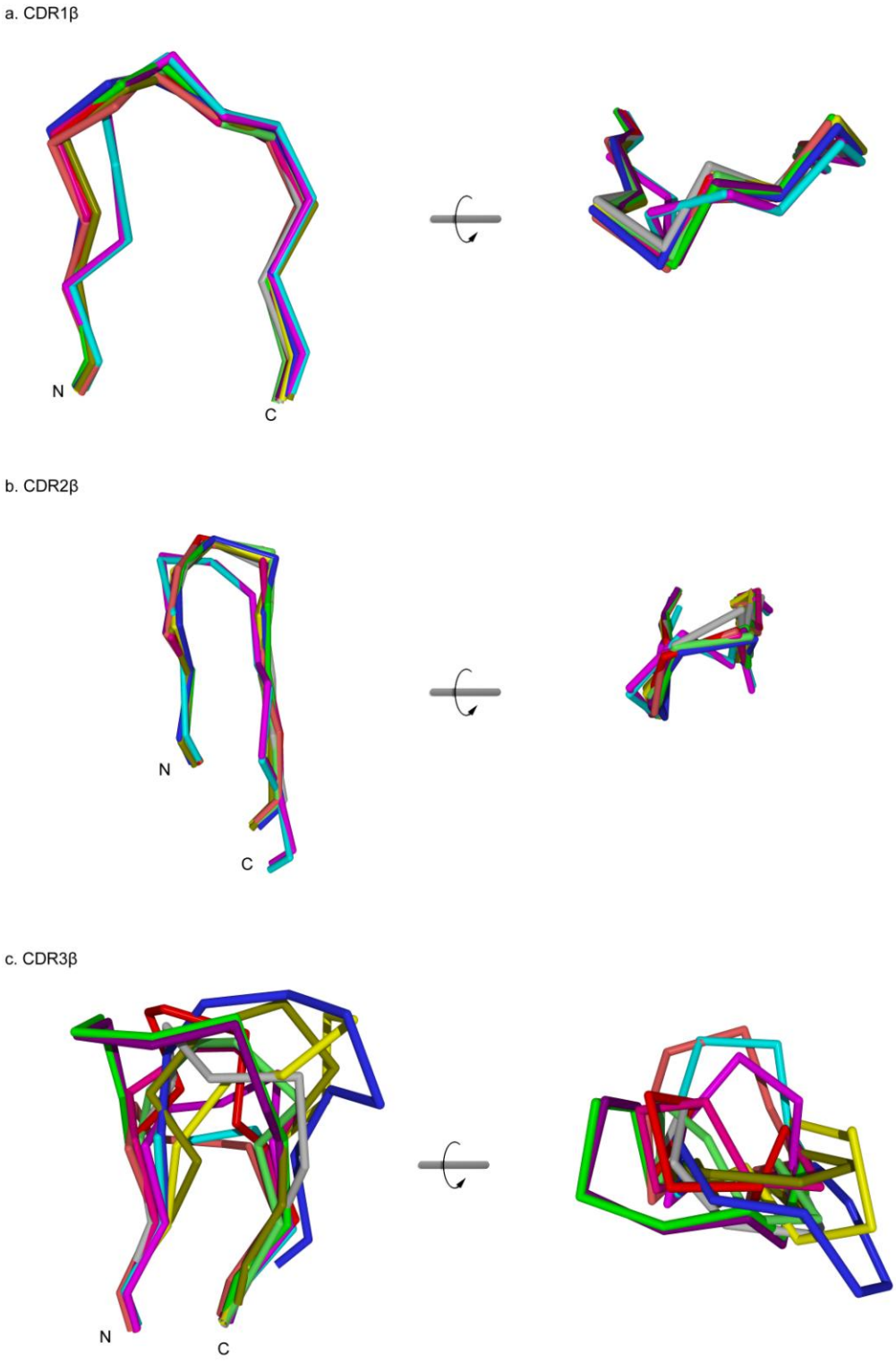


Figure 3.2.14 Alignment of T cell receptor CDR beta regions
The CDR loops of the TCR structures from Figure 3.2.8 are shown in isolation. Protein chains are coloured as before. Both the CDR1β and 2β loops exhibit very few structural differences in the main-chain conformation. The CDR3β loops show the most diversity between the structures.

3.2.6 CD1d recognition surface

The fact that the CDR loops of the NKTCRs conform to the canonical main-chain conformations seen for class I MHC recognising TCRs implies that the engagement of CD1d/lipid by these proteins would occur in a similar fashion to that of pMHC complexes. With the CDR loops making contact with the α1 and α2 helices of the antigen binding site and the CDR3 loops making the most intimate contact with the antigen [222].

The CD1d recognition surface of the three proteins is quite similar for 5E TCR and *i*NKTCR, with 5B TCR having more positively charged residues on its surface (Figure 3.2.12). All structures exhibit a surface cavity (calculated using Surfnet as described in Chapter 2.2.1) between the CDR 3 loops of the alpha and beta chains, which may accommodate the interaction with the sugar head group of α-GalCer (Figure 3.2.15 green wire-frame) and the extended sugar head of the ligand iGb3.

A scheme for the binding of CD1d to the NKTCRs has been previously suggested [197] (shown as shaded triangles in Figure 3.2.15). Specific residues involved in TCR recognition on the Cd1d binding surface have been identified by mutagenesis. The mutation of R79 to glutamic acid and D80 to alanine (and in some cases S76 to glycine) all serve to abrogate TCR binding, because these residues reside outside of the α-GalCer binding pocket they most likely make contacts with the TCRβ chain. The CDR3α loop of the T cell receptors may make some direct contact to the α-GalCer molecule; the conserved R96 in this loop is suggested to be a likely candidate for interactions with the polar-head-group of the Cd1d ligand.

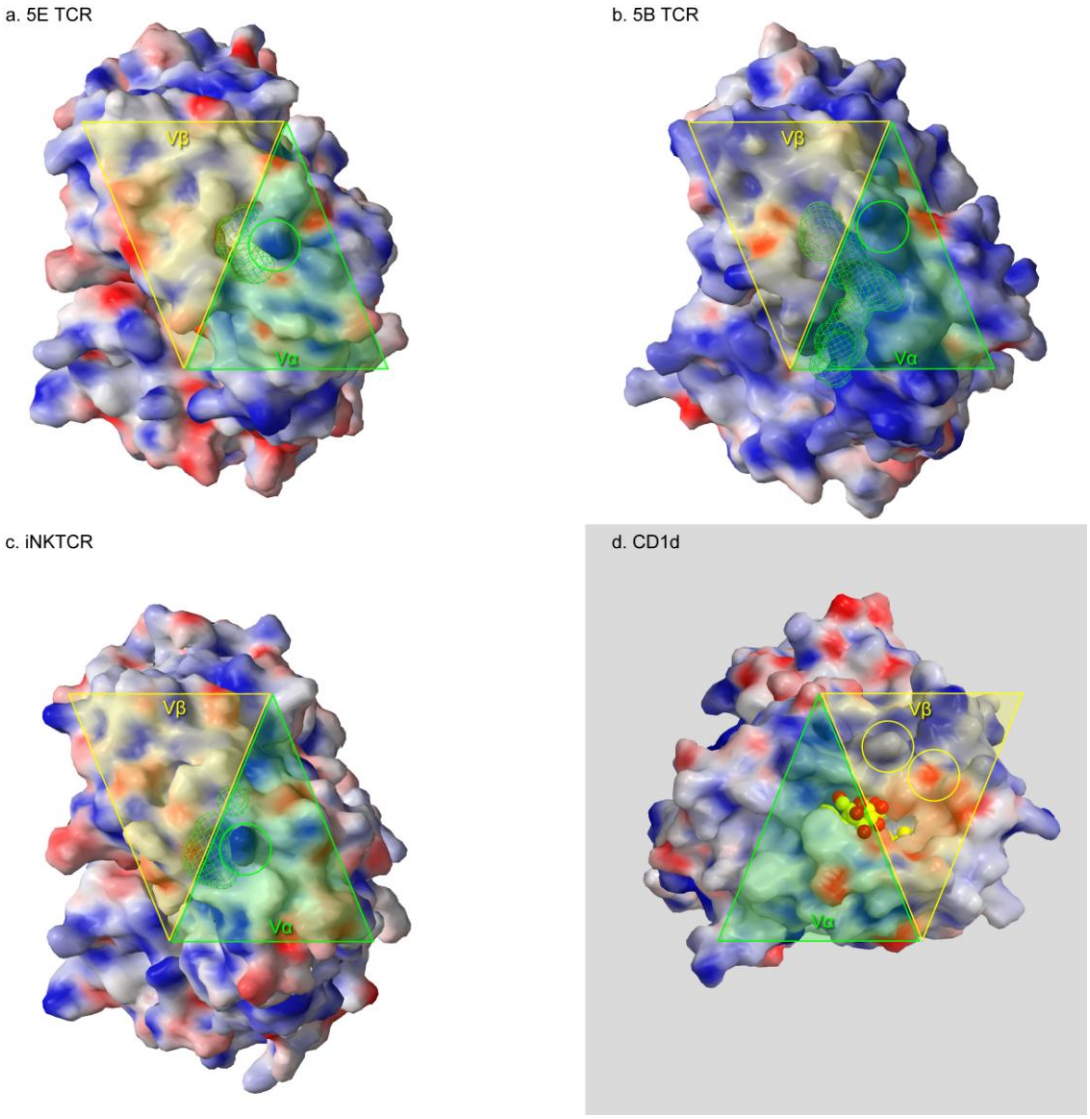


Figure 3.2.15 CD1 recognition surfaces of T cell receptors
The CD1 recognition surfaces of the three NKTCRs and the ligand-binding surface of CD1d. The molecular surface of each receptor is shaded in with +10 kT in blue and -10 kT in red, surface cavities between CDR3 loops were calculated using Surfnet and are shown in green wire-frame. The structure of α-GalCer in CD1d is shown in space-filling representation with carbon coloured yellow and oxygen in red. The proposed model of interaction between the TCRs and CD1d is shown with shaded triangles corresponding to the orientation of the Vα (green) and Vβ (yellow) on the surface of CD1d. Residues thought be involved in TCR/CD1d complex formation are circled, R96 on the surface of the TCRs and R79/D80 on the surface of CD1d.

3.3 Conclusions

3.3.1 NKTCR Structures

The structures of three CD1d binding NK TCRs are presented here. The structure of the conserved regions of all three proteins are remarkably similar, as would be expected from the sequence conservation and the CDR loop architecture for the α1/2 and β1/2 loops conforms closely to that described previously [218] even though the structures were not refined on this basis. This demonstrates a solid argument for the basis of these canonical structures in reality and the fact that CD1 recognising TCRs are not fundamentally different from those that bind MHC presented antigen.

3.3.2 CD1d binding mechanism

The CD1d binding mechanism suggested by the structures of the NK TCRs and previous biochemical data [212] has the TCR chains binding in a similar fashion to the binding of peptide-MHC, with the protein interacting with CD1d across the antigen binding site. The flexibility seen in the CDR3 loops between the three structures in the absence of ligand interaction is quite striking, this movement may allow for a clamping action upon engagement of CD1d presented antigen and the recognition of molecules with multiple sugar groups. The three TCRs may show differences in affinity for CD1d because of the different residues and resulting charge differences seen in the ligand-binding surface of the receptors. From the visualisation of the CD1d binding surfaces of the three receptors the *i*NKTCR and 5E TCR will most likely show a similar binding affinity for their ligand, with 5B TCR showing a markedly different affinity.

3.3.3 Future perspectives

Finally, the structure of the TCR/CD1d/antigen complex, when it is solved, will show the degree of interaction between the CDR3 loops and CD1d and answer questions of how CD1d specificity is achieved by related TCRs. The structures presented here are the first CD1d binding TCRs solved and their structures imply that the engagement of CD1d by TCRs will not be markedly different from the general scheme of MHC binding. Because there are so few different CD1 isoforms and the fact that only a very small antigenic surface is accessible, the orientation of the TCR complex over the antigen presentation surface of CD1 may show less variation than has been found for TCR/pMHC complexes. The limited potential for interaction between the TCR and the solvent exposed portion of antigen implies that in the case of CD1d/antigen recognition the CDR3 loops will probably be the only regions of the TCR to make direct contact with the ligand, with other CDR loops being specific to particular CD1 isoforms.

Summary of Thesis

The OPPF high-throughput protein production pipeline was successfully applied to the production of a number of recombinant expression products for the TNF and TNFR superfamilies. The rapid dilution method of protein refolding was applied to those targets that produced inclusion bodies in *E.coli*, this protocol was successful for both the TNF ligand target TRAIL and DR6 and may be applied as a protocol of choice for the production of the cysteine-rich receptor molecules for structural studies.

The refolded DR6 protein consistently produced crystals in a number of crystallization conditions. These crystals were used for x-ray crystallographic studies, but the solution of the model structure for DR6 remained elusive due to limited resolution of the data and lack of amenability of this protein to both the molecular replacement and sulphur SAD phasing techniques.

A single TNF ligand-like target, VEGI, produced soluble protein, which when purified and set up in crystallization trials produced a number of crystals. These crystals gave extremely anisotropic diffraction, but serve to illuminate the potential for future optimisation of crystallization conditions and the dehydration of crystals before x-ray data collection.

The expression of the LTβR protein construct in a eukaryotic system serves to provide an alternative pathway to the production of TNF receptor-like molecules in a prokaryotic system and the refolding that has been necessary in that system. Overall yieds of this protein were low and a number of problems with peptide cleavage were

encountered, but a number of crystallization trials have shown that the protein does form crystals in a number of related crystallization conditions.

The potential for small molecule inhibition of receptor binding to TNF ligand superfamily members has been explored using the structures of a number of TNF ligand/receptor pairs. In a medical context, TNF-targeting therapies are of great interest, but currently rest on the use of recombinant proteins. The application of a number of strategies to the design, or discovery, of small molecules that bind to specific TNF ligands has been explored, with the *a priori* molecule creation of Ligbuilder compared to the methods of database mining and automated docking. Database mining, when coupled with automated docking techniques, seems to present the most conceptually satisfying and potentially productive method for the discovery of drug-leads for the TNF ligand superfamily of molecules. *A priori* design methods are still in their infancy and the implementation of Ligbuilder seems more suited to the rational optimization of drug targets than the initial design of them.

The solution of the crystal structures of three NK restricted T cell receptors that recognise CD1d, has shed light on the processes of CD1d presented lipid antigen recognition by these proteins and has shown that the engagement of CD1d by TCR proteins is most likely very similar to that of MHC recognition. The fact that the CDR loops of these TCR proteins conform to the canonical structures described for these regions is further strong evidence of the conservation of these structural characteristics across the different TCR proteins. Structural features of the CD1d receptor-binding region of the three NK TCRs imply that they will show differential binding affinities for CD1d.

Appendices

Appendix A: The TNF ligand superfamily

Ligand	Receptor partners	Tissue Distribution	Functions	Reference
Lymphotoxin α	TNFR1/2, HVEM, LTβR	Secreted by lymphocytes	Cytotoxic to tumours	[223]
TNFα	TNFR1/2	Secreted by macrophages	Induction of inflammation and cell death	[224]
Lymphotoxin β	LTβR	Macrophages	Cytotoxic and as cell anchor for LTα	[225]
OX40 Ligand	OX40	Activated T-cells	T-cell proliferation and cytokine production	[226]
CD40 Ligand	CD40	T-cells	B-cell proliferation and IgE secretion	[227]
FAS Ligand	FAS	Lymphoid tissue, testes	Peripheral tolerance	[228]
CD27 Ligand	CD27	B-cell cells	T-cell proliferation	[229]
CD30 Ligand	CD30	Activated T-cells	Negative selection of T-cells	[230]
4-1 BB Ligand	CD137	Activated T-cells	Proliferation of active T-cells	[231]
TRAIL	DR5, DR4, Dc-R1, Dc-R2	Wide, predominant in spleen, lung, prostate	Apoptosis of cancer cells	[232]
RANKL	RANK, OPG	Peripheral lymph nodes	Activation of osteoclasts, mediator of t-cell proliferation	[47]
TWEAK	DR3	Muscle tissue, peripheral lymphocytes	Angiogenesis and epithelial proliferation	[233]
APRIL	BCMA, TACI	Monocytes and macrophages	Regulation of tumour cell growth, macrophage proliferation	[234]
BAFF	BCMA, TACI, BAFFR	Peripheral blood leukocytes and T-cells	B-cell maturation	[235]
LIGHT	HVEM	Spleen, brain	T-cell proliferation	[236]
VEGI	Not known	Endothelial cells	Inhibits vascular endothelial growth	[74]
EDA-A1	EDAR	Ectodermal cells	Mesenchymal signalling during ectodermal development	[237]
EDA-A2	XEDAR			
GITRL	GITR	Endothelial cells	Peripheral T-cell survival	[238]

Appendix B: The TNF receptor superfamily

Receptor	Number of CRDs	Death domain	Ligand partner	Tissue Distribution	Functions	Reference
TNFR1	4	Yes	TNFα, LTα	Hematopoetic cells	Apoptisis and peyer's patch formation	[239]
TNFR2	4	No	TNFα, LTα	Hematopoetic cells, myeloid cells	Inflammatory apoptosis	[103]
LTβR	4	No	LTβ	Lymphoid tissues	Apoptosis, Lymphoid organ development	[240]
OX40	4	No	OX40L	Activated T-cells	B-cell proliferation, T-cell co-stimulation	[241]
CD40	4	No	CD40L	B-cells, macrophages, dendritic cells	B-cell activation and cytokine production	[242]
FAS	3	Yes	FASL	Wide spread	Apoptosis, peripheral tolerance	[243]
DcR3	4	No	FASL	Wide spread – over expressed in colon cancer	FAS decoy	[27]
CD27	3	No	CD27L	T-cells, NK-Cells, memory B-cells	T-cell proliferation and co-stimulation	[244]
CD30	2 x 3	No	CD30L	T-cells, activated B-cells, Lymph nodes	T and B-cell proliferation	[245]
CD137	3	No	4-1BBL	Activated T-cells	T-cell activation and differentiation	[231]
DR4	3	Yes	TRAIL	Widespread, spleen, peripheral leukocytes	Apoptosis	[246]
DR5	3	Yes	TRAIL	Widespread, over expressed in tumour cell lines	Apoptosis	[247]
DcR1	3	No	TRAIL	Lymphocytes, spleen, skeletal muscle	TRAIL decoy	[248]
DcR2	3	Non-functional	TRAIL	Widespread	TRAIL decoy	[249]
RANK	4	No	RANKL	Widespread	Osteoclastogenesis, T-cell stimulation	[47]

Receptor	Number of CRDs	Death domain	Ligand partner	Tissue Distribution	Functions	Reference
OPG	4	Yes - 2	RANKL	Widespread, high in bone marrow and foetal tissues	RANKL decoy – mediates osteoclastogenesis	[250]
TACI	2	No	APRIL, BAFF	Spleen, leukocytes, resting B-cells, activated T-cells	T and B-cell stimulation, regulation of humoral immunity	[251]
BAFFR	1	No	BAFF	Spleen, lymph nodes, resting B-cells	Promotes B-cell survival	[41]
HVEM	3	No	LIGHT, HSV glycoprotein D, LTα	Widespread	Lymphocyte activation Mediates HSV-1 entry	[252]
NGFR	4	Yes	Nerve Growth Factor	Bone marrow stroma, follicular dendritic cells	Peripheral sensory nervous development	[253]
BCMA	1	No	APRIL, BAFF	Mature B-cells	Promotes B-cell survival	[254]
GITR	3	No	GITR	Lymph nodes, activated T-cells	Modulation of T-cell survival in peripheral tissues	[238]
TROY	3	No	Not known	Prostate, peripheral blood leukocytes	Mediates caspase independent cell death	[255]
DR6	4	Yes	Not known	Heart, brain, skeletal muscle	NF-κb activation	[256]
EDAR	3	No	EDA-A1	Ectodermal cells	Promotes caspase independent cell death	[257]
XEDAR	3	No	EDA-A2	Ectodermal cells	Mesenchymal signalling during ectodermal development	[258]
DR3	4	Yes	Not Known	Thymus, lymphocytes	Apoptosis in lymphocyte homeostasis	[259]
RELT	1	No	Not known	Spleen, lymph node, peripheral leukocytes	NF-κb activation	[260]
Fn14	1	No	TWEAK	Heart, placenta, kidney	Angiogenesis, endothelial proliferation	[261]

Appendix Reagents

Prokaryotic cell media

Luria-Bertani media

10 g tryptone, 5 g yeast extract, 10 g NaCl

Dissolve in 1 litre of deionised water and autoclave

For LB-agar, add 15 g bacto-agar

Inclusion body preparation and refolding

Triton wash buffer

50mM Tris pH 8.0, 100 mM NaCl, 0.5 % triton X-100

Denaturing buffer

6 M guanidine, 50 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA, 10 mM DTT

Refolding buffer

200 mM Tris pH 8.0, 1 M L-arginine, 10 mM EDTA, 6.5 mM cysteamine, 3.7 mM cystamine

Protein purification

Nickel binding buffer

20 mM $NaPO_4$ pH 7.4, 500mM NaCl, 10 mM imidazole

Nickel elution buffer

20 mM $NaPO_4$ pH 7.4, 500mM NaCl, 500 mM imidazole

Gel filtration buffer

50 mM Tris pH 8, 150 mM NaCl

Gel filtration buffer 2

50mM MES pH6, 150 mM NaCl

Solutions for SDS-PAGE

10 ml, 15 % gel

Lower gel 4X buffer	2.5 ml
Water	4.56 ml
Acrylamide solution	4.69 ml
10 % APS	50 µl
TEMED	5 µl
10% SDS	100 µl

10 ml Stacking gel

Upper gel 4x buffer	2.5 ml
Water	6.6 ml
Acrylamide solution	0.8 ml
10% APS	100 µl
TEMED	10 µl
10% SDS	100 µl

Solution compositions

Acrylamide solution

38.9 g acrylamide, 1.1 g bis-acrylamide in 100 ml water

Upper gel 4x buffer

0.5 M TRIS, pH 6.8

Lower gel buffer

1.5 M TRIS, pH 8.8

Gel running buffer

25 mM TRIS, 192 mM glycine 0.1 % SDS, pH 8.3

Appendix TNFR: sequences and primer design for TNF-receptor constructs

LTβR

Construct: Extracellular region without signal sequence

HindIIICDM7-Fc - POI - BamH1CDM7-Fc

Construct sequence

```
TCGCAGCCCCAGGCGGTGCCTCCATATGCGTCGGAGAACCAGACCTGCAGGGACCAGGAAAAGGAATACTATGAGCCCCAGCACCG
CATCTGCTGCTCCCGCTGCCCGCCAGGCACCTATGTCTCAGCTAAATGTAGCCGCATCCGGGACACAGTTTGTGCCACATGTGCCG
AGAATTCCTACAACGAGCACTGGAACTACCTGACCATCTGCCAGCTGTGCCGCCCCTGTGACCCAGTGATGGGCCTCGAGGAGATT
GCCCCCTGCACAAGCAAACGGAAGACCCAGTGCCGCTGCCAGCCGGGAATGTTCTGTGCTGCCTGGGCCCTCGAGTGTACACACTG
CGAGCTACTTTCTGACTGCCCGCCTGGCACTGAAGCCGAGCTCAAAGATGAAGTTGGGAAGGGTAACAACCACTGCGTCCCCTGCA
AGGCCGGGCACTTCCAGAATACCTCCTCCCCCAGCGCCCGCTGCCAGCCCCACACCAGGTGTGAGAACCAAGGTCTGGTGGAGGCA
GCTCCAGGCACTGCCCAGTCCGACACAACCTGCAAAAATCCATTAGAG
```

Construct length: 564 bases

Forward Primer Sequence

```
gcgcgcaagcttTCGCAGCCCCAGGCGGTG
```

HindIIICDM7-Fc

Reverse Primer Sequence

```
gcgcgcggatccCTCTAATGGATTTTTGCAGGTTGTGTCGG
```

BamH1CDM7-Fc

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
SQPQAVPPYASENQTCRDQEKEYYEPQHRICCSRCPPGTYVSAKCSRIRDTVCATCAENSYNEHWNYLTICQLCRPCDPVMGLEEI
APCTSKRKTQCRCQPGMFCAAWALECTHCELLSDCPPGTEAELKDEVGKGNNHCVPCKAGHFQNTSSPSARCQPHTRCENQGLVEA
APGTAQSDTTCKNPLE
```

Untagged Protein Length: 188 aa

LTβR (modified)

Construct: Extracellular domain without leader sequence for cloning into pHLsec

pHLsec_Age_F - POI - KpnI_pHLsec

Construct sequence

```
CAGGCGGTGCCTCCATATGCGTCGGAGAACCAGACCTGCAGGGACCAGGAAAAGGAATACTATGAGCCCCAGCACCGCATCTGCTG
CTCCCGCTGCCCGCCAGGCACCTATGTCTCAGCTAAATGTAGCCGCATCCGGGACACAGTTTGTGCCACATGTGCCGAGAATTCCT
ACAACGAGCACTGGAACTACCTGACCATCTGCCAGCTGTGCCGCCCCTGTGACCCAGTGATGGGCCTCGAGGAGATTGCCCCCTGC
ACAAGCAAACGGAAGACCCAGTGCCGCTGCCAGCCGGGAATGTTCTGTGCTGCCTGGGCCCTCGAGTGTACACACTGCGAGCTACT
TTCTGACTGCCCGCCTGGCACTGAAGCCGAGCTCAAAGATGAAGTTGGGAAGGGTAACAACCACTGCGTCCCCTGCAAGGCCGGGC
ACTTCCAGAATACCTCCTCCCCCAGCGCCCGCTGCCAGCCCCACACCAGGTGTGAGAACCAAGGTCTGGTGGAGGCAGCTCCAGGC
ACTGCCCAGTCCGACACAACCTGCAAAAATCCATTAGAG
```

Construct length: 555 bases

Forward Primer Sequence

```
gatataataccggtCAGGCGGTGCCTCCATATGC
```

pHLsec_Age_F

Reverse Primer Sequence

```
cgcggtaccCTCTAATGGATTTTTGCAGGTTG
```

KpnI_pHLsec

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QAVPPYASENQTCRDQEKEYYEPQHRICCSRCPPGTYVSAKCSRIRDTVCATCAENSYNEHWNYLTICQLCRPCDPVMGLEEIAPC
TSKRKTQCRCQPGMFCAAWALECTHCELLSDCPPGTEAELKDEVGKGNNHCVPCKAGHFQNTSSPSARCQPHTRCENQGLVEAAPG
TAQSDTTCKNPLE
```

Untagged Protein Length: 185 aa

CD40

Construct: N/C truncated (aa23-187)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CCCACTGCATGCAGAGAAAAACAGTACCTAATAAACAGTCAGTGCTGTTCTTTGTGCCAGCCAGGACAGAAACTGGTGAGTGACTG
CACAGAGTTCACTGAAACGGAATGCCTTCCTTGCGGTGAAAGCGAATTCCTAGACACCTGGAACAGAGAGACACACTGCCACCAGC
ACAAATACTGCGACCCCAACCTAGGGCTTCGGGTCCAGCAGAAGGGCACCTCAGAAACAGACACCATCTGCACCTGTGAAGAAGGC
TGGCACTGTACGAGTGAGGCCTGTGAGAGCTGTGTCCTGCACCGCTCATGCTCGCCCGGCTTTGGGGTCAAGCAGATTGCTACAGG
GGTTTCTGATACCATCTGCGAGCCCTGCCCAGTCGGCTTCTTCTCCAATGTGTCATCTGCTTTCGAAAAATGTCACCCTTGGACAA
GCTGTGAGACCAAAGACCTGGTTGTGCAACAGGCAGGCACAAACAAGACTGATGTTGTCTGTGGT
```

Construct length: 495 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcacca
ccaccatcacCCCACTGCATGCAGAGAAAAACAG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaACCACAGACAACATCAGTCTTGTTTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
PTACREKQYLINSQCCSLCQPGQKLVSDCTEFTETECLPCGESEFLDTWNRETHCHQHKYCDPNLGLRVQQKGTSETDTICTCEEG
WHCTSEACESCVLHRSCSPGFGVKQIATGVSDTICEPCPVGFFSNVSSAFEKCHPWTSCETKDLVVQQAGTNKTDVVCG
```

Untagged Protein Length: 165 aa

FAS

Construct: N/C truncated (aa47-170)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CAGAACTTGGAAGGCCTGCATCATGATGGCCAATTCTGCCATAAGCCCTGTCCTCCAGGTGAAAGGAAAGCTAGGGACTGCACAGT
CAATGGGGATGAACCAGACTGCGTGCCCTGCCAAGAAGGGAAGGAGTACACAGACAAAGCCCATTTTTCTTCCAAATGCAGAAGAT
GTAGATTGTGTGATGAAGGACATGGCTTAGAAGTGGAAATAAACTGCACCCGGACCCAGAATACCAAGTGCAGATGTAAACCAAAC
TTTTTTTGTAACTCTACTGTATGTGAACACTGTGACCCTTGCACCAAATGTGAACATGGAATCATCAAGGAATGCACACTCACCAG
CAACACCAAGTGCAAAGAGGAAGGATCC
```

Construct length: 372 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcacca
ccaccatcacCAGAACTTGGAAGGCCTGCATCATG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGGATCCTTCCTCTTTGCACTTGGTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QNLEGLHHDGQFCHKPCPPGERKARDCTVNGDEPDCVPCQEGKEYTDKAHFSSKCRRCRLCDEGHGLEVEINCTRTQNTKCRCKPN
FFCNSTVCEHCDPCTKCEHGIIKECTLTSNTKCKEEGS
```

Untagged Protein Length: 124 aa

DcR3

Construct: N/C truncated (aa33-198)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
ACACCCACCTACCCCTGGCGGGACGCAGAGACAGGGGAGCGGCTGGTGTGCGCCCAGTGCCCCCCAGGCACCTTTGTGCAGCGGCC
GTGCCGCCGAGACAGCCCCACGACGTGTGGCCCGTGTCCACCGCGCCACTACACGCAGTTCTGGAACTACCTGGAGCGCTGCCGCT
ACTGCAACGTCCTCTGCGGGGAGCGTGAGGAGGAGGCACGGGCTTGCCACGCCACCCACAACCGTGCCTGCCGCTGCCGCACCGGC
TTCTTCGCGCACGCTGGTTTCTGCTTGGAGCACGCATCGTGTCCACCTGGTGCCGGCGTGATTGCCCCGGGCACCCCCAGCCAGAA
CACGCAGTGCCAGCCGTGCCCCCCAGGCACCTTCTCAGCCAGCAGCTCCAGCTCAGAGCAGTGCCAGCCCCACCGCAACTGCACGG
CCCTGGGCCTGGCCCTCAATGTGCCAGGCTCTTCCTCCCATGACACCCTGTGCACCAGCTGCACTGGC
```

Construct length: 498 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcacca
ccaccatcacACACCCACCTACCCCTGGCG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGCCAGTGCAGCTGGTGCACAG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
TPTYPWRDAETGERLVCAQCPPGTFVQRPCRRDSPTTCGPCPPRHYTQFWNYLERCRYCNVLCGEREEEARACHATHNRACRCRTG
FFAHAGFCLEHASCPPGAGVIAPGTPSQNTQCQPCPPGTFSASSSSSEQCQPHRNCTALGLALNVPGSSSHDTLCTSCTG
```

Untagged Protein Length: 166 aa

CD27

Construct: N/C truncated (aa24-141)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CCCAAGAGCTGCCCAGAGAGGCACTACTGGGCTCAGGGAAAGCTGTGCTGCCAGATGTGTGAGCCAGGAACATTCCTCGTGAAGGA
CTGTGACCAGCATAGAAAGGCTGCTCAGTGTGATCCTTGCATACCGGGGGTCTCCTTCTCTCCTGACCACCACACCCGGCCCCACT
GTGAGAGCTGTCGGCACTGTAACTCTGGTCTTCTCGTTCGCAACTGCACCATCACTGCCAATGCTGAGTGTGCCTGTCGCAATGGC
TGGCAGTGCAGGGACAAGGAGTGCACCGAGTGTGATCCTCTTCCAAACCCTTCGCTGACCGCTCGGTCGTCTCAGGCCCTGAGCCC
ACACCCTCAG
```

Construct length: 354 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCCCAAGAGCTGCCCAG
AGAGGC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaCTGAGGGTGTGGGCTCAGGGC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
PKSCPERHYWAQGKLCCQMCEPGTFLVKDCDQHRKAAQCDPCIPGVSFSPDHHTRPHCESCRHCNSGLLVRNCTITANAECACRNG
WQCRDKECTECDPLPNPSLTARSSQALSPHPQ
```

Untagged Protein Length: 118 aa

CD137

Construct: N/C truncated (aa17-160)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
AACTTTGAGAGGACAAGATCATTGCAGGATCCTTGTAGTAACTGCCCAGCTGGTACATTCTGTGATAATAACAGGAATCAGATTTG
CAGTCCCTGTCCTCCAAATAGTTTCTCCAGCGCAGGTGGACAAAGGACCTGTGACATATGCAGGCAGTGTAAAGGTGTTTTCAGGA
CCAGGAAGGAGTGTTCCTCCACCAGCAATGCAGAGTGTGACTGCACTCCAGGGTTTCACTGCCTGGGGGCAGGATGCAGCATGTGT
GAACAGGATTGTAAACAAGGTCAAGAACTGACAAAAAAAGGTTGTAAAGACTGTTGCTTTGGGACATTTAACGATCAGAAACGTGG
CATCTGTCGACCCTGGACAAACTGTTCTTTGGATGGAAAGTCTGTGCTTGTGAATGGGACGAAGGAGAGGGACGTGGTCTGTGGAC
CA
```

Construct length: 432 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacAACTTTGAGAGGACAA
GATCATTGCAGG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTGGTCCACAGACCACGTCCCTC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
NFERTRSLQDPCSNCPAGTFCDNNRNQICSPCPPNSFSSAGGQRTCDICRQCKGVFRTRKECSSTSNAECDCTPGFHCLGAGCSMC
EQDCKQGQELTKKGCKDCCFGTFNDQKRGICRPWTNCSLDGKSVLVNGTKERDVVCGP
```

Untagged Protein Length: 144 aa

DR4

Construct: N/C truncated (aa107-232)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
TCAGCTGCAACCATCAAACTTCATGATCAATCAATTGGCACACAGCAATGGGAACATAGCCCTTTGGGAGAGTTGTGTCCACCAGG
ATCTCATAGATCAGAACATCCTGGAGCCTGTAACCGGTGCACAGAGGGTGTGGGTTACACCAATGCTTCCAACAATTTGTTTGCTT
GCCTCCCATGTACAGCTTGTAAATCAGATGAAGAAGAGAGAAGTCCCTGCACCACGACCAGGAACACAGCATGTCAGTGCAAACCA
GGAACTTTCCGGAATGACAATTCTGCTGAGATGTGCCGGAAGTGCAGCAGAGGGTGCCCCAGAGGGATGGTCAAGGTCAAGGATTG
TACGCCCTGGAGTGACATCGAGTGTGTCCACAAA
```

Construct length: 378 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacTCAGCTGCAACCATCA
AACTTCATG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTTTGTGGACACACTCGATGTCACTCC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
SAATIKLHDQSIGTQQWEHSPLGELCPPGSHRSEHPGACNRCTEGVGYTNASNNLFACLPCTACKSDEEERSPCTTTRNTACQCKP
GTFRNDNSAEMCRKCSRGCPRGMVKVKDCTPWSDIECVHK
```

Untagged Protein Length: 126 aa

DR5

Construct: N/C truncated (aa57-181)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

ACCCAACAAGACCTAGCTCCCCAGCAGAGAGCGGCCCCACAACAAAAGAGGTCCAGCCCCTCAGAGGGATTGTGTCCACCTGGACA
CCATATCTCAGAAGACGGTAGAGATTGCATCTCCTGCAAATATGGACAGGACTATAGCACTCACTGGAATGACCTCCTTTTCTGCT
TGCGCTGCACCAGGTGTGATTCAGGTGAAGTGGAGCTAAGTCCCTGCACCACGACCAGAAACACAGTGTGTCAGTGCGAAGAAGGC
ACCTTCCGGGAAGAAGATTCTCCTGAGATGTGCCGGAAGTGCCGCACAGGGTGTCCCAGAGGGATGGTCAAGGTCGGTGATTGTAC
ACCCTGGAGTGACATCGAATGTGTCCACAAA

Construct length: 375 bases

Forward Primer Sequence

ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacACCCAACAAGACCTAG
CTCCCCAGC

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

ggggaccactttgtacaagaaagctgggtctcaTTTGTGGACACATTCGATGTCACTCC

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

TQQDLAPQQRAAPQQKRSSPSEGLCPPGHHISEDGRDCISCKYGQDYSTHWNDLLFCLRCTRCDSGEVELSPCTTTRNTVCQCEEG
TFREEDSPEMCRKCRTGCPRGMVKVGDCTPWSDIECVHK

Untagged Protein Length: 125 aa

DcR2

Construct: N/C truncated (aa 57-1810

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

ACCATCCCCCGGCAGGACGAAGTTCCCCAGCAGACAGTGGCCCCACAGCAACAGAGGCGCAGCCTCAAGGAGGAGGAGTGTCCAGC
AGGATCTCATAGATCAGAATATACTGGAGCCTGTAACCCGTGCACAGAGGGTGTGGATTACACCATTGCTTCCAACAATTTGCCTT
CTTGCCTGCTATGTACAGTTTGTAAATCAGGTCAAACAAATAAAAGTTCCTGTACCACGACCAGAGACACCGTGTGTCAGTGTGAA
AAAGGAAGCTTCCAGGATAAAAACTCCCCTGAGATGTGCCGGACGTGTAGAACAGGGTGTCCCAGAGGGATGGTCAAGGTCAGTAA
TTGTACGCCCCGGAGTGACATCAAGTGCAAA

Construct length: 375 bases

Forward Primer Sequence

ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacACCATCCCCCGGCAGG
AC

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

ggggaccactttgtacaagaaagctgggtctcaTTTGCACTTGATGTCACTCCGGG

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged

)TIPRQDEVPQQTVAPQQQRRSLKEEECPAGSHRSEYTGACNPCTEGVDYTIASNNLPSCLLCTVCKSGQTNKSSCTTTRDTVCQC
EKGSFQDKNSPEMCRTCRTGCPRGMVKVSNCTPRSDIKCK

Untagged Protein Length: 125 aa

Osteoprotegerin

Construct: N/C truncated (aa26-188)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CCAAAGTACCTTCATTATGACGAAGAAACCTCTCATCAGCTGTTGTGTGACAAATGTCCTCCTGGTACCTACCTAAAACAACACTG
TACAGCAAAGTGGAAGACCGTGTGCGCCCCTTGCCCTGACCACTACTACACAGACAGCTGGCACACCAGTGACGAGTGTCTATACT
GCAGCCCCGTGTGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGCAATCGCACCCACAACCGCGTGTGCGAATGCAAGGAAGGGCGC
TACCTTGAGATAGAGTTCTGCTTGAAACATAGGAGCTGCCCTCCTGGATTTGGAGTGGTGCAAGCTGGAACCCCAGAGCGAAATAC
AGTTTGCAAAAGATGTCCAGATGGGTTCTTCTCAAATGAGACGTCATCTAAAGCACCCTGTAGAAAACACACAAATTGCAGTGTCT
TTGGTCTCCTGCTAACTCAGAAAGGAAATGCAACACACGACAACATATGTTCCGGAAAC
```

Construct length: 489 bases

Forward Primer sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCCAAAGTACCTTCATT
ATGACGAAGAAACC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGTTTCCGGAACATATGTTGTCGTGTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
PKYLHYDEETSHQLLCDKCPPGTYLKQHCTAKWKTVCAPCPDHYYTDSWHTSDECLYCSPVCKELQYVKQECNRTHNRVCECKEGR
YLEIEFCLKHRSCPPGFGVVQAGTPERNTVCKRCPDGFFSNETSSKAPCRKHTNCSVFGLLLTQKGNATHDNICSGN
```

Untagged Protein Length: 163 aa

TROY

Construct: N/C truncated (aa26-162)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
AAAGTGACTTGTGAATCAGGAGACTGTAGACAGCAAGAATTCAGGGATCGGTCTGGAAACTGTGTTCCCTGCAACCAGTGTGGGCC
AGGCATGGAGTTGTCTAAGGAATGTGGCTTCGGCTATGGGGAGGATGCACAGTGTGTGACGTGCCGGCTGCACAGGTTCAAGGAGG
ACTGGGGCTTCCAGAAATGCAAGCCCTGTCTGGACTGCGCAGTGGTGAACCGCTTTCAGAAGGCAAATTGTTCAGCCACCAGTGAT
GCCATCTGCGGGGACTGCTTGCCAGGATTTTATAGGAAGACGAAACTTGTCGGCTTTCAAGACATGGAGTGTGTGCCTTGTGGAGA
CCCTCCTCCTCCTTACGAACCGCACTGTGCCAGCAAGGTCAACCTCGTGAAGATCGCGTCCACGGCC
```

Construct length: 411 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacAAAGTGACTTGTGAAT
CAGGAGACTG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGGCCGTGGACGCGATCTTC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
KVTCESGDCRQQEFRDRSGNCVPCNQCGPGMELSKECGFGYGEDAQCVTCRLHRFKEDWGFQKCKPCLDCAVVNRFQKANCSATSD
AICGDCLPGFYRKTKLVGFQDMECVPCGDPPPPYEPHCASKVNLVKIASTA
```

Untagged Protein Length: 137 aa

DR6

Construct: N/C truncated

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CAGCCAGAACAGAAGGCCTCGAATCTCATTGGCACATACCGCCATGTTGACCGTGCCACCGGCCAGGTGCTAACCTGTGACAAGTG
TCCAGCAGGAACCTATGTCTCTGAGCATTGTACCAACACAAGCCTGCGCGTCTGCAGCAGTTGCCCTGTGGGGACCTTTACCAGGC
ATGAGAATGGCATAGAGAAATGCCATGACTGTAGTCAGCCATGCCCATGGCCAATGATTGAGAAATTACCTTGTGCTGCCTTGACT
GACCGAGAATGCACTTGCCCACCTGGCATGTTCCAGTCTAACGCTACCTGTGCCCCCCATACGGTGTGTCCTGTGGGTTGGGGTGT
GCGGAAGAAAGGGACAGAGACTGAGGATGTGCGGTGTAAGCAGTGTGCTCGGGGTACCTTCTCAGATGTGCCTTCTAGTGTGATGA
AATGCAAAGCATACACAGACTGTCTGAGTCAGAACCTGGTGGTGATCAAGCCGGGGACCAAGGAGACAGACAACGTCTGTGGCACA
CTC
```

Construct length: 519 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCAGCCAGAACAGAAGG
CCTCG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGAGTGTGCCACAGACGTTGTCTGTC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QPEQKASNLIGTYRHVDRATGQVLTCDKCPAGTYVSEHCTNTSLRVCSSCPVGTFTRHENGIEKCHDCSQPCPWPMIEKLPCAALT
DRECTCPPGMFQSNATCAPHTVCPVGWGVRKKGTETEDVRCKQCARGTFSDVPSSVMKCKAYTDCLSQNLVVIKPGTKETDNVCGT
L
```

Untagged Protein Length: 173 aa

DcR1

Construct: N/C truncated (aa30-150)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CGGCAGGAGGAAGTTCCCCAGCAGACAGTGGCCCCACAGCAACAGAGGCACAGCTTCAAGGGGGAGGAGTGTCCAGCAGGATCTCA
TAGATCAGAACATACTGGAGCCTGTAACCCGTGCACAGAGGGTGTGGATTACACCAACGCTTCCAACAATGAACCTTCTTGCTTCC
CATGTACAGTTTGTAAATCAGATCAAAAACATAAAAGTTCCTGCACCATGACCAGAGACACAGTGTGTCAGTGTAAAGAAGGCACC
TTCCGGAATGAAAACTCCCCAGAGATGTGCCGGAAGTGTAGCAGGTGCCCTAGTGGGGAAGTCCAAGTCAGTAATTGTACGTCCTG
GGATGATATCCAGTGTGTT
```

Construct length: 363 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCGGCAGGAGGAAGTTC
CCCAG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequenc

```
eggggaccactttgtacaagaaagctgggtctcaAACACACTGGATATCATCCCAGGACG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
RQEEVPQQTVAPQQQRHSFKGEECPAGSHRSEHTGACNPCTEGVDYTNASNNEPSCFPCTVCKSDQKHKSSCTMTRDTVCQCKEGT
FRNENSPEMCRKCSRCPSGEVQVSNCTSWDDIQCV
```

Untagged Protein Length: 121 aa

GITR

Construct: N/C truncated (aa 33-155)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GGGTGCGGCCCTGGGCGCCTCCTGCTTGGGACGGGAACGGACGCGCGCTGCTGCCGGGTTCACACGACGCGCTGCTGCCGCGATTA
CCCGGGCGAGGAGTGCTGTTCCGAGTGGGACTGCATGTGTGTCCAGCCTGAATTCCACTGCGGAGACCCTTGCTGCACGACCTGCC
GGCACCACCCTTGTCCCCCAGGCCAGGGGGTACAGTCCCAGGGGAAATTCAGTTTTGGCTTCCAGTGTATCGACTGTGCCTCGGGG
ACCTTCTCCGGGGGCCACGAAGGCCACTGCAAACCTTGGACAGACTGCACCCAGTTCGGGTTTCTCACTGTGTTCCCTGGGAACAA
GACCCACAACGCTGTGTGCGTCCCA
```

Construct length: 369 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGGGTGCGGCCCTGGGC
GC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTGGGACGCACACAGCGTTGTGG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
GCGPGRLLLGTGTDARCCRVHTTRCCRDYPGEECCSEWDCMCVQPEFHCGDPCCTTCRHHPCPPGQGVQSQGKFSFGFQCIDCASG
TFSGGHEGHCKPWTDCTQFGFLTVFPGNKTHNAVCVP
```

Untagged Protein Length: 123 aa

CD30

Construct: N/C truncated (aa28-151)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
ACCTGTCATGGAAACCCCAGCCACTACTATGACAAGGCTGTCAGGAGGTGCTGTTACCGCTGCCCCATGGGGCTGTTCCCGACACA
GCAGTGCCCACAGAGGCCTACTGACTGCAGGAAGCAGTGTGAGCCTGACTACTACCTGGATGAGGCCGACCGCTGTACAGCCTGCG
TGACTTGTTCTCGAGATGACCTCGTGGAGAAGACGCCGTGTGCATGGAACTCCTCCCGTGTCTGCGAATGTCGACCCGGCATGTTC
TGTTCCACGTCTGCCGTCAACTCCTGTGCCCGCTGCTTCTTCCATTCTGTCTGTCCGGCAGGGATGATTGTCAAGTTCCCAGGCAC
GGCGCAGAAGAACACGGTCTGTGAGCCG
```

Construct length372 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacACCTGTCATGGAAACC
CCAGCC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaCGGCTCACAGACCGTGTTCTTCTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
TCHGNPSHYYDKAVRRCCYRCPMGLFPTQQCPQRPTDCRKQCEPDYYLDEADRCTACVTCSRDDLVEKTPCAWNSSRVCECRPGMF
CSTSAVNSCARCFFHSVCPAGMIVKFPGTAQKNTVCEP
```

Untagged Protein Length: 124 aa

HVEM

Construct: GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CCCTGCTACGCCCCAGCTCTGCCGTCCTGCAAGGAGGACGAGTACCCAGTGGGCTCCGAGTGCTGCCCCAAGTGCAGTCCAGGTTA
TCGTGTGAAGGAGGCCTGCGGGGAGCTGACGGGCACAGTGTGTGAACCCTGCCCTCCAGGCACCTACATTGCCCACCTCAATGGCC
TAAGCAAGTGTCTGCAGTGCCAAATGTGTGACCCAGCCATGGGCCTGCGCGCGAGCCGGAACTGCTCCAGGACAGAGAACGCCGTG
TGTGGCTGCAGCCCAGGCCACTTCTGCATCGTCCAGGACGGGGACCACTGCGCCGCGTGCCGCGCTTACGCCACCTCCAGCCCGGG
CCAGAGGGTGCAGAAGGGAGGCACCGAGAGTCAGGACACCCTGTGTCAGAACTGCCCCCCGGGGACCTTCTCTCCCAATGGGACCC
TGGAGGAATGTCAGCACCAGACCAAGTGCAGCTGGCTGGTGACGAAGGCCGGAGCTGGGACCAGCAGCTCCCAC
```

Construct length: 504 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCCCTGCTACGCCCCAG
CTC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaGTGGGAGCTGCTGGTCCCAGC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
PCYAPALPSCKEDEYPVGSECCPKCSPGYRVKEACGELTGTVCEPCPPGTYIAHLNGLSKCLQCQMCDPAMGLRASRNCSRTENAV
CGCSPGHFCIVQDGDHCAACRAYATSSPGQRVQKGGTESQDTLCQNCPPGTFSPNGTLEECQHQTKCSWLVTKAGAGTSSSH
```

Untagged Protein Length: 168 aa

Appendix TNF Ligands: sequences and primer design for TNF-ligand constructs

Lymphotoxin alpha

Construct: Full length CDS w/o signal sequence

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CTCCCTGGTGTTGGCCTCACACCTTCAGCTGCCCAGACTGCCCGTCAGCACCCCAAGATGCATCTTGCCCACAGCACCCTCAAACC
TGCTGCTCACCTCATTGGAGACCCCAGCAAGCAGAACTCACTGCTCTGGAGAGCAAACACGGACCGTGCCTTCCTCCAGGATGGTT
TCTCCTTGAGCAACAATTCTCTCCTGGTCCCCACCAGTGGCATCTACTTCGTCTACTCCCAGGTGGTCTTCTCTGGGAAAGCCTAC
TCTCCCAAGGCCACCTCCTCCCCACTCTACCTGGCCCATGAGGTCCAGCTCTTCTCCTCCCAGTACCCCTTCCATGTGCCTCTCCT
CAGCTCCCAGAAGATGGTGTATCCAGGGCTGCAGGAACCCTGGCTGCACTCGATGTACCACGGGGCTGCGTTCCAGCTCACCCAGG
GAGACCAGCTATCCACCCACACAGATGGCATCCCCCACCTAGTCCTCAGCCCTAGTACTGTCTTCTTTGGAGCCTTCGCTCTG
```

Construct length: 513 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCTCCCTGGTGTTGGCC
TCACAC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaCAGAGCGAAGGCTCCAAAGAAGACAG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
LPGVGLTPSAAQTARQHPKMHLAHSTLKPAAHLIGDPSKQNSLLWRANTDRAFLQDGFSLSNNSLLVPTSGIYFVYSQVVFSGKAY
SPKATSSPLYLAHEVQLFSSQYPFHVPLLSSQKMVYPGLQEPWLHSMYHGAAFQLTQGDQLSTHTDGIPHLVLSPSTVFFGAFAL
```

Untagged Protein Length: 171 aa

CD27 ligand

Construct: N truncated (aa56-194)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GACGTAGCTGAGCTGCAGCTGAATCACACAGGACCTCAGCAGGACCCCAGGCTATACTGGCAGGGGGGCCCAGCACTGGGCCGCTC
CTTCCTGCATGGACCAGAGCTGGACAAGGGGCAGCTACGTATCCATCGTGATGGCATCTACATGGTACACATCCAGGTGACGCTGG
CCATCTGTTCCTCCACGACGGCCTCCAGGCACCACCCCACCACCCTGGCCGTGGGAATCTGCTCTCCCGCCTCCCGTAGCATCAGC
CTGCTGCGTCTCAGCTTCCACCAAGGTTGTACCATTGCCTCCCAGCGCCTGACGCCCCTGGCCCGAGGGGACACACTCTGCACCAA
CCTCACTGGGACACTTTTGCCTTCCCGAAACACTGATGAGACCTTCTTTGGAGTGCAGTGGGTGCGCCCCTGA
```

Construct length: 417 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGACGTAGCTGAGCTGC
AGCTGAATC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTCAGGGGCGCACCCACTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
DVAELQLNHTGPQQDPRLYWQGGPALGRSFLHGPELDKGQLRIHRDGIYMVHIQVTLAICSSTTASRHHPTTLAVGICSPASRSIS
LLRLSFHQGCTIASQRLTPLARGDTLCTNLTGTLLPSRNTDETFFGVQWVRP
```

Untagged Protein Length: 138 aa

TRAIL

Construct: N truncated (aa114-282)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GTGAGAGAAAGAGGTCCTCAGAGAGTAGCAGCTCACATAACTGGGACCAGAGGAAGAAGCAACACATTGTCTTCTCCAAACTCCAA
GAATGAAAAGGCTCTGGGCCGCAAAATAAACTCCTGGGAATCATCAAGGAGTGGGCATTCATTCCTGAGCAACTTGCACTTGAGGA
ATGGTGAACTGGTCATCCATGAAAAAGGGTTTTACTACATCTATTCCCAAACATACTTTCGATTTCAGGAGGAAATAAAAGAAAAC
ACAAAGAACGACAAACAAATGGTCCAATATATTTACAAATACACAAGTTATCCTGACCCTATATTGTTGATGAAAAGTGCTAGAAA
TAGTTGTTGGTCTAAAGATGCAGAATATGGACTCTATTCCATCTATCAAGGGGGAATATTTGAGCTTAAGGAAAATGACAGAATTT
TTGTTTCTGTAACAAATGAGCACTTGATAGACATGGACCATGAAGCCAGTTTTTTCGGGGCCTTTTTAGTTGGCTAA
```

Construct length: 507 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGTGAGAGAAAGAGGTC
CTCAGAGAG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTTAGCCAACTAAAAAGGCCCCG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
VRERGPQRVAAHITGTRGRSNTLSSPNSKNEKALGRKINSWESSRSGHSFLSNLHLRNGELVIHEKGFYYIYSQTYFRFQEEIKEN
TKNDKQMVQYIYKYTSYPDPILLMKSARNSCWSKDAEYGLYSIYQGGIFELKENDRIFVSVTNEHLIDMDHEASFFGAFLVG
```

Untagged Protein Length: 168 aa

APRIL

Construct: N truncated (aa105-248)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GCAGTGCTCACCCAAAAACAGAAGAAGCAGCACTCTGTCCTGCACCTGGTTCCCATTAACGCCACCTCCAAGGATGACTCCGATGT
GACAGAGGTGATGTGGCAACCAGCTCTTAGGCGTGGGAGAGGCCTACAGGCCCAAGGATATGGTGTCCGAATCCAGGATGCTGGAG
TTTATCTGCTGTATAGCCAGGTCCTGTTTCAAGACGTGACTTTCACCATGGGTCAGGTGGTGTCTCGAGAAGGCCAAGGAAGGCAG
GAGACTCTATTCCGATGTATAAGAAGTATGCCCTCCCACCCGGACCGGGCCTACAACAGCTGCTATAGCGCAGGTGTCTTCCATTT
ACACCAAGGGGATATTCTGAGTGTCATAATTCCCCGGGCAAGGGCGAAACTTAACCTCTCTCCACATGGAACCTTCCTGGGACTTT
GA
```

Construct length: 432 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGCAGTGCTCACCCAAA
AACAGAAG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTCAAAGTCCCAGGAAGGTTCCATGTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
AVLTQKQKKQHSVLHLVPINATSKDDSDVTEVMWQPALRRGRGLQAQGYGVRIQDAGVYLLYSQVLFQDVTFTMGQVVSREGQGRQ
ETLFRCIRSMPSHPDRAYNSCYSAGVFHLHQGDILSVIIPRARAKLNLSPHGTFLGL
```

Untagged Protein Length: 143 aa

LIGHT

Construct: N truncated (aa 66-178)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

GGAGAGATGGTCACCCGCCTGCCTGACGGACCTGCAGGCTCCTGGGAGCAGCTGATACAAGAGCGAAGGTCTCACGAGGTCAACCC
AGCAGCGCATCTCACAGGGGCCAACTCCAGCTTGACCGGCAGCGGGGGGCCGCTGTTATGGGAGACTCAGCTGGGCGTGGCCTTCC
TGAGGGGCCTCAGCTACCACGATGGGGCCCTTGTGGTCACCAAAGCTGGCTACTACTACATCTACTCCAAGGTGCAGCTGGGCGGT
GTGGGCTGCCCGCTGGGCCTGGCCAGCACCATCACCCACGGCCTCTACAAGCGCACACCCCGCTACCCCGAGGAGCTGTAG

Construct length: 339 bases

Forward Primer Sequence

ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGGAGAGATGGTCACCC
GCCTG

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

ggggaccactttgtacaagaaagctgggtctcaCTACAGCTCCTCGGGGTAGCGG

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

GEMVTRLPDGPAGSWEQLIQERRSHEVNPAAHLTGANSSLTGSGGPLLWETQLGVAFLRGLSYHDGALVVTKAGYYYIYSKVQLGG
VGCPLGLASTITHGLYKRTPRYPEEL

Untagged Protein Length: 112 aa

CD30 Ligand

Construct: N truncated (aa66-235)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

GACTCCATTCCCAACTCACCTGACAACGTCCCCCTCAAAGGAGGAAATTGCTCAGAAGACCTCTTATGTATCCTGAAAAGGGCTCC
ATTCAAGAAGTCATGGGCCTACCTCCAAGTGGCAAAGCATCTAAACAAAACCAAGTTGTCTTGGAACAAAGATGGCATTCTCCATG
GAGTCAGATATCAGGATGGGAATCTGGTGATCCAATTCCCTGGTTTGTACTTCATCATTTGCCAACTGCAGTTTCTTGTACAATGC
CCAAATAATTCTGTCGATCTGAAGTTGGAGCTTCTCATCAACAAGCATATCAAAAAACAGGCCCTGGTGACAGTGTGTGAGTCTGG
AATGCAAACGAAACACGTATACCAGAATCTCTCTCAATTCTTGCTGGATTACCTGCAGGTCAACACCACCATATCAGTCAATGTGG
ATACATTCCAGTACATAGATACAAGCACCTTTCCTCTTGAGAATGTGTTGTCCATCTTCTTATACAGTAATTCAGACTGA

Construct length: 510 bases

Forward Primer Sequence

ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGACTCCATTCCCAACT
CACCTGAC

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

ggggaccactttgtacaagaaagctgggtctcaTCAGTCTGAATTACTGTATAAGAAGATG

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

DSIPNSPDNVPLKGGNCSEDLLCILKRAPFKKSWAYLQVAKHLNKTKLSWNKDGILHGVRYQDGNLVIQFPGLYFIICQLQFLVQC
PNNSVDLKLELLINKHIKKQALVTVCESGMQTKHVYQNLSQFLLDYLQVNTTISVNVDTFQYIDTSTFPLENVLSIFLYSNSD

Untagged Protein Length: 169 aa

GIRTL

Construct: GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CAATTAGAGACTGCTAAGGAGCCCTGTATGGCTAAGTTTGGACCATTACCCTCAAAATGGCAAATGGCATCTTCTGAACCTCCTTG
CGTGAATAAGGTGTCTGACTGGAAGCTGGAGATACTTCAGAATGGCTTATATTTAATTTATGGCCAAGTGGCTCCCAATGCAAACT
ACAATGATGTAGCTCCTTTTGAGGTGCGGCTGTATAAAAACAAAGACATGATACAAACTCTAACAAACAAATCTAAAATCCAAAAT
GTAGGAGGGACTTATGAATTGCATGTTGGGGACACCATAGACTTGATATTCAACTCTGAGCATCAGGTTCTAAAAAATAATACATA
CTGGGGTATCATTTTACTAGCAAATCCCCAATTCATCTCCTAG
```

Construct length: 387 bases

Forward Primer sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCAATTAGAGACTGCTA
AGGAGCCCTGTATGG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaCTAGGAGATGAATTGGGGATTTGCTAG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QLETAKEPCMAKFGPLPSKWQMASSEPPCVNKVSDWKLEILQNGLYLIYGQVAPNANYNDVAPFEVRLYKNKDMIQTLTNKSKIQN
VGGTYELHVGDTIDLIFNSEHQVLKNNTYWGIILLANPQFIS
```

Untagged Protein Length: 128 aa

VEGI

Construct: N truncated (aa90-252)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GCAGACGGAGATAAGCCAAGGGCACACCTGACAGTTGTGAGACAAACTCCCACACAGCACTTTAAAAATCAGTTCCCAGCTCTGCA
CTGGGAACATGAACTAGGCCTGGCCTTCACCAAGAACCGAATGAACTATACCAACAAATTCCTGCTGATCCCAGAGTCGGGAGACT
ACTTCATTTACTCCCAGGTCACATTCCGTGGGATGACCTCTGAGTGCAGTGAAATCAGACAAGCAGGCCGACCAAACAAGCCAGAC
TCCATCACTGTGGTCATCACCAAGGTAACAGACAGCTACCCTGAGCCAACCCAGCTCCTCATGGGGACCAAGTCTGTATGCGAAGT
AGGTAGCAACTGGTTCCAGCCCATCTACCTCGGAGCCATGTTCTCCTTGCAAGAAGGGGACAAGCTAATGGTGAACGTCAGTGACA
TCTCTTTGGTGGATTACACAAAAGAAGATAAAACCTTCTTTGGAGCCTTCTTACTATAG
```

Construct length: 489 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGCAGACGGAGATAAGC
CAAGGGC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaCTATAGTAAGAAGGCTCCAAAGAAGG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
ADGDKPRAHLTVVRQTPTQHFKNQFPALHWEHELGLAFTKNRMNYTNKFLLIPESGDYFIYSQVTFRGMTSECSEIRQAGRPNKPD
SITVVITKVTDSYPEPTQLLMGTKSVCEVGSNWFQPIYLGAMFSLQEGDKLMVNVSDISLVDYTKEDKTFFGAFLL
```

Untagged Protein Length: 162 aa

4-1BB ligand

Construct: N truncated (aa 85-255)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
TTGGACCTGCGGCAGGGCATGTTTGCGCAGCTGGTGGCCCAAAATGTTCTGCTGATCGATGGGCCCCTGAGCTGGTACAGTGACCC
AGGCCTGGCAGGCGTGTCCCTGACGGGGGGCCTGAGCTACAAAGAGGACACGAAGGAGCTGGTGGTGGCCAAGGCTGGAGTCTACT
ATGTCTTCTTTCAACTAGAGCTGCGGCGCGTGGTGGCCGGCGAGGGCTCAGGCTCCGTTTCACTTGCGCTGCACCTGCAGCCACTG
CGCTCTGCTGCTGGGGCCGCCGCCCTGGCTTTGACCGTGGACCTGCCACCCGCCTCCTCCGAGGCTCGGAACTCGGCCTTCGGTTT
CCAGGGCCGCTTGCTGCACCTGAGTGCCGGCCAGCGCCTGGGCGTCCATCTTCACACTGAGGCCAGGGCACGCCATGCCTGGCAGC
TTACCCAGGGCGCCACAGTCTTGGGACTCTTCCGGGTGACCCCCGAAATCCCAGCCGGACTCCCTTCACCGAGGTCGGAATAA
```

Construct length: 513 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacTTGGACCTGCGGCAGG
GC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTTATTCCGACCTCGGTGAAGGG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
LDLRQGMFAQLVAQNVLLIDGPLSWYSDPGLAGVSLTGGLSYKEDTKELVVAKAGVYYVFFQLELRRVVAGEGSGSVSLALHLQPL
RSAAGAAALALTVDLPPASSEARNSAFGFQGRLLHLSAGQRLGVHLHTEARARHAWQLTQGATVLGLFRVTPEIPAGLPSPRSE
```

Untagged Protein Length: 170 aa

CD40 ligand

Construct: N truncated (aa116-262)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
GGTGATCAGAATCCTCAAATTGCGGCACATGTCATAAGTGAGGCCAGCAGTAAAACAACATCTGTGTTACAGTGGGCTGAAAAAGG
ATACTACACCATGAGCAACAACTTGGTAACCCTGGAAAATGGGAAACAGCTGACCGTTAAAAGACAAGGACTCTATTATATCTATG
CCCAAGTCACCTTCTGTTCCAATCGGGAAGCTTCGAGTCAAGCTCCATTTATAGCCAGCCTCTGCCTAAAGTCCCCCGGTAGATTC
GAGAGAATCTTACTCAGAGCTGCAAATACCCACAGTTCCGCCAAACCTTGCGGGCAACAATCCATTCACTTGGGAGGAGTATTTGA
ATTGCAACCAGGTGCTTCGGTGTTTGTCAATGTGACTGATCCAAGCCAAGTGAGCCATGGCACTGGCTTC
ACGTCCTTTGGCTTACTCAAACTCTGA
```

Construct length: 441 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacGGTGATCAGAATCCTC
AAATTGCGG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTCAGAGTTTGAGTAAGCCAAAGGACGTG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
GDQNPQIAAHVISEASSKTTSVLQWAEKGYYTMSNNLVTLENGKQLTVKRQGLYYIYAQVTFCSNREASSQAPFIASLCLKSPGRF
ERILLRAANTHSSAKPCGQQSIHLGGVFELQPGASVFVNVTDPSQVSHGTGFTSFGLLKL
```

Untagged Protein Length: 146 aa

RANKL

Construct: N truncated (157-318)

GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
AGGAGCAAGCTTGAAGCTCAGCCTTTTGCTCATCTCACTATTAATGCCACCGACATCCCATCTGGTTCCCATAAAGTGAGTCTGTC
CTCTTGGTACCATGATCGGGGTTGGGCCAAGATCTCCAACATGACTTTTAGCAATGGAAAACTAATAGTTAATCAGGATGGCTTTT
ATTACCTGTATGCCAACATTTGCTTTCGACATCATGAAACTTCAGGAGACCTAGCTACAGAGTATCTTCAACTAATGGTGTACGTC
ACTAAAACCAGCATCAAAATCCCAAGTTCTCATACCCTGATGAAAGGAGGAAGCACCAAGTATTGGTCAGGGAATTCTGAATTCCA
TTTTTATTCCATAAACGTTGGTGGATTTTTTAAGTTACGGTCTGGAGAGGAAATCAGCATCGAGGTCTCCAACCCCTCCTTACTGG
ATCCGGATCAGGATGCAACATACTTTGGGGCTTTTAAAGTTCGAGATATAGATTGA
```

Construct length: 486 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacAGGAGCAAGCTTGAAG
CTCAGCC
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTCAATCTATATCTCGAACTTTAAAAGCC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
RSKLEAQPFAHLTINATDIPSGSHKVSLSSWYHDRGWAKISNMTFSNGKLIVNQDGFYYLYANICFRHHETSGDLATEYLQLMVYV
TKTSIKIPSSHTLMKGGSTKYWSGNSEFHFYSINVGGFFKLRSGEEISIEVSNPSLLDPDQDATYFGAFKVRDID
```

Untagged Protein Length: 161 aa

Lymphotoxin beta

Construct: GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CAGGATCAGGGAGGACTGGTAACGGAGACGGCCGACCCCGGGGCACAGGCCCAGCAAGGACTGGGGTTTCAGAAGCTGCCAGAGGA
GGAGCCAGAAACAGATCTCAGCCCCGGGCTCCCAGCTGCCCACCTCATAGGCGCTCCGCTGAAGGGGCAGGGGCTAGGCTGGGAGA
CGACGAAGGAACAGGCGTTTCTGACGAGCGGGACGCAGTTCTCGGACGCCGAGGGGCTGGCGCTCCCGCAGGACGGCCTCTATTAC
CTCTACTGTCTCGTCGGCTACCGGGGCCGGGCGCCCCCTGGCGGCGGGGACCCCCAGGGCCGCTCGGTCACGCTGCGCAGCTCTCT
GTACCGGGCGGGGGGCGCCTACGGGCCGGGCACTCCCGAGCTGCTGCTCGAGGGCGCCGAGACGGTGACTCCAGTGCTGGACCCGG
CCAGGAGACAAGGGTACGGGCCTCTCTGGTACACGAGCGTGGGGTTCGGCGGCCTGGTGCAGCTCCGGAGGGGCGAGAGGGTGTAC
GTCAACATCAGTCACCCCGATATGGTGGACTTCGCGAGAGGGAAGACCTTCTTTGGGGCCGTGATGGTGGGGTGA
```

Construct length: 591 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCAGGATCAGGGAGGAC
TGGTAACGG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTCACCCCACCATCACGGCC
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QDQGGLVTETADPGAQAQQGLGFQKLPEEEPETDLSPGLPAAHLIGAPLKGQGLGWETTKEQAFLTSGTQFSDAEGLALPQDGLYY
LYCLVGYRGRAPPGGGDPQGRSVTLRSSLYRAGGAYGPGTPELLLEGAETVTPVLDPARRQGYGPLWYTSVGFGGLVQLRRGERVY
VNISHPDMVDFARGKTFFGAVMVG
```

Untagged Protein Length: 196 aa

FAS Ligand

Construct: GWF-RBS-MetAlaHis6 - POI - Stop-g-GWR

Construct sequence

```
CAGATGCACACAGCATCATCTTTGGAGAAGCAAATAGGCCACCCCAGTCCACCCCCTGAAAAAAAGGAGCTGAGGAAAGTGGCCCA
TTTAACAGGCAAGTCCAACTCAAGGTCCATGCCTCTGGAATGGGAAGACACCTATGGAATTGTCCTGCTTTCTGGAGTGAAGTATA
AGAAGGGTGGCCTTGTGATCAATGAAACTGGGCTGTACTTTGTATATTCCAAAGTATACTTCCGGGGTCAATCTTGCAACAACCTG
CCCCTGAGCCACAAGGTCTACATGAGGAACTCTAAGTATCCCCAGGATCTGGTGATGATGGAGGGGAAGATGATGAGCTACTGCAC
TACTGGGCAGATGTGGGCCCGCAGCAGCTACCTGGGGGGCAGTGTTCAATCTTACCAGTGCTGATCATTTATATGTCAACGTATCTG
AGCTCTCTCTGGTCAATTTTGAGGAATCTCAGACGTTTTTCGGCTTATATAAGCTCTAA
```

Construct length: 489 bases

Forward Primer Sequence

```
ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggcacatcaccaccaccatcacCAGATGCACACAGCAT
CATCTTTGG
```

GWF-RBS-MetAlaHis6

Reverse Primer Sequence

```
ggggaccactttgtacaagaaagctgggtctcaTTAGAGCTTATATAAGCCGAAAAACG
```

Stop-g-GWR

Restriction sites: not any of BamHI, EcoRI, HindIII, NcoI, NdeI, NotI, SalI, XhoI

Amino Acid Sequence (Untagged)

```
QMHTASSLEKQIGHPSPPPEKKELRKVAHLTGKSNSRSMPLEWEDTYGIVLLSGVKYKKGGLVINETGLYFVYSKVYFRGQSCNNL
PLSHKVYMRNSKYPQDLVMMEGKMMSYCTTGQMWARSSYLGAVFNLTSADHLYVNVSELSLVNFEESQTFFGLYKL
```

Untagged Protein Length: 162 aa

Appendix XTAL: published structures of TNF ligand and receptors

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Reference
Ligand						
APRIL	1U5X	0.1 M citric acid, 1.0 M licl, pH 5.0	1.8	I 2 3	21.4 (25.6)	[262]
APRIL	1U5Y	0.1 M TRIS pH 8.0, 1M licl	2.3	C 1 2 1	20.1 (26.8)	[262]
APRIL	1U5Z	0.1 m TRIS, pH 8.5, 0.5 m KSCN	2.4	I 2 3	21.9 (26.5)	[262]
BAFF	1JH5	5 mM b-mercaptoethanol, 50 mM bicine, ph 9.0, 150 mM NaCl, 35%-38% dioxane	3.0	P 6_3 2 2	23.6 (25.2)	[23]
BAFF	1KD7	8% PEG 4000, 0.1 M sodium acetate pH 4.5	2.8	P 6_5	21.7 (25.0)	[263]
BAFF	1KXG	25% dioxane, 25 mM $MgCl_2$	2.0	P 6_5	18.9 (20.9)	[264]
CD40L	1ALY	1.4 M Na citrate, 50 mM HEPES pH 7.5	2.0	H 3	22.3 (29.5)	[265]
EDA-A1	1RJ7	20% PEG 3350, 0.2m di-sodium hydrogen phosphate, pH 6.5	2.3	P 1 2_1 1	19 (n/a)	[141]
EDA-A2	1RJ8	0.2M NaCl, 0.1M HEPES, 25% PEG 3350, pH 7.5	2.23	P 1 2_1 1	20.0 (24.1)	[141]
TNF		Ammonium sulphate	2.9	P 3_1 2 1	22.2 (n/a)	[18]
RANKL	1JTZ	11-14% PEG 4000, 16 Mm calcium chloride, 80 Mm HEPES, pH 7.5	2.6	P 2_1 2_1 2_1	23.5 (28.6)	[266]

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Reference
TNFα	4TSV	1.4 M sodium formate	1.8	H3	20.0 (26.2)	[267]
TNFα	5TSW	20% PEG 4000, 0.1 M sodium citrate pH 5.6	2.5	P 2_1 2_1 2_1	20.7 (29.9)	[267]
TNFα R31D mutant	1A8M	Crystallized from 88% $MgSO_4$, 1-2% peg 400, 0.2 M MES, pH 5.5	2.3	P 4_1 2_1 2	21.8 (24.0)	[268]
TNFα	1TNF		2.6	P 4_1 2_1 2	23.0 (n/a)	[269]
TNFβ (LTα)						[270]
TRAIL	1D2Q	Peg 550 MME, BICINE pH 9.0, cadmium chloride	2.8	P 63	20.9 (28.2)	[271]
TRAIL	1DG6	Well 32% MPD; drop 20MM TRIS HCl pH 7.5, 1% MPD	1.3	H 3 2	14.1 (19.7)	[20]
Receptor						
BAFF-R	1OSX	n/a	n/a	n/a	Solution structure	[28]
HVEM	1JMA	1.8M ammonium sulphate, 5% PEG 400, TRIS-HCl 100 mM PH 8.0	2.65	P 3_1 2 1	23.6 (25.9)	[272]
TACI (CRD2)	1XUT	n/a	n/a	n/a	Solution structure	[139]
TNF-R1	1NCF		2.25	P 4_1 2_1 2	19.3 (24.9)	[273]
TNF-R1	1EXT	pH 3.7, $MgSO_4$	1.85	P $_{21}$ 2_1 2_1	20.3 (24.3)	[274]
TNF-R1	1FT4	MPD, ammonium acetate, TRIS, pH 8.5	2.9	P 4_1 2_1 2	27.4 (n/a)	[275]

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Reference
Ligand/Receptor Complex						
APRIL/BCMA	1XU1	70% MPD, 0.1 M HEPES, pH 7.5	1.9	P 2_1 2_1 2_1	16.7 (20.3)	[139]
APRIL/TACI	1XU2	0.1M MES, 5% PEG 8000, 10% PEG 1000, pH 5.0	2.35	P 6_1	17.8 (21.3)	[139]
BAFF/BCMA	1OQD	Dioxane, ph 9.0	2.6	P 6_3 2 2	24.0 (26.0)	[30]
BAFF/BAFF-R	1OQE	Dioxane, ph 9.0	2.5	P 6_3 2 2	24.4 (25.9)	[30]
BAFF/BAFF-R hairpin	1OSG	20% PEG 3350, 1.0 M licl, 0.1M sodium cacodylate, pH 6.7	3.0	P 6_5	21.3 (28.6)	[28]
BAFF/BAFF-R	1OTZ/1P0T	Sodium formate, pH 5.5	3.3	P 1 2_1 1	21.8 (22.6)	[29]
TNF/TNF-R1	1TNR		2.85	I 2 3	16.0 (n/a)	[12]
TRAIL/DR5	1D4V	25% ethylene glycol, 0.1% n-octyl-beta-glucoside	2.2	P 3 2 1	22.1 (27.0)	[13]
TRAIL/DR5	1D0G	15% PEG 8000, 10% Ethylene Glycol, 0.2M Ammonium sulphate, 0.1M Tris-HCl, pH 7.5	2.4	P 2_1 2_1 2_1	22.2 (26.7)	[39]
TRAIL/DR5	1DU3					[38]

Appendix PCR

All PCR reactions were carried out using the KOD-HiFi polymerase and standard conditions as described below.

PCR reactions were carried out with the following reagents:

28.6 µl PCR Grade Water

5 µl 10X Buffer #1 For KOD HiFi DNA Polymerase
5 µl dNTPs (final concentration 0.2 mM)
2 µl $MgCl_2$ (final concentration 1 mM)
1 µl template DNA
4 µl 5' primer (5 pmol/µl, final concentration 0.4 µM)
4 µl 3' primer (5 pmol/µl, final concentration 0.4 µM)
0.4 µl KOD HiFi DNA Polymerase (2.5 U/µl) 50

(50 µl total reaction mixture)

The following thermal cycler programmes were used for amplification of target material:

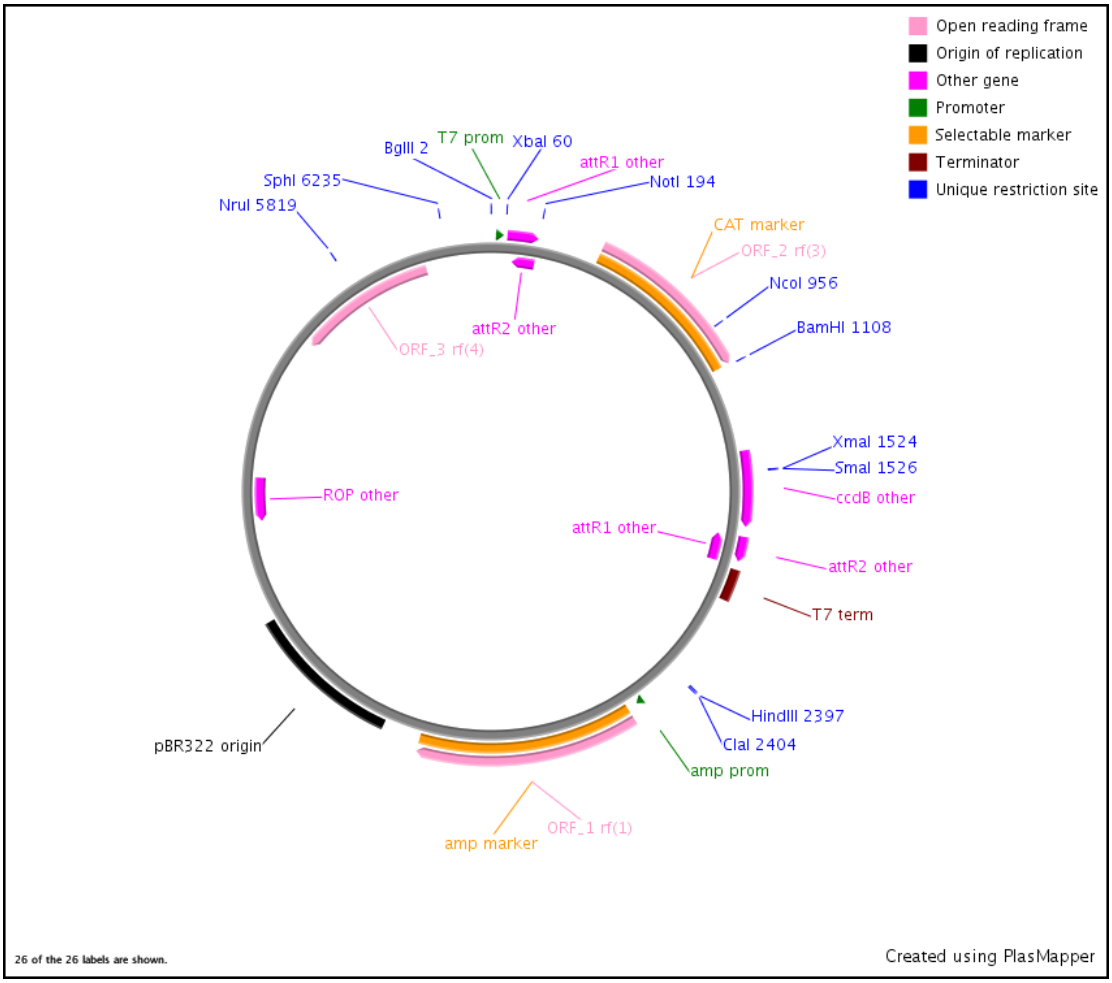
Phage and plasmid DNA templates				
Cycling parameters	0.5 kbp target DNA	1–2 kbp target DNA	3–4 kbp target DNA	5–6 kbp target DNA
Denature	15 sec 98 °C	15 sec 98 °C	15 sec 98 °C	30 sec 94 °C
Anneal	1–30 sec 68 °C	2 sec (Tm-5) °C	5 sec (Tm-5) °C	30 sec (Tm-5) °C
Extend	none[a]	20 sec 72 °C	40 sec 72 °C	60 sec 72 °C
No. cycles	25	25	25–30	30

	Genomic DNA templates	**cDNA templates**
Cycling Parameters	Up to 2 kbp	Up to 2 kbp
Denature	15 sec 95°C	20 sec 98°C
Anneal	30 sec (T_m-5)°C	30 sec 68°C
Extend	30–60 sec 72°C	none[a]
No. cycles	30	30

Appendix Vectors

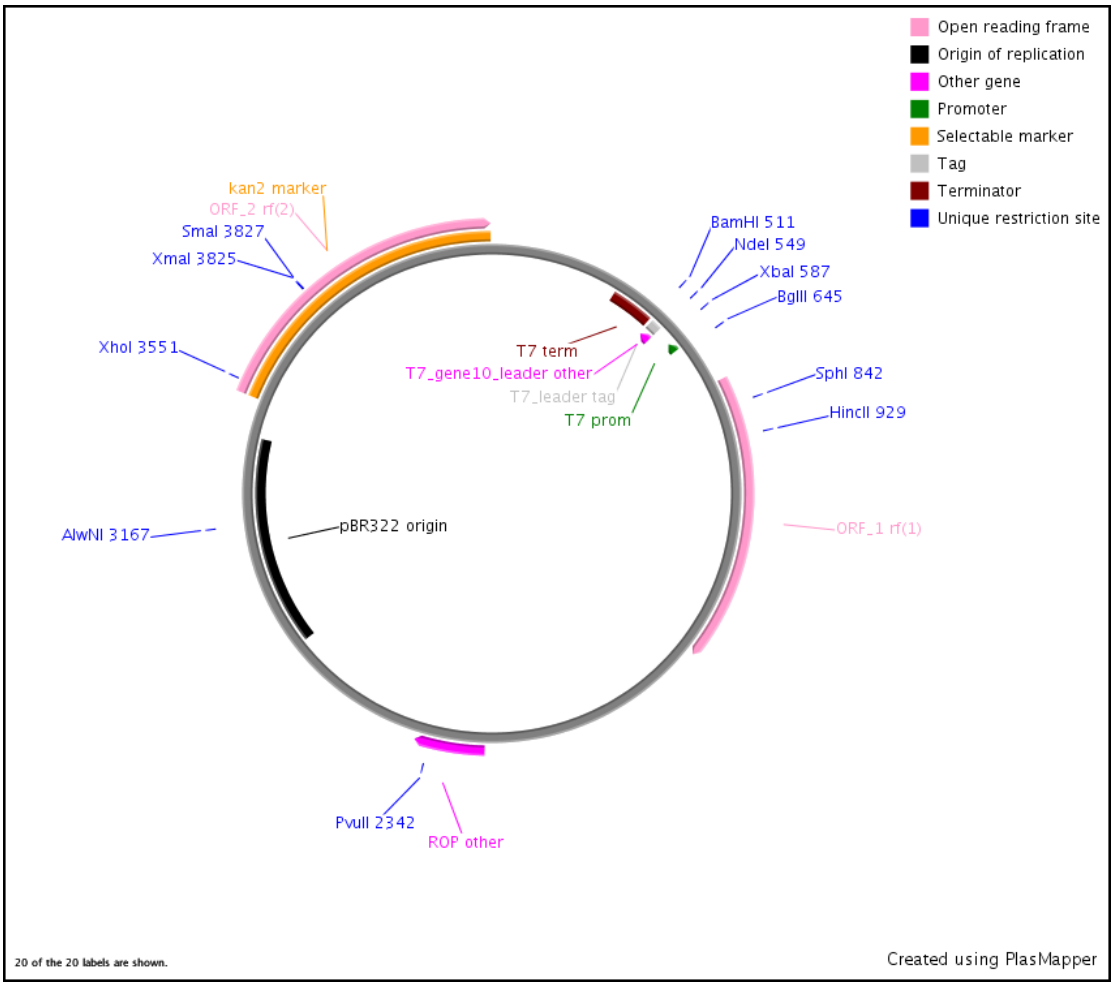
pDEST14

This is a standard T7 promoter based Gateway™ destination vector. This vector was used for all TNF targets that were to be produced in *E.coli*.



pET9c

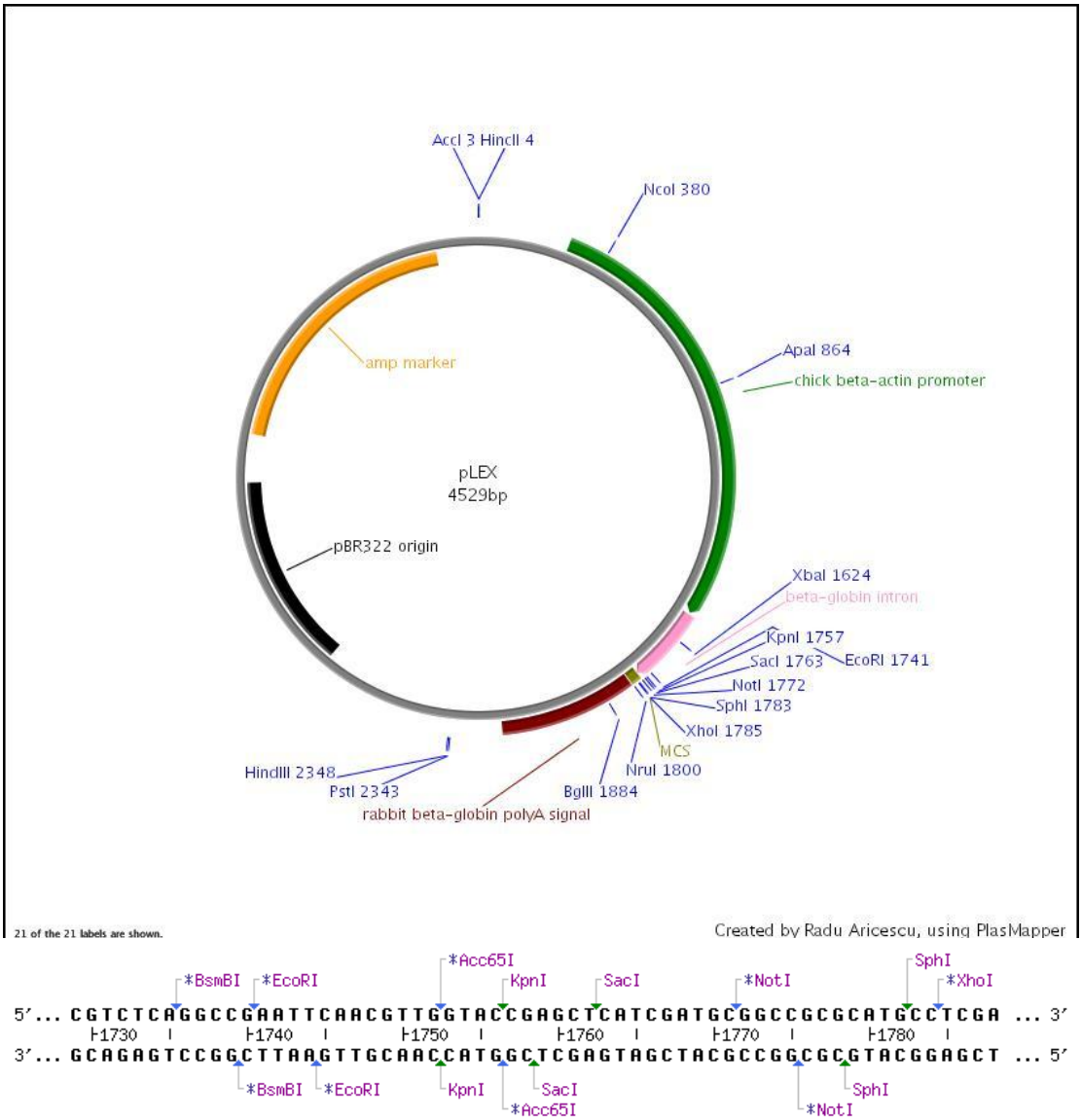
This vector was used for the TRAIL construct, which was a gift of J. Mongkolsapaya.



pHLsec

This vector was used for the second LTβR construct to be expressed in 293-T cells. The pHLsec vector is based on the pLEX backbone with the following changes to the multiple cloning site for use with secreted protein products as designed by A. Radu Aricescu:

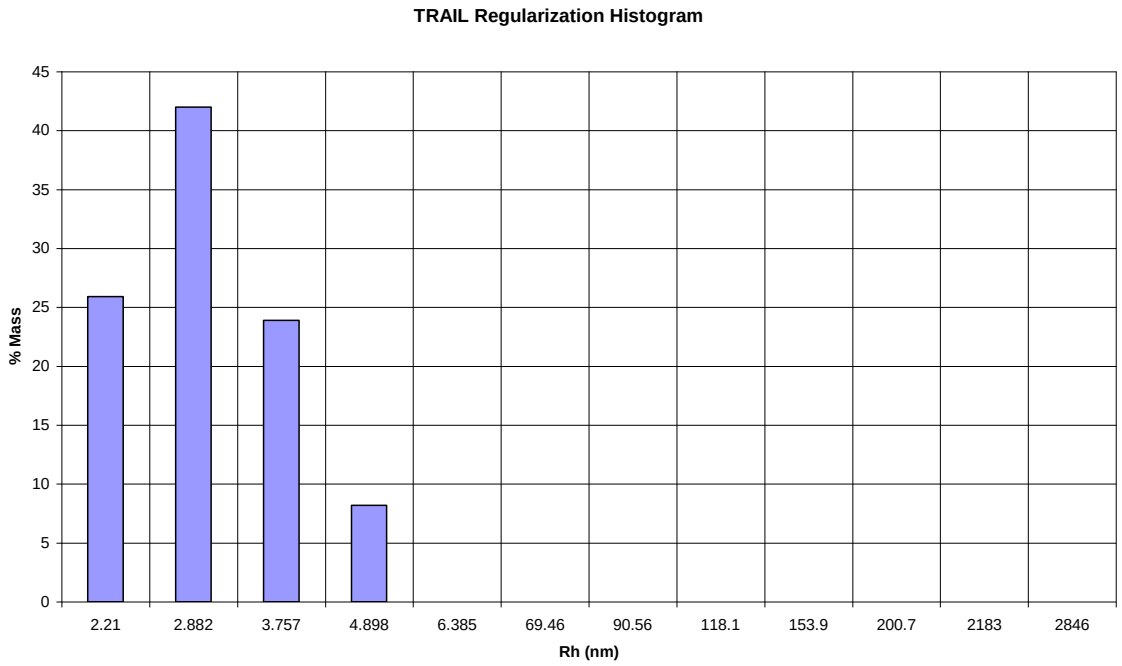
1. A Kozak sequence; 2. A secretion signal sequence; 3. A C-terminal K_His6 tag. Constructs can be cloned into this vector, preserving and making use of the features listed above, by using the AgeI (compatible with XmaI, BspEI, NgoM IV) and KpnI (isoschizomer Acc65I is compatible with BsiWI and BsrGI) sites. A map of this vector with the multiple cloning site is shown below.



Appendix DLS

All DLS experiments were performed at 1 mg ml^{-1} protein concentration in 50 mM Tris-HCl, 150 mM NaCl buffer, at 21 ºC using a Dynapro Light Scattering instrument (Chapter 1.2.11). The regularization histogram from each experiment, indicating the mass distribution as a function of the calculated radius of hydration, of the particles in solution is shown with the estimated molecular weight of the top mass peak.

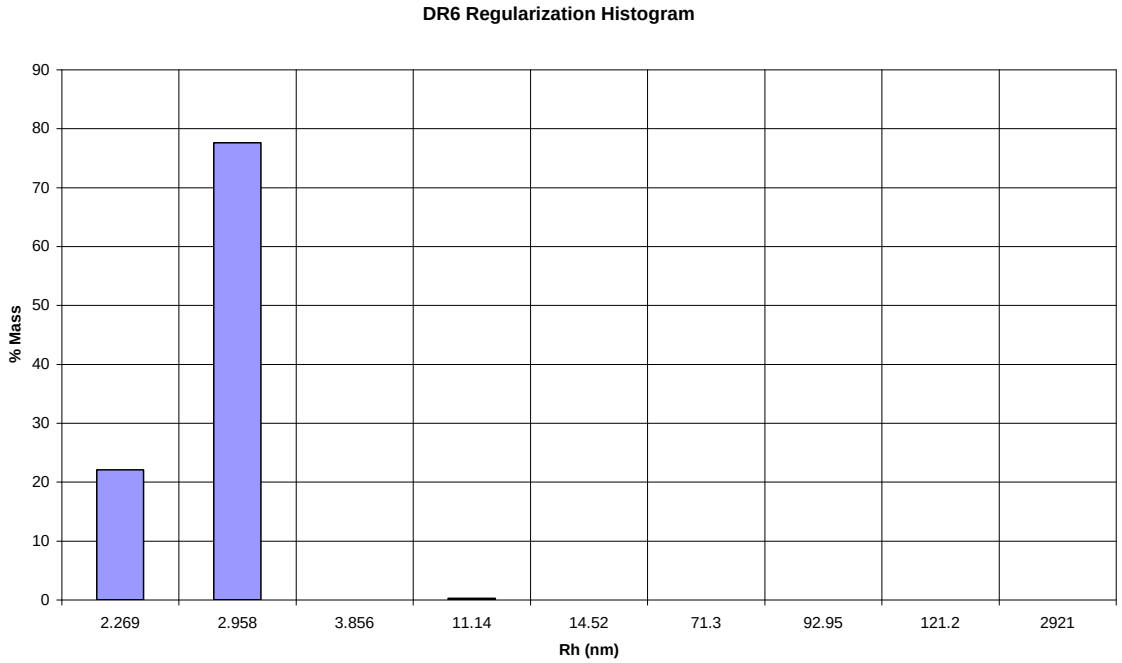
TRAIL



Peak 2.882, MW ≈ 71.7 kDa

Polydispersity = 23.6 %

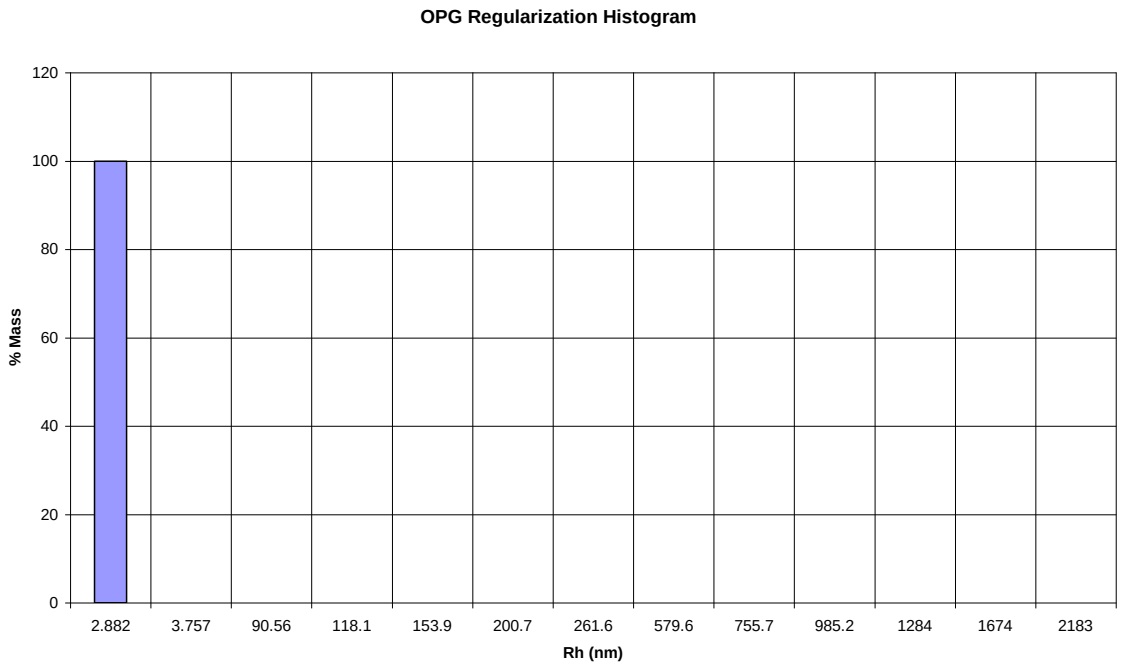
DR6



Peak 2.958, MW ≈ 40 kDa

Polydispersity = 7.6 %

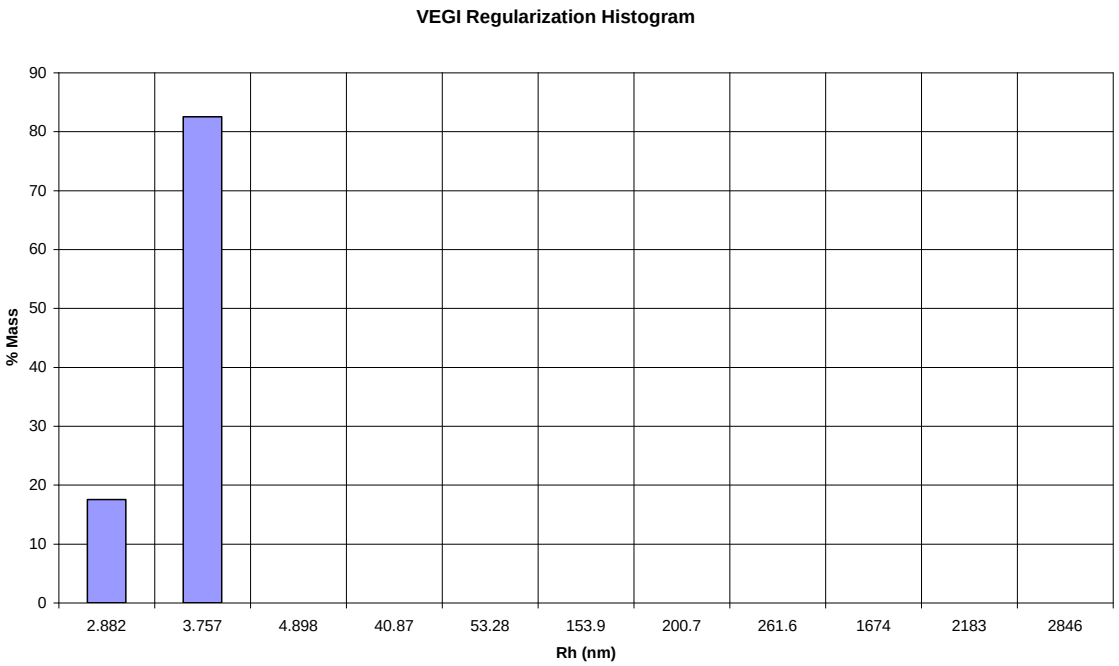
OPG



Peak 2.882, MW ≈ 40.6 kDa

Polydispersity = 0 %

VEGI

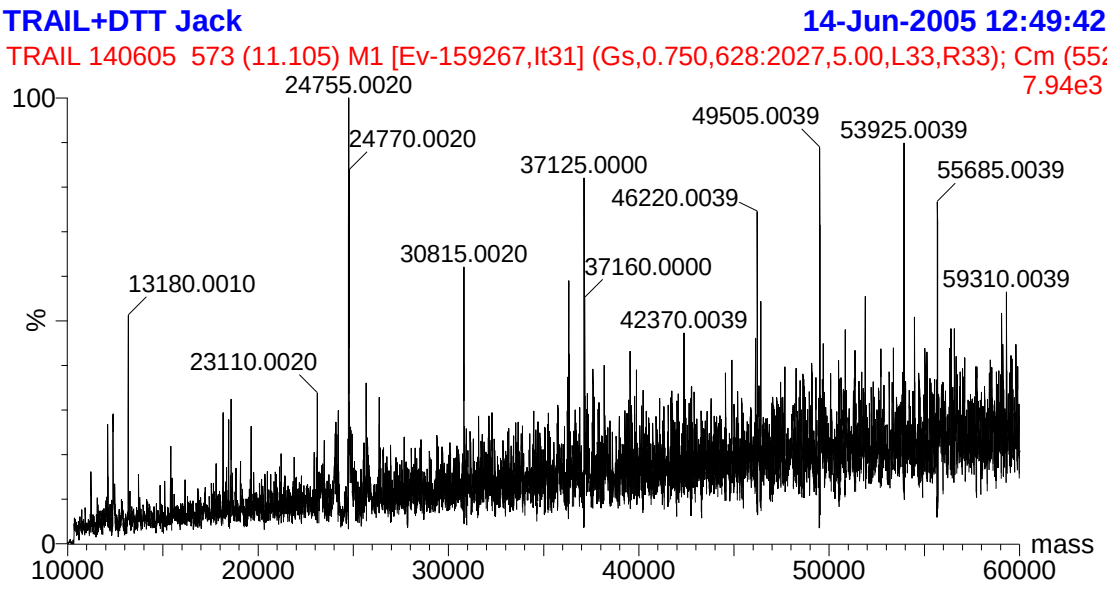


Peak 3.757, MW ≈ 71.0

Polydispersity = 6.7 %

Appendix MS: Mass spectroscopy data for TNF targets

TRAIL

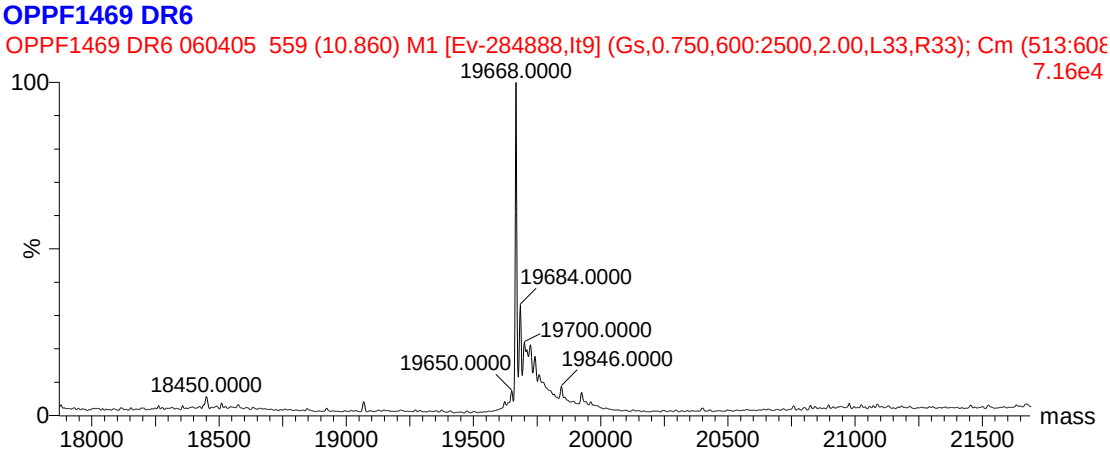


Calculated molecular weight (- N-term Met) - 20,386.72 Da

Experimentally determined molecular weight - 24,755 Da

Potential dimer species at 49,505 Da.

DR6



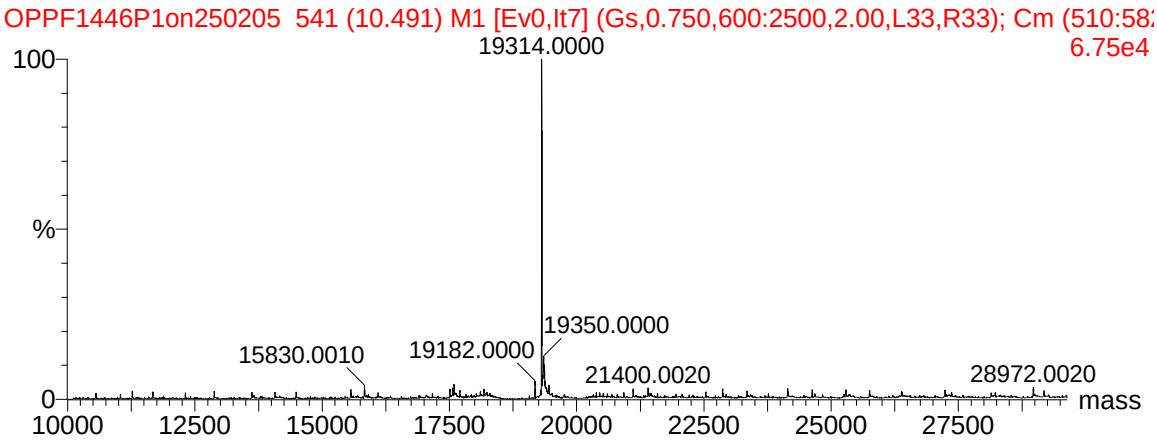
Calculated molecular weight (- N-term Met) – 19,686.33 Da

Experimentally determined molecular weight – 19, 668 Da

Discrepancy due to formation of nine cysteine bridges

VEGI

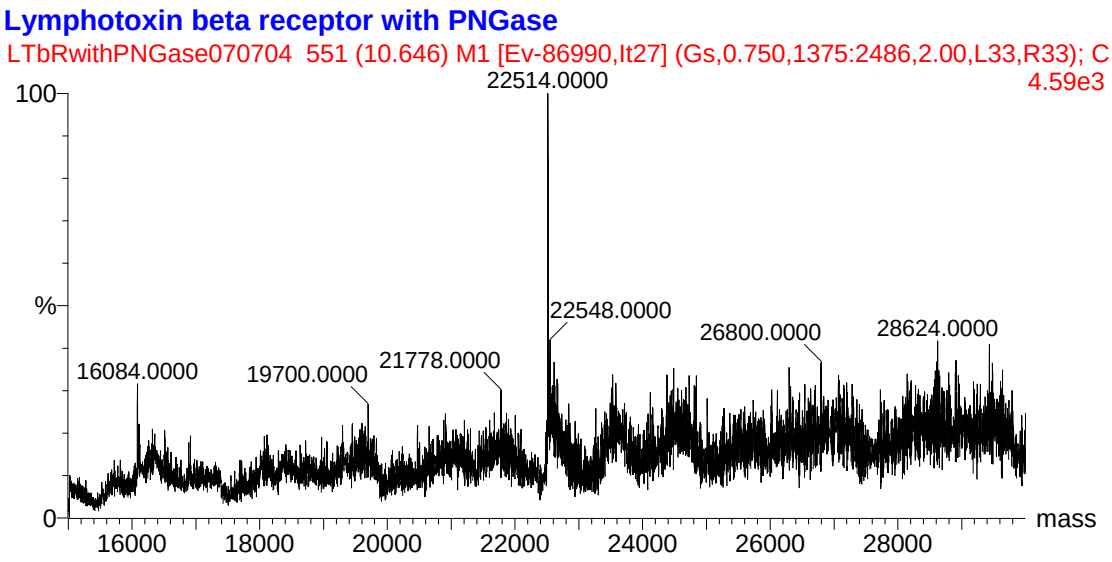
OPPF1446 P1



Calculated molecular weight (- N-term Met) – 19,315.86 Da

Experimentally determined molecular weight – 19, 314 Da

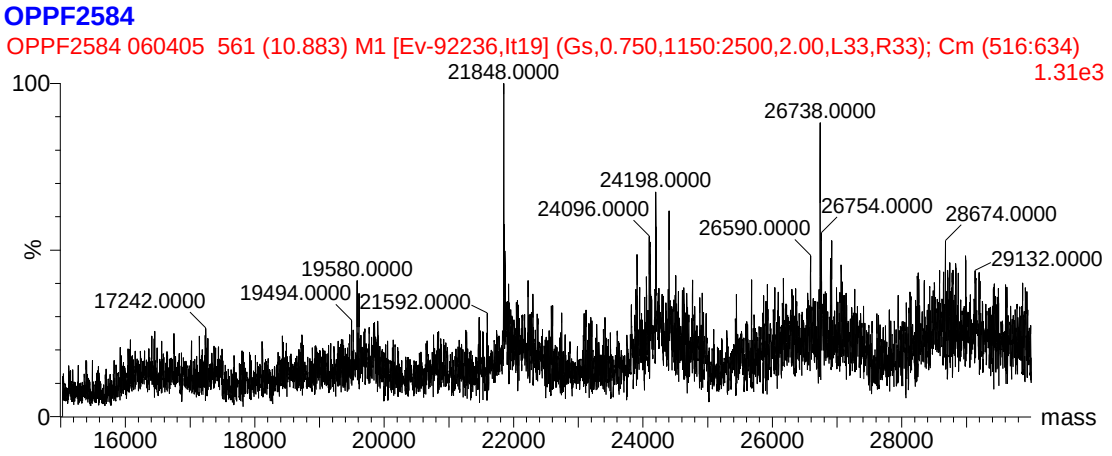
LTβR



Calculated molecular weight (-N-term Met) – 22,516 Da

Experimentally determined molecular weight – 22,514 Da

Mass spectroscopy analysis of LTβR crystals:

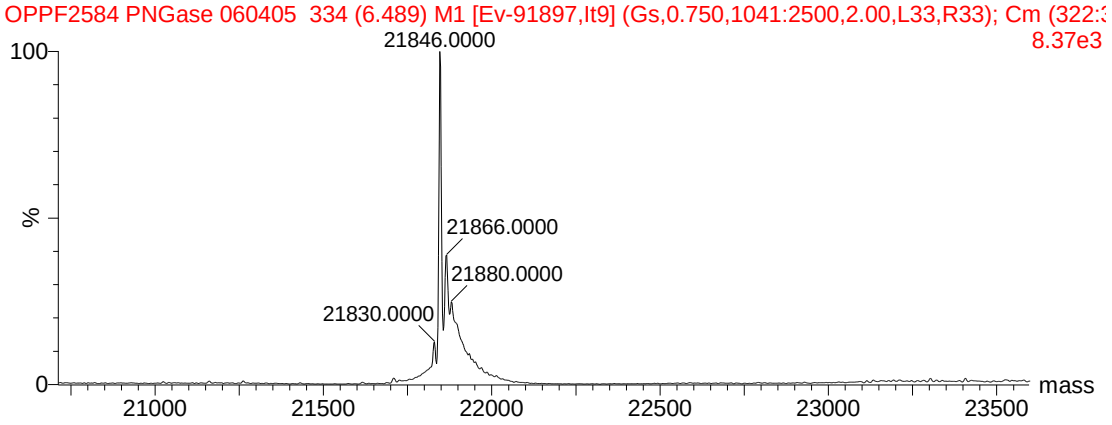


Calculated molecular weight (- N-term Met) – 21,866.4 Da

Experimentally determined molecular weight – 21, 848 Da

The 18 Da discrepancy is due to formation of nine cysteine bridges. Other peaks at 26,738 Da correspond to glycosylated product.

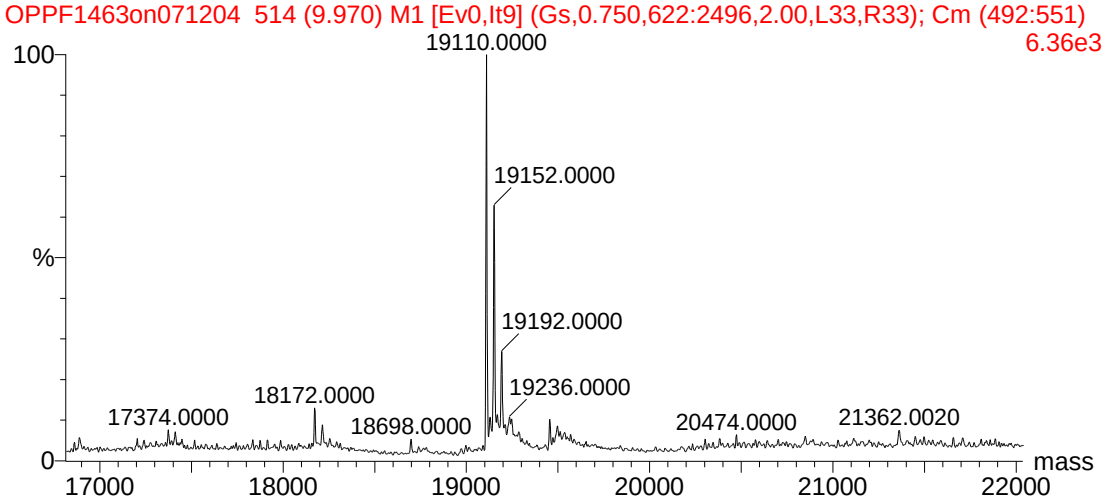
OPPF2584 PNGase



With PNG-ase treatment the determined mass is 21, 846 Da which implies there are two major sites of glycosylation on the protein.

Osteoprotegerin

OPPF1463



Calculated molecular weight (- N-term Met) – 19,469 Da

Experimentally determined molecular weight – 19,110 Da

Discrepancy of 359 Da is most likely due to the loss of amino acids from N or C-terminus.

Appendix TCR: Published structures of T-Cell Receptors and T-Cell Receptor/MHC complexes

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Average B-factor	Reference
TCR α							
1934.4	1AC6	5 M Sodium Formate, pH 7.7	2.3	P 2_1 2_1 2_1	16.1 (27.9)	24.1	[276]
2C Vα	1I9E	Di-Sodium Tartrate, NaCl, Imidazole/HCl	2.5	I 4_1 2 2	20.3 (23.3)	25.87	[277]
Vα 2.6, Jα38	1B88	Sodium Formate, PH 7.0	2.5	P 2_1 2_1 2_1	22.4 (32.4)	35.8	[278]
Vα85.33	1H5B	100 mM Sodium Citrate/HCl, 1.4-1.7 M Lithium Chloride, pH 5.0-6.0	1.85	P 3_2 2 1	22.4 (24.5)	39.6	[279]
TCR β							
Vβ8.2, Jβ2.1, Cβ1	1BEC	n/a	1.7	C2	19.8 (n/a)	23.9	[280]
14.3 TCRβ, Sec3	1JCK	20% PEG1000, 0.1 M $MgCl^2$, 0.025% Agarose, 0.1M TRIS-HCl, pH 6.5	3.5	P 1 2_1 1	23.8 (32.8)	33.4	[220]
Vβ2.1, Dβ2.1, Jβ2.3, Cβ2, TCR β chain, hVβ 2.1	1KTK	15 % PEG 8000, 0.2 M $MgCl_2$, 0.1 M TRIS pH 8	3	P1 2_1 1	32.6 (33.4)	76.1	[281]
14.3 D TCR β chain	1L0X,1L0Y	15 % PEG 8000, 0.2 M $MgCl_2$, 0.1 M TRIS pH 8	2.8	P 1 2_1 1	23.2 (27.8)	40.0	[281]
14.3 TCR β	1SBB	n/a	2.4	P 1 2_1 1	22.8 (30.9)	35.8	[282]
TCR δ							
Vδ3, Dδ2, Jδ1	1TVD	0.65-0.75 M Sodium Citrate 0.1M, Sodium Cacodylate, pH 6.5	1.9	$P2_1$ 2_1 2_1	16.1 (22.4)	14.8	[283]
TCR αβ							
Vα3, Jα58, Vβ8.2, Dβ2, Jβ2.4,Cβ2	1TCR	n/a	2.5	C1 2 1	20.8 (28.5)	35.8	[219]
Vα8, Jα19, Vβ5.2, Dβ2, Jβ2.7, Cβ2	1NFD	8-10% PEG 8000, 0.2 M KCl, 0.1 M Acetate, pH5.5	2.8	P1 2_1 1	30.9 (36.9)	26.59	[284]
Vα2.3, Jα10, Vβ2, Dβ2, Jβ2.3	1KB5	15% PEG 6000, 100 mM HEPES, pH 6.9-7.5, 200 mM NaCl, 0.1% NaN_3	2.5	$P2_1$ 2_1 2	21.9 (31.5)	48.9	[285]
LC13 TCR	1KCG	12-15 % PEG 8000, 200 mM NaCl, pH 6.5	1.5	P 2_1 2_1 2_1	21.1 (23.4)	23.8	[286]

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Average B-factor	Reference
TCR Cα, Cβ	2BNU	n/a	1.4	P 1 2_1 1	24.5 (n/a)	24.3	[287]
TCR γδ							
Vγ9, Vδ2	1HXM	8% PEG 4000, 0.2 M Li_2SO_4, 0.1 M TRIS-HCl, 20% Glycerol	3.12	P 1 2_1 1	29 (34.3)	42.3	[288]
TCR/MHC							
Vα3, Jα58, Vβ8.2, Dβ2, Jβ2.4,Cβ2 MHC H-2K^b Peptide – dEV8	2CKB	0.2 M TRIS Acetate pH 7.2, 0.1 M Sodium Chloride, 12 % PEG 4000	3.2	$P2_1$ 2_1 2	22.1 (32.2)	43.2	[181]
Vα2.3, Jα24, Vβ12.3, Dβ2.1, Jβ2.7, Cβ2, HLA-A2, HTLV	1A07	10% PEG 8000, 100 mM Mg Acetate, 50 mM Na Cacodylate, pH 6.5	2.6	C1 2 1	24.5 (32)	42.4	[179]
Vβ8.2, Vα2, MHCII-I-A^k	1D9K	PEG 8000, Sodium chloride, TRIS	3.2	P 2_1 2_1 2	24.7 (29.3)	62.6	[289]
Vβ17, Vα10.2, HLA-A2	1OGA	14% PEG8000, 50MM MES PH 6.5	1.4	P 1 2_1 1	21.8 (23.1)	31.5	[184]
A6-TCR, HLA-A2, HTLV 1	1QRN	50 mM MOPS pH 7.0, 75 mM $MgSO_4$, 13% PEG 8000	2.8	C 1 2 1	21.6 (27.3)	50.1	[290]
A6-TCR, HLA-A2, HTLV 1	1QSE	50 mM MOPS, pH 7.0, 75 mM $MgSO_4$,13% PEG 8000	2.8	C 2	24.3 (29.0)	66.1	[290]
A7-TCR, HLA-A2, HTLV 1	1BD2	12% PEG 8000, 20 mM MOPS, 100 mM Magnesium Acetate, pH 7.1	2.5	P 2_1 2_1 2_1	23.8 (31.2)	41.1	[182]
BM3.3 TCR, H-2K^b, pbm1	1FO0	10 % PEG 6000, 0.1 M HEPES, pH 7.0 0.25 M Magnesium Acetate, 0.25 M NaCl	2.5	P 2_1 2_1 2	21.8 (27.6)	57.93	[291]
HA1.7 αβ, HLA-DR1, HA	1FYT	11% PEG 8000, 1 M NaCl, 100 mM HEPES	2.6	C 1 2 1	22.1 (25.5)	38.7	[292]
2C TCR, HLA-2K^b, SIYR	1G6R	0.2 M TRIS/Acetate, 10% PEG 6000	2.8	P 2_1 2_1 2	29.8 (32.7)	63.1	[293]
HA1.7, HLA-DR4, HA	1J8H	13% PEG 8000, 1 M NaCl, 100 mM HEPES	2.4	C 1 2 1	21.1 (24.6)	46.1	[221]

	PDB Accession Code	Crystallization conditions	Resolution (Å)	Space-Group	R-factor (R-free) (%)	Average B-factor	Reference
KB5-C20 TCR, HLA-2K^b, PKB1	1KJ2	17-19% PEG 6000, 0.1 M MES, 0.1 M NaCl, 0.1M Magnesium Acetate	2.7	P 1 2_1 1	22.0 (27.8)	45.65	[294]
AHIII 12.2 TCR, HLA-A2, p1049	1LP9	PEG 3350, 25 mM MES, pH 6.5	2	P 1 2_1 1	22.1 (25.3)	14.2	[295]
TCR LC13, HLA-B8, FLRGRAYGL	1MI5	PEG, Ammonium Acetate, pH 7.5	2.5	P 2_1 2_1 2_1	22.6 (28.8)	39.6	[296]
2C/H-2K(bm3)-dEV8	1MWA	0.1 M TRIS Acetate, 12% PEG 4000, 18% glycerol, pH 6.8	2.4	P 2_1 2_1 2	28.4 (31.3)	47.5	[297]
BM3.3 TCR, HLA-2K^b, VSV8	1NAM	13- 17 % PEG 6000, Mg Acetate 0.1 M, NaCl 0.1 M, HEPES 0.1 M, pH 7	2.7	P 4_3 2_1 2	23.0 (29.8)	56.81	[298]
172.10 TCR, I-A^u MHC, MBP	1U3H	21 % PEG 3350, 0.1 M HEPES, 0.2 M $LiSO_4$, pH 7.5	2.42	C 2 2 2_1	23.2 (27.4)	61.7	[299]
OB.1A12 TCR, HLA-DR2, MBP	1YMM	Sodium Tartrate, Ammonium Formate, HEPES, pH 7.0	3.5	F 2 2	27.4 (31.8)	71.5	[195]
TCR γδ G8, MHC-1B T22	1YPZ	n/a	3.4	P 2_1 2_1 2_1	26.4 (33.0)	103.7	[300]
TCR, HLA-A2, SLLMWITQC	2BNQ	n/a	1.7	P 1 2_1 1	22.6 (25.3)	32.1	[287]
TCR, HLA-A2	2BNR	n/a	1.9	P 1 2_1 1	23.1 (n/a)	33.0	[287]

Bibliography

1. Coley, W.B. The Treatment of malignant tumours by repeated inoculations of erysipelas: with a report of ten original cases. *Am. J. Med. Sci.* **105**, 487-511 (1893).
2. Granger, G.A., Williams, T.W. Lymphocyte cytotoxicity in vitro:adaptation and release of a cytotoxic factor. *Nature* **218**, 1253-1254 (1968).
3. Carswel, E.A., Old, L.J., Kassel, R.L. Green, S., Fiore, N., Williamson, B. An endotoxin induced serum factor that causes necrosis of tumours. *Proc. Nat. Acad. Sci.* **72**, 3666-3670 (1975).
4. Screaton, G., Xu, X-N. T cell life and death signalling via TNF-receptor family members. *Curr. Op. Immunol.* **12**, 316-322 (2000).
5. Wallach, D. et al. How are the regulators regulated? The search for mechanisms that impose specificity on induction of cell death and NF-kappaB activation by members of the TNF/NGF receptor family. *Arthritis Res* **4 Suppl 3**, S189-96 (2002).
6. Jones, E.Y. The tumour necrosis factor family: life or death choices. *Curr. Op. Struc. Biol.* **10**, 644-648 (2000).
7. Thiel, L.E., Boyle, W.J. & Penninger, J.M. RANK-L and RANK: T cells, bone loss and mammalian evolution. *Annu. Rev. Immunol.* **20**, 795-823 (2002).
8. Gravestein, L.A. & Borst, J. Tumor necrosis factor receptor family members in the immune system. *Semin Immunol* **10**, 423-34 (1998).
9. Iyer, S., Yamauchi, P. & Lowe, N.J. Etanercept for severe psoriasis and psoriatic arthritis: observations on combination therapy. *Br J Dermatol* **146**, 118-21 (2002).

10. Komata, T., Tsuchiya, N., Matsushita, M., Hagiwara, K. & Tokunaga, K. Association of tumor necrosis factor receptor 2 (TNFR2) polymorphism with susceptibility to systemic lupus erythematosus. *Tissue Antigens* **53**, 527-33 (1999).

11. Bodmer, J.L., Schneider, P. & Tschopp, J. The molecular architecture of the TNF superfamily. *Trends in Biochemical Sciences* **27**, 19-26 (2002).

12. Banner, D.W. et al. Crystal structure of the soluble human 55 kd TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* **73**, 431-45 (1993).

13. Mongkolsapaya, J. et al. Structure of the TRAIL-DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nat Struct Biol* **6**, 1048-53 (1999).

14. Chan, F.K. et al. A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling. *Science* **288**, 2351-4 (2000).

15. Locksley, R.M., Killeen, N., Lenardo M.J. The TNF and TNF Receptor Superfamilies: Integrating Mammalian Biology. *Cell* **104**, 487-501 (2001).

16. Chen, Y. Mutations within a furin consensus sequence block proteolytic release of ectodysplasin-A and cause X-liked hypohidrotic ectodermal dysplasia. *Proc. Nat .Acad. Sci.* **98**, 7128-7223 (2001).

17. Tanaka, M. Downregulation of Fas ligand by shedding. *Nat. Med.* **4**, 31-36 (1998).

18. Jones, E.Y., Stuart, D.I., Walker, N.P.C. Structure of tumour necrosis factor. *Nature* **338**, 225-22 (1989).

19. Liu, Y., Xu, L., Opalka, N., Kappler, J., Shu, H-B., Zhang, G. Crystal Structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* **108**, 383-394 (2002).

20. Hymowitz, S.G., O'Connell, M. P., Ultsch, M. H., De Vos, A. M., Kelley, R. F. A Unique Zinc-Binding Site Revealed by a High Resolution X-Ray Structure of Homotrimeric Apo2L/Trail. *Biochemistry*, 633 (2000).

21. Mackay, F. & Browning, J.L. BAFF: a fundamental survival factor for B cells. *Nat Rev Immunol* **2**, 465-75 (2002).

22. Zhukovsky, E.A. et al. TNF ligands: is TALL-1 a trimer or a virus-like cluster? *Nature* **427**, 413-4; discussion 414 (2004).

23. Liu, Y. et al. Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* **108**, 383-94 (2002).

24. Kishore, U. & Reid, K.B. C1q: structure, function, and receptors. *Immunopharmacology* **49**, 159-70 (2000).

25. Chirpich, T.P. Rates of protein evolution: a function of amino acid composition. *Science* **188**, 1022-3 (1975).

26. Idriss, H.T. & Naismith, J.H. TNF alpha and the TNF receptor superfamily: structure-function relationship(s). *Microsc Res Tech* **50**, 184-95 (2000).

27. Pitti, R.M. et al. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**, 699-703 (1998).

28. Gordon, N.C. et al. BAFF/BLyS receptor 3 comprises a minimal TNF receptor-like module that encodes a highly focused ligand-binding site. *Biochemistry* **42**, 5977-83 (2003).

29. Kim, H.M. et al. Crystal structure of the BAFF-BAFF-R complex and its implications for receptor activation. *Nat Struct Biol* **10**, 342-8 (2003).

30. Liu, Y. et al. Ligand-receptor binding revealed by the TNF family member TALL-1. *Nature* **423**, 49-56 (2003).

31. Naismith, J.H., Devine, T.Q., Kohno, T., Sprang, S.R. Structures of the extracellular domain of the type I tumor necrosis factor receptor. *Structure* **4**, 1251-1262 (1996).

32. Yue, T.L. et al. TL1, a novel tumor necrosis factor-like cytokine, induces apoptosis in endothelial cells. Involvement of activation of stress protein kinases (stress-activated protein kinase and p38 mitogen-activated protein kinase) and caspase-3-like protease. *J Biol Chem* **274**, 1479-86 (1999).

33. Tunggal, P. Laminins: structure and genetic regulation. *Microscope Research Techniques* **51**, 214-227 (2000).

34. Bieri, S., Djordjevic, J.T., Daly, N.L., Smith, R. & Kroon, P.A. Disulfide bridges of a cysteine-rich repeat of the LDL receptor ligand-binding domain. *Biochemistry* **34**, 13059-65 (1995).

35. Daly, N.L., Scanlon, M.J., Djordjevic, J.T., Kroon, P.A. & Smith, R. Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proc Natl Acad Sci U S A* **92**, 6334-8 (1995).

36. Lin, S.L. & Nussinov, R. A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. *Nat Struct Biol* **2**, 835-7 (1995).

37. Ward, C.W., Hoyne, P.A. & Flegg, R.H. Insulin and epidermal growth factor receptors contain the cysteine repeat motif found in the tumor necrosis factor receptor. *Proteins* **22**, 141-53 (1995).

38. Cha, S.-S., Sung, B.J., Kim, Y.A., Song, Y.L., Kim, H.J, Kim, S., Lee, M.S., Oh, B.H. Crystal structure of TRAIL-DR5 complex identifies a critical role of

the unique frame insertion in conferring recognition specificity. *J. Biol. Chem.* **275**, 31171-31177 (2000).

39. Hymowitz, S.G. et al. Triggering cell death: the crystal structure of Apo2L/TRAIL in a complex with death receptor 5. *Mol Cell* **4**, 563-71 (1999).
40. Park, Y.C., Burkitt, V., Villa, A.R., Tong, L. & Wu, H. Structural basis for self-association and receptor recognition of human TRAF2. *Nature* **398**, 533-8 (1999).
41. Thompson, J.S. et al. BAFF-R, a newly identified TNF receptor that specifically interacts with BAFF. *Science* **293**, 2108-11 (2001).
42. Yan, M. et al. Identification of a novel receptor for B lymphocyte stimulator that is mutated in a mouse strain with severe B cell deficiency. *Curr Biol* **11**, 1547-52 (2001).
43. Schulze-Osthoff, K., Ferrari, D., Los, M., Wesselborg, S. & Peter, M. Apoptosis signaling by death receptors. *Eur J Biochem* **254**, 439-459 (1998).
44. Wallach, D. et al. Cell death induction by receptors of the TNF family: towards a molecular understanding. *FEBS Lett* **410**, 96-106 (1997).
45. Chen, G. & Goeddel, D.V. TNF-R1 signaling: a beautiful pathway. *Science* **296**, 1634-5 (2002).
46. Grell, M. et al. Induction of cell death by tumour necrosis factor (TNF) receptor 2, CD40 and CD30: a role for TNF-R1 activation by endogenous membrane-anchored TNF. *Embo J* **18**, 3034-43 (1999).
47. Anderson, D.M. et al. A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175-9 (1997).
48. Lacey, D.L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165-76 (1998).

49. Papoff, G. et al. An N-terminal domain shared by Fas/Apo-1 (CD95) soluble variants prevents cell death in vitro. *J Immunol* **156**, 4622-30 (1996).

50. Ni, C.Z. et al. Key molecular contacts promote recognition of the BAFF receptor by TNF receptor-associated factor 3: implications for intracellular signaling regulation. *J Immunol* **173**, 7394-400 (2004).

51. Ni, C.Z. et al. Molecular basis for CD40 signaling mediated by TRAF3. *Proc Natl Acad Sci U S A* **97**, 10395-9 (2000).

52. Naismith, J.H., Devine, T. Q., Brandhuber, B. J., Sprang, S. R. Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor. *J. Biol. Chem.* **270**, 13303 (1995).

53. Livnah, O. et al. Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987-90 (1999).

54. Remy, I., Wilson, I.A., Michnick, S.W. Erythropoietin Receptor Activation by a Ligand-Induced Conformation Change. *Science* **283**, 990-993 (1999).

55. Chan, F.K. The pre-ligand binding assembly domain: a potential target of inhibition of tumour necrosis factor receptor function. *Ann Rheum Dis* **59 Suppl 1**, i50-3 (2000).

56. Siegel, R., Freferiksen, J.K., Zacharias, D.A., Chan, F.K-M., Johnson, M., Lynch, D., Tsien, R.Y., Lenardo, M.J. Fas Preassociation Required for Apoptosis Signalling and Dominant Inhibition by Pathogenic Mutations. *Science* **288**, 2354-2357 (2000).

57. Puck, J.M. & Sneller, M.C. ALPS: an autoimmune human lymphoproliferative syndrome associated with abnormal lymphocyte apoptosis. *Semin Immunol* **9**, 77-84 (1997).

58. Jiang, Y., Woronicz, J.D., Liu, W. & Goeddel, D.V. Prevention of constitutive TNF receptor 1 signaling by silencer of death domains. *Science* **283**, 543-6 (1999).

59. Irmler, M. et al. Inhibition of death receptor signals by cellular FLIP. *Nature* **388**, 190-5 (1997).

60. Algeciras-Schimnich, A., Griffith, T.S., Lynch, D.H. & Paya, C.V. Cell cycle-dependent regulation of FLIP levels and susceptibility to Fas-mediated apoptosis. *J Immunol* **162**, 5205-11 (1999).

61. Fulda, S. et al. Activation of the CD95 (APO-1/Fas) pathway in drug- and gamma-irradiation-induced apoptosis of brain tumor cells. *Cell Death Differ* **5**, 884-93 (1998).

62. Tschopp, J., Irmler, M. & Thome, M. Inhibition of fas death signals by FLIPs. *Curr Opin Immunol* **10**, 552-8 (1998).

63. Kataoka, T. et al. FLIP prevents apoptosis induced by death receptors but not by perforin/granzyme B, chemotherapeutic drugs, and gamma irradiation. *J Immunol* **161**, 3936-42 (1998).

64. Strausberg, R.L., Feingold, E.A., Klausner, R.D. & Collins, F.S. The mammalian gene collection. *Science* **286**, 455-7 (1999).

65. SRS7. www.sanger.ac.uk.

66. Apweiler, R. et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-9 (2004).

67. Walhout, A.J. et al. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* **328**, 575-92 (2000).

68. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **58**, 913-49 (1989).

69. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol* **226**, 735-45 (1992).

70. Mayo, C.J. et al. Benefits of automated crystallization plate tracking, imaging, and analysis. *Structure (Camb)* **13**, 175-82 (2005).

71. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr* **50**, 414-23 (1994).

72. Stewart, L., Clark, R. & Behnke, C. High-throughput crystallization and structure determination in drug discovery. *Drug Discov Today* **7**, 187-96 (2002).

73. Wooh, J.W., Kidd, R.D., Martin, J.L. & Kobe, B. Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallogr D Biol Crystallogr* **59**, 769-72 (2003).

74. Zhai, Y. et al. VEGI, a novel cytokine of the tumor necrosis factor family, is an angiogenesis inhibitor that suppresses the growth of colon carcinomas in vivo. *Faseb J* **13**, 181-9 (1999).

75. Clark, E.D.B. Refolding of recombinant proteins. *Curr Opin Biotechnol* **9**, 157-63 (1998).

76. Middelberg, A.P. Preparative protein refolding. *Trends Biotechnol* **20**, 437-43 (2002).

77. Yoshii, H. et al. Refolding of denatured/reduced lysozyme at high concentration with diafiltration. *Biosci Biotechnol Biochem* **64**, 1159-65 (2000).

78. Machold, C., Schlegl, R., Buchinger, W. & Jungbauer, A. Matrix assisted refolding of proteins by ion exchange chromatography. *J Biotechnol* **117**, 83-97 (2005).

79. Langenhof, M., Leong, S.S., Pattenden, L.K. & Middelberg, A.P. Controlled oxidative protein refolding using an ion-exchange column. *J Chromatogr A* **1069**, 195-201 (2005).

80. Rogl, H., Kosemund, K., Kuhlbrandt, W. & Collinson, I. Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett* **432**, 21-6 (1998).

81. Xie, Y. & Wetlaufer, D.B. Control of aggregation in protein refolding: the temperature-leap tactic. *Protein Sci* **5**, 517-23 (1996).

82. Burgess, R.R. Purification of overproduced Escherichia coli RNA polymerase sigma factors by solubilizing inclusion bodies and refolding from Sarkosyl. *Methods Enzymol* **273**, 145-9 (1996).

83. Willis, M.S. et al. Investigation of protein refolding using a fractional factorial screen: A study of reagent effects and interactions. *Protein Sci* (2005).

84. Buckle, A.M. et al. The matrix refolded. *Nat Methods* **2**, 3 (2005).

85. Brown, J. et al. Structure of a functional IGF2R fragment determined from the anomalous scattering of sulfur. *Embo J* **21**, 1054-62 (2002).

86. Zhang, Z. & Henzel, W.J. Signal peptide prediction based on analysis of experimentally verified cleavage sites. *Protein Sci* **13**, 2819-24 (2004).

87. Love, C.A. et al. The ligand-binding face of the semaphorins revealed by the high-resolution crystal structure of SEMA4D. *Nat Struct Biol* **10**, 843-8 (2003).

88. Walter, T.S. et al. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **61**, 651-7 (2005).

89. Kantardjieff, K.A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-71 (2003).

90. Matthews, B.W. Solvent content of protein crystals. *J. Mol. Biol.* **33**(1968).

91. Claude, J.B., Suhre, K., Notredame, C., Claverie, J.M. & Abergel, C. CaspR: a web server for automated molecular replacement using homology modelling. *Nucleic Acids Res* **32**, W606-9 (2004).

92. Poirot, O., Suhre, K., Abergel, C., O'Toole, E. & Notredame, C. 3DCoffee@igs: a web server for combining sequences and structures into a multiple sequence alignment. *Nucleic Acids Res* **32**, W37-40 (2004).

93. Notredame, C., Higgins, D.G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17 (2000).

94. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).

95. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367-72 (2001).

96. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-21 (1998).

97. Read, R. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D* **57**, 1373-1382 (2001).

98. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallographica Section D* **61**, 458-464 (2005).

99. Rossmann, M.G. & Blow, D.M. *Acta Crystallogr. D.* **15**, 24-31 (1962).

100. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D* **57**, 1367-1372 (2001).

101. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458-64 (2005).

102. Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**, 307-326 (1997).

103. Kohno, T. et al. A second tumor necrosis factor receptor gene product can shed a naturally occurring tumor necrosis factor inhibitor. *Proc Natl Acad Sci U S A* **87**, 8331-5 (1990).

104. Ehrlich, S., Infante-Duarte, C., Seeger, B. & Zipp, F. Regulation of soluble and surface-bound TRAIL in human T cells, B cells, and monocytes. *Cytokine* **24**, 244-53 (2003).

105. Kuhn, P., Wilson, K., Patch, M.G. & Stevens, R.C. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr Opin Chem Biol* **6**, 704-10 (2002).

106. Macchiarulo, A. et al. The role of electrostatic interaction in the molecular recognition of selective agonists to metabotropic glutamate receptors. *Proteins* **50**, 609-19 (2003).

107. Pawson, T. & Nash, P. Protein-protein interactions define specificity in signal transduction. *Genes Dev* **14**, 1027-47 (2000).

108. Cappello, V., Tramontano, A. & Koch, U. Classification of proteins based on the properties of the ligand-binding site: the case of adenine-binding proteins. *Proteins* **47**, 106-15 (2002).

109. Koh, J.T. Engineering selectivity and discrimination into ligand-receptor interfaces. *Chem Biol* **9**, 17-23 (2002).

110. Kangas, E. & Tidor, B. Charge optimization leads to favorable electrostatic binding free energy. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **59**, 5958-61 (1999).

111. Drewe, E., McDermott, E.M., Powell, P.T., Isaacs, J.D. & Powell, R.J. Prospective study of anti-tumour necrosis factor receptor superfamily 1B fusion protein, and case study of anti-tumour necrosis factor receptor superfamily 1A fusion protein, in tumour necrosis factor receptor associated periodic syndrome (TRAPS): clinical and laboratory findings in a series of seven patients. *Rheumatology (Oxford)* **42**, 235-9 (2003).

112. Bleumink, G.S., ter Borg, E.J., Ramselaar, C.G. & Ch Stricker, B.H. Etanercept-induced subacute cutaneous lupus erythematosus. *Rheumatology (Oxford)* **40**, 1317-9 (2001).

113. Lorenz, H.M. & Kalden, J.R. Perspectives for TNF-alpha-targeting therapies. *Arthritis Res* **4 Suppl 3**, S17-24 (2002).

114. Blundell, T.L. Structure-based drug design. *Nature* **384**, 23-6 (1996).

115. Rockey, W.M. & Elcock, A.H. Progress toward virtual screening for drug side effects. *Proteins* **48**, 664-71 (2002).

116. Salemme, F.R., Spurlino, J. & Bone, R. Serendipity meets precision: the integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure* **5**, 319-24 (1997).

117. Burbaum, J. & Tobal, G.M. Proteomics in drug discovery. *Curr Opin Chem Biol* **6**, 427-33 (2002).

118. Anderson, S. & Chiplin, J. Structural genomics: shaping the future of drug design? *Drug Discov Today* **7**, 105-7 (2002).

119. Taylor, R. Life-science applications of the Cambridge Structural Database. *Acta Cryst. D* **58**, 879-888 (2002).

120. Milne, G.W.A., Nicklaus, M.C., Driscoll, J.S., Wang, S. and Zaharevitz. D. The NCI Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **34**, 1219-1224 (1994).

121. Shoichet, B.K., McGovern, S.L., Wei, B. & Irwin, J.J. Lead discovery using molecular docking. *Curr Opin Chem Biol* **6**, 439-46 (2002).

122. Graffner-Nordberg, M., Kolmodin, K., Aqvist, J., Queener, S.F. & Hallberg, A. Design, synthesis, computational prediction, and biological evaluation of ester soft drugs as inhibitors of dihydrofolate reductase from Pneumocystis carinii. *J Med Chem* **44**, 2391-402 (2001).

123. Ridderstrom, M., Zamora, I., Fjellstrom, O. & Andersson, T.B. Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J Med Chem* **44**, 4072-81 (2001).

124. Costantino, G., Macchiarulo, A., Camaioni, E. & Pellicciari, R. Modeling of poly(ADP-ribose)polymerase (PARP) inhibitors. Docking of ligands and quantitative structure-activity relationship analysis. *J Med Chem* **44**, 3786-94 (2001).

125. Wu, S.Y. et al. Discovery of a novel family of CDK inhibitors with the program LIDAEUS: structural basis for ligand-induced disordering of the activation loop. *Structure (Camb)* **11**, 399-410 (2003).

126. Milne, G.W., Nicklaus, M.C., Driscoll, J.S., Wang, S. & Zaharevitz, D. National Cancer Institute Drug Information System 3D database. *J Chem Inf Comput Sci* **34**, 1219-24 (1994).

127. Wang, R., Gao, Y., Lui, L. Ligbuilder: A Multipurpose Program for Structure-Based Drug Design. *J. Mol. Model.* **6**, 498-516 (2000).

128. Holbeck, S.L. Update on NCI in vitro drug screen utilities. *Eur J Cancer* **40**, 785-93 (2004).

129. Naccess http://wolf.bms.umist.ac.uk/naccess/.

130. Laskowski, R.A. SURFNET: A program for visualizing molecular surfaces, cavities and intermolecular interactions. *J. Mol. Graph.* **13**, 323-330 (1995).

131. Guex, N. & Peitsch, M.C. Swiss-Model and the Swiss-Pdb viewer: an environment for comparitive protein modelling. *Electrophoresis* **18**, 2714-2723 (1997).

132. Schuttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63 (2004).

133. Fechner, U., Franke, L., Renner, S., Schneider, P. & Schneider, G. Comparison of correlation vector methods for ligand-based similarity searching. *J Comput Aided Mol Des* **17**, 687-98 (2003).

134. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-74 (2004).

135. Goodsell, D.S., Morris, G.M. & Olson, A.J. Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **9**, 1-5 (1996).

136. Cactus Browser. http://cactus.nci.nih.gov/ncidb2/. (2003).

137. Carlson, H.A. Protein flexibility and drug design: how to hit a moving target. *Curr Opin Chem Biol* **6**, 447-52 (2002).

138. Kallblad, P. & Dean, P.M. Efficient Conformational Sampling of Local Side-chain Flexibility. *J Mol Biol* **326**, 1651-65 (2003).

139. Hymowitz, S.G. et al. Structures of APRIL-receptor complexes: like BCMA, TACI employs only a single cysteine-rich domain for high affinity ligand binding. *J Biol Chem* **280**, 7218-27 (2005).

140. Garcia, K.C., Teyton, L. & Wilson, I.A. Structural basis of T cell recognition. *Annu Rev Immunol* **17**, 369-97 (1999).

141. Hymowitz, S.G. et al. The crystal structures of EDA-A1 and EDA-A2: splice variants with distinct receptor specificity. *Structure (Camb)* **11**, 1513-20 (2003).

142. Lill, M.A., Vedani, A. & Dobler, M. Raptor: combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: application toward the simulation of structurally diverse ligand sets. *J Med Chem* **47**, 6174-86 (2004).

143. Kastenholz, M.A., Pastor, M., Cruciani, G., Haaksma, E.E. & Fox, T. GRID/CPCA: a new computational tool to design selective ligands. *J Med Chem* **43**, 3033-44 (2000).

144. Hendlich, M., Bergner, A., Gunther, J. & Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J Mol Biol* **326**, 607-20 (2003).

145. Hanessian, S., Parthasarathy, S., Mauduit, M. & Payza, K. The power of visual imagery in drug design. Isopavines as a new class of morphinomimetics and their human opioid receptor binding activity. *J Med Chem* **46**, 34-48 (2003).

146. Guner, O., Clement, O. & Kurogi, Y. Pharmacophore modeling and three dimensional database searching for drug design using catalyst: recent advances. *Curr Med Chem* **11**, 2991-3005 (2004).

147. Schneider, G. & Bohm, H.J. Virtual screening and fast automated docking methods. *Drug Discov Today* **7**, 64-70 (2002).

148. Shoichet, B.K. & Kuntz, I.D. Matching chemistry and shape in molecular docking. *Protein Eng* **6**, 723-32 (1993).

149. Li, C., Xu, L., Wolan, D.W., Wilson, I.A. & Olson, A.J. Virtual screening of human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. *J Med Chem* **47**, 6681-90 (2004).

150. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).

151. Siani, M.A., Weininger, D. & Blaney, J.M. CHUCKLES: a method for representing and searching peptide and peptoid sequences on both monomer and atomic levels. *J Chem Inf Comput Sci* **34**, 588-93 (1994).

152. Openbabel. http://openbabel.sourceforge.net/.

153. De Martino, G. et al. Novel 1-[2-(Diarylmethoxy)ethyl]-2-methyl-5-nitroimidazoles as HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors. A Structure-Activity Relationship Investigation. *J Med Chem* **48**, 4378-4388 (2005).

154. Gadek, T.R., Burdick, D.J., McDowell, R.S., Stanley, M.S., Marsters, J.C., Paris, K.J., Oare, D.A., Reynolds, M.E., Ladner, C., Zioncheck, K.A., Lee, W.P., Gribling, P., Dennis, M.S., Skelton, N.J., Tumas, D.B., Clark, K.R., Keating, S.M., Beresini, M.H., Tilley, J.W., Presta, L.G., Bodary, S.C. Generation of an LFA-1 Antagonist by the Transfer of the ICAM-1 Immunoregulatory Epitope to a Small Molecule. *Science* **295**, 1086-1089 (2002).

155. Rockey, W.M. & Elcock, A.H. Rapid computational identification of the targets of protein kinase inhibitors. *J Med Chem* **48**, 4138-52 (2005).

156. Olsen, L., Pettersson, I., Hemmingsen, L., Adolph, H.W. & Jorgensen, F.S. Docking and scoring of metallo-beta-lactamases inhibitors. *J Comput Aided Mol Des* **18**, 287-302 (2004).

157. Liu, G. et al. Inhibitory mode of indole-2-carboxamide derivatives against HLGPa: molecular docking and 3D-QSAR analyses. *Bioorg Med Chem* **12**, 4147-57 (2004).

158. Davis, M.M. et al. Ligand recognition by alpha beta T cell receptors. *Annu Rev Immunol* **16**, 523-44 (1998).

159. Rhijn, I.V., Zajonc, D.M., Wilson, I.A. & Moody, D.B. T-cell activation by lipopeptide antigens. *Curr Opin Immunol* **17**, 222-9 (2005).

160. Purtic, B., Pitcher, L.A., van Oers, N.S. & Wulfing, C. T cell receptor (TCR) clustering in the immunological synapse integrates TCR and costimulatory signaling in selected T cells. *Proc Natl Acad Sci U S A* **102**, 2904-9 (2005).

161. Corbin, G.A. & Harty, J.T. T cells undergo rapid ON/OFF but not ON/OFF/ON cycling of cytokine production in response to antigen. *J Immunol* **174**, 718-26 (2005).

162. Calabi, F. & Milstein, C. A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* **323**, 540-3 (1986).

163. Porcelli, S.A. The CD1 family: a third lineage of antigen-presenting molecules. *Adv Immunol* **59**, 1-98 (1995).

164. Porcelli, S., Morita, C.T. & Brenner, M.B. CD1b restricts the response of human CD4-8- T lymphocytes to a microbial antigen. *Nature* **360**, 593-7 (1992).

165. Beckman, E.M. et al. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. *Nature* **372**, 691-4 (1994).

166. Godfrey, D.I. & Zlotnik, A. Control points in early T-cell development. *Immunol Today* **14**, 547-53 (1993).

167. Hoffman, E.S. et al. Productive T-cell receptor beta-chain gene rearrangement: coincident regulation of cell cycle and clonality during development in vivo. *Genes Dev* **10**, 948-62 (1996).

168. Davis, M.M. & Bjorkman, P.J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395-402 (1988).

169. Takeshita, S., Toda, M. & Yamagishi, H. Excision products of the T cell receptor gene support a progressive rearrangement model of the alpha/delta locus. *Embo J* **8**, 3261-70 (1989).

170. Surh, C.D. & Sprent, J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature* **372**, 100-3 (1994).

171. Alam, S.M. et al. T-cell-receptor affinity and thymocyte positive selection. *Nature* **381**, 616-20 (1996).

172. Jameson, S.C. & Bevan, M.J. T-cell selection. *Curr Opin Immunol* **10**, 214-9 (1998).

173. Sebzda, E. et al. Selection of the T cell repertoire. *Annu Rev Immunol* **17**, 829-74 (1999).

174. Arden, B., Clark, S.P., Kabelitz, D. & Mak, T.W. Human T-cell receptor variable gene segment families. *Immunogenetics* **42**, 455-500 (1995).

175. Matsui, K. et al. Low affinity interaction of peptide-MHC complexes with T cell receptors. *Science* **254**, 1788-91 (1991).

176. Karjalainen, K. High sensitivity, low affinity--paradox of T-cell receptor recognition. *Curr Opin Immunol* **6**, 9-12 (1994).

177. Chothia, C., Boswell, D.R. & Lesk, A.M. The outline structure of the T-cell alpha beta receptor. *Embo J* **7**, 3745-55 (1988).

178. Arstila, T.P. et al. A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286**, 958-61 (1999).

179. Garboczi, D.N. et al. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134-41 (1996).

180. Teng, M.K. et al. Identification of a common docking topology with substantial variation among different TCR-peptide-MHC complexes. *Curr Biol* **8**, 409-12 (1998).

181. Garcia, K.C. et al. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* **279**, 1166-72 (1998).

182. Ding, Y.H. et al. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* **8**, 403-11 (1998).

183. van der Merwe, P.A. & Davis, S.J. Molecular interactions mediating T cell antigen recognition. *Annu Rev Immunol* **21**, 659-84 (2003).

184. Stewart-Jones, G.B., McMichael, A.J., Bell, J.I., Stuart, D.I. & Jones, E.Y. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* **4**, 657-63 (2003).

185. Rudolph, M.G. & Wilson, I.A. The specificity of TCR/pMHC interaction. *Curr Opin Immunol* **14**, 52-65 (2002).

186. Rudolph, M.G., Luz, J.G. & Wilson, I.A. Structural and thermodynamic correlates of T cell signaling. *Annu Rev Biophys Biomol Struct* **31**, 121-49 (2002).

187. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol Today* **19**, 395-404 (1998).

188. Dascher, C.C. & Brenner, M.B. Evolutionary constraints on CD1 structure: insights from comparative genomic analysis. *Trends Immunol* **24**, 412-8 (2003).

189. Kinjo, Y. et al. Recognition of bacterial glycosphingolipids by natural killer T cells. *Nature* **434**, 520-5 (2005).

190. Mattner, J. et al. Exogenous and endogenous glycolipid antigens activate NKT cells during microbial infections. *Nature* **434**, 525-9 (2005).

191. Batuwangala, T. et al. The crystal structure of human CD1b with a bound bacterial glycolipid. *J Immunol* **172**, 2382-8 (2004).

192. Moody, D.B. et al. T cell activation by lipopeptide antigens. *Science* **303**, 527-31 (2004).

193. Moody, D.B., Zajonc, D.M. & Wilson, I.A. Anatomy of CD1-lipid antigen complexes. *Nat Rev Immunol* **5**, 387-99 (2005).

194. Gadola, S.D. et al. Structure of human CD1b with bound ligands at 2.3 A, a maze for alkyl chains. *Nat Immunol* **3**, 721-6 (2002).

195. Hahn, M., Nicholson, M.J., Pyrdol, J. & Wucherpfennig, K.W. Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor. *Nat Immunol* **6**, 490-6 (2005).

196. Zeng, Z. et al. Crystal structure of mouse CD1: An MHC-like fold with a large hydrophobic binding groove. *Science* **277**, 339-45 (1997).

197. Koch, M. et al. The crystal structure of human CD1d with and without alpha-galactosylceramide. *Nat Immunol* (2005).

198. Zajonc, D.M. et al. Structure and function of a potent agonist for the semi-invariant natural killer T cell receptor. *Nat Immunol* (2005).

199. Zajonc, D.M., Elsliger, M.A., Teyton, L. & Wilson, I.A. Crystal structure of CD1a in complex with a sulfatide self antigen at a resolution of 2.15 A. *Nat Immunol* **4**, 808-15 (2003).

200. Zajonc, D.M. et al. Molecular mechanism of lipopeptide presentation by CD1a. *Immunity* **22**, 209-19 (2005).

201. Lawton, A.P. & Kronenberg, M. The Third Way: Progress on pathways of antigen processing and presentation by CD1. *Immunol Cell Biol* **82**, 295-306 (2004).

202. Kawano, T. et al. CD1d-restricted and TCR-mediated activation of valpha14 NKT cells by glycosylceramides. *Science* **278**, 1626-9 (1997).

203. Benlagha, K., Weiss, A., Beavis, A., Teyton, L. & Bendelac, A. In vivo identification of glycolipid antigen-specific T cells using fluorescent CD1d tetramers. *J Exp Med* **191**, 1895-903 (2000).

204. Matsuda, J.L. et al. Tracking the response of natural killer T cells to a glycolipid antigen using CD1d tetramers. *J Exp Med* **192**, 741-54 (2000).

205. Kronenberg, M. & Gapin, L. The unconventional lifestyle of NKT cells. *Nat Rev Immunol* **2**, 557-68 (2002).

206. Godfrey, D.I. & Kronenberg, M. Going both ways: immune regulation via CD1d-dependent NKT cells. *J Clin Invest* **114**, 1379-88 (2004).

207. Porcelli, S.A. & Modlin, R.L. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu Rev Immunol* **17**, 297-329 (1999).

208. Bendelac, A., Killeen, N., Littman, D.R. & Schwartz, R.H. A subset of CD4+ thymocytes selected by MHC class I molecules. *Science* **263**, 1774-8 (1994).

209. Wilson, S.B. et al. Extreme Th1 bias of invariant Valpha24JalphaQ T cells in type 1 diabetes. *Nature* **391**, 177-81 (1998).

210. Oishi, Y. et al. Selective reduction and recovery of invariant Valpha24JalphaQ T cell receptor T cells in correlation with disease activity in patients with systemic lupus erythematosus. *J Rheumatol* **28**, 275-83 (2001).

211. Mieza, M.A. et al. Selective reduction of V alpha 14+ NK T cells associated with disease development in autoimmune-prone mice. *J Immunol* **156**, 4035-40 (1996).

212. Gadola, S.D., Dulphy, N., Salio, M. & Cerundolo, V. Valpha24-JalphaQ-independent, CD1d-restricted recognition of alpha-galactosylceramide by human CD4(+) and CD8alphabeta(+) T lymphocytes. *J Immunol* **168**, 5514-20 (2002).

213. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).

214. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

215. Morris, R.J., Perrakis, A. & Lamzin, V.S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**, 229-44 (2003).

216. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **57**, 122-33 (2001).

217. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).

218. Al-Lazikani, B., Lesk, A.M. & Chothia, C. Canonical structures for the hypervariable regions of T cell alphabeta receptors. *J Mol Biol* **295**, 979-95 (2000).

219. Garcia, K.C. et al. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science* **274**, 209-19 (1996).

220. Fields, B.A. et al. Crystal structure of a T-cell receptor beta-chain complexed with a superantigen. *Nature* **384**, 188-92 (1996).

221. Hennecke, J. & Wiley, D.C. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J Exp Med* **195**, 571-81 (2002).

222. Godfrey, D.I., McCluskey, J. & Rossjohn, J. CD1d antigen presentation: treats for NKT cells. *Nat Immunol* **6**, 754-6 (2005).

223. Nedospasov, S.A. et al. Tandem arrangement of genes coding for tumor necrosis factor (TNF-alpha) and lymphotoxin (TNF-beta) in the human genome. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 611-24 (1986).

224. Nedwin, G.E. et al. Human lymphotoxin and tumor necrosis factor genes: structure, homology and chromosomal localization. *Nucleic Acids Res* **13**, 6361-73 (1985).

225. Browning, J.L. et al. Lymphotoxin beta, a novel member of the TNF family that forms a heteromeric complex with lymphotoxin on the cell surface. *Cell* **72**, 847-56 (1993).

226. Miura, S. et al. Molecular cloning and characterization of a novel glycoprotein, gp34, that is specifically induced by the human T-cell leukemia virus type I transactivator p40tax. *Mol Cell Biol* **11**, 1313-25 (1991).

227. Graf, D., Korthauer, U., Mages, H.W., Senger, G. & Kroczek, R.A. Cloning of TRAP, a ligand for CD40 on human T cells. *Eur J Immunol* **22**, 3191-4 (1992).

228. Minuto, F. et al. Study of bioassayable somatomedin on renal artery and vein. *J Endocrinol Invest* **2**, 305-6 (1979).

229. Goodwin, R.G. et al. Molecular and biological characterization of a ligand for CD27 defines a new family of cytokines with homology to tumor necrosis factor. *Cell* **73**, 447-56 (1993).

230. Smith, C.A. et al. CD30 antigen, a marker for Hodgkin's lymphoma, is a receptor whose ligand defines an emerging family of cytokines with homology to TNF. *Cell* **73**, 1349-60 (1993).

231. Alderson, M.R. et al. Molecular and biological characterization of human 4-1BB and its ligand. *Eur J Immunol* **24**, 2219-27 (1994).

232. Wiley, S.R. et al. Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* **3**, 673-82 (1995).

233. Chicheportiche, Y. et al. TWEAK, a new secreted ligand in the tumor necrosis factor family that weakly induces apoptosis. *J Biol Chem* **272**, 32401-10 (1997).

234. Hahne, M. et al. APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth. *J Exp Med* **188**, 1185-90 (1998).

235. Shu, H.B., Hu, W.H. & Johnson, H. TALL-1 is a novel member of the TNF family that is down-regulated by mitogens. *J Leukoc Biol* **65**, 680-3 (1999).

236. Mauri, D.N. et al. LIGHT, a new member of the TNF superfamily, and lymphotoxin alpha are ligands for herpesvirus entry mediator. *Immunity* **8**, 21-30 (1998).

237. Kere, J. et al. X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nat Genet* **13**, 409-16 (1996).

238. Gurney, A.L. et al. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr Biol* **9**, 215-8 (1999).

239. Schall, T.J. et al. Molecular cloning and expression of a receptor for human tumor necrosis factor. *Cell* **61**, 361-70 (1990).

240. Crowe, P.D. et al. A lymphotoxin-beta-specific receptor. *Science* **264**, 707-10 (1994).

241. Latza, U. et al. The human OX40 homolog: cDNA structure, expression and chromosomal assignment of the ACT35 antigen. *Eur J Immunol* **24**, 677-83 (1994).

242. Stamenkovic, I., Clark, E.A. & Seed, B. A B-lymphocyte activation molecule related to the nerve growth factor receptor and induced by cytokines in carcinomas. *Embo J* **8**, 1403-10 (1989).

243. Itoh, N. et al. The polypeptide encoded by the cDNA for human cell surface antigen Fas can mediate apoptosis. *Cell* **66**, 233-43 (1991).

244. Camerini, D., Walz, G., Loenen, W.A., Borst, J. & Seed, B. The T cell activation antigen CD27 is a member of the nerve growth factor/tumor necrosis factor receptor gene family. *J Immunol* **147**, 3165-9 (1991).

245. Durkop, H. et al. Molecular cloning and expression of a new member of the nerve growth factor receptor family that is characteristic for Hodgkin's disease. *Cell* **68**, 421-7 (1992).

246. Pan, G. et al. The receptor for the cytotoxic ligand TRAIL. *Science* **276**, 111-3 (1997).

247. Walczak, H. et al. TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL. *Embo J* **16**, 5386-97 (1997).

248. Degli-Esposti, M.A. et al. Cloning and characterization of TRAIL-R3, a novel member of the emerging TRAIL receptor family. *J Exp Med* **186**, 1165-70 (1997).

249. Marsters, S.A. et al. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr Biol* **7**, 1003-6 (1997).

250. Simonet, W.S. et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309-19 (1997).

251. von Bulow, G.U. & Bram, R.J. NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. *Science* **278**, 138-41 (1997).

252. Montgomery, R.I., Warner, M.S., Lum, B.J. & Spear, P.G. Herpes simplex virus-1 entry into cells mediated by a novel member of the TNF/NGF receptor family. *Cell* **87**, 427-36 (1996).

253. Johnson, D. et al. Expression and structure of the human NGF receptor. *Cell* **47**, 545-54 (1986).

254. Laabi, Y. et al. A new gene, BCM, on chromosome 16 is fused to the interleukin 2 gene by a t(4;16)(q26;p13) translocation in a malignant T cell lymphoma. *Embo J* **11**, 3897-904 (1992).

255. Eby, M.T., Jasmin, A., Kumar, A., Sharma, K. & Chaudhary, P.M. TAJ, a novel member of the tumor necrosis factor receptor family, activates the c-Jun N-terminal kinase pathway and mediates caspase-independent cell death. *J Biol Chem* **275**, 15336-42 (2000).

256. Pan, G. et al. Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. *FEBS Lett* **431**, 351-6 (1998).

257. Monreal, A.W. et al. Mutations in the human homologue of mouse dl cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nat Genet* **22**, 366-9 (1999).

258. Yan, M. et al. Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**, 523-7 (2000).

259. Kitson, J. et al. A death-domain-containing receptor that mediates apoptosis. *Nature* **384**, 372-5 (1996).

260. Sica, G.L. et al. RELT, a new member of the tumor necrosis factor receptor superfamily, is selectively expressed in hematopoietic tissues and activates transcription factor NF-kappaB. *Blood* **97**, 2702-7 (2001).

261. Feng, S.L. et al. The Fn14 immediate-early response gene is induced during liver regeneration and highly expressed in both human and murine hepatocellular carcinomas. *Am J Pathol* **156**, 1253-61 (2000).

262. Wallweber, H.J., Compaan, D.M., Starovasnik, M.A. & Hymowitz, S.G. The crystal structure of a proliferation-inducing ligand, APRIL. *J Mol Biol* **343**, 283-90 (2004).

263. Karpusas, M. et al. Crystal structure of extracellular human BAFF, a TNF family member that stimulates B lymphocytes. *J Mol Biol* **315**, 1145-54 (2002).

264. Oren, D.A. et al. Structural basis of BLyS receptor recognition. *Nat Struct Biol* **9**, 288-92 (2002).

265. Karpusas, M., Hsu, Y. M., Wang, J. H., Thompson, J., Lederman, S., Chess, L., Thomas, D. 2 angstrom crystal structure of an extracellular fragment of human CD40 ligand. *Structure* **3**, 1426 (1995).

266. Lam, J., Nelson, C.A., Ross, F.P., Teitelbaum, S.L. & Fremont, D.H. Crystal structure of the TRANCE/RANKL cytokine reveals determinants of receptor-ligand specificity. *The Journal of Clinical Investigation* **108**, 971-979 (2001).

267. Cha, S.S., Kim, J. S., Cho, H. S., Shin, N. K., Jeong, W., Shin, H. C., Kim, Y. J., Hahn, J. H., Oh, B. H. High resolution crystal structure of a human tumor necrosis factor-a mutant with low systemic toxicity. *J. Biol. Chem* **273**, 2153 (1998).

268. Reed, C. et al. Crystal structure of TNF-alpha mutant R31D with greater affinity for receptor R1 compared with R2. *Protein Eng* **10**, 1101-7 (1997).

269. Eck, M.J. & Sprang, S.R. The structure of tumor necrosis factor-alpha at 2.6 A resolution. Implications for receptor binding. *J Biol Chem* **264**, 17595-605 (1989).

270. Eck, M.J., Ultsch, M., Rinderknecht, E., de Vos, A.M. & Sprang, S.R. The structure of human lymphotoxin (tumor necrosis factor-beta) at 1.9-A resolution. *J Biol Chem* **267**, 2119-22 (1992).

271. Cha, S.-S., Oh, B.-H. 2.8 Å Resolution Crystal Structure of Human Trail, a Cytokine with Selective Antitumor Activity. *Immunity* **11**, 253 (1999).

272. Carfi, A., Willis, S. H., Whitbeck, J. C., Krummenacker, C., Cohen, G. H., Eisenberg, R. J., Wiley, D. C. Crystal Structure of the Herpes Simplex Virus Glycoprotein D Bound to the Cellular Receptor Hvea/Hvem. *Mol. Cell.* **8**, 169 (2001).

273. Naismith, J.H., Devine, T.Q., Brandhuber, B.J. & Sprang, S.R. Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor. *J Biol Chem* **270**, 13303-7 (1995).

274. Naismith, J.H., Devine, T.Q., Kohno, T. & Sprang, S.R. Structures of the extracellular domain of the type I tumor necrosis factor receptor. *Structure* **4**, 1251-62 (1996).

275. Carter, P.H. et al. Photochemically enhanced binding of small molecules to the tumor necrosis factor receptor-1 inhibits the binding of TNF-alpha. *Proc Natl Acad Sci U S A* **98**, 11879-84 (2001).

276. Li, H., Lebedeva, M.I., Ward, E.S. & Mariuzza, R.A. Dual conformations of a T cell receptor V alpha homodimer: implications for variability in V alpha V beta domain association. *J Mol Biol* **269**, 385-94 (1997).

277. Rudolph, M.G., Huang, M., Teyton, L. & Wilson, I.A. Crystal structure of an isolated V(alpha) domain of the 2C T-cell receptor. *J Mol Biol* **314**, 1-8 (2001).

278. Plaksin, D., Chacko, S., Navaza, J., Margulies, D.H. & Padlan, E.A. The X-ray crystal structure of a Valpha2.6Jalpha38 mouse T cell receptor domain at 2.5 A resolution: alternate modes of dimerization and crystal packing. *J Mol Biol* **289**, 1153-61 (1999).

279. Machius, M., Cianga, P., Deisenhofer, J. & Ward, E.S. Crystal structure of a T cell receptor Valpha11 (AV11S5) domain: new canonical forms for the first and second complementarity determining regions. *J Mol Biol* **310**, 689-98 (2001).

280. Bentley, G.A., Boulot, G., Karjalainen, K. & Mariuzza, R.A. Crystal structure of the beta chain of a T cell antigen receptor. *Science* **267**, 1984-7 (1995).

281. Sundberg, E.J. et al. Structures of two streptococcal superantigens bound to TCR beta chains reveal diversity in the architecture of T cell signaling complexes. *Structure (Camb)* **10**, 687-99 (2002).

282. Li, H. et al. Three-dimensional structure of the complex between a T cell receptor beta chain and the superantigen staphylococcal enterotoxin B. *Immunity* **9**, 807-16 (1998).

283. Li, H. et al. Structure of the Vdelta domain of a human gammadelta T-cell antigen receptor. *Nature* **391**, 502-6 (1998).

284. Wang, J. et al. Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. *Embo J* **17**, 10-26 (1998).

285. Housset, D. et al. The three-dimensional structure of a T-cell antigen receptor V alpha V beta heterodimer reveals a novel arrangement of the V beta domain. *Embo J* **16**, 4205-16 (1997).

286. Kjer-Nielsen, L. et al. The 1.5 A crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. *Structure (Camb)* **10**, 1521-32 (2002).

287. Chen, J.L. et al. Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J Exp Med* **201**, 1243-55 (2005).

288. Allison, T.J., Winter, C.C., Fournie, J.J., Bonneville, M. & Garboczi, D.N. Structure of a human gammadelta T-cell antigen receptor. *Nature* **411**, 820-4 (2001).

289. Reinherz, E.L. et al. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* **286**, 1913-21 (1999).

290. Ding, Y.H., Baker, B.M., Garboczi, D.N., Biddison, W.E. & Wiley, D.C. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity* **11**, 45-56 (1999).

291. Reiser, J.B. et al. Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nat Immunol* **1**, 291-7 (2000).

292. Hennecke, J., Carfi, A. & Wiley, D.C. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *Embo J* **19**, 5611-24 (2000).

293. Degano, M. et al. A functional hot spot for antigen recognition in a superagonist TCR/MHC complex. *Immunity* **12**, 251-61 (2000).

294. Reiser, J.B. et al. A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity* **16**, 345-54 (2002).

295. Buslepp, J., Wang, H., Biddison, W.E., Appella, E. & Collins, E.J. A correlation between TCR Valpha docking on MHC and CD8 dependence: implications for T cell selection. *Immunity* **19**, 595-606 (2003).

296. Kjer-Nielsen, L. et al. A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* **18**, 53-64 (2003).

297. Luz, J.G. et al. Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing V(beta) Interactions. *J Exp Med* **195**, 1175-86 (2002).

298. Reiser, J.B. et al. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol* **4**, 241-7 (2003).

299. Maynard, J. et al. Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: insights into MHC bias and antigen specificity. *Immunity* **22**, 81-92 (2005).

300. Adams, E.J., Chien, Y.H. & Garcia, K.C. Structure of a gammadelta T cell receptor in complex with the nonclassical MHC T22. *Science* **308**, 227-31 (2005).