Engineering IgE antibodies and CD23 for therapeutic discovery

Ilkow, Veronica Franciszka

Awarding institution:
King's College London

ENGINEERING IGE ANTIBODIES AND CD23 FOR THERAPEUTIC DISCOVERY

Veronica F. Ilkow

A thesis submitted in fulfilment of the requirements for the degree Doctor of Philosophy at King's College London.

Randall Division of Cell and Molecular Biophysics,

MRC & Asthma UK Centre in Allergic Mechanisms of Asthma,

School of Basic and Medical Bioscience,

Faculty of Life Sciences and Medicine,

King's College London

Abstract

Immunoglobulin E (IgE) is fundamental to the allergic response and the functions of IgE are mediated by its Fc region binding to two receptors, FcεRI and CD23 (FcεRII). The interaction of IgE with other proteins have complicated our investigations of the unique role each receptor plays. To solve this, a small-scale library of IgE-Fc proteins was designed with two key positions, one at each receptor-binding site mutated. The unpredictable allosteric nature of IgE prevents rational engineering approaches, thus the design of a membrane-bound IgE-Fc-GFP-tagged protein allowed for the generation of a membrane-surface display library of stable cell lines. A FACS selection assay identified IgE-Fc proteins with weakened binding to a single IgE-receptor, which serves as a proof-of-principle for this concept.

Additional studies into human CD23 and the differences between it and murine CD23 revealed additional levels of regulation for IgE-binding not seen in other species and this is due to its unique properties. Human CD23 is an unusual antibody receptor, being a calcium dependent (C-type) lectin that has lost its carbohydrate binding capability. Ca^{2+} binds to and increases CD23's affinity for IgE, and one of two Ca^{2+} binding sites usually present in C-type lectins is absent in human but present in murine CD23. To understand if the loss of the second Ca^{2+} binding site has led to a regulatory gain/loss of function in human CD23, a panel of CD23 mutant proteins with increasingly 'mouse-like' sequences was generated.

The insertion of the second Ca^{2+} binding site was verified by HSQC-NMR whilst molecular dynamic simulations provided a means of understanding the flexibility of the proteins. It revealed that binding of two Ca^{2+} ions tethers the soluble CD23 loops into position in the most mouse-like mutant protein, limiting possible conformations for IgE binding. Complementary Biacore experiments indicated that higher calcium binding affinity may have come at a cost of weakened IgE binding, as data in the presence and absence of Ca^{2+} showed decreased binding affinities of the proteins for human IgE. This regulatory difference between murine and human soluble CD23 could inform the development of CD23/IgE inhibitor therapeutics for the treatment of allergy.

Publications

Poster Abstracts

<u>Veronica Ilkow</u>, Balvinder Dhaliwal, Anthony Keeble, James McDonnell, Brian Sutton, Andrew Beavil. (Engineering and Molecular Dynamics Simulations of Calcium Binding Site in the low affinity IgE Receptor CD23). Presented at the FEBS-IUBMB Biointeractomics conference Seville, Spain, 17th-20th May 2016.

Awarded the FEBS Letters International Poster Prize 2016.

<u>Veronica Ilkow</u>, Balvinder Dhaliwal, Anthony Keeble, James McDonnell, Brian Sutton, Andrew Beavil. (Engineering and Molecular Dynamics Simulations of Calcium Binding Site in the low affinity IgE Receptor CD23). Presented at the 19th IUPAB congress and 11th ESBA congress, Edinburgh, Scotland, 16th-20th July 2017.

Table of Contents

5

Table of Figures

11

15

Table of Tables

Table of Equations

Acknowledgements

I would like to thank my primary and secondary supervisors, Dr Andrew Beavil and Professor Brian Sutton along with my funders, MRC and Asthma UK for providing the opportunity to pursue this doctoral degree. I would also like to extend my gratitude to Professor Jim McDonnell, Dr Balwinder Dhaliwal, Dr Rebecca Beavil and Dr Anthony Keeble for their scientific advice. My thanks go to past and present members of the Randall Asthma and Allergy group for their conversations.

From the wider Randall community, I would like to thank Roksana for our cake-trade and Vicky for her enthusiasm. I really enjoyed being a PhD student representative and meeting all the PhD students in the Centre. I hope the mentoring scheme and the KCL blog I co-founded with the other PhD representatives, Justin and Federico, continues to grow. From Guy's hospital, I would also like to thank Ren and Seran for their support and for granting me access to their department's equipment.

Special thanks to the 'Furry Ladies', Caramella and Mu. I cat-sitted for Michelle while writing my thesis and Mu, in particular, took great interest in my thesis. I think they deserve a degree for all the times I tried explaining my figures to them.

Many thanks to my boyfriend Mateen, for believing in me and his constant encouragement and strength. Finally, I would like to thank my parents, for their never-ending love, faith and support.

Abbreviations

^{1}H - ^{15}N HSQC Spectroscopy	Two-dimensional proton and Nitrogen Heteronuclear Single Quantum Coherence Spectroscopy
^{1}H-NMR Spectroscopy	One dimensional proton Nuclear Magnetic Resonance
ADAM10	A Disintegrin And Metalloproteinase Domain-containing protein 10
αγ-fusion protein	the soluble fragment of the high-affinity IgE receptor α-chain fused to the Fc region of IgG$_4$ (Shi et al. 1997)
AIT	Allergen Immunotherapy
APCs	Antigen Presenting Cells
Apo-derCD23	Metal-free derCD23 protein
BCR	B-cell Receptor
B$_{max}$	Maximum binding response
bp	Nucleic Acid base pairs
C-type	Calcium dependent
CR2	Complement Receptor 2
CSR	Class-Switch Recombination
CTLD	C-type lectin Domain
CTLs	C-type lectins
Cε2	The second constant domains of IgE
Cε3	The third constant domains of IgE
D.O.L	Degree of Labelling
derCD23	A soluble fragment of CD23 cleaved by the house dust mite protease Der p I
ΔH	enthalpy
DMEM	Dulbecco's Modified Eagles Medium
ΔS	entropy
EMPD	Extracellular Membrane Proximal Domain
Fab	Fragment-Antigen Binding
FAP	Facilitated Antigen Presentation
FBS	Fetal Bovine Serum
Fc	Fragment Crystallisable
FcRn	neonatal Fc Receptor for Immunoglobulin G

FcεRI	the high affinity receptor for IgE
FcεRII	the low affinity receptor for IgE
FF	Force Field
FRT	Flp Recombinant Target
HEK293	Human Embryonic Kidney 293 cells
HIC	Hydrophobic Interaction Chromatography
HSQCs	Heteronuclear Single Quantum Coherence Spectroscopy
Ig	Immunoglobulin
IgE	Immunoglobulin E
IgE-Fc	The Fc portion of immunoglobulin E
IgE-Fc$_{2-4}$	dimer of the second, third and forth constant domains of IgE
IgE-Fc$_{3-4}$	dimer of the third and forth constant domains of IgE
IgG	Immunoglobulin G
ITAMS	Immunoreceptor tyrosine-based activation motifs
ITC	Isothermal Titration Calorimetry
K$_B$	Binding Constant
K$_D$	Equilibrium Dissociation Constant
LB	Luria Broth
MD	Molecular Dynamics
memIgE-Fc	surface membrane IgE-Fc
MFI	Mean Flourescence Intensity
MHC	Major Histocompatability Complex
MWCO	Molecular Weight Cut-Off
n	stoichiometry
NMR	Nuclear Magnetic Resonance
P.O.I	Protien of Interest
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PCT	Pre-Crystallisation Test
pIgR	polyimmunoglobulin receptor
QCC	Quick Quality Control
Rg	Radius of Gyration

RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
RU	Resonance Units
SDS-PAGE	Sodium dodecyl-sulphate polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
sIgE	soluble IgE
SPR	Surface Plasmon Resonance
TCR	T-cell Receptor
Th2 cells	T helper type 2 cells
triCD23	soluble trimeric CD23
UV	Ultraviolet light

N.B: The convention that has been adopted in this thesis for referring to mutants, is by its amino acid mutation designation e.g. referring to asparagine 225 in a protein that has been mutated to aspartic acid is designated N225D.

Chapter 1 Introduction

1.1 Allergy, asthma and its prevalence in the world

Allergy is a defined as a range of conditions following an immune system response to an otherwise innocuous antigen. It is a growing problem and the prevalence of allergic diseases has continued to rise in the industrialised world for more than 50 years (Pawankar et al., 2011). Across the world, school children are experiencing sensitization rates of 40-50% to one or more common allergens. Worldwide, the incidence of allergic asthma is plateauing in high income countries, whilst still rising in prevalence within low and middle income countries, as they adopt a more Western-type lifestyle (Pawankar et al., 2011).

Asthma is particularly problematic in the United Kingdom, which has the highest prevalence rates of childhood asthma worldwide (Asthma UK)(Anandan et al., 2010). This corresponds to one in eleven children suffering from asthma in the UK, and has considerable impact on the quality of life of people with asthma. School children with asthma miss 254 days of school per 100 children (Asthma UK)(Mukherjee et al., 2016). For adults, asthma accounts for 79 days taken off work per 1000 adults, rendering dire financial implications for companies, with the total cost of asthma in the UK estimated to be about £5 billion (Mukherjee et al., 2016; Nunes et al., 2017). Better care is required for all these people, especially vulnerable patients with severe asthma. It was been estimated that better routine care could have prevented 46% of asthma deaths (Levy, 2015).

The pharmacological management of asthma has not changed much over recent years for most people with asthma. Newly diagnosed patients are prescribed inhalers containing corticosteroids and, if symptoms worsen, they are prescribed combination inhalers, which contain corticosteroids combined with inhalable long-acting β_2 agonists. If the symptoms still cannot be controlled, the doses are increased until the patient is given oral steroids. At this point, biologics such as monoclonal antibody treatment may be suggested as a steroid sparing treatment (British Thoracic Society, 2016).

1.2 The role of Immunoglobulin E (IgE) in asthma

Allergen-specific IgE has been recognised to play a central role in allergic asthma, with most asthma cases due to an IgE-mediated reaction after sensitisation to an inhaled allergen. IgE is a glycosylated protein belonging to the immunoglobulin family of proteins, and was the fifth and final antibody class to be identified in humans (Stanworth, 1993). IgE can either be expressed as a membrane-bound part of the B-cell receptor (BCR) as one of five B-cell receptors isotypes or be secreted as a soluble protein by plasma cells. In the serum of healthy individuals, the concentration of IgE is limited to ~50-200 ng/ml, which is relatively low compared to other antibody isotypes found at ~1-10 mg/ml (Dullaers et al., 2012).

Despite knowing the importance of IgE in asthma and allergies, the role of IgE in normal physiology is not fully understood. High levels of IgE are produced in humans in response to parasitic infections such as *Ascaris lumbricoides* (Lynch et al., 1998; McSharry et al., 1999) and *Schistosoma mansoni* (Rihet et al., 1991) although IgE is not critical for their removal (Harris and Gause, 2011). High IgE levels were also found in patients less likely to die from acute mycocardial infarction which was associated with delayed thrombin production in blood clotting (Szczeklik et al., 1993; Szczeklik and Dawson, 2000). A role for IgE in tumour surveillance has also been suggested in which IgE directs an immune attack on tumours by antibody-dependent cell-mediated cytotoxicity (Karagiannis et al., 2003; Josephs et al., 2014). Insights into the workings of this rare antibody are best understood in allergic asthma.

Asthmatic patients, when asked about their condition, would mostly describe their clinical symptoms such as wheezing, coughs and chest tightening. These symptoms are presented as the 'tip of the iceberg' depicted in Figure 1-1, while the pathophysiological role of IgE in allergic asthma is presented in the rest of the iceberg. IgE can be seen to influence every layer of the iceberg at every stage of the allergic cascade (Palomares et al., 2017). High levels of total serum IgE are associated with a high asthma severity score, lung function abnormalities, a higher risk of hospital admission and a high dose of inhaled corticosteroids (Carroll et al., 2006; Tanaka et al., 2014). Targeting IgE has been proven to be efficient in patients with severe asthma (Lowe et al., 2009; Kulus et al., 2010).

As indicated in Figure 1-1, there are two stages in the provocation or immediate hypersensitivity response; the sensitisation stage, and the re-exposure stage and IgE is involved in both. In the sensitisation stage, an inhaled allergen is taken up by an antigen presenting cell such as a dendritic cell or a B cell in the epithelium lining of the lungs or nose. The antigen presenting cells internalise the allergen, process it and present the processed allergen to other immune cells, in particular T cells, which then acquire a T helper type 2 (Th2) cell phenotype. The Th2 cells will interact specifically with B-cells through the B cell major histocompatibility complex (MHC) class II and co-stimulatory molecules, thereby causing the B-cell to undergo class-switch recombination (CSR) to induce the production of IgE antibodies (Galli and Tsai, 2012).

Further future encounters with the same allergen lead to an exacerbation that can present in two phases: the early acute phase (that occurs within minutes), and the late-phase reaction (that occurs within hours). In the early acute phase, allergen-caused crosslinking of IgE-bound to its receptor, FcεRI, located on mast cells and basophils, results in degranulation of the cell, and the release of preformed mediators such as histamine, inflammatory cytokines and chemokines, and leukotrienes. These mediators induce the constriction of bronchial smooth muscle, thereby narrowing the pulmonary airways. This results in patients presenting with clinical symptoms which include wheezing, sneezing and rhinorrhea and/or mucus secretion.

Mediators released by the early phase recruit inflammatory granulocytes and Th2 cells to the site of inflammation and continue into the late-phase reaction. These cells release toxic proteins such as eosinophilic cationic protein, major basic protein etc. and cytokines, resulting in increased vascular permeability, bronchial oedema and bronchospasms. If the immune reaction occurs without any resolution or if the allergen persists, the constant inflammation can evolve into the chronic phase. This phase is associated with airway remodelling, eosinophilia and further IgE production. IgE has been implicated in airway remodelling in humans by increasing the proliferation of airway smooth muscle, and depositing pro-inflammatory collagens and fibronectin (Roth et al., 2013; Redhu et al., 2013). For the patient, this results in progressive loss of lung function and fixed airway obstruction.
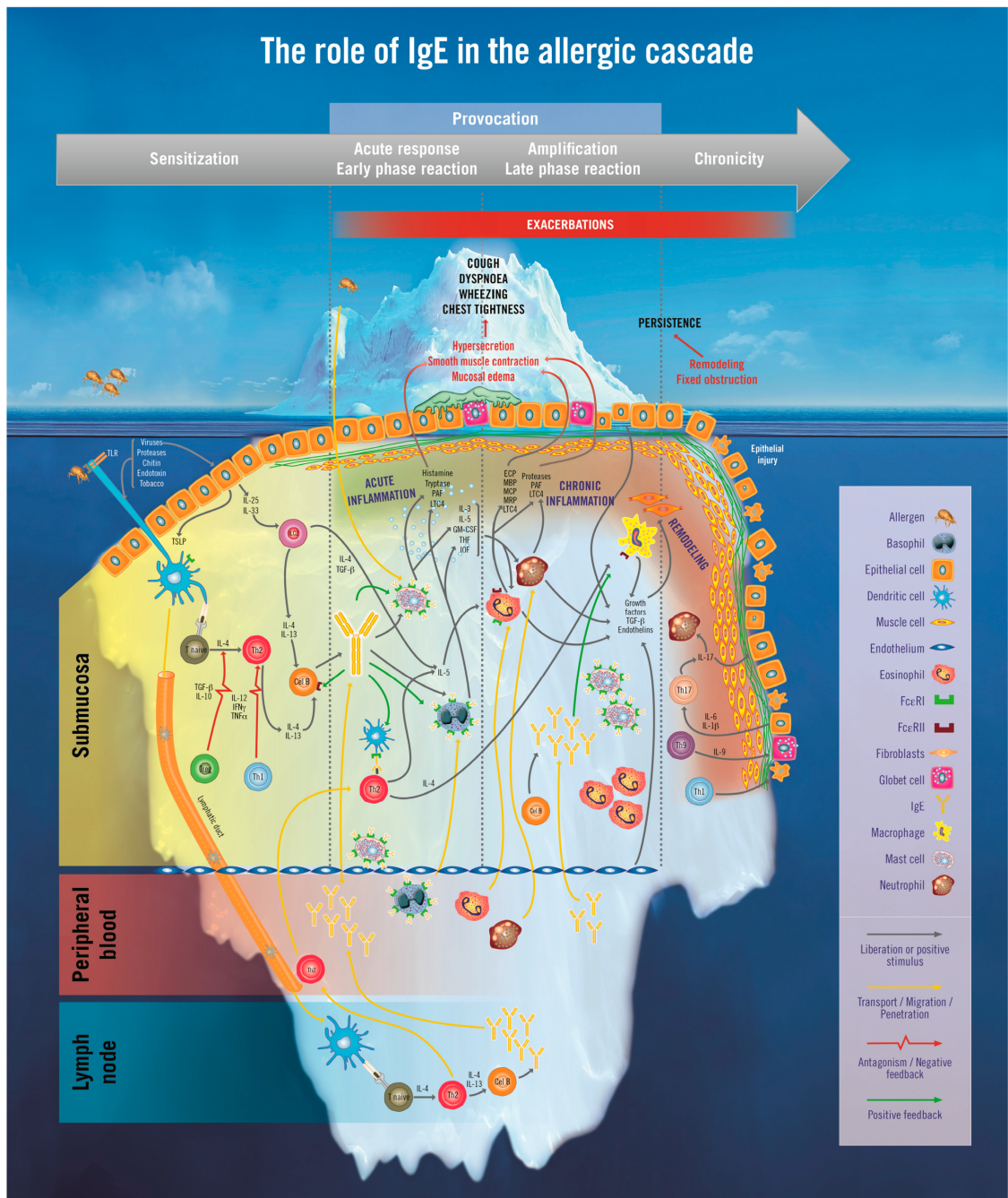
Figure 1-1 **The Iceberg model of IgE in allergic asthma**. The symptoms of allergic asthma are indicated by the portion of the iceberg above water; whilst the pathophysiological changes of the disease occur in the depths of the iceberg. Changes in the submucosa, peripheral blood vessels and lymph nodes occurs from early onset of the disease. IgE in yellow, is present at all stages of the diseases and at al layers of the iceberg. ECP = eosinophil cationic protein; FcεRI = high affinity IgE receptor; FcεRII = low affinity IgE; GM-CSF = granulocyte-macrophage colony-stimulating factor; IFNγ = interferon gamma; IgE = immunoglobulin E; IL = interleukin; ILC = innate lymphoid cells; LTC4 = leukotriene C4; MBP = major basic protein; MCP = monocyte chemotactic protein; MRP = myeloid related proteins; PAF = platelet-activating factor; TGFβ = transforming growth factor beta; Th = T helper cells, TLR: toll-like receptors; TNFα = tumour necrosis factor alpha; Treg = regulatory T cells; TSLP: thymic stromal lymphopoietin. Image reproduced from Palomares et al. 2017.

IgE plays a role in every stage of allergic asthma as depicted in Figure 1-2. It acts with a variety of cells from the sensitisation stage through to the late phase of the allergic response, and even in the chronic phase of the disease. It exerts these immunomodulatory effects via binding to its two receptors, the structure and function of which are discussed in the next sections.
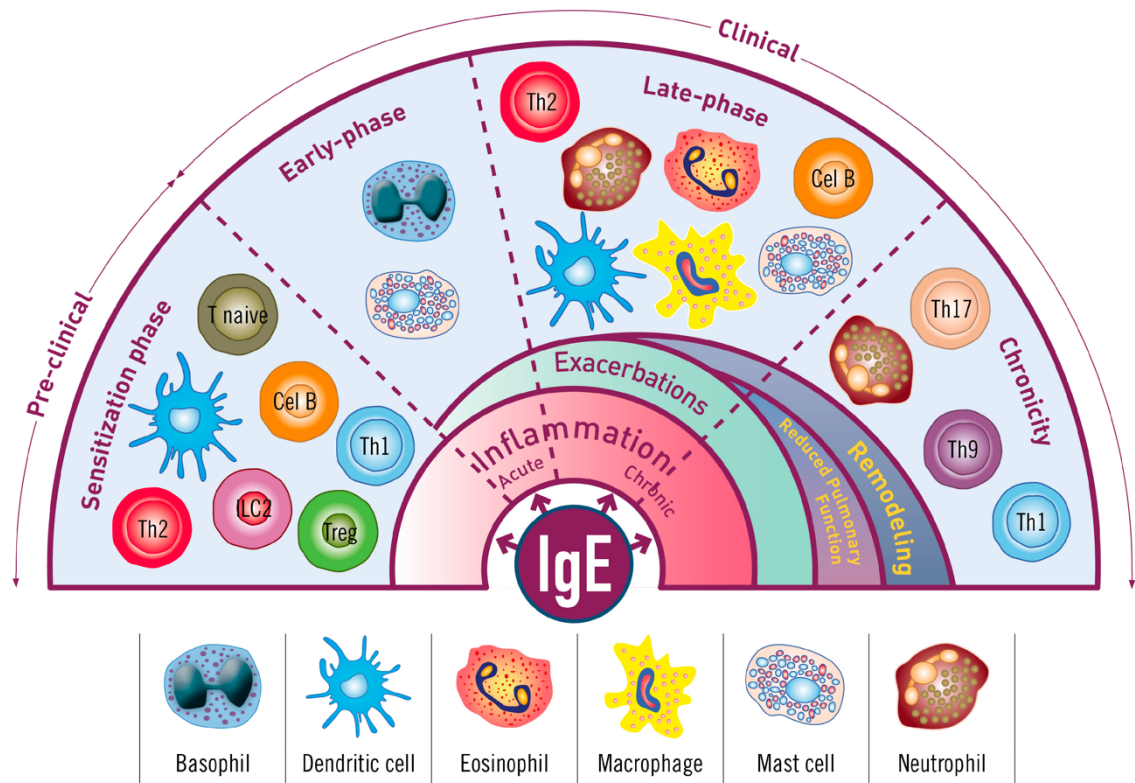
Figure 1-2 **The role of IgE in all phases of allergic asthma.** IgE plays a central role in the sensitisation, early-acute, and late phases, as well the chronic consequences of non-controlled allergic asthma. Image reproduced from Palomares et al. 2017.

1.3 The Structure of IgE

IgE is the most glycosylated antibody of all the immunoglobulin classes (Plomp et al., 2014; Arnold et al., 2004; Dorrington and Bennich, 1978) as shown in Figure 1-3. It has a molecular weight of 190 kDa, of which about 170 kDa is protein and around 20 kDa is carbohydrate. Structurally, it shares the same molecular Y-shape as all antibodies, consisting of two identical heavy chains and two identical light chains. IgE differs from immunoglobulin G (IgG), in that the heavy ε chain contains an additional domain (Cε2) in place of the hinge region of IgG. Unlike IgG, the native structure of IgE is not extended but asymmetrically bent (Wan et al., 2002).

The fragment-antigen binding (Fab) region of IgE contains the variable domains, as depicted in Figure 1-3, which bind to antigens and confers the epitope-specificity of the antibody. The constant or fragment crystallisable (Fc) regions of IgE impart its effector functions (Figure 1-3). IgE mediates its immune functions by binding through the Cε3 domain of the Fc region to its receptors: the high affinity receptor, FcεRI, and the low affinity receptor FcεRII or CD23. Simultaneous engagement of both receptors by IgE is prevented by an intrinsic allosteric mechanism (Dhaliwal et al., 2012; Borthakur et al., 2012).
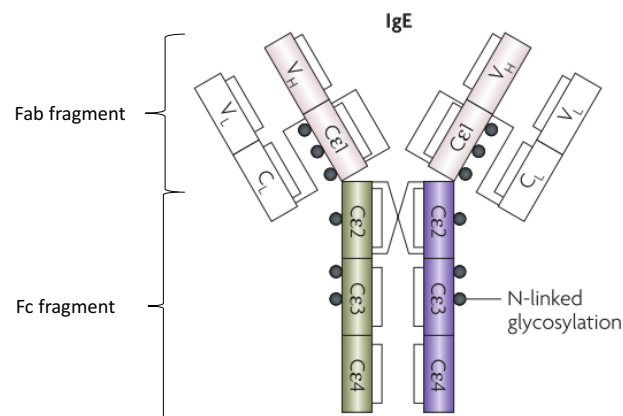
Figure 1-3 **The structure of IgE.** A cartoon representation of the domain structures of IgE showing the inter- and intra-molecular disulphide bridges, and the sites of N-linked glycosylation. Image reproduced from Gould and Sutton 2008.

1.3.1 Membrane IgE

Comparing membrane-bound IgE to soluble IgE, additional domains are observed in the membrane-bound structure as illustrated in Figure 1-4. These additional domains include the extracellular membrane proximal domain (EMPD), the transmembrane domain, and the cytoplasmic tail, which are all found complexed with the signalling chains Igα and Igβ; together they form the B cell receptor (BCR). The EMPD is found between the Cε4 domain and the transmembrane domain. In humans, the EMPD has two isoforms, which differ in length by 52 amino acids, which are involved in stabilising inter ε chain disulphide bridges. The short isoform has an EMPD length of 14 amino acids, whilst the long isoform has an EMPD length of 66 amino acids, rendering the latter isoform at least twice as long as the EMPD from any other immunoglobulin class (Achatz et al., 2008). The EMPD of the long isoform also

contains 4 cysteines, unlike other antibodies, which only contain one cysteine. Furthermore, it is proposed to have a different disulphide bonding pattern, in which two cysteines form an intramolecular bond, while the remaining cysteines form intermolecular bonds as expected (Bestagno et al., 2001).
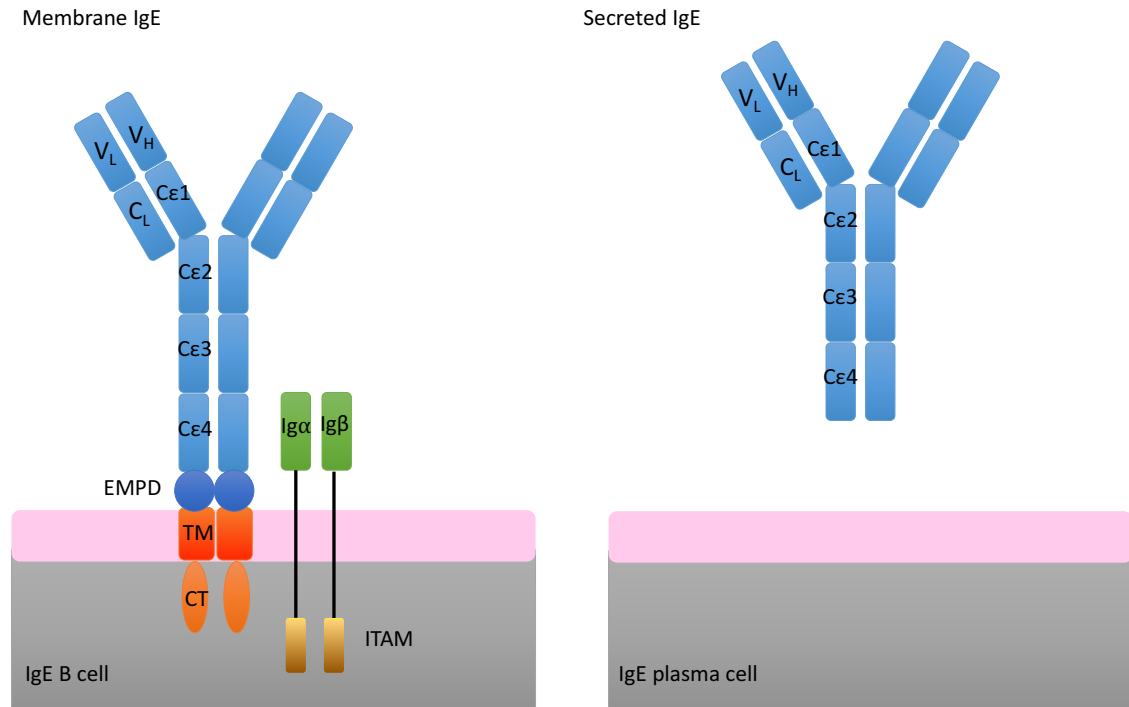
Figure 1-4 **Membrane IgE as part of the BCR and soluble IgE.** V = variable domain, Cε = epsilon constant domain, EMPD = extracellular membrane proximal domain, TM = transmembrane domain, CT = cytoplasmic tail. The two forms of IgE are synthesised by different cells.

The EMPD isoforms bind to differently glycosylated Igα proteins; the long form associates with the completely glycosylated Igα protein, whereas the short form associates with a glycoform which is partially sensitive to endoglycosidase H (Batista et al., 1996). The cell membrane expression changes too, with the short isoform transported to the cell surface at a higher rate, while transport of the long isoform is less efficient (Batista et al., 1996). The signalling pattern of the two isoforms also differs, with the kinetics of protein tyrosine phosphorylation varying, although still phosphorylating the same substrate proteins. Cross-linking of the short isoform leads to inhibition of B-cell growth, whereas signalling through the long form does not affect cell proliferation (Batista et al., 1996).

The importance of the long EMPD is not fully understood since it is not found in mice. It has been proposed to act as a sensor during apoptotic signalling, involving antigen

30

engagement of an IgE BCR (Poggianella et al., 2006). Its length might give an advantage to the conformational space the IgE-Fabs can explore, or the type of allergens it can bind to. Alternatively, the EMPD may play a role in the IgE repertoire determination or in the formation of memory IgE cells (Zuidscherwoude and van Spriel, 2012; Haniuda et al., 2016). The domain has been targeted therapeutically to prevent the production of soluble IgE, and is being considered as a second generation anti-IgE candidate (Feichtner et al., 2008; Vigl et al., 2017). The exact role of the EMPD in allergic disease is however, yet to be fully explored, with its three-dimensional structure not yet known.

The hydrophobic transmembrane region of membrane-bound IgE is similar in sequence to other immunoglobulins, with the cytoplasmic tail of IgE not expected to signal without its chaperone proteins Igα and Igβ. However, it has been suggested that the cytoplasmic tail may contribute to signalling (Yang et al., 2016).

1.3.2 IgE as a flexible protein

IgE is a very flexible protein that can bend such that the Cε2 domains fold back over the Cε3-4 domains or extend all the Fc domains (Drinkwater et al., 2014). In solution or when bound to its receptors, IgE is found in a symmetrically bent conformation (Zheng et al., 1992). IgE can, however, extend from its bent conformation, and 'flip' to a bent conformation on its other side. A proposed timeline of IgE unbending, determined by molecular dynamics, can be observed in Figure 1-5. This movement is not a simple hinge movement, as it consists of the $(Cε2)_2$ domains twisting relative to the Cε3-4 domains, and may involve local disorder or unfolding of domains or linker regions to accommodate such large-scale motions (Drinkwater et al., 2014).

The reason behind IgE's flexibility can perhaps be understood in terms of allergen-binding. A bent IgE structure, bound to FcεRI, is optimally positioned for allergen capture, and crosslinking, which triggers the allergic response. In constrast, an extended IgE molecule is beneficial for allergen capture when IgE is part of the BCR because the Fab arms can sample a greater range of conformational space, thus ensuring B-cell survival and proliferation (Drinkwater et al., 2014).
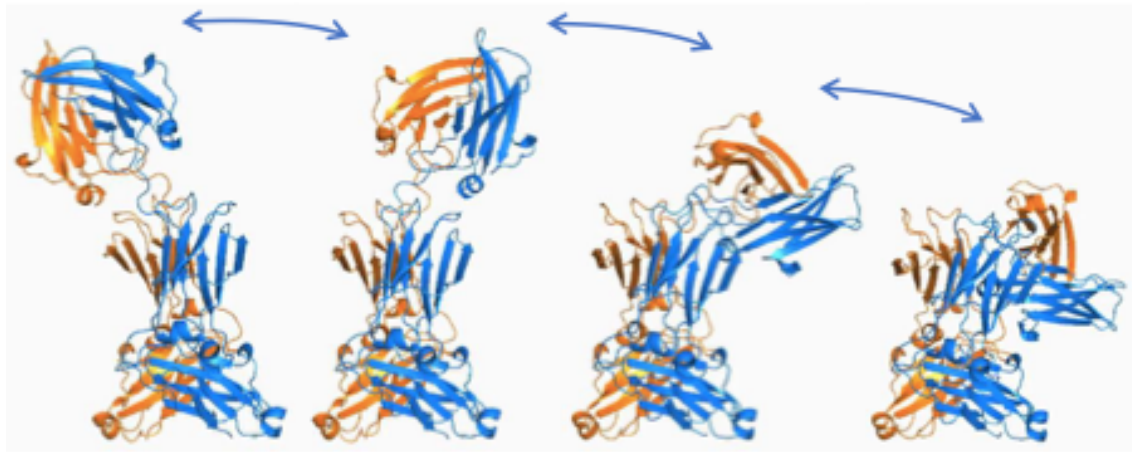
Figure 1-5 **IgE-Fc unbending**. Theoretical conformations of IgE-Fc found in energy minima of molecular dynamics simulations illustrate a possible pathway for the IgE-Fc unbending process. This process involves twisting and bending in an asymmetrical fashion. Image adapted from Drinkwater et al. 2014.

1.4 IgE and its Receptors: The high affinity receptor, FcεRI

FcεRI, when activated by allergen cross-linked IgE, causes mast cell and basophil degranulation as shown in Figure 1-6A. This interaction is fundamental to the allergic response; as seen in FcεRI deficient mice, local or systemic anaphylactic responses were abolished when compared to wild type (WT) mice (Dombrowicz et al., 1993; Kinet, 1999). The FcεRI protein is a membrane-bound complex expressed as a tetramer on mast cells and basophils. It is composed of one α subunit, one β subunit and two γ subunits as depicted in Figure 1-6B. The alpha subunit contains one transmembrane domain, and two Ig-like domains (D1 and D2) (Garman et al., 1998; Kinet, 1999). D2 contains two asymmetric binding sites for one IgE-Fc molecule, so that each Cε3 domain of IgE-Fc binds to one FcεRI receptor (Garman et al., 2000). The β and γ subunits contain Immunoreceptor tyrosine-based activation motifs (ITAMs) involved in signal transduction. The human high affinity receptor can also be expressed without the β subunit as a trimeric αγ2 receptor, which is expressed on many more immune cells such as monocytes, macrophages, eosinophils, dendritic cells and platelets (Kinet, 1999).
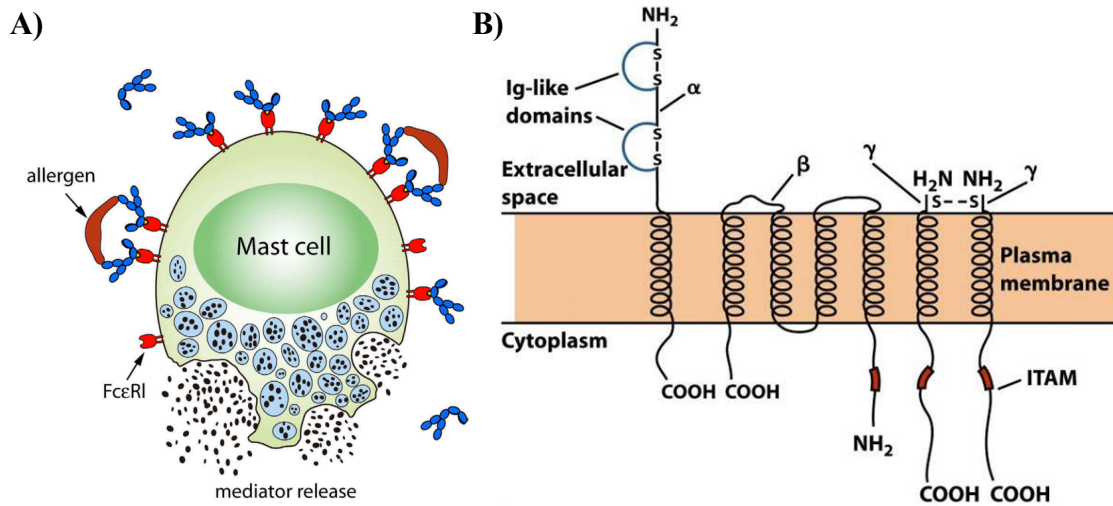
Figure 1-6 **The role of FcεRI and its arrangement of subunits. A)** FcεRI is involved in mast-cell degranulation. Image reproduced from Wright et al. 2015. **B)** A schematic diagram of the FcεRI tetramer (Kindt et al., 2007). Image reproduced from Kuby Immunology, sixth edition, 2007.

The D1 and D2 domains of FcεRI fold back to reveal a hydrophobic 'ridge', which is the IgE binding site, allowing for a 1:1 binding stoichiometry. There are in fact, two hydrophobic sub-sites on Cε3 of IgE that bind to FcεRI with high affinity ($K_a \sim 10^{10}$ M^{-1}) (Garman et al., 2000) as depicted in Figure 1-7. The high affinity coupled with the slow dissociation rate ($k_{off} \sim 10^{-5}$ s^{-1}) (Holdom et al., 2011) ensures that FcεRI present on mast cells is saturated with bound IgE, even in the absence of antigen. Combined with local production of IgE in the tissues and rebinding of IgE, the IgE stays associated for most of the cell's life span so that the cells are always primed for the Th2 cell response. These characteristics of IgE are unique among the antibodies classes; IgG has as lower affinity and faster k_{off} rate constant for its receptors when compared to IgE (Maenaka et al., 2001).

Although the extracellular domains of the α subunit contain seven N-linked glycosylation sites, and the Cε3 domain of IgE is also glycosylated, the interaction between FcεRI and IgE-Fc is strictly protein based. The key IgE residue involved in FcεRI binding in subsite 1 is R334, involved in a salt bridge and hydrogen bonding at the interface as shown in Figure 1-7A. The R334S mutation has a pronounced effect on the affinity of IgE-Fc for FcεRI, reducing affinity by 50-fold (Henry et al., 1997; Hunt et al., 2008) The second subsite shown in Figure 1-7B, is characterised by a 'proline sandwich'. Pro426 of IgE-Fc is placed between Try87 and Trp110 of the α subunit D2 domain, a conserved feature found in IgG-FcγR complexes (Holdom et al., 2011).
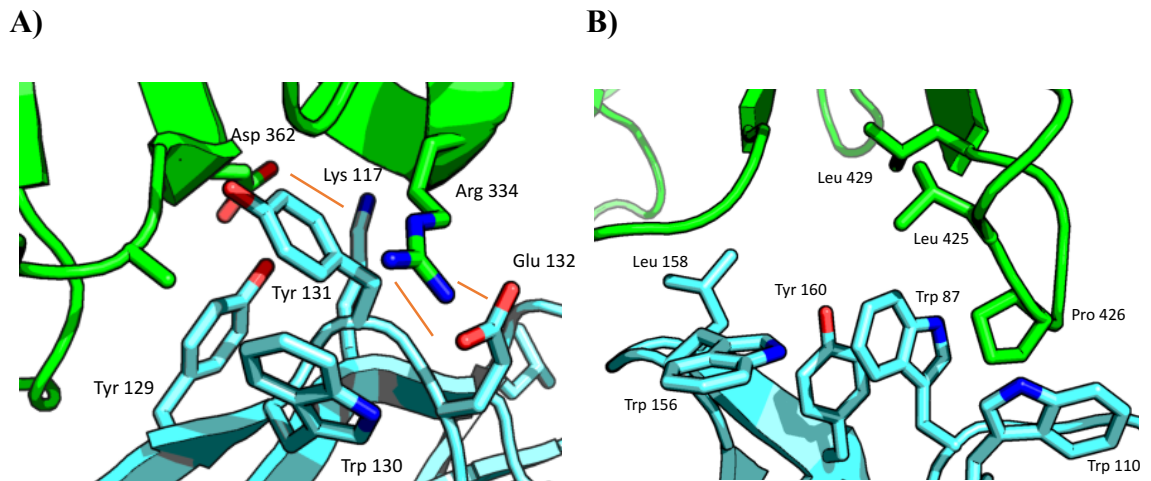
Figure 1-7 **IgE binding to FcεRI; interactions at both subsites.** Green = IgE-Fc, blue = FcεRI, orange = hydrogen bonds. **A)** Cε3 domain binds via two salt bridges (Arg334-Glu132 and Asp362-Lys117) to a single chain of FcεRIα. **B)** The proline sandwich at site 2, with Pro426 of Cε3 packed between Trp87 and Trp110 of the receptor (blue). PDB file: 1F6A.

Inhibiting the IgE-FcεRI interaction has been proven to be clinically efficacious as demonstrated by the first generation anti-IgE biologic Xolair® (Omalizumab), further discussed in section 1.6.

1.5 IgE and its Receptors: The low affinity receptor, FcεRII or CD23

CD23 differs from most other immunoglobulin receptors, as it belongs to the C-type (calcium-dependent) lectin superfamily. As a membrane-bound protein, CD23 is a member of the type II transmembrane glycoprotein family, and as a trimer composed of three lectin domains, followed by a triple α-helical coiled-coil stalk, as illustrated in Figure 1-8A. The stalk is distanced from the membrane by a short extracellular sequence before the single transmembrane domain, while the intracellular part of the protein contains a short N-terminal cytoplasmic tail. The stalk region is susceptible to proteolysis by proteases such as A Disintegrin And Metalloproteinase Domain-containing protein 10 (ADAM10), and the major house dust mite protease allergen *Der p* I, to release soluble fragments pictured in Figure 1-8B & C (Lemieux et al., 2007; Shakib et al., 1998). *Der p* I cleaves CD23 at two sites: one site is located in the stalk at Ser155-Ser156, and the second site is located in the C-terminal tail that binds to CD21 (CR2), at Glu298-Ser299, as depicted in Figure 1-8A (Schulz et al., 1997). The cleavage results in a 16 kDa fragment known as derCD23, as shown in Figure 1-8C.

DerCD23 retains its binding activity to CD21, which is a membrane glyco-protein expressed on B-cells (Nadler et al., 1981). The ectodomain of CD21 is composed of 15 (Weis et al., 1988; Gilbert et al., 2006) or 16 (Pochon et al., 1992) short consensus repeat domains with a total length of 38 nm. CD21 on B cells upregulates IgE production by binding to soluble CD23 (Aubry and Pochon, 1992), as discussed further in section 1.5.3.
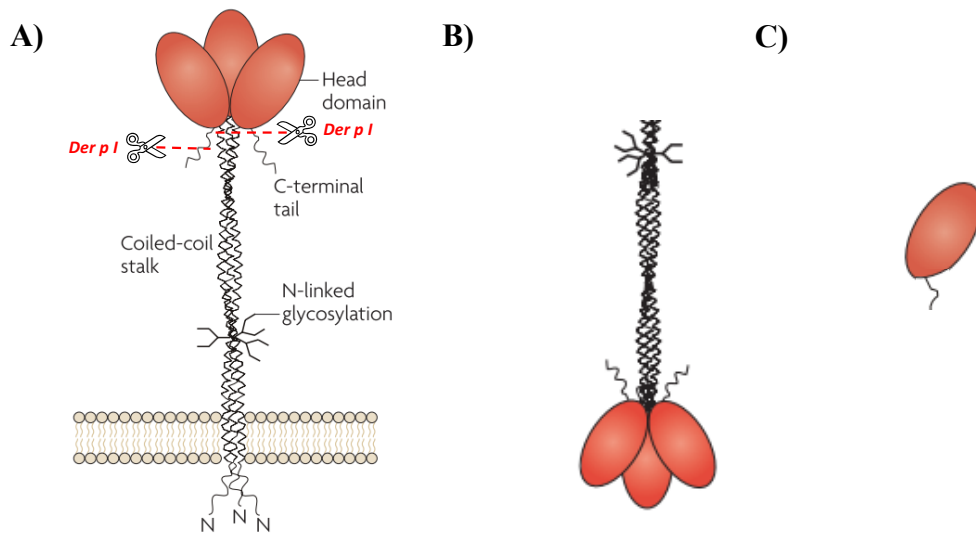
Figure 1-8 M... le forms. A) Membrane-bound CD23, the scissors denote where the pro... . B) Soluble trimeric CD23 C) Soluble derCD23. Images adapted from Gould and Sutton 2008.

Human membrane CD23 is expressed on a range of cells, including T and B cells (Kikutani et al., 1989; Armitage et al., 1989) monocytes (Vercelli et al., 1988), follicular dendritic cells, intestinal epithelial cells (Yu et al., 2003), and bone marrow stromal cells. In mice, membrane CD23 expression is limited to B cells, follicular dendritic cells (Richards and Katz, 1990) and enterocytes (Yu et al., 2003). The coiled-coil stalk of membrane CD23 is protected from proteolysis when bound to soluble IgE, which leads to a reduction in the synthesis of IgE, as shown in Figure 1-9. Soluble trimeric CD23 fragments through interactions with CD21, increase IgE synthesis, as shown in Figure 1-9, whereas soluble monomeric CD23 reduces IgE synthesis. These multiple roles of CD23, dependent on the oligomerisation and soluble or membrane-bound state of the protein, demonstrate the importance of CD23 in regulating IgE homeostasis (McCloskey et al., 2007; Cooper et al., 2012).

C-type lectins usually bind to carbohydrate, and CD23 is unusual in this respect, in that it binds to its ligands via protein-protein interactions. IgE is the principal ligand for CD23, however, human CD23 also binds to domains 1 and 2 of the complement receptor CD21 (CR2) (Aubry and Pochon, 1992), MHC class II proteins (Kijimoto-Ochiai, 2002), the vitronectin receptor ($\alpha_V\beta_3$-integrin) (Hermann et al., 1999) and the CD11b, CD11c and $\alpha_4\beta_5$-integrins, interactions that are crucial for human B cell survival (Acharya et al., 2010). A putative carbohydrate binding site to CD23 on CD21 domains 5-8 has been described (Aubry et al., 1994) but this has not been confirmed subsequently (Hibbert, 2005) and no crystal structure of the CD21-CD23 complex exists.
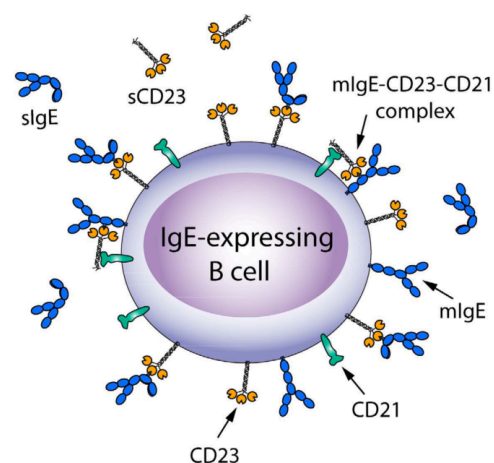
Figure 1-9 **The roles of membrane-bound and soluble CD23**. CD23 regulates soluble IgE (sIgE) production. Membrane CD23 binds to sIgE to reduce IgE synthesis. Soluble CD23 interacts with membrane IgE and CD21 to upregulate IgE synthesis. Image reproduced from Wright et al. 2015.

As previously mentioned, CD23 has various roles dependent on its oligomerisation state, and this also has an impact on the affinity for its interactions with its ligands. A single head domain of CD23 binds to IgE with a K_A of $\sim 10^6$-10^7 M^{-1}. Due to their low binding affinities, C-type lectins often enhance their overall affinity for a ligand by oligomerising, to create an avidity affect (Taylor and Drickamer, 2003). The affinity of trimeric CD23 for IgE had a K_A of $\sim 10^8$-10^9 M^{-1}, approaching the affinity for FcεRI for IgE with a K_A of $\sim 10^{10}$ M^{-1} (Gould and Sutton, 2008).

1.5.1 The Structure of derCD23 as a C-type lectin

C-type lectins (CTLs), found on a plethora of cells, characteristically bind carbohydrates in a calcium-dependent manner. The functional versatility of this protein fold, depicted in Figure 1-10A, enables a wide range of interactions; from cell adhesion molecules to pathogen recognition receptors in innate immunity (Cambi et al., 2005; Zelensky and Gready, 2005; Drickamer and Taylor, 2015). The head domain of CD23 contains the C-type lectin-like domain (CTLD), as shown in Figure 1-10B. This domain has four conserved cysteine residues labelled C1-C4 in Figure 1-10, two of which form a cysteine bridge at the base of the long-loop region (Zelensky and Gready, 2003). The long-loop region begins and terminates at the same point of the core domain and encompasses loop 1 to loop 4 of the CTLD structure (Zelensky and Gready, 2005). It is also the region involved in calcium-dependent carbohydrate binding.

Calcium ions tend to bind to the long-loop region at two locations, namely the primary and auxiliary binding sites. Structures of CTLDs reveal that they are capable of binding to zero, one, two or three calcium ions. The primary site is involved in carbohydrate binding (Zelensky and Gready, 2005). CD23 in many species is predicted to bind to two calcium ions based on sequence analysis. Human CD23 is unusual in that it only binds to one calcium ion at the primary site with an affinity of 1.5 mM (Hibbert et al., 2005; Yuan et al., 2013). The residues in the auxiliary binding site of human CD23 are positively charged, as shown in Figure 1-11, unlike the residues in CD23 of other species, meaning that they are unlikely to co-ordinate a calcium ion.

The calcium-dependent carbohydrate interactions of CTLs can be divided into two groups, based on their glycan-specificity. CTLs containing a QPD tripeptide motif preceding loop 4, bind to galactose and N-acetylgalactosamine, while CTLs with an EPN motif are specific binders for mannose, N-acetylglucosamine, glucose and L-fucose (Lee et al., 2011). Human CD23 lacks both motifs, however murine CD23 contains the EPN motif. In human CD23, the asparagine of the EPN has been substituted with a threonine residue, causing the loss of the carbonyl group that coordinates calcium ions for glycan binding. In fact, human CD23 is unable to bind to

carbohydrates, and binds to its primary ligand, IgE, via carbohydrate-independent interactions only.
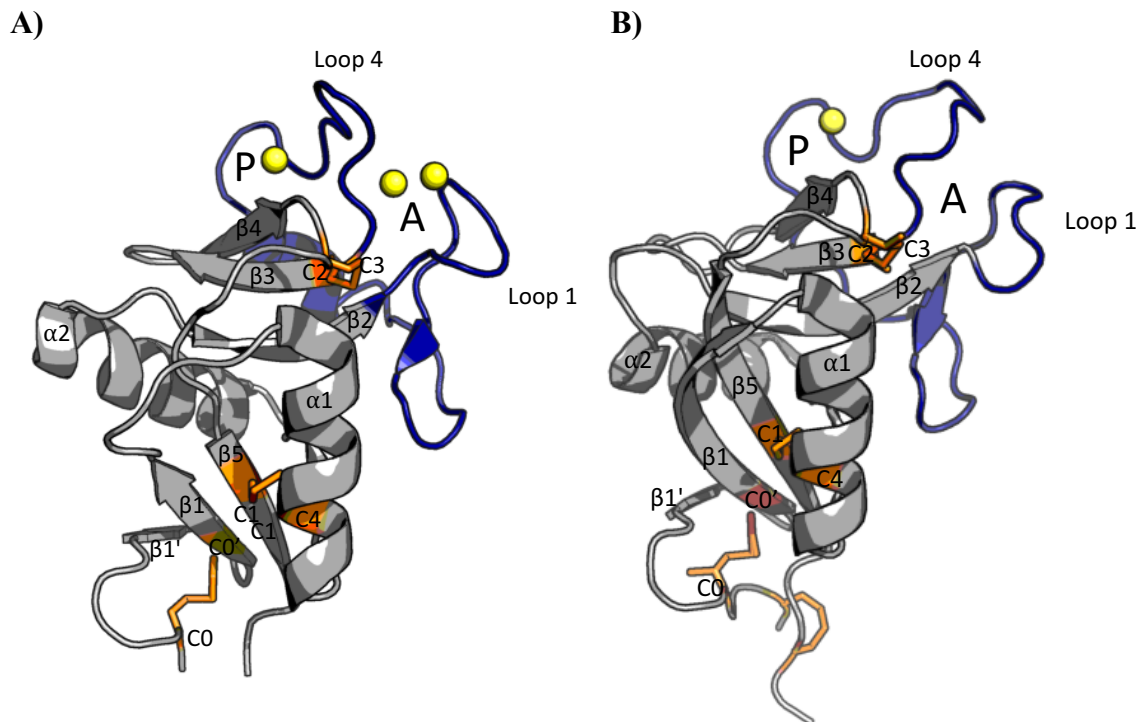
Figure 1-10 **Structural comparison of a typical CTLD and WT human derCD23**. Yellow spheres = calcium ions, orange = cysteine bridges, blue = long loop region. P = primary calcium binding site, A = auxiliary calcium binding site. α helices and β strands are numbered according to convention. **A)** A cartoon representation of a typical CTLD, DC-SIGN, pdb file: 1K9I. **B)** A cartoon representation of WT human derCD23, pdb fie: 4G9A.

The loss of carbohydrate binding in human CD23, coupled with the loss of calcium binding in the "auxiliary" binding site, may have freed those co-ordinating residues to make additional contacts with other ligands, such as IgE. This may have resulted in stronger binding to IgE, which in the context of allergy, may be a factor in determining why humans are more prone to allergies whilst mice do not present these symptoms.
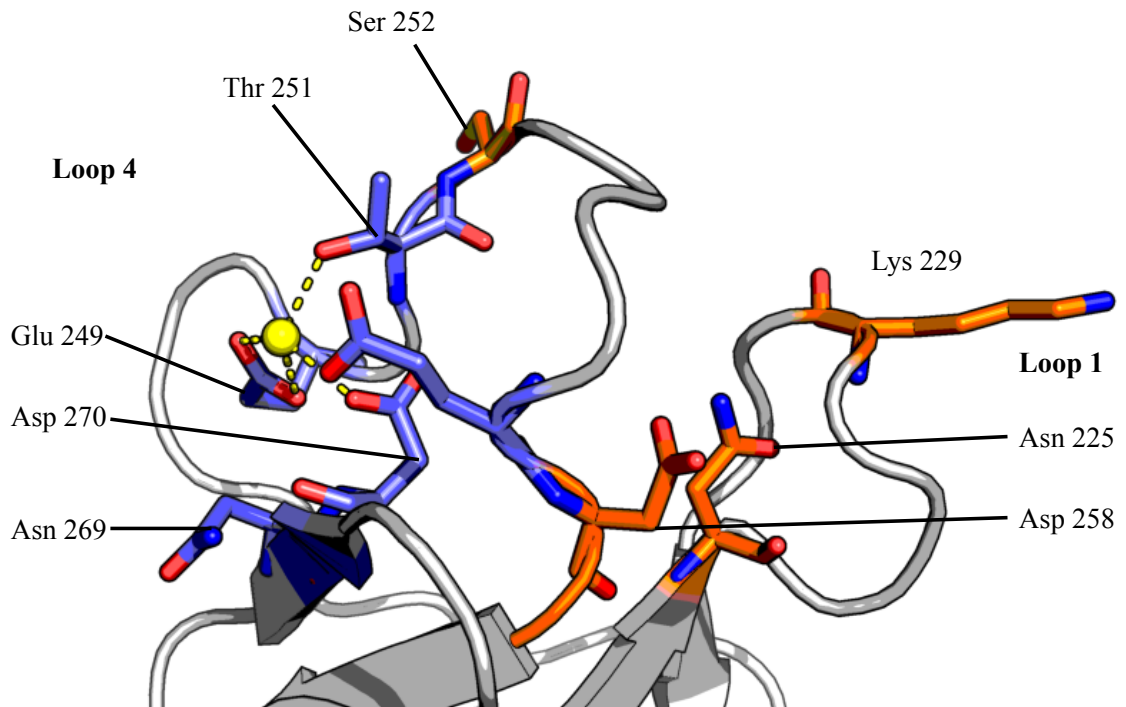
Figure 1-11 **The calcium binding site of human derCD23**. The principal calcium-ion-binding site (blue) and the "auxiliary" site (orange) in wild-type human CD23. Calcium ion shown in yellow, residues and loops labelled, pdb file: 4G9A.

1.5.2 CD23 interactions with IgE

Calcium is not essential for CD23 binding to human IgE-Fc, but enhances the affinity of derCD23 for IgE-Fc by 30-fold (Hibbert et al., 2005; Yuan et al., 2013). The protein interface of the derCD23/Cε3-4 interaction is illustrated in Figure 1-12. CD23 binds to IgE with a 2:1 stoichiometry, in contrast to the 1:1 stoichiometry seen between IgE and FcεRI. Loops 1 and 4 contribute to IgE binding; loop 1 consists of derCD23 residues Leu226 to Glu231 while loop 4 consists of residues Arg253 to Glu257. The poor surface complementarity means that the interaction is dominated by salt bridges, and hydrogen bonds (Dhaliwal et al., 2012). Binding of IgE-Fc (consisting of the Cε2, Cε3 and Cε4 domains) to derCD23 involves the same residues as that of the derCD23/Cε3-4 interaction, but with an additional hydrogen bond between His216 of derCD23 and Thr260 of Cε2 (Dhaliwal et al., 2017).
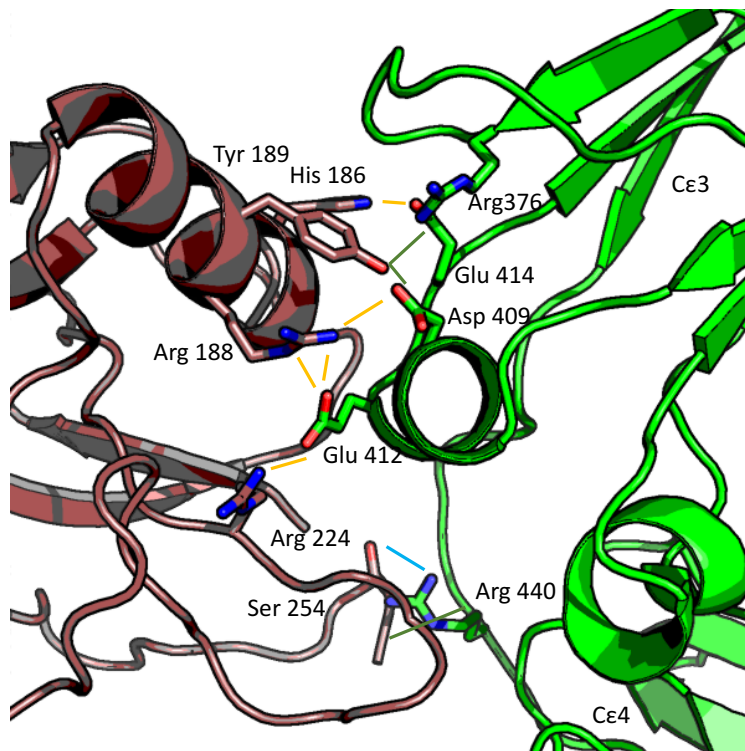
Figure 1-12 **Calcium-free IgE binding to derCD23.** Red = derCD23, green = IgE, orange = salt bridges, green = hydrogen bonds, blue = a hydrogen bond found in 5 out 6 molecules in the crystallographic unit cells. PDB file: 4EZM.

In the presence of calcium, loop 4 of derCD23 makes additional contacts with Fcε3-4, as depicted in Figure 1-13. Two new salt bridges form between Asp227(derCD23) and Arg440 of Fcε3-4, and Asp258(derCD23) and Arg440, as well as two new hydrogen bonds between Gln255(derCD23) and Arg440, and Cys273 (derCD23) and Ser437 of Fcε3-4, despite the two hydrogen bonds between Ser254(derCD23) and Arg440 being lost (Yuan et al., 2013). Thermodynamically, without calcium, the derCD23/Fcε3-4 interaction is highly temperature-dependent, however, in the presence of calcium, it less so. The binding affinity of derCD23 for Fcε3-4 changes by two-fold with vs without calcium present at 5°C, while at 35°C there is a 30-fold difference (Yuan et al., 2013).
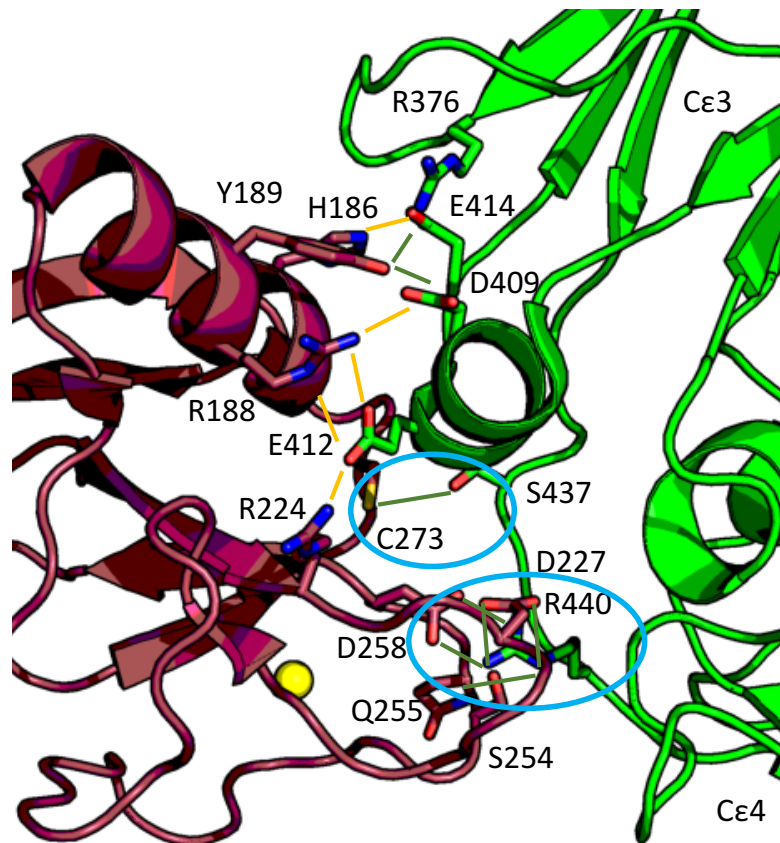
Figure 1-13 **IgE binding to calcium-bound derCD23.** Red = derCD23, green = IgE, orange = salt bridges, green = hydrogen bonds, blue circle = areas of difference, compare with Figure 1-12. PDB file: 4GKO.

The FcεRI and CD23 binding sites on IgE are allosterically linked (Dhaliwal et al., 2012). This means that binding of one receptor to IgE-Fc is transmitted through the protein to the other functional binding site, allowing for the regulation of IgE activity. This activity might have evolved to prevent CD23 crosslinking IgE-bound to FcεRI on mast cells and basophils, which would result in systemic anaphylaxis (Gould and Sutton, 2008). The different orientations of the Cε3 domains and the clashes that arise when the other receptor is modelled onto the complex can be seen in Figure 1-14.
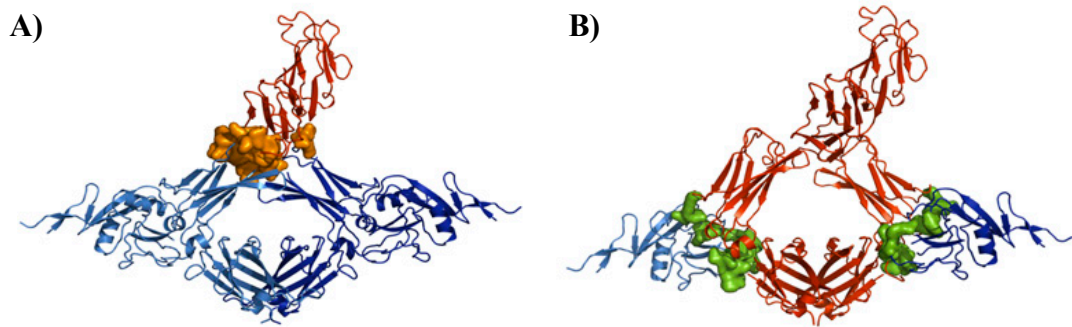
Figure 1-14 **The steric clashes that arise when both IgE receptors are modelled to bind to Fcε3-4.** Orange spheres = clashes when FcεRI is modelled onto the derCD23-Fcε3-4 complex. Green spheres = when derCD23 is modelled onto the FcεRI - Fcε3-4 complex. **A)** The Cα trace of the derCD23-Fcε3-4 'open' complex shown in blue with FcεRI in red. **B)** The Cα trace of the FcεRI-Fcε3-4 'closed' complex in red with derCD23 in blue. Image reproduced from Dhaliwal et al. 2012.

1.5.3 The Role of CD23 in IgE homeostasis

Membrane and soluble CD23 play two distinctly differing roles in regulating IgE synthesis. As a trimeric soluble protein, human CD23 binds to membrane-bound IgE, and CD21 on B cells, thereby increasing IgE synthesis as depicted in the left-hand side of Figure 1-15. When CD23 is membrane bound, it is capable of binding to allergen-IgE complexes, which results in the reduction of IgE synthesis, as seen in the right-hand side of Figure 1-15. CD23 also has a role in buffering against the accumulation of soluble IgE; when bound to the antibody, the binding site to FcεRI is blocked by CD23, rendering IgE incapable of binding to mast cells, and thus avoiding triggering the allergic cascade (Gould and Sutton, 2008).
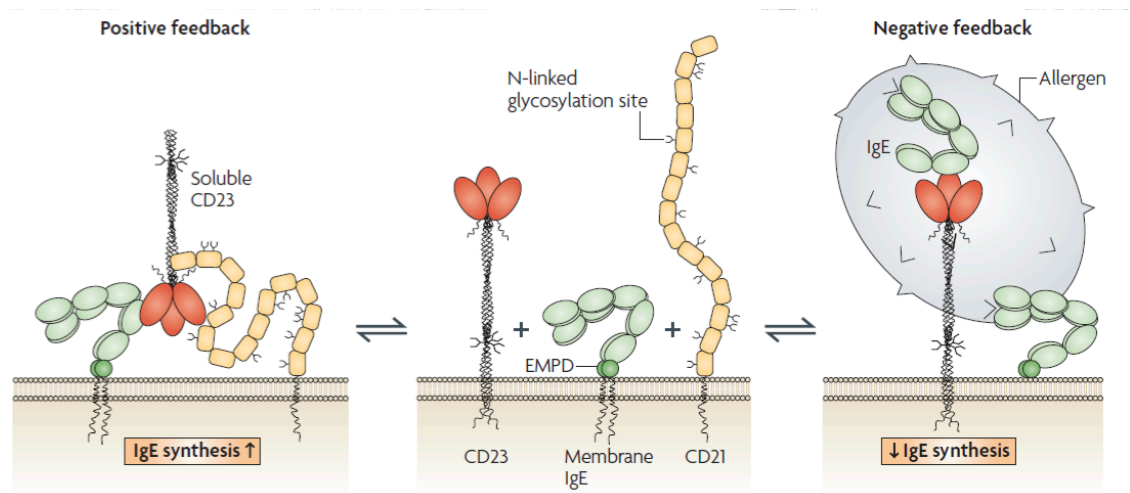
Figure 1-15 **Positive and negative mechanisms of IgE regulation by CD23.** The co-ligation of membrane IgE, CD21 and soluble CD23 upregulates IgE synthesis. Membrane CD23 and IgE co-ligates with IgE-allergen complexes to negatively regulate IgE synthesis. Homeostasis (in the middle panel) is achieved by CD21 and CD23 competing for membrane IgE. Image reproduced from Gould and Sutton 2008.

This dual regulatory mechanism only appears to operate in humans, since in mice, CD23 is unable to bind to CD21. Without a CD21 homotypic interaction, only the negative-feedback mechanism is likely to function in mice, and this agrees with data showing the up-regulation of IgE synthesis in CD23-deficient mice (Yu et al., 1994). These precise mechanisms are still being debated, and could be resolved if the correct tools could be produced to probe these long-standing questions (Conrad et al., 2007).

1.5.4 The Role of CD23 in receptor-mediated endocytosis

Receptor-mediated endocytosis is a cellular method of selectively internalising ligands such as nutrients, pathogens, toxins and many other ligands as illustrated in Figure 1-16. Once internalised, the ligand and receptor dissociate, allowing the receptor to be recycled and returned to the plasma membrane, while the ligand is processed in the lysosomal pathway. A number of different protein families can act as receptors for this process, including C-type lectins such as membrane CD23, which traffics IgE-allergen complexes for allergen presentation (Carlsson et al., 2007). The CTL endocytic receptors all contain a common model for calcium-dependent binding, in which the co-ordination of a calcium ion by a set of amino acids is of vital importance (Andersen and Moestrup, 2014).
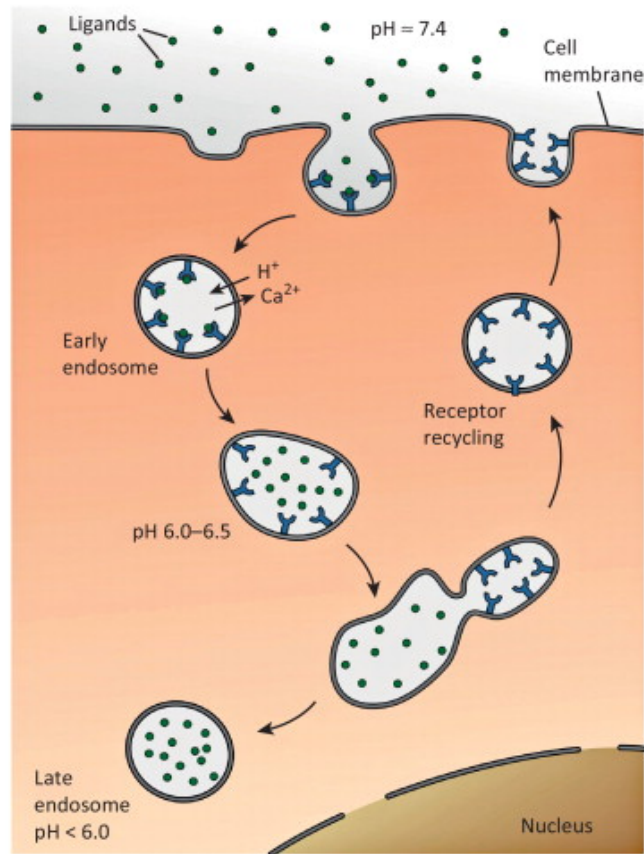
Figure 1-16 **A model of receptor-mediated endocytosis.** Ligands bound to membrane-bound receptors are internalised by a clathrin-dependent pathway. In the early endosome, calcium efflux and proton influx causes ligand dissociation. The receptors cluster to one region of the endosome and pinch off to be recycled back to the plasma membrane. The sequestered ligands fuse with late endosomes/lysosomes and the decrease in pH results in ligand processing or digestion by lysosomal enzymes. Image reproduced from Andersen & Moestrup 2014.

Upon internalisation, the ligand dissociates from the receptor due to a rapid decrease in calcium concentrations of about 100-fold or more and a drop in pH (Gerasimenko et al., 1998). This point of the pathway has the lowest calcium concentrations. The ligand and receptor are segregated into different lysosomal compartments, and the part containing the ligand fuses with a late-endosome/lysosome, leading to further acidification, and an decrease in calcium concentrations followed by lysosomal enzymes digesting the ligand. In this context, calcium can be considered a switch for endocytic receptors. Calcium turns endocytic receptors, such as CD23, 'on' upon binding at high calcium concentrations at the plasma membrane. It then, turns endocytic receptors 'off' by dissociating when the calcium rapidly drops in the endosomal pathway (Andersen and Moestrup, 2014). It is not clear if the low calcium concentration may also induce

aggregation or influence the recycling of endocytic receptors (Andersen and Moestrup, 2014).

1.5.5 The Role of CD23 in Facilitated Antigen Presentation

The Th2 cell response in allergy is well documented, especially during the late-phase reaction. These cells are activated by MHC class II, and co-stimulatory molecules found on antigen presenting cells (APCs) such as dendritic cells, macrophages and B cells. These APCs sample extracellular antigens from the environment by three main methods: phagocytosis, pinocytosis and receptor-mediated endocytosis. A variation of receptor-mediated endocytosis, unique to allergy and involving CD23, is known as facilitated antigen presentation (FAP).

FAP is a more specific method of T-cell antigen presentation, enabling T-cell activation at antigen concentrations 100-1000 fold lower than regular antigen uptake by APCs (van der Heijden et al., 1995; van Neerven et al., 2006). During this process, an allergen-bound IgE binds to membrane-bound CD23, which initiates receptor-mediated endocytosis and results in the internalisation of the complex (Engeroff et al., 2017). Inside the endosome, an efflux of calcium and a change in pH causes the allergen-IgE complex to dissociate, and the allergen is proteolytcically cleaved into peptides. The processed allergen peptides are then loaded onto MHC class II molecules for presentation at the B-cell surface as illustrated in Figure 1-17. The receptor, CD23, is then recycled to the cell membrane. It has also been observed that CD23 can co-express with MHC class II molecules on the B-cell membrane, and it is then recycled alongside them to the cell membrane to enhance efficiency (Karagiannis et al., 2001).
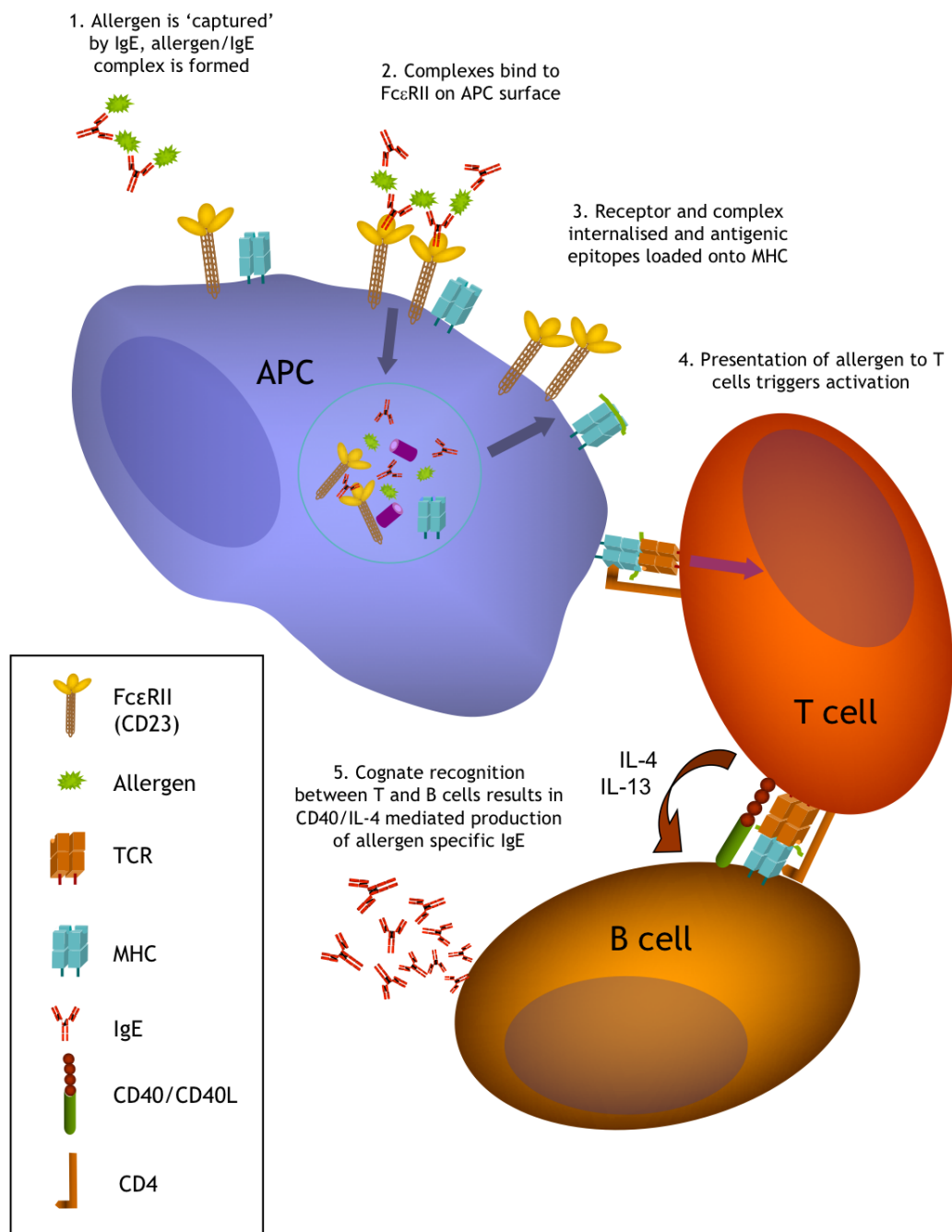
Figure 1-17 **CD23-mediated IgE-facilitated allergen presentation.** The cartoon depicts the process of capture of allergen-IgE complexes by membrane CD23 on the surface of B cells and the resulting presentation of the processed allergen to the allergen-specific T-cell. This process results in the further production of allergen-specific IgE. Image reproduced from Wilcock et al. 2007.

FAP has been demonstrated *in vivo* in both mice (Heyman et al., 1993; van der Heijden et al., 1995) and humans (Klunker et al., 2007; Poulsen, 2008), and studies have explored methods of inhibiting the process. Van Neerven and colleagues have demonstrated that blocking antibodies induced during immunotherapy were capable of inhibiting IgE FAP-mediated T-cell activation (van Neerven et al., 1999). In this paper, the authors suggest that blocking antibodies (allergen-specific IgG antibodies)

prevented the allergen-IgE complex binding to CD23, thus halting the FAP process. This observation was confirmed by Wachholz and colleagues (Wachholz et al., 2003).

FAP is also related to the phenomenon of epitope spreading, in which a B cell simultaneously processes an unrelated antigen or a different epitope of the same antigen, and presents it to a T-cell, leading to the production of antibodies directed against both antigens, as illustrated in Figure 1-18. It has been proposed that this mechanism might be partially responsible for allergy expansion in which a patient originally allergic to one allergen becomes allergic to multiple allergens (Mudde et al. 1995b; Mudde et al. 1995a).
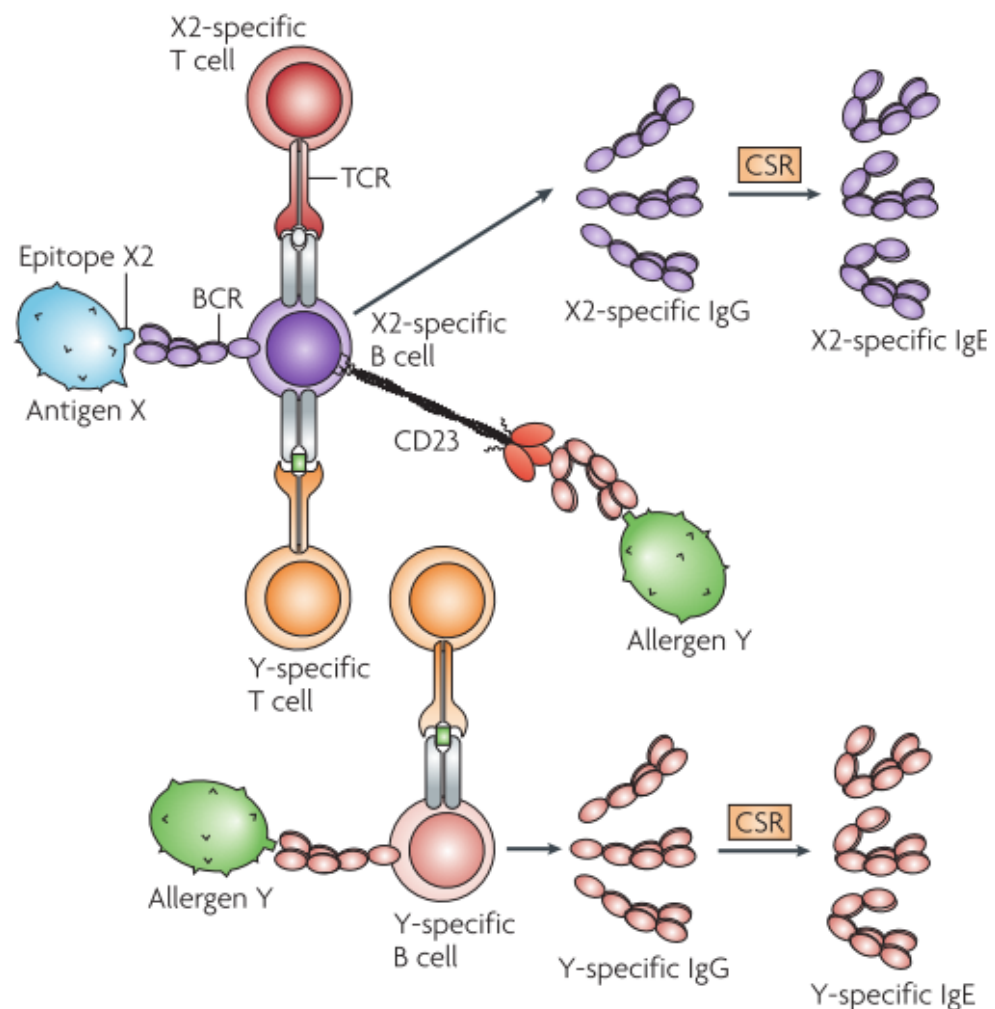
Figure 1-18 **CD23–dependent epitope spreading.** B cells recognize and present specific antigen X to specific T-cells. IgE-allergen-Y complexes bound to CD23 are internalised via endocytosis along with MHC class II molecules. Peptides generated from the allergen processing are loaded onto MHC II, and presented to T cells via the T-cell receptor (TCR). The allergen-specific T cells can differentiate to Th2 cells and together with IL-4 production can induce class switch recombination (CSR) to IgE in the B cell leading the B cells to secrete new IgE antibodies. Image reproduced from Gould & Sutton, 2008.

1.5.6 Epitope Spreading and Allergen Immunotherapy

A possible method of limiting epitope spreading is via a treatment called allergen immunotherapy (AIT). AIT is the administration of increasing doses of an allergen to an allergic patient in a controlled environment, thereby inducing immunological tolerance. This treatment provides people with allergy with an alternative to allergen avoidance or pharmacotherapy, furthermore it has long-lasting effects with demonstrable clinical efficacy (Durham et al., 1999; Fitzhugh and Lockey, 2011). During AIT, IgE production is suppressed, whilst IgG_4 production is increased, which is beneficial due to the non-inflammatory properties of IgG_4 and its inability to activate mast cells and basophils (Fujita et al., 2012). The role of IgG_4 is not limited to the early phase allergic response, and it might also be involved in inhibiting the late phase response by blocking FAP (Wilcock et al., 2006).
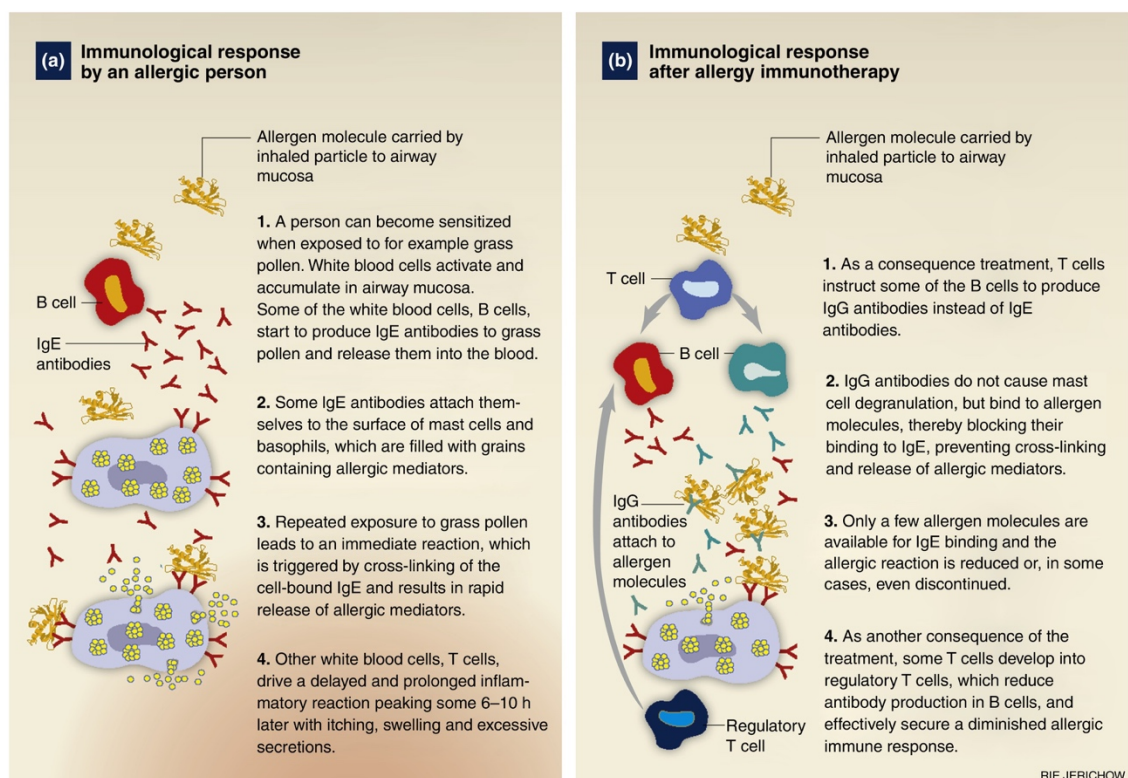
Figure 1-19 **The allergic response is modified by allergy immunotherapy. A)** The standard allergic response. **B)** The immunological response modified by AIT. Immunological tolerance is induced through the synthesis of IgG blocking antibodies and regulatory T cells. Image reproduced from Larsen et al. 2016.

The mechanisms behind epitope spreading and AIT are not yet fully understood, and the roles of the IgE receptors in these processes need to be further elucidated. Better

understanding of these processes can lead to more efficient and safer treatments. The first step in unravelling the roles of the IgE receptors would be to create appropriate tools to probe the possible mechanisms. Tools which target FcεRI or CD23 would be beneficial in determining the individual roles of each IgE receptor. An entire protein tool-box could be created by designing an IgE surface display library, which could be screened and selected for IgE proteins which bind to one receptor but not the other, to explore these pathways.

One instance in which the new tools could be used, could be to trigger protective immunity with AIT. An engineered IgE that cannot bind to FcεRI will not induce mast cell degranulation and thus limit the most dangerous side-effect of anaphylaxis. However, the IgE could still bind to allergen and CD23 and thus the complex could be processed and presented on APCs. This might instruct T cells to activate B cells so that they produce more IgG_4 blocking antibodies, thus enhancing the effect of AIT and regulating the immune system by increasing levels of T regulatory cells.

Harnessing epitope spreading to produce IgG_4 blocking antibodies provides an opportunity for therapeutic intervention in AIT. Controlling epitope spreading is also of interest in IgE-mediated anti-cancer therapies. In this context, it would be desirable for epitope spreading to be limited so that it cannot induce the spreading of tumour antigens. IgE has potent effector functions and is proven to be a safe biologic (Jensen-Jarolim et al., 2010) being developed in the field of Allergo-oncology.

The nascent field of Allergo-oncology attempts to harness the Th2-domainted allergic reaction, described in section 1.2, to eradicate cancerous cells by establishing an 'allergic-response' to the tumour (Singer and Jensen-Jarolim, 2014). The mechanisms of cancer cell death are dependent on IgE-Fc to trigger antibody-dependent cell-mediated cytotoxicity and antibody-dependent cell-mediated phagocytosis (Karagiannis et al., 2007; Jensen-Jarolim et al., 2010; Karagiannis et al., 2012). The absence of an IgE-equivalent of the inhibitory FcγRIIb receptor also provides IgE with an additional advantage, in that it avoids the suppressive effects experienced by IgG in the tumour microenvironment as well as being shown to be safe and efficacious (Karagiannis et al., 2007).

A modified IgE able to bind to FcεRI but not CD23 would be unable to induce epitope spreading, and may provide a safer substitute for IgE-mediated anti-cancer therapeutics (Josephs et al., 2014). It would prevent the immune system reacting to other epitopes similar to the cancer antigen or to unrelated cancer antigens delivered in the same 'allergic' context.

1.5.7 The role of CD23 in receptor-mediated transcytosis

The lung epithelium is the first defensive barrier to the external environment and inhaled allergens. It is also immunologically active, and can secrete a wide range of cytokines and chemokines to recruit immune cells upon sensing allergens, as well as secreting host-defence proteins such as lysozyme and defensins. Dysregulation of the epithelial barrier can increase its permeability to allergens because tight junctions connecting the epithelial cells are disrupted (Mattila et al., 2011). Tight junctions can also be bypassed by a form of receptor-mediated immunoglobulin transport known as transcytosis, illustrated in Figure 1-20.

CD23 is constitutively expressed in human airway epithelial cells and can bind to IgE-allergen complexes in the airways. Furthermore, it is able to bypass the epithelial cell tight junctions to transport the complexes into the mucosa, where the IgE-allergen complexes are released (Palaniyandi et al., 2011). In the mucosa, the IgE-allergen complexes are then able to bind subepithelial immune effector cells such as mast cells, causing degranulation and allergic inflammation. This transport is bidirectional, when CD23 is found on the basal side of the epithelial cells, it binds to IgE and creates a positive feedback loop to support an influx of inflammatory cells (Palaniyandi et al., 2015) . This mechanism of allergen uptake is believed to explain the immediate onset of clinical symptoms after allergen contact.

These insights also have an impact on food allergies, as CD23 is also expressed on intestinal epithelial cells (Tu et al., 2005). IgE is also synthesised locally within the gastrointestinal tract, and the pathogenesis of food allergies is described in Figure 1-20. This finding has been identified in both humans (H. Li et al., 2006) and mice (Yu et al., 2003), highlighting the importance of CD23 and soluble IgE in mucosal immunology. Usually immunoglobulin A and its receptor, the polyimmunoglobulin receptor (pIgR), are associated with mucosal defence, however, CD23's role in transcytosis is similar to

that of pIgR. CD23 can also be compared to the neonatal Fc receptor for immunoglobulin G (FcRn), which transports IgG across epithelia in a bidirectional manner (Baker et al. 2010).
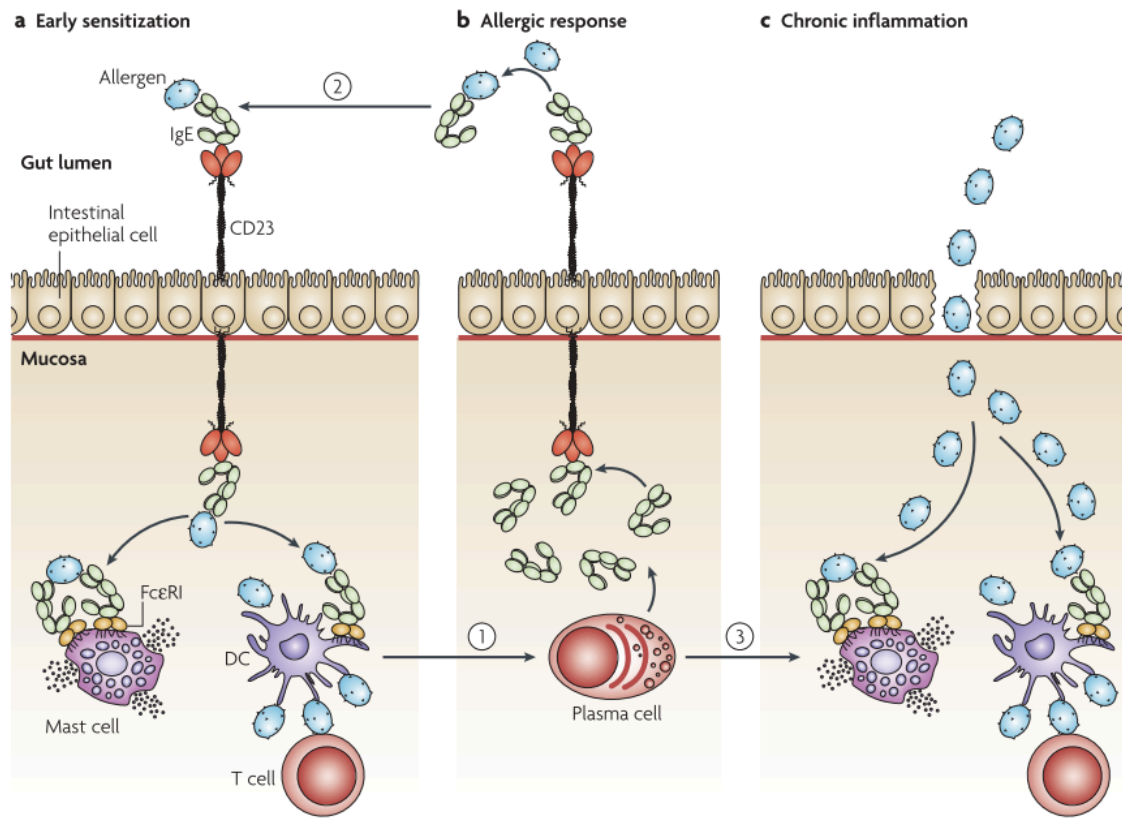
Figure 1-20 **The role of CD23 on intestinal epithelial cells in the pathogenesis of food allergies. A)** IgE-allergen complexes are captured by membrane CD23 expressed on intestinal epithelial cells. The captured complexes are transported across the epithelial membrane into the mucosa, where they are released, binding to FεRI on mast cells and dendritic cells, resulting in activation of the early phase of the allergic response. **B)** IgE antibodies are secreted by plasma cells in the mucosa which are then transported into the gut lumen by CD23. In the gut lumen, the IgE is free to bind to allergens, thus perpetuating the allergic cycle. **C)** The tight-junctions of the epithelial cells are disrupted by the allergic inflammation mediated by IgE. Allergens from the lumen can freely pass into the mucosa and crosslink IgE-bound to FcεRI on mast cells and dendritic cells, exacerbating the food allergy. Image reproduced from Gould and Sutton 2008.

1.5.8 Other biological roles of calcium in human CD23 transport

An alternative pathway of vesicular trafficking involving IgE-allergen complexes and CD23 is the exosome pathway. Exosomes are tiny membrane bound particles ranging from 30 − 150 nm in size (Li et al., 2006; Keller et al., 2006). B cell derived exosomes (bexosomes) contain ADAM10, CD23 and IgE-allergen complexes, and can transfer

these components from B cells to dendritic cells, as illustrated in Figure 1-21 (Martin et al., 2014). This too is a calcium sensitive process, just like receptor-mediated endocytosis, since calcium enhances the release of exosomes (Mathews et al., 2010).
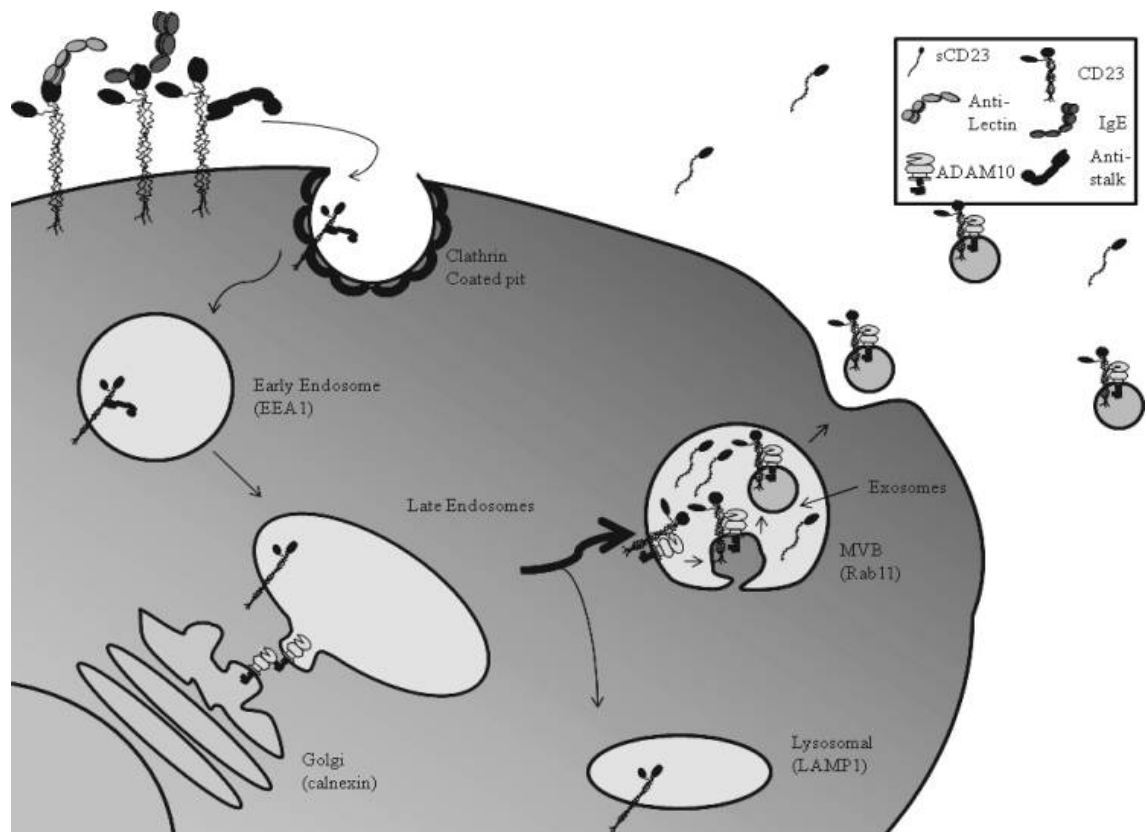
Figure 1-21 **A model of CD23 endosomal trafficking and exosome formation**. IgE-allergen complexes bound to membrane CD23 are internalised in a clathrin-dependent pathway, and trafficked into the endosomal pathway. The increasing acidity grants ADAM10 greater accessibility to CD23, resulting in cleavage and/or sorting into exosomes. The multivesicular bodies (MVB) fuse with the plasma membrane to release the contained exosome into the external environment. Image reproduced from Mathews et al. 2010.

1.5.9 Differences between human and murine CD23

Mice are the experimental tool of choice for most immunologists, and the study of their immune responses has yielded many discoveries, and greater understanding of the human immune system. However, there are several differences between human and murine CD23 that must be considered when using mice as preclinical models of human allergic disease. In this section, a few of these differences are highlighted.

Human and murine derCD23 have a 70% protein sequence identity, and murine CD23 is expected to have a very similar protein fold to human derCD23 based on Swiss-model predictions (since no structure of murine CD23 exists). Murine derCD23 refers to the equivalent range residues in murine CD23 as *der p I* is not able to cleave murine CD23 since the Ser155-156 cleavage site is absent (Schulz et al., 1997; Bettler et al., 1989). However, murine CD23 can still be cleaved to form soluble CD23 (Bartlett et al., 1995). Structurally, one key difference is the tail region of the CTLD. This region in human derCD23 binds to CD21, however, those residues are absent in murine CD23. There is no similar adhesion between CD23 and CD21 in mice. Furthermore, murine derCD23 is also unable to bind to αv integrins due to lacking the Arg172 equivalent residue, which is critical for αv integrin binding of human CD23 (Edkins et al., 2012).

Murine CD23 is expected to bind to two calcium ions based on amino acid sequence analysis and comparisons to related CTLs. This is in contrast to human CD23, which has been demonstrated to bind to only one calcium ion per CTLD (Wurzburg et al., 2006; Yuan et al., 2013). In addition, murine CD23 has been shown to bind to murine IgE, however both calcium ions must be present for binding to occur, otherwise no interaction will be achieved (Richards and Katz, 1990). This contrasts with human derCD23, which can bind to human IgE without calcium, whilst in the presence of calcium, its affinity is further enhanced by 30-fold (Hibbert et al., 2005; Yuan et al., 2013).

Murine CD23 contains the 'EPN' carbohydrate binding motif, and other CTLs with this motif have sugar-binding specificities for mannose, N-acetylglucosamine, glucose and I-fucose (Lee et al., 2011). Indeed, carbohydrate interactions with mannose, N-acetylglucosamine and galactose have been reported (Richards and Katz, 1990). This contrasts with human CD23, which despite earlier studies, was linked to mannose and galactose binding (Kijimoto-Ochiai and Toshimitsu, 1995; Wurzburg et al., 2006). More sensitive assays conducted since have not detected any specific binding for human CD23 to glycans (Hibbert 2005; http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp - Glycan Array: 634).

If the allergic response is considered an abnormal immune response, could these differences between mice and humans, summarised in Table 1-1, make *Homo sapiens* more susceptible to allergies? Could these differences also be targets of therapeutic worth?

Murine CD23	Feature	Human CD23
2	Number of Ca^{2+} ions binding	1
✔	IgE binding Ca^{2+} dependent?	✗ - but Ca^{2+} enhances binding
✗	CD21 binding	✔
✔	Carbohydrate binding	✗
✗	Integrin binding	✔

Table 1-1 **Summary of the differences between murine and human CD23.**

1.5.10 Calcium as a regulator of CD23

As mentioned in section 1.5.4, calcium can act as a switch for endocytic receptors. In the context of CD23, which regulates IgE synthesis, could calcium be 'fine-tuning' CD23's response to IgE? (Yuan et al., 2013). The extracellular concentration of calcium ranges between 1 and 1.5 mM, and with derCD23's affinity for calcium at 1.5 mM, it is most likely that calcium is bound to CD23, thus increasing its affinity for IgE-Fc by 30-fold. This might profoundly influence interactions between soluble trimeric CD23 and membrane IgE, influencing IgE synthesis in B cells (McCloskey et al., 2007; Cooper et al., 2012). Furthermore, it might also enhance binding of allergen-IgE complexes to membrane CD23, thus promoting facilitated antigen presentation, as described in section 1.5.5.

Could this calcium 'switch' sensitivity be unique to humans? The requirement for calcium in murine CD23 differs, as detailed in section 1.5.9, therefore the role of CD23 in allergy may have been a recent development in evolution.

1.6 Targeting IgE therapeutically

Despite the central role of IgE in allergic diseases, it took over two decades to develop a licensed anti-IgE treatment, marketed as Xolair® (Omalizumab) (Incorvaia and Mauro, 2015). The difficulty in producing anti-IgE agents came in the selection of molecules that do not cross-link IgE bound to FcεRI on mast cells and basophils. It is of vital importance that crosslinking of IgE is avoided so that the allergic response is not triggered, since anaphylaxis is a risk that may culminate in death. With IgE-Fc shown to bind to both of it receptors at the Cε3 region, it was known that this domain could be targeted to sterically hinder binding to FcεRI.

Solving the structure of IgE, FcεRI and CD23 gave greater insight into the mechanisms of IgE binding as well as revealing the allosteric control behind these interactions that prevented simultaneous binding of both receptors to IgE (Borthakur et al., 2012; Wright et al., 2015; Davies et al., 2017). The large conformational changes that occur in IgE, as described in section 1.3.2 , alongside the allosteric control of the receptor binding sites (section 1.5.2) present an arduous task when attempting rational engineering of IgE.

1.7 Aims of this thesis

This thesis presents two main projects on the theme of protein engineering of IgE and its receptor, CD23.

The first two results chapters address CD23 and its auxiliary nonfunctional calcium binding site. To examine the mutations that have occurred during evolution and their impact on IgE binding, the "auxiliary" calcium binding site residues were selectively mutated back to those found in murine CD23, in order to restore calcium binding. Probing these mutant proteins may determine whether the loss of the second calcium ion conferred a gain-of-function, such as increased affinity for IgE, in human CD23.

To achieve these aims, the following objectives were set:

- Generate a panel of human derCD23 mutants that may restore calcium binding to the "auxiliary" binding site by site-directed mutagenesis.

- Express the recombinant human derCD23 proteins, refold and purify them.

- Analyse the calcium stoichiometry and calcium-binding thermodynamic parameters of the mutated human derCD23 proteins.

- Analyse the effects of the auxiliary binding site mutations on IgE binding affinity.

- Determine the geometry and structural basis of additional calcium binding by solving the structures of the mutant proteins.

- Build a molecular dynamics system to explore the conformations of the mutant proteins and how calcium affects them.

The third results chapter focuses on the development of an IgE surface display library, similar to phage display but using mammalian cells. This provides a range of mutant IgE proteins to overcome designing and mutating individual or sets of residues. Screening assays were designed to select IgE-Fc mutants that can bind to one receptor and not the other, and vice versa. For this project, the following objectives were set:

- Establish a mammalian transfection protocol for the mammalian cell display system.

- Identify key residues to substitute in IgE-Fc and design a proof-of-principle IgE-Fc surface display library.

- Design and optimise assays to screen the IgE-Fc surface display library.

It is anticipated that these mutant IgE-Fc proteins may be used to probe the role of each IgE receptor in functional activities such as antigen presentation and 'epitope spreading'.

Chapter 2 Methods and Materials

2.1 DNA Manipulation Techniques

2.1.1 PCR materials and software

The low-copy pET5a plasmid contained the gene for derCD23, which coded for the amino acids Ser156 to Glu298 (using the Swiss-Prot accession code: P06734 numbering scheme). In addition the pET151-GA-tri vector coded for a trimeric triCD23 construct (McCloskey et al., 2007), which included a 30-residue synthetic stalk sequence and CD23 residues 136-321 separated by the *Hind*III cleavage site (with the same numbering scheme). Both plasmids contained an ampicillin resistance gene for the selection of positively transformed host cells and were kindly provided by Dr Rebecca Beavil.

2.1.2 DNA primers

All oligonucleotides were synthesised by MWG Eurofins or Integrated DNA Technologies. Vector maps and sequences were created in Snapgene Viewer (Snapgene software v 3.0.3 from GSL Biotech).

2.1.3 Gene Synthesis

The membrane IgE construct as detailed in Appendix A, was manufactured as a double stranded 1923 bp DNA gBlock® gene fragment (Integrated DNA Technologies).

2.1.4 DNA Sequencing

15 μl of purified DNA sample (between 50 ng/μl and 100 ng/μl) had 2 μl of T7 forward primer added (5'-TAATACGACTCACTATAGGG-3' at 10 μM) and sequenced using the Mix2Seq (Eurofins) service. Sequencing results were aligned by ClustalW or Clustal Omega software (EMBL-EBI).

2.1.5 DNA purification and quantification

Plasmid DNA was purified using the Monarch® Nucleic Acid Purification Kit (NEB) as per manufacturer's instructions. The concentration of eluted DNA was quantified by a NanoDrop ND-1000 spectrophotometer at 260 nm. Plasmid DNA was stored at -20°C.

2.1.6 DNA purification of the membrane IgE library

The membrane IgE library plasmids were purified using the Plasmid Maxi Kit (Qiagen) as per manufacturer's instructions. The concentration of the extracted DNA was quantified by the NanoDrop ND-1000 spectrophotometer at 260 nm. The pooled plasmid DNA was stored at -20°C.

2.1.7 Site-Directed Mutagenesis of derCD23

Initially the Quickchange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) was used to perform site-directed mutagenesis of derCD23 to produce the first mutation (N225D) as per manufacturer's instructions. However, optimisation with the polymerase Phusion™ Flash High Fidelity PCR MM (Thermo) replaced the Quikchange lightning kit for the subsequent steps, whilst using the same cycling conditions as described in Table 2-1. The primers used are listed in Table 2-2 while the recipe for both set mutagenesis reactions is listed in Table 2-3 respectively.

Cycles	Temperatures (°C)	Time
1	95	2 mins
	95	20 seconds
18	60	10 seconds
	68	2.5 mins
1	68	5 mins

Table 2-1 **Cycling conditions for site-directed mutagenesis.**

N225D F	CCTGGATTGGCCTTCGGgAtTTGGACCTGAAGGGAGAG
N225D R	CTCTCCCTTCAGGTCCAAcTaCCGAAGGCCAATCCAGGAGCC
K229E F	GGCCTTCGGgatTTGGACCTGgAaGGAGAGTTTATCTGGGTGG
K229E R	CCACCCAGATAAACTCTCCtTcCAGGTCCAAatcCCGAAGGC
S252N F	CTGGGCTCCAGGGGAGCCCACCAaCCGGAGCCAGGGCGAGGAC
S252N R	CTGGGCTCCAGGGGAGCCCACCAaCCGGAGCCAGGGCGAGGAC
T251N & S252N F	CTGGGCTCCAGGGGAGCCCAaCAaCCGGAGCCAGGGCGAGGACTG
T251N & S252N R	CAGTCCTCGCCCTGGCTCCGGtTGtTGGGCTCCCCTGGAGCCCAG
R253G & S252G F	GGGCTCCAGGGGAGCCCAaCAaCgGcgGCCAGGGCGAGGACTGC
R252G & S252G R	GCAGTCCTCGCCCTGGCcgccGtTGtTGGGCTCCCCTGGAGCCC

Table 2-2 **A list of primers used for site-directed mutagenesis.** Mutations are shown in lower case. F = forward primer. R = reverse primer. Primer sequences listed in the 5' to 3' format.

A)		B)	
Quantity	**Reagent**	**Quantity**	**Reagent**
5 μl	10x reaction buffer	5 μl	Phusion Master Mix
10 – 100 ng	dsDNA template	0.5 μM	Primer 1
125 ng	Primer 1	0.5 μM	Primer 2
125 ng	Primer 2	20 ng	dsDNA template
1.5 μl	dNTP mix	Up to 10 μl	ddH$_2$0
1 μl	QuikSolution reagent		
Up to 50 μl	ddH$_2$0		
Followed by 1	QuikChange		
μl	Lightning Enzyme		

Table 2-3 **Reagents required for the derCD23 mutagenesis PCR. A**) QuikChange Lightening Site-Directed Mutagenesis kit. **B)** The optimised PCR.

2.1.8 Dpn I digestion of the amplification products

Template DNA was removed by *Dpn* I treatment; briefly: 1 μl of Fast digest *Dpn* I (Thermo), was added to 2 μl of the PCR product, 2 μl 10x Fast Digest buffer (Thermo) and 15 μl dH$_2$O. The reaction was mixed well and incubated at 37°C for 10 min.

2.1.9 Agarose Gel Electrophoresis

DNA products were analysed by agarose gel electrophoresis. 0.8% agarose gels were made by dissolving 0.4 g agarose (Bioline) in 50 ml of 1 x TBE running buffer (90 mM Tris, 90 mM Boric acid, 2 mM EDTA pH8). The gel was cooled until hand hot before adding 5 μl of Gel red nucleic acid stain (Biotum). The solution was poured into a casting tray with a well comb added and allowed to set. The set gel was transferred to an electrophoresis tank and covered in x1 TBE running buffer. The gel comb was

removed and 10 µl GeneRuler 1kb DNA ladder (Fermentas) molecular weight marker was loaded. DNA samples were prepared by adding 5x loading dye solution (Fermentas) and mixing well before loading onto the gel. The gel was run at 100 V until the desired separation and visualized with UV light in the Biorad Gel Dock™ XR+ Gel System (Biorad) and a digital photograph was taken of the image.

2.1.10 Competent Cell Transformation

10 µl NEB 10-beta competent *E. coli* cells (NEB) or One Shot® Top10 chemically competent *E.coli* cells (Thermo) had 0.5 µl of DNA added, with the tube flicked to gently mix the solution. The solution was incubated on ice for 30 mins before heat shocking the cells at 42°C for 30s. The cells were then returned to ice for 2 mins before adding 200 µl of SOC media (Invitrogen). After 1 hr of incubating the cells at 37°C with shaking at 200 rpm, the cells were plated onto LB agar petri dishes containing (100 µg/ml) ampicillin and left to propagate at 37°C overnight.

2.1.11 Zero blunt® TOPO® PCR cloning

Due to limited stock of the designed membrane IgE construct (section 2.1.3, Figure 5-5), the gBlock® was first cloned into a pCR-blunt-II-TOPO vector (Invitrogen) so that it could be replicated and sub-cloned easily. The reagents for the reaction were provided in the Zero blunt® TOPO® PCR cloning kit (Invitrogen). Briefly, 0.5 µl of gBlock® DNA, 1 µl salt solution, 1 µl pCR-blunt-II-TOPO vector with 3.5 µl dH$_2$O was mixed gently. The reagents were incubated for 5 mins at room temperature and 1 µl was used to transform One Shot® Top10 chemically competent *E.coli* cells (Invitrogen). The *E.coli* cells were plated onto a 50 µg/ml kanamycin agar plate.

2.1.12 GeneArt® Seamless Cloning and Assembly Kit

The GeneArt® Seamless Cloning and Assembly Kit (Thermo) was used for the insertion of the designed membrane IgE construct (Plasmid map shown in Figure 5.5) into the pcDNA™5/FRT (Thermofisher) vector. The pcDNA™5/FRT vector for mammalian cell transfection contained an ampicillin resistance gene for vector propagation and a hygromycin resistance gene for cell transfection selection. The membrane IgE construct was manufactured as a gBlock® gene fragment (Integrated DNA Technologies, sequence provided in Appendix A). The primers were designed based on manufacturer's recommendations and are listed in Table 2-4. The reagents for both the insert and vector PCR included 0.5 μl of vector (~20 ng) or insert DNA, 10 μl of Phusion Master Mix, 0.2 ng of forward primer, 0.2 ng of reverse primer and 7.5 μl of dH$_2$O. The cycling conditions differed for the two PCRs with the vector PCR having the following cycling conditions: (98°C initial denaturation for 10 s, followed by 30 cycles of denaturation at 98°C for 1 s, annealing at 72°C for 5 s and extension at 72°C for 75 s, followed by 1 cycle of final extension at 72°C for 1 second), whilst the insert PCR consisted of the following conditions: 98°C initial denaturation for 10 s, followed by denaturation at 98°C for 1 s, annealing at 58°C for 5 s and extension at 72°C for 30 s followed by 1 cycle of final extension at 72°C for 1 second.

4 μl of the PCR was mixed with loading gel to be checked on a 0.8% agarose gel and the remaining PCR was cleaned-up with the Monarch PCR-clean-up kit (NEB). For the GeneArt reaction, 4 μl of the purified vector reaction and 8 μl of the purified insert reaction were combined with 2 μl of dH$_2$O, 4 μl 5x reaction buffer and 2 μl of 10x enzyme mix. The reaction was incubated at room temperature for 30 mins, and 1 μl of the reaction was used to transform 10 μl of One Shot® Top10 chemically competent

E.coli (Thermo) as described in section 2.1.10. Colonies were checked by colony PCR prior to sequencing as described in the next section, 2.1.13.

Vector F	GCAGATATTTAAGGTTGGTACCGAGCTCGGAT
Vector R	AATTCCAGGTTTAAACGCTAGCCAGCTTGGGT
Insert F	GTTTAAACCTGGAATTCGCCCTTATGGACTGG
Insert R	AAGCTTAAATATCTGCAGAATTCGCCCTTCTA

Table 2-4 **A list of primers used for the GeneArt Seamless Cloning and Assembly Kit.** F = forward primer. R = reverse primer. Primer sequences listed in the 5' to 3' format.

2.1.13 Colony PCR

Transformed *E.coli* colonies on agar plates were tested prior to sequencing to see if they contained the desired plasmid DNA sequence. For the membrane-IgE constructs, colony PCR primers specific for the insert are listed in Table 2-5 and PCR conditions are detailed in Table 2-6. When all the solutions of Table 2-7 had been added to the PCR tube, the chosen *E.coli* colony on the agar plate was lightly brushed with a pipette tip and then mixed well into the PCR reagents. This method left most of the *E.coli* colony intact and available for future experiments. The resulting PCR product was run on a 0.8% agarose gel to visualise results.

| TM Colony | ACTCTTCCTGCTCAG |
| M13 Reverse | CAGGAAACAGCTATGAC |

Table 2-5 **A list of primers used for colony PCR**. Primer sequences listed in the 5' to 3' format.

Cycles	Temperature (°C)	Time (mins)
1	94	5
	94	1
30	52	2
	72	5
1	72	10

Table 2-6 **Cycling conditions for colony PCR.**

Quantity	Reagent
5 μl	2x ReadyMix Taq PCR mix (Sigma)
4 μl	ddH$_2$0
10 μM	Primer 1
10 μM	Primer 2
+ Brush of a colony	

Table 2-7 **Reagents required for colony PCR.**

2.1.14 Mutant IgE-Fc Library Generation

The membrane IgE-Fc library was made using the construct made in section 2.1.12. The plasmid had mutations introduced at two sites, R334 and P426. To make sure that all the amino acids were evenly coded for at those sites, and to avoid stop codons, the '22-c trick' was used. The '22-c trick' uses degenerate primers that represent the 22 unique codons for all the amino acids and no stop codons. The degenerate primers, listed in Table 2-8, were mixed in a 12 NDT : 9 VHG : 1 TGG primer ratio (Kille et al., 2013). For the PCR, the (10 pmoles) of each forward and reverse primer mixture were mixed with 20 ng of plasmid pcDNA™5/FRT/memIgE template and 10 μl of Phusion™ Flash High Fidelity PCR Master Mix (Thermo) in a total of 20 μl of dH$_2$O. The PCR

conditions used are listed in Table 2-9. The PCR product was treated with *Dpn I* as described in section 2.1.8 before transforming *E.coli* cells as described in section 2.1.10.

Pro426_1_F   CACCCCCACCTGNDTAGGGCCCTCATG

Pro426_1_R   CATGAGGGCCCTAHNCAGGTGGGGGTG

Pro426_2_F   CCCACCTGVHGAGGGCCCTCATG

Pro426_2_R   CATGAGGGCCCTCDBCAGGTGGGG

Pro426_3_F   CACCCCCACCTGTGGAGGGCCCTCATGCG

Pro426_3_R   GCATGAGGGCCCTCCACAGGTGGGGGTGG

Arg334_1_F   ATTCCAACCCGNDTGGGGTGAGCGC

Arg334_1_R   GCGCTCACCCCAHNCGGGTTGGAATC

Arg334_2_F   AGATTCCAACCCGVHGGGGGTGAGCGC

Arg334_2_R   GCGCTCACCCCCDBCGGGTTGGAATC

Arg334_3_F   ATTCCAACCCGTGGGGGGTGAGCGC

Arg334_3-R   AGGCGCTCACCCCCCACGGGTTGGAATC

Table 2-8 **A list of degenerate primers used for the generation of the mutant IgE-Fc library.** F = forward primer. R = reverse primer. Primer sequences listed in the 5' to 3' format. Degenerate bases highlighted in blue, follow the International Union of Pure and Applied Chemistry (IUPAC) designation. H = A or C or T. B = C or G or T. V = A or C or G. D = A or G or T. N = A or C or G or T.

Cycles	Temperatures (°C)	Time
1	95	2 mins
	95	20 seconds
18	60	10 seconds
	68	2.5 mins
1	68	5 mins

Table 2-9 **Cycling conditions for the memIgE library plasmid generation.**

2.1.15 Quick Quality Control (QCC)

10 colonies were randomly picked from each agar plate and their plasmid DNA was purified as described in section 2.1.5 before being sent for sequencing as detailed in section 2.1.4. The nucleotide base distribution of each mutated base was estimated by comparing the sequenced results to the original WT bases. The values were converted to pie charts using Microsoft Excel (v 14.7.1, Microsoft Corporation) and compared to the expected values. The 22-c trick method expects 27%(A) + 27%(C) + 27%(G) + 19%(T) at the first nucleotide position, 32%(A) + 18%(C) + 18%(G) + 32%(T) at the second, and 45%(G) + 55%(T) at the third (Kille et al., 2013).

2.2 Protein Methods

2.2.1 Protein expression of derCD23 proteins in *E. coli*

BL21 DE3 competent *E. coli* cells (Invitrogen) were thawed slowly on ice, prior to the addition of 0.5 µl of plasmid DNA and transformed as described earlier in section 2.1.10. One pipette tip's streak of cells from the plate was used to inoculate a 20 ml starter colony of LB and ampicillin 100 µg/ml; ~5 hours later this colony was poured into 0.5 L of ZYP-5052 rich autoinduction medium (928 ml of ZY, 1 x 5052, 1 x NPS and 50 µg/ml of ampicillin as described in Table 2-10, Studier 2005) in 2L Erlenmeyer baffled flasks. The cultures were grown shaking at 37°C at 200 rpm for ~ 48 hours. To confirm expression of the desired protein, 250 µl samples were extracted for SDS-PAGE analysis. The cells were harvested by centrifugation at 4000 g for 20 mins, and the supernatant was discarded. The pellet was left in the -80°C freezer over night to for ice crystals to form to encourage cell lysis.

Name	Reagent	Amount
ZY	N-Z-amine AS	10g
	Yeast Extract	5g
	dH$_2$O	925mL
50 x 5052	Glycerol	250 g
	Glucose	25 g
	α-Lactose	100 g
	dH$_2$O	730 ml
20 x NPS	(NH$_4$)$_2$SO$_4$	0.5 M
	KH$_2$PO$_4$	1 M
	Na$_2$HPO$_4$	1 M
	dH$_2$O	Up to 1 L

Table 2-10 **ZYP-5052 autoinduction medium − 1L stock solutions.** Reagents for the 50x 5052 and the 20x NPS were added to the dH$_2$O in the order listed (Studier, 2005).

2.2.2 ^{15}N labelled protein expression of derCD23 in *E.coli*

Performed the same way as in section 2.2.1 but with ^{15}N ammonium sulphate for the P-5052 minimal autoinduction medium (Studier, 2005).

2.2.3 **Protein Expression of triCD23 in *E. coli***

BL21 DE3 competent *E. coli* cells (Invitrogen) were thawed slowly on ice, prior to the addition of 0.5 µl of plasmid pET151-GA-tri DNA and transformed as described earlier in section 2.1.10. Cells were grown and harvested as explained in section 2.2.1.

2.2.4 SDS-PAGE gel electrophoresis analysis

Protein purity was assessed by a 16% Tris-SDS mini-gel, prepared as stipulated in Table 2-11 (Laemmli, 1970). Typically, $2 - 20$ µg of pure protein or 20 µl of sample was mixed with 5x non-reducing sample buffer (5x non-reducing SDS sample buffer: 0.313 M Tris-HCl pH 6.8, 10% SDS, 50% glycerol and 0.05% bromophenol blue), with 1 µl of β-mercaptoethanol added for reducing gels. This sample was then boiled for 3 mins. 5 µl SeeBlue Plus2 Pre-stained Protein Standard (Invitrogen) was loaded into the first well followed by the samples. The mini-gel was run at 120 V in SDS running buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS, pH 8.3) until the blue dye had reached the end of the gel. Gels were extracted and the stacking gel removed to allow for staining with Instant Blue Comassie Stain (Expedeon) for 30 mins until the protein bands were clearly visible. Stained gels were washed with dH$_2$O before capturing the image on the Bio-Rad gel dock and the Bio-Rad ImageLab software (Bio-Rad v5.2.1).

'Separating Gel'	Volume (µl)
Lower Tris	2000
Acrylamide	2500
dH$_2$O	5300
10 % ammonium persulphate	50
TEMED	10
'Stacking Gel'	
Upper Tris	1250
Acrylamide	1000
dH$_2$O	3180
10 % ammonium persulphate	50
TEMED	10

Table 2-11 **SDS-PAGE gel recipe.** TEMED = Tetramethylethylenediamine.

2.2.5 Extraction of inclusion bodies

derCD23 is expressed in inclusion bodies, not as a soluble protein, to extract the inclusion bodies from the *E.coli*, the cell pellet was first removed from the -80 °C freezer and thawed prior to resuspension in 20 ml Phosphate buffered saline (PBS, pH 7.4: 0.14 M NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4, pH 7.4) with added protease inhibitor cocktail tablet (Roche). The solution was passed through a cell disruptor twice at a pressure of 20 KPSI (CF Range, Constant Systems Ltd.) with the product collected on ice. The cooled product was centrifuged at 20450 rcf for 20 min to separate the inclusion bodies, with the supernatant discarded. The inclusion bodies were resuspended in inclusion body wash buffer (50 mM Tris pH 7.5, 100 mM NaCl, 0.5% octylphenoxypolyethoxyethanol (Igepal), 0.1% Sodium-azide) and centrifuged at 20450 rcf for 20 mins with the supernatant discarded. The washing with the inclusion body wash buffer was repeated for a total of 5 times.

2.2.6 Solubilisation of the inclusion bodies

After the final centrifugation step at 20450 rcf for the inclusion body wash process, the supernatant was discarded and the inclusion body pellet was resuspended in 5 ml of 20 mM Tris, 6M Guanidinium hydrochloride, pH 7.5 and left rotating at room temperature overnight at 200 rpm. Insoluble contaminants were removed by centrifuging the solution at 20450 rcf. The concentration of the solubilised protein was quantified by a NanoDrop ND-1000 spectrophotometer at 280 nm and the absorbance was divided by 2.786 (the absorbance per cm for 1 mg/ml at 280nm) to gain the actual mg/ml value, before storing the sample in the -80°C freezer.

2.2.7 Extraction and solubilisation of triCD23 from inclusion bodies

TriCD23 inclusion bodies were washed as described in section 2.2.5, however, after the final wash, the inclusion bodies were solubilised in 100 ml of 6 M Guanidine

hydrochloride, 20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole and 10 mM DTT, pH 7.4. The solubilised protein was quantified by a NanoDrop ND-1000 spectrophotometer at 280 nm and the absorbance was divided by 1.65 (the absorbance per cm for 1 mg/ml at 280nm) to gain the actual mg/ml value. The protein was stored in a -80°C freezer as 90 mg aliquots.

2.2.8 DerCD23 protein refolding

To recover biologically active recombinant protein a protocol based on Taylor et al., 1992 was followed. 20 mg of the solubilised derCD23 protein was reduced with Dithiothreitol (DTT), added to a final concentration of 1 mM and incubated at room temperature for 1 hour. The sample was then diluted ten-fold into 6 M Guanidine, 0.5 M Tris-Acetate pH 8.6 and 100 mM oxidised glutathione and incubated at 4°C, stirring for 24 hours. This sample was then slowly diluted into 2 L of ice-cold refolding buffer 100 mM Tris-Acetate pH 8.6, 3 mM cysteine and 2 mM $CaCl_2$ using a peristaltic pump at a setting of 0.5 (~ 100 μl/min) to avoid precipitation. The protein continued to refold at 4°C with constant slow stirring, in the dark for 48 hours.

2.2.9 TriCD23 protein refolding

One aliquot of frozen solubilised triCD23 from section 2.2.7, was thawed and spun briefly to remove debris. To refold the protein, a protocol modified from section 2.2.8 was used. The solubilised protein was slowly diluted into 1L of refold buffer (1 M Guanidine hydrochloride, 0.4 M arginine, 50 mM CHES, 1:2 ratio of reduced to oxidised glutathione, pH 9) using a peristaltic pump at a setting of 4.0 (~800 μl/min). The protein continued to refold at 4°C with constant slow stirring, in the dark overnight.

2.2.10 Protein purification by hydrophobic interaction chromatography (HIC)

The refolded derCD23 proteins were purified by HIC. First, ammonium sulphate precipitation was performed to remove unwanted protein. 400 g (approx. 1.5 M) of ammonium sulphate was slowly added to 2 L of the refolded protein with constant stirring and allowed to dissolve fully. Then the solution was filtered through a 0.45 micron filter fitted with a pre-filter disk by vacuum filtration. The filtered solution was connected to an ÄKTA Prime System (Amersham) fitted with a 10 ml Phenyl Sepharose 6 Fast Flow (high sub) column equilibrated in buffer A (1.5 M $(NH_4)_2SO_4$, 25 mM Tris-Cl pH 7.5, 2 mM $CaCl_2$). The experiment ran at a flow rate of 10 ml/min. The derCD23 proteins were eluted into 10 ml fractions into buffer B (25 mM Tris- HCl pH7.5, 2 mM $CaCl_2$), with a gradient elute program The column flow-through was kept with more ammonium sulphate added to it (to a total of approx. 2 M), before filtering and passing the solution over the HIC column again. A maximum of three ammonium sulphate cuts were performed (max. of 2.5 M ammonium sulphate added).

2.2.11 Protein Purification by Immobilised Metal Ion Affinity Chromatography (IMAC)

The refolded his-tagged triCD23 protein were purified by IMAC. First, 200 ml of the refold solution was diluted to 1 L with dH_2O to reduce the pH of the refold solution to pH 7.4. A 1 ml HisTrap FF crude column (GE Healthcare Life Sciences) was connected to an ÄKTA Prime System (Amersham) and equilibriated with binding buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole pH 7.4). The experiment ran at a flow rate of 1 ml/min using the ion exchange gradient elute method program. The protein was eluted into twenty 1 ml fractions in a 0-500 mM imidazole gradient using the elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole pH 7.4).

2.2.12 Protein concentrating

DerCD23 proteins were concentrated using a 5000 MWCO centrifugal concentrator (Vivaspin) spun at 3901 rcf at 4˚C for 20 min intervals. TriCD23 was concentrated using an Amicon 10 000 MWCO centrifugal concentrator (Millipore) spun at 3901 rcf at 4˚C for 5 min intervals to about 4 mg/ml. The concentrator tube was inverted three times between spins with constant monitoring of the protein concentration throughout. Flow-through was checked for protein content to check the membrane was not leaking and discarded if fine. Concentrated protein was decanted and stored in 1.5 ml microcentrifuge tubes at -20 ˚C.

2.2.13 Analytical Gel Filtration

To assess the integrity of the protein, a small sample was passed through an analytical gel filtration Superdex 200 10/300 GL column (GE Healthcare) to measure impurities or/and aggregation. 100 μg of protein was diluted into 100 μl of running buffer (Tris Buffered Saline TBS, pH 7.5: 25 mM Tris-HCl pH 7.5, 125 mM NaCl, 0.05% azide w/v for derCD23 proteins, and PBS, pH 7.4: 0.14 M NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4, pH 7.4, 0.05 % azide w/v for all other proteins). The Gilson HPLC system was equilibrated in the same buffer the protein was diluted in. A flow rate of 0.75 ml/min was used and protein was detected by absorbance at 280 nm. Chromatograms were analysed with Trilution®LC software (Gilson).

2.2.14 Size Exclusion Chromatography (SEC)

The concentrated triCD23 protein was further purified by SEC on a Superdex 200 column 10/300 GL column (GE Healthcare). The column was equilibrated in 0.5 M Tris, 100 mM NaCl pH 7.5 and the protein was loaded onto the system in 500 μl

aliquots and run at 0.75 ml/min. The protein was eluted into 200 μl fractions collected

between 10 and 30 mins in Tris buffer (0.5 M Tris, 100 mM NaCl pH 7.5).

2.2.15 Buffer exchange

Samples were dialysed by placing the concentrated protein sample into 3.5 kDa MWCO

dialysis tubing (Fisher Scientific Ltd) pre-rinsed in dH_2O. The tubing was transferred

into a large beaker containing the end buffer of choice. Samples were dialysed

overnight at 4°C with stirring, and carefully removed from the dialysis tubing with

pipettes. Protein samples had their concentration determined before storage at 4°C for

further use.

2.2.16 Co-transfection of pcDNA5/FRT/Fcε-GFP-V5-His vectors and pOG44 Flp-recombinase vector

As recommended for the transfection reagent FuGene (Promega), briefly: the day before

the transfection, half a 24-well plate (Flat round well, tissue culture 24-well Nunc™

plate, Thermo) was plated at a density of 8 x 10^4 cells/well with FlpIn HEK293 cells,

and the second half at a density of 4 x 10^4 cells/well in 500μl of complete growth

medium (DMEM + 10% Fetal Bovine Serum). For a single protein transfection in

duplicate, 0.11 ug of pcDNA5-FRT vector and 0.99 ug of pOG44 DNA (Thermofisher

lot no. 1542822 and 417309) were added in a combined total volume of 52 μl of sterile

deionized water. pcDNA™5/FRT/CAT is the positive control (Thermofisher lot no.

468373) and is also co-transfected with pOG44. Using a 3:1 Fugene to DNA ratio, 3.3

μl of FuGene was carefully added. This was achieved by avoiding touching the sides of

the 1.5ml microcentrifuge tube with the tip of the pipette. The solution was vortexed for

20 seconds, and spun-down to recover all the solution in the base of the 1.5 ml

microcentrifuge tube. After 10 minutes incubation at room temperature, 25 μl of

complex was added per well (one at the higher density and one at the lower cell density)

and mixed thoroughly. Cells were returned to the incubator and after 48 hours the cells were split into 6-well plates containing complete media with 50 μg/ml hygromycin (Invitrogen hygromycin B in PBS, lot no. HY068-L11) for selection. A month later successfully transfected cells form loci in the wells which can be expanded.

2.2.17 Large-scale transfection of FlpIn cells for the creation of the memIgE-Fc library

As recommended for FuGene (Promega), briefly: the day before the transfection, a T75 flask was seeded at a density of 8 x 10^5 cells/well with FlpIn HEK293 cells in 22 ml of complete growth medium (DMEM + 10% Fetal Bovine Serum). For a single transfection of the whole library, 2.5 ug of the pooled pcDNA5-FRT-IgE-Fc-GFP mutant vectors and 22.5 ug of pOG44 DNA (Thermofisher lot no. 1542822 and 417309) were added in a combined total volume of 1163 μl of sterile deionized water. Using a 3:1 Fugene to DNA ratio, 74 μl of FuGene was carefully added. This was achieved by avoiding touching the sides of the 1.5ml microcentrifuge tube with the tip of the pipette. The solution was vortexed for 20 seconds, and spun-down to recover all the solution in the base of the 1.5 ml microcentrifuge tube. After 10 minutes incubation at room temperature, 1125 μl of complex was added to the flask and swirled to mix thoroughly. Cells were returned to the incubator and after 48 hours the cells were split into a T175 flask containing complete media with 50 μg/ml hygromycin (Invitrogen hygromycin B in PBS, lot no. HY068-L11) for selection. A month later successfully transfected cells form loci in the wells which were pooled to create the library.

2.2.18 Cell Lines

Cell Line Name	Species	Cell Type	Growth Medium
FlpIn™ HEK293 Lot no: 1735580 & 534478	Human	Embryonic Kidney	DMEM + 10% FBS + Zeocin Post-transfection: DMEM + 10% FBS + 1% PSG + 50 μg/ml hygromycin (no Zeocin)
WEHI cells	Mouse	B-cells	DMEM + 10% FBS + Geneticin + β-mercaptoethanol

Table 2-12 **Cell lines used in the experiments.** DMEM = Dulbecco's Modified Eagle's Medium. FBS = fetal bovine serum, PSG = penicillin-streptomycin-glutamine.

2.2.19 Thawing Cell lines

Cells were extracted from the liquid nitrogen store and thawed quickly in a water bath at 37°C and were immediately transferred to a T25 flask containing 5 ml of pre-warmed medium (DMEM and 20% FBS).

2.2.20 Passaging Cell lines

Medium from adherent cells was decanted and the cells were subsequently washed twice with Gibco™ PBS (Thermo) before 5 ml of trypsin was added to the T175 flask. The trypsin-covered flask was left to incubate at 37°C for 2 – 5 mins until cells started to detach. At this point, the trypsin was neutralised with media and split accordingly into new flasks. New T175 flasks, which contained fresh media and correct antibiotics were mixed well by pipetting the cells up and down before leaving the cells to grow in the incubator.

Suspension cells (For WEHI cells): A few mls of cells was removed directly from the flask and transferred into a new flask with fresh media at a 1:4 ratio.

2.2.21 Freezing Cells

Medium from cells was removed and washed twice with Gibco™ PBS (Thermo) and trypsinsed. Cells were counted and resuspended at a density of at least 3×10^6 cells/ml in freezing media (FBS + 10% DMSO). 1 ml aliquots were divided into freezing nunc vials and were stored in the Mr Frosty™ Freezing Container (Fisher Scientific) at -80˚C for 24 hours to allow slow freezing before being moved into liquid nitrogen storage.

2.2.22 Determination of Cell Numbers

10 μl of cells were extracted and mixed with 10 μl of trypan blue before adding 10 μl of the mix to a haemocytometer. An average was taken of the cells in the quarters of the haemacytometer and multiplied by the dilution factor to define the cells/ml density.

2.2.23 Antibiotic Cell Culture Kill Curve

The optimum concentration of hygromycin antibiotic used for selection of a stably transfected FlpIn HEK293 cell line expressing the protein of interest was determined by a kill curve. FlpIn HEK293 cells were plated at a density of 8×10^4 cells/well for half a plate, and 4×10^4 cells/well for the second half of the plate in 500 μl of complete growth medium (DMEM + 10% Fetal Bovine Serum). The hygromycin antibiotic (Invitrogen hygromycin B in PBS, lot no. HY068-L11) was added at a range of concentrations: 0, 25, 50, 100, 150 and 200 μg/ml and monitored for 2 weeks. 50 μg/ml was found to be the optimum concentration of hygroymcin used for selection.

2.2.24 Epifluorescence Microscopy

Transfected cells grown in plates were checked for GFP fluorescence by an Axiovert 200M inverted microscope (Zeiss). Tissue culture plates were placed on the microscope stage platform and cells were examined at 10x optical zoom. Cells were located in the brightfield channel before switching to the GFP channel (LEJ EBQ 100 isolated mercury discharge lamp power supply, GFP filter – EX BP 469/38) with images captured with the AxioVision software v4.4. Images were processed with ImageJ v1.49 (Wayne Rasband, NIH, USA).

2.2.25 Flow Cytometry

Cells were harvested, washed twice with Gibco™ PBS (Thermo), and re-suspended at 1 x 10^6 cells/ml in FACS buffer (Gibco™ PBS pH7.4 + 10% FBS) and 100 μl was aliquoted into FACS tubes. The cells were stained with either antibody, or labelled protein reagent (triCD23-Alexa488, triCD23-Alexa647 or αγ-fusion protein-Alexa647) at 1 ug/ml and incubated on ice for 30 mins. The cells were then washed twice and re-suspended in 1 ml of FACS buffer for analysis. Cells were analysed either on the Attune NxT Acoustic Focusing Cytometer (Lasers: BRVX) or Guava Easy Cyte Plus™ CytoSoft 5.2.7. For the Attune, the voltage parameters were kept constant as follows: FSC 170 V, SSC 330 V, BL1 250 V, BL3 400 V, RL1 350 V. Flow cytometry data was analysed in FlowJo version 10.2.

Antibodies: For the positive control: anti-human IgE-PerCP Cy5.5, clone MHE-18, (BioLegend, lot no: B218089 and B222669).

2.2.26 Protein Fluorophore labelling

Proteins were labelled with either Alexa647 (Monolith NT™ Protein Labelling Kit) or Alexa488 (AlexaFluor™ 488, Thermo) according to manufacturer's instructions. Briefly, proteins were first dialysed into a primary amine-free buffer e.g. PBS (PBS pH 7.4: 0.14 M NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4, pH 7.4). The fluorophore was reconstituted, and added to the protein and mixed slowly. The protein and fluorophore were incubated rotating at room temperature and covered in tin foil to block out any light for 30 mins. To remove free dye, the mix was passed over an equilibrated PD-10 desalting column and eluted into fresh dialysis buffer into 0.5ml fractions. The degree of labelling (D.O.L) was determined using Equation 1.

$$D.O.L = \frac{A_{max} \times DF}{[protein]\varepsilon_{dye}}$$

Equation 1 **Degree of labelling equation.** D.O.L = degree of labelling, A_{max} = the absorbance of dye solution measured at the wavelength maximum, DF = dilution factor, ε_{dye} = molar extinction coefficient of the fluorescent dye.

As fluorescent dyes also absorb at 280 nm, to adjust for the contribution of the dye to the A_{280}, a correction factor was used for each dye as described in Equation 2. The correction factors used for alexa647 were 0.03 and 0.11 for Alexa488.

$$Protein\ concentration\ (M) = \frac{A_{280} - (A_{max} \times CF)}{\varepsilon} \times DF$$

Equation 2 **Correction factors used to calculate protein molarity.** A_{280} = the absorbance of the protein-dye at 280 nm. A_{max} = the absorbance of the dye solution measured at the wavelength maximum, CF = correction factor, DF = dilution factor, ε = molar extinction coefficient of the protein.

2.2.27 Protein Biotinylation

IgE-Fc was biotinylated with the EZ-link® Sulfo-NHS-LC-Biotin, No-weigh™ (Thermo) kit as per manufacturer's recommendation. Briefly, biotin was resuspended in dH$_2$O to 10 mM and added to IgE-Fc in at a 1:1 molar ratio. The mixture was incubated on ice for 2 hours and purified by SEC on the Superdex 200 10/300 GL column (GE Healthcare) at a flow rate of 0.75 ml/min. The protein was eluted into 10 mM HEPES, 150 mM NaCl, pH 7.4, 0.05% Na-Az. The 0.5 ml fractions containing the biotinylated protein were pooled and stored at 4°C.

2.2.28 Other protein materials/Gifted proteins

αγ-fusion protein; the soluble fragment of the high-affinity IgE receptor α-chain fused to the Fc region of IgG$_4$ (Shi et al., 1997), and FcεRI were kindly provided by Dr Rebecca Beavil.

IgE-Fc$_{3-4}$ and IgE-Fc$_{2-4}$ was kindly provided by UCB, Celltech.

2.3 Biophysical Methods

2.3.1 Isothermal Titration Calorimetry (ITC)

ITC is a technique that provides experimental details about the energetic components of protein interactions. ITC complements structural studies to yield parameters such as binding constants (K_B), stoichiometry (n), enthalpy (ΔH) and entropy (ΔS) of binding. It measures the change in heat associated with a chemical reaction in relation to the reference cell, triggered by the mixing of 2 reagents, such as a titration of a ligand into a protein solution (Pierce et al., 1999). The technique and data output is illustrated in Figure 2-1.
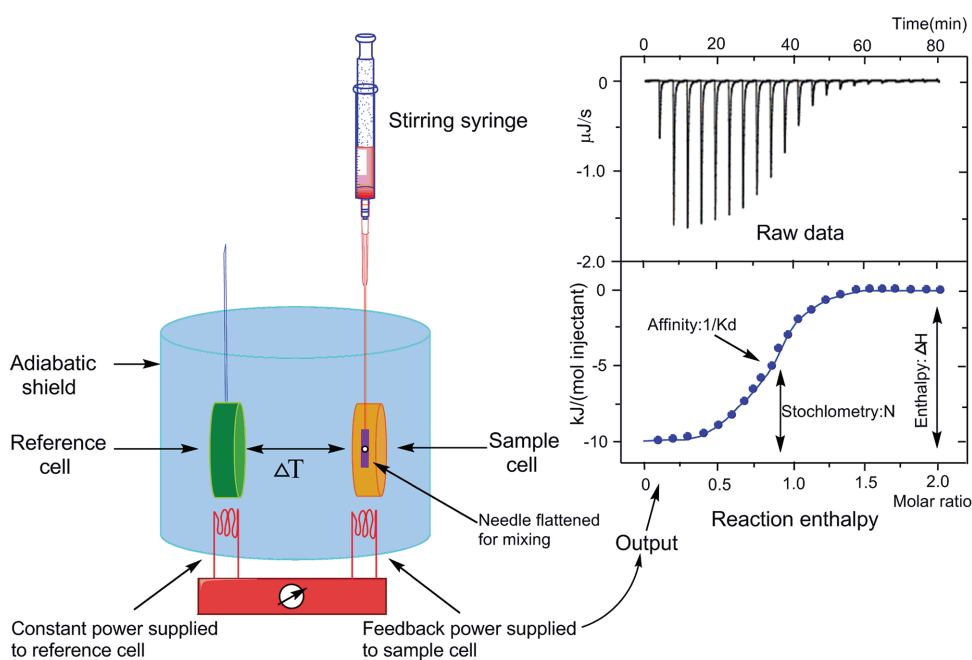
Figure 2-1 **The basic principle of Isothermal Titration Calorimetry.** A schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment and analysis (right). The instrument consists of two cells in an insulated jacket and a power supply to maintain the temperature between the two cells. Injections of the ligand (calcium) from the stirring syringe into the sample cell (protein) cause a change in the temperature, which is measured and recorded as a titration thermogram (upper right). The raw data is integrated (blue circles) and plotted against the molar ratio to be fitted with a curve that corresponds to a model which considers n binding sites and gives thermodynamic properties of the reaction. Image reproduced from (Chengcheng et al., 2015).

When studying protein interactions, choosing the experimental protein concentration is dependent on the c-value as defined by Equation 3.

$$c = \frac{nP_t}{K_d}$$

Equation 3 **The ITC parameter C**. n = number of binding sites per protein molecule, P_t = protein concentration in the measurement cell, K_D = The dissociation constant.

This c value affects the shape of the binding isotherm in the processed data, as shown in Figure 2-2. To obtain a sigmoidal shaped binding isoform, necessary for good estimates of n, c is required to be between 20 and 100. If c is below 20, then thermodynamic parameters can only be obtained if one is fixed during the curve fitting process (Dutta et al., 2015). The following ITC experiments were conducted using historic experimental settings as detailed in (Yuan et al., 2013) which had low c-values.
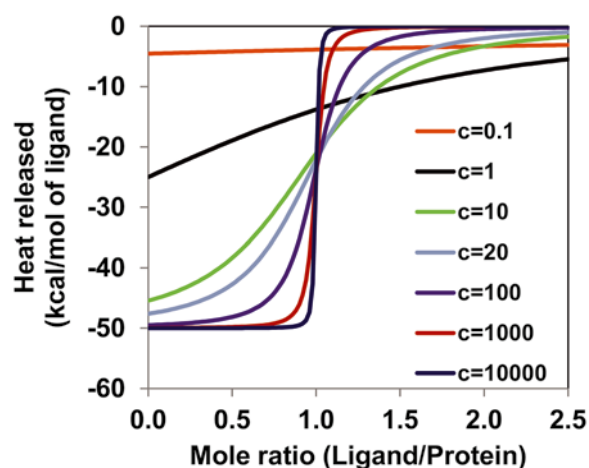
Figure 2-2 **The importance of the c value for binding isotherms.** Simulated binding isotherms for a 1:1 system were generated. For low c values (c < 5), the binding isotherm loses the S-shape and thus thermodynamic parameters cannot be determined. At high c values (c > 1000), some thermodynamic parameters can be determined such as the stoichiometry (n) and ΔH. In optimum values of c (20 < c < 100), all thermodynamic parameters such as K_D, ΔH, ΔS and n can be obtained from the sigmoidal binding isotherm. Image reproduced from Dutta et al. 2015.

2.3.2 Isothermal Titration Calorimetry Measurements

WT human derCD23, mutant B and mutant C were initially dialyzed against 10 mM Tris pH 6.8, 125 mM NaCl and 10 mM EDTA to remove the bound Ca^{2+} present in the refold buffer. This was followed by extensive dialysis against 50g/L Chelex-100-pretreated 10 mM Tris pH6.8, 125 mM NaCl buffer. The 10 mM Tris pH6.8, 125 mM NaCl buffer was filtered with a 0.22 micron filter spun at 8736 rcf and was used to dilute the protein (if required) and to establish the baseline for the iTC200 microcalorimeter (Microcal, GE Healthcare) by titrating 20 mM $CaCl_2$ into the buffer. For the protein experiment, 20 mM $CaCl_2$ diluted in buffer was titrated into 400 µl of 50 or 500 µM derCD23 protein, measured at 25°C. Data was analysed using the MicroCal ITC-ORIGIN Analysis Software (Microcal, GE Healthcare).

2.3.3 Surface Plasmon Resonance (SPR)

SPR is a spectroscopic technique that measures the binding between molecules in solution (the analyte) to an immobilised molecule (the ligand) in real time. The immobilised molecule is bound to a planar metal surface, usually gold. Gold can be regarded as plasma, it is a medium with an equal concentration of negative and positive charges, of which at least one charge type is mobile. Plasma has a natural oscillation frequency, the plasma frequency, and plasma oscillations are known as plasmons. A plasma oscillation along the surface of the metal is a known as a surface plasmon as illustrated by k_{sp} in Figure 2-3. A surface plasmon can couple with a photon (k_x), from the polarised light shone at the gold layer. The polarised light encounters an interface (between the gold layer and the analyte solution, which is of low refractive index (n_2)) at a specific angle (θ), and is totally reflected within the prism. The electromagnetic component of the polarised light combined with a surface plasmon of the same

wavevector, leads to the enhancement of the evanescent wave (E). This results in a dip of light intensity observed (Ritzefeld and Sewald, 2012).

A change in the refractive angle results in an angular shift of the position of the light dip echoed in the reflective curve, as shown in Figure 2-4, the change between phase 2 to 3. These shifts in the refractive index are directly proportional to the mass being immobilised and can be used to measure kinetic parameters and the affinity (K_D).
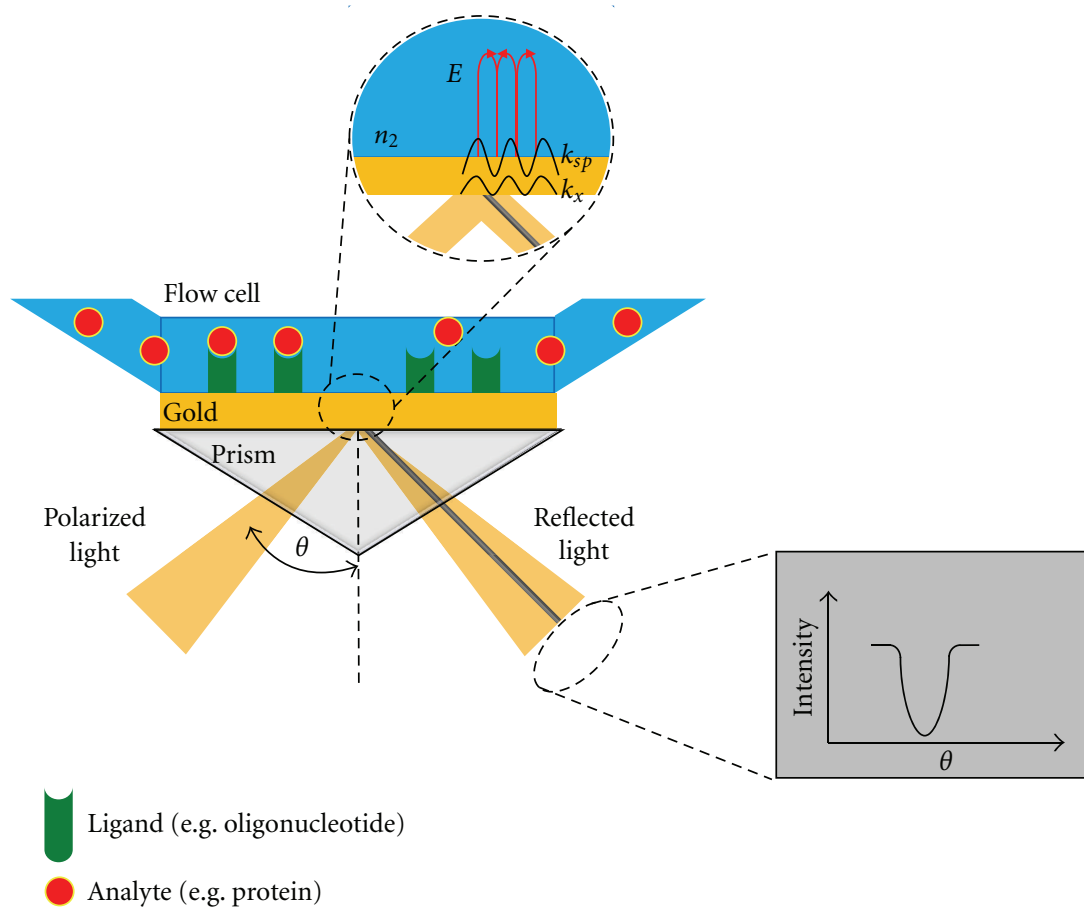
Figure 2-3 **The principle of Surface Plasmon Resonance.** E = evanescent field amplitude, n_2 = refractive index of medium with lower refractive index, k_{sp} = wavevector of surface plasmon, k_x = wavevector of photon. Insert = the change in refractive index. Image reproduced from Ritzefeld and Sewald 2012.

85

SPR machines are built using microfluidics with various numbers of channels to allow parallelisation of the experiments. One channel is always the reference channel in which no molecules are immobilised to it, this allows for changes caused by analyte flow to be substracted from the experimental channels. The difference between the two channels is plotted on a sensorgram, depicted in Figure 2-4. The process is repeated with different concentrations of analyte to obtain K_D values.
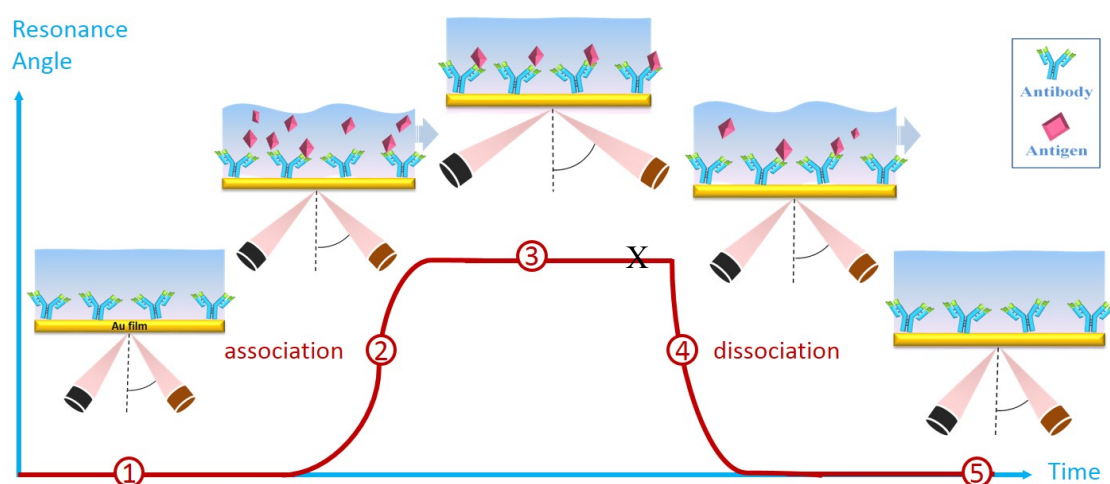
Figure 2-4 **The phases of a Surface Plasmon Resonance experiment.** The red line shows a typical SPR sensorgram divided into five phases. In this example, the interactions between an antibody and an antigen were measured. **1**) The antibody is bound to the gold film and the baseline is established. **2)** The analyte (the antigen) is flowed over the film, some of the antigen binds to the antibody in the association phase. This change in the refractive index, causes a variation of the specific angle (θ) corresponding to change of mass of the surface. **3)** In the equilibrium phase, the association and dissociation of the antigen from the antibody occur at equal rates. **4)** In the dissociation phase, the analyte flow is stopped and the antigen dissociates from the antibody. During this process, a regeneration buffer is added to remove all the analyte from the antibody. **5)** The surface has re-generated and has returned to its baseline. This process is repeated for different concentrations of analyte. Measurements taken during the steady-state of the equilibrium phase, (X), are plotted against the analyte concentration to give a binding curve, used to determine the K_D. Image reproduced from Weistron 2014.

2.3.4 Preparation of the Surface Plasmon Resonance Chip

The CM5 sensor chip (GE Healthcare) and the streptavidin chip (GE Healthcare) were prepared at 25°C. The carboxymethyl groups on the CM5 chip in all flow channels were activated by 0.4M N-ethylN'-(3-diethyl-aminopropyl)- carbodiimide (EDC) and 0.1M

N-hydroxysuccinimide (NHS) in a 1:1 ratio. The derCD23 proteins had been dialysed into HBS/Ca^{2+} (HEPES pH 7.4, 150 mM NaCl, 4 mM CaCl$_2$) and were diluted into 10 mM Na-acetate pH 4.5 with 0.005% (v/v) surfactant P-20 for amine coupling to the measurement channels. The coupling density was limited to ~70-100 RU. Free ester groups in all channels were blocked with 1 M ethanolamine-HCl pH 8. Biotinylated IgE-Fc was dialysed into HBS/Ca^{2+} (HEPES pH 7.4, 150 mM NaCl, 4 mM CaCl$_2$) and immobilised on a streptavidin chip by biotin covalent linkage at two densities (~400 RU for the lower density and ~800 RU for the higher density).

2.3.5 Surface Plasmon Resonance Measurements

The CD23 proteins and IgE-Fc were dialysed into either HBS (HEPES pH 7.4, 150 mM NaCl, 10 mM EDTA) or HBS/Ca (HEPES pH 7.4, 150 mM NaCl, 4 mM CaCl$_2$) which were also used as the running buffer with 0.005% (v/v) surfactant P-20 added. Experiments were performed on the Biacore T200 instrument (GE Healthcare) at 25°C with the flow rate set to 20 μl/mins using conditions previously optimised for derCD23, (Yuan et al., 2013). Binding of Fcε3-4 to immobilised derCD23 was collected at a range of concentrations: 8000, 4000, 2000, 1000, 500, 250, 125, 62.5 and 31.2 nM with a 90 s association phase followed by a 5 min dissociation phase. Binding of derCD23 proteins to immobilised IgE-Fc were collected at a different range of concentrations: 40, 20, 10, 5, 2.5, 1.25, 0.63 and 0.32 μM respectively. The raw data was processed with standard double referencing data subtraction methods before analysing the rates of equilibrium in Microcal Origin (OriginLab).

2.3.6 Nuclear Magnetic Spectroscopy (NMR)

NMR relies on the observation that atoms with an odd number of protons or neutrons spin about an axis and generate their own magnetic field. Such atoms that magnetic

moment include hydrogen (^{1}H), carbon (^{13}C) and nitrogen (^{15}N) and are found in biological samples as shown in Figure 2-5A. An NMR machine consists of a very strong magnet and when a biological sample is placed inside the magnet, the atoms with magnetic moment adopt one of two states; they either align with or against the magnetic field (Figure 2-5B). A specific radiofrequency wave is applied to the sample and the atoms with magnetic moment 'flip' to their other state (Figure 2-5C). When the radiofrequency wave is discontinued, the atoms relax and release radiofrequency waves which provide information on the location and chemical environment of the atoms (Figure 2-5D) (Bothwell and Griffin, 2011).

The data collected is processed and analysed to give a spectrum (one dimensional) (Figure 2-5E). An additional radiofrequency wave (e.g. one with a wavelength targeted for ^{1}H and the second to target ^{15}N) will provide data for a two-dimensional spectrum.
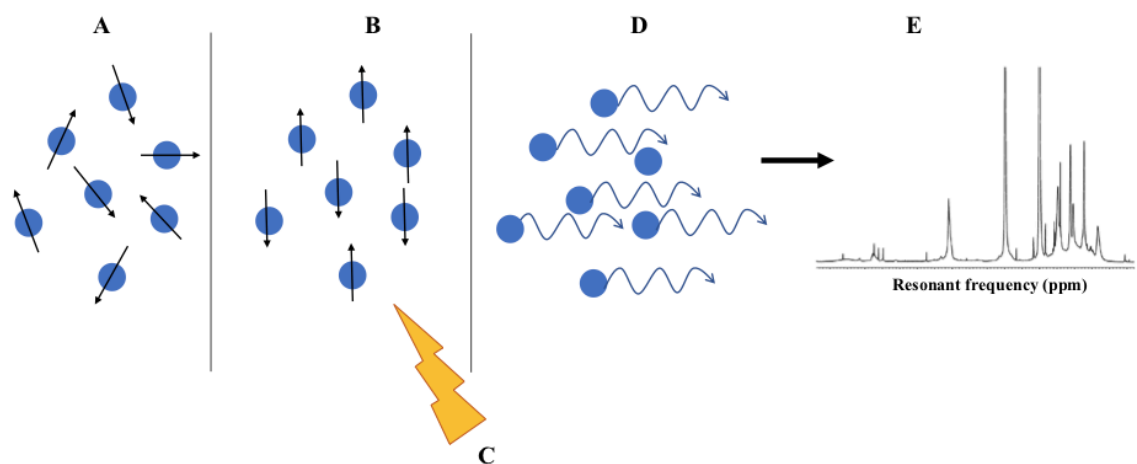
Figure 2-5 **An example of an NMR experiment. A)** Nuclei (blue circles) in a sample have random magnetic moment in a sample. **B)** The sample is placed inside the NMR magnetic and the nuclei either align with the magnetic field (up arrow) or against the magnetic field (down arrow). **C)** A radiofrequency is applied to the experiment. **D)** The nuclei relax and emit electromagnetic radiation (wavy lines). **E)** The radiation is summed, analysed and plotted as a spectrum in which intensity is plotted against resonant frequency.

2.3.7 ^{1}H-NMR Spectroscopy

A pure protein sample, of at least 90 μM in TBS, pH 7.5 (25 mM Tris-HCl pH 7.5, 125 mM NaCl), had 5% D_2O was added, to confirm correct protein refolding. Measurements for one-dimensional ^{1}H-NMR analysis were taken at a proton frequency of 500 MHz at 35°C with all experiments performed using a 500 MHz NMR machine (Bruker DMX500) using conditions previously optimised for derCD23 (Hibbert et al., 2005).

2.3.8 ^{1}H - ^{15}N HSQC NMR Spectroscopy Titrations

A 0.5ml sample of ^{15}N labelled protein (please refer to section 2.2.2) with 5% D_2O added, was used to collect titration spectra. Protein samples with concentrations between 350 μM and 400 μM were stored in a buffer containing 25 mM Tris, 125 mM NaCl, pH 6.8. Samples were spun at 15000 g for 10 mins to remove debris, and were transferred to Shigemi NMR tubes (Shigemi Corporation). Measurements were taken at 35°C with all experiments performed using a 500 and 700 MHz NMR machine (Bruker DMX500, DMX700). For calcium titration experiments, $CaCl_2$ was added at step-wise increments until saturation was seen. For carbohydrate titration experiments, mannose was added at step-wise increments until saturation was seen. Data was processed in TopSpin3.5pl6 (Bruker) and analysed in SPARKY (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

2.3.9 Glycan Array

The RayBio Glycan Array 100 [4 sample kit] (RayBiotech, Inc. lot no: 042817 30056) was performed as per manufacturer's instructions. The PBS buffer pH 8.0, was also prepared according to the manufacturer's instructions as stated here: 0.6 g KCl, 24 g NaCl, 0.6 g KH_2PO_4 and 3.45 g Na_2HPO_4 were dissolved in 2500 ml of ddH_2O with the pH adjusted to pH 8 and the final volume adjusted to 3000 ml with ddH_2O. In brief, the

calcium-free wild type derCD23, mutant B and mutant C proteins were initially dialyzed into 1x PBS, pH8.0 buffer with 3 buffer exchanges, and biotinylated by incubating the reaction solution at room temperature with gentle rocking for 30 min. The reaction solution was mixed every 5 min by gentle tapping. 3 μl of stop solution was added to each reaction solution tube, and immediately dialysed into the same 1X PBS, pH8.0 buffer.

The printed glycan array glass slide was air-dried and subsequently blocked with kit sample diluent E for 30 mins at room temperature. This was followed by incubation with the protein samples and dH_2O as a negative control, with gentle rocking overnight at 4°C. Samples were decanted from the wells, and each well was washed 5x 5 mins with 800 μl of 1x Wash buffer I, and 2x 5 mins with 800 μl of Wash buffer II, with both sets of washes at room temperature with gentle shaking. To each well, 400 μl of Cy3 equivalent dye-conjugated streptavidin was added. The whole glass slide was covered with plastic adhesive strips, and incubated in aluminium foil to avoid exposure to light at room temperature for 1 hour with gentle rocking.

Next, the samples were decanted from the wells and washed 5x 5 mins with 800 μl of 1x Wash buffer I at room temperature with gentle shaking. The slide assembly was disassembled and gently placed into the slide washer tube, and covered with 1x Wash buffer I and incubated at room temperature for 15 mins with gentle rocking. This was followed by a single wash with 1x Wash buffer II for 5 mins. The glass slide was washed with dH_2O for 5 mins and the signals were acquired and processed using the LuxScan™ 10K Microarray scanner (Capital Bio) and the LuxScan 10K software. The RayBio Analysis Tool for GA-Glycan-100 (RayBiotech, Inc.) was used to analyse the results. A list of the glycans printed on the array can be found in Appendix B.

To repeat the results in the presence of calcium, the glycan array was washed first with PBS-Tween at room temperature with gentle rocking for 20 mins. Next the array was washed with Wash Buffer 1 and then with Wash buffer II, both at room temperature with gentle shaking for 20 mins. The glycan array was air-dried and repeated, starting at the blocking step. The array was incubated with the same protein samples, except with 4 mM $CaCl_2$ added to the samples and incubated overnight at 4°C with gentle rocking. Samples were decanted and the protocol followed as already described.

2.4 X-ray Crystallography

In order to determine the 3D structure of a protein at atomic level, samples and data must be processed in a different manner to light microscopy. First samples are prepared by growing protein crystals from pure protein, the individual protein molecules are aligned in a repeating three-dimensional pattern to create a lattice. To collect data, X-rays (which are close in wavelength to the distance of an atomic bond) are shone at the protein crystal and diffraction is collected.

The diffraction pattern represents scattering constructive waves off a crystal lattice, described as waves which are an integer apart with the same amplitude. These conditions are known as Bragg's Law as described in Equation 4, and Figure 2-6 displays the conditions required for diffraction. The diffraction of measurable X-rays from the minimum spacing (d) of crystal lattice planes gives the definition of resolution which is measured in Å. The resolution is often used as a measure of quality and describes the minimum distance between neighbouring structural features in an electron-density map. The higher the resolution, (the lower the value in Å), the smaller the spacing (d) between features in an electron-density map which results in more independent reflections and a better-defined structure.

$$n \lambda = 2d \sin\Theta$$

Equation 4 **Bragg's Law of diffraction.** n = an integer, λ = the wavelength of the incident X-rays, d = the interplanar spacing of the crystal and Θ = the angle of incidence.
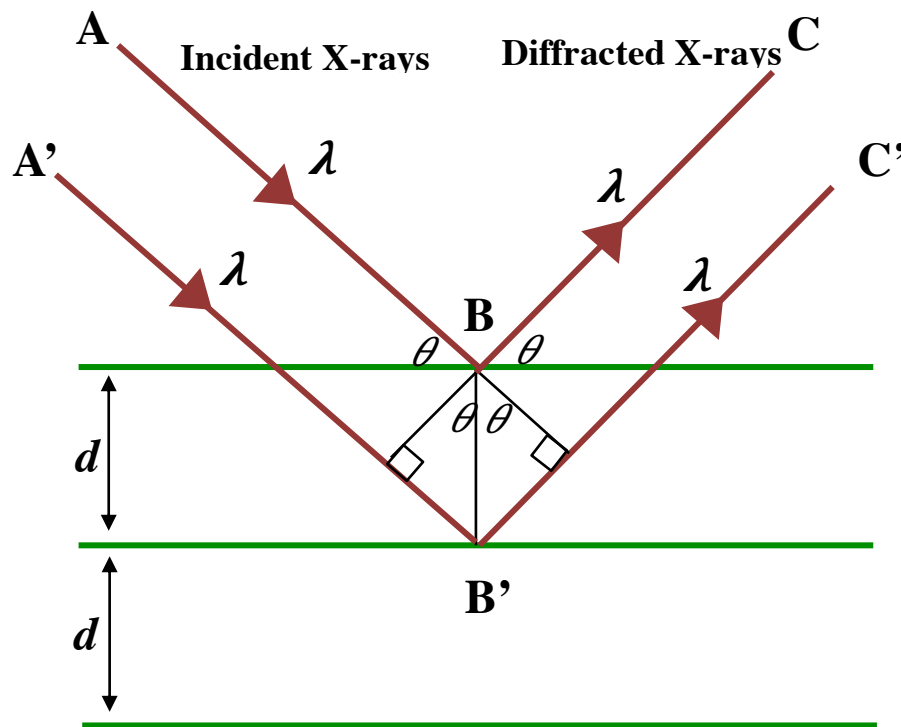
Figure 2-6 **Bragg's Law.** Diffraction is created when diffracted X-rays exhibit constructive interference when the distance between paths ABC and A'B'C' differs by an integer number of wavelengths (λ).

In a diffraction pattern, the amplitude of a wave can be derived from experimentally-measured intensities of the diffracted beams. The phase of the wave cannot be determined and thus in macromolecular X-ray crystallography various techniques are used to overcome this 'phase problem'. The most commonly used technique is called molecular replacement, in which a similar solved structure is used to approximate a set of phases which is then compared to the experimental observations and further refined.

2.4.1 Pre-crystallisation trials

The pre-crystallisation test (PCT) (Hampton research) determines if the protein concentration for crystallisation screening is appropriate. Trials were run as per manufacturer's instructions.

2.4.2 Crystallisation Screens

Commercial screens were used to investigate the conditions required for protein crystallisation. The Hampton Index HT (Hampton Research), PEG and JCSG (Qiagen) screens contain 96 different reservoir solutions, differing in precipitant, salt and buffer. The screens were prepared in polystyrene MRC 96-well plates in a sitting-drop vapour diffusion method with 100 nl or 200 nl of concentrated protein (stored in TBS buffer, 25 mM Tris-HCl pH 7.5, 125 mM NaCl, 0.05% azide w/v) dispensed into the wells with the Mosquito Liquid Handling robot (TTP Labtech). 100 nl or 200 nl was then removed from the reservoir and added to the protein well also with the Mosquito Liquid Handling robot (TTP Labtech). The plates were incubated at 18°C, and were checked for crystal growth regularly. Photographs of the crystals were taken using NIS elements software (Nikon).

2.4.3 Salt detection with the PX scanner

Grown crystals can either be salt or protein based. To discriminate between the two, their X-ray diffraction qualities were assessed. Sealed crystallisation plates were inserted into the PX-scanner (Agilent Technologies) and exposed to X-rays to quantify their diffraction properties. Results were visualised in CrystalEyes software.

2.4.4 Crystallisation Optimisation

The initial crystal growing conditions for mutant B were found in the commercial crystallisation screen JSCG: 0.1M sodium-cacodylate pH 6.5 and 1M tri-sodium-citrate. These conditions were systematically changed and optimised in a 6 by 4 grid format. The grid conditions ranged from 0.8 - 1.3 M tri-sodium-citrate and pH 6-6.75 sodium-cacodylate, and in the second round of optimisation, this was refined to $0.8 - 1.1$ M tri-sodium-citrate, with a duplicate grid containing an added 5% glycerol. Plates were incubated at 18°C and took a week to grow.

2.4.5 Cryo-cooling crystals in cryoprotectant

To suppress ice formation, cyroprotectant was added to the protein crystal well, before mounting the crystals onto nylon cryo-loops for data collection. For most of the crystal growing conditions, the cryoprotectant was the reservoir solution + 20% glycerol, alternatively if the crystals had been soaked with calcium first, the solution would also have worked as a cryoprotectant. Once the protein crystals were mounted onto the cryo-loop, the crystal was flash-cooled in liquid nitrogen (Teng, 1990), and stored in liquid nitrogen until data collection.

2.4.6 Calcium soaking of crystals

The crystal plate seal was broken briefly to add $CaCl_2$ directly to the protein well in a drop-wise fashion and immediately re-sealed. The plate was left to incubate at 18°C for 4 days before cryo-cooling the crystals as described above. Three different calcium-containing solutions were used: 1.1 M tri-sodium-acetate, 0.1 M Na-cacodylate pH 7.1, 10 mM $CaCl_2$ with either 20% or 16.7% glycerol or 1.35 M tri-sodium-acetate, 0.1 M Na-cacodylate pH 7.1, 10 mM $CaCl_2$.

2.4.7 Data Collection

Data was collected at Diamond Light Source (Synchrotron) using X-ray radiation at beamline I04 for the WT human derCD23, beamline I04-1 for mutant B and beamline I02 for mutant C. All data was collected at cryo-temperatures of 100 K, maintained by a stream of nitrogen.

2.4.8 Data processing

The raw data collected at the Synchrotron was first inspected with iMosflm (Leslie et al., 2011). The data was indexed and the unit cell parameters defined, and refined before integration with the accurate cell parameters. Pointless was run on the data to determine

the space group before scaling and merging the data in SCALA. The protein structures were solved by molecular replacement using PDB file: 4J6J, crystal form A as the search model. Refinement was conducted in PHENIX (Adams et al., 2010) and alternated with manual model-building in COOT (Emsley et al., 2010). Structures were validated using the POLYGON tool in PHENIX. Data is measured in Å; the lower the number of Å, the higher the resolution, which provides more detail about the structure of the protein

2.4.9 Graphics software used for structure visualisation

Figures were generated using PyMol (The PyMol Molecular Graphics System, v 1.7.4.3, Schrodinger, LLC) with specific codes for visualising clashes (show_bumps.py written by Thomas Holder) and for the Porcupine plots (modevectors.py written by Sean Law and Srinivasa, Michigan State University). Codes were accessed from the PyMol wiki (www.pymolwiki.org).

Figures exhibiting electron density were created in CCP4mg molecular graphics software (McNicholas et al., 2011).

2.4.10 Graph graphics software

Graphs were drawn in either Prism (GraphPad v 7.0b for Mac OS X) or QtiPlot (v 0.9.9-rc9).

2.5 Molecular Dynamics Simulations

2.5.1 System Preparation

All MD simulations were set up using the AMBER14 software suite (Case et al., 2014); coordinates for a monomer of derCD23 were chosen from chain A of the asymmetric unit of the crystal structure PDB 4G9A. This model was also used as the base for the mutant proteins generated in PyMol. All the structures were then processed by pdb4amber (Case et al., 2014) to remove water molecules, build disulphide bonds and assign amber specific residue identifiers. The online tool H++ (http://biophysics.cs.vt.edu/H++ version 3.2) (Anandakrishnan et al., 2012; Gordon et al., 2005) was used to determine the protonation state of the histidine residues, which were manually renamed in the pdb text file. Another amber14 tool; reduce (Word, et. al. 1999) added protons to the model and determined optimum geometry of each residue. The previous tools remove all non-covalently bound ions so calcium was modelled using Yasara v16.3.8 (Elmar Krieger). The calcium ions were modelled on the positions of calcium ions in another C-type lectin protein, DC-SIGN, pdb file: 1SL4.

Parameter files were generated using AMBER14SB and force fields frcmod.ionsjc_tip3p and frcmod.ionslm_1264_tip3p for the ions. The protein was solvated in the TIP3PBOX model with a 10 Å layer of water. The total charge of the protein was neutralized with counterions and the average size of the monomer simulations is ~ 22700 atoms which is ~7000 residues in a 58.6 Å by 60.4 Å by 82.9Å box with ~6800 water molecules. This process is summarised in Figure 2-7.

Structure is retrieved from the PDB or generated by SwissModel

All non-protein atoms are removed

Hydrogens are added

A periodic box of water is added

Ions are added and the system in neutralised
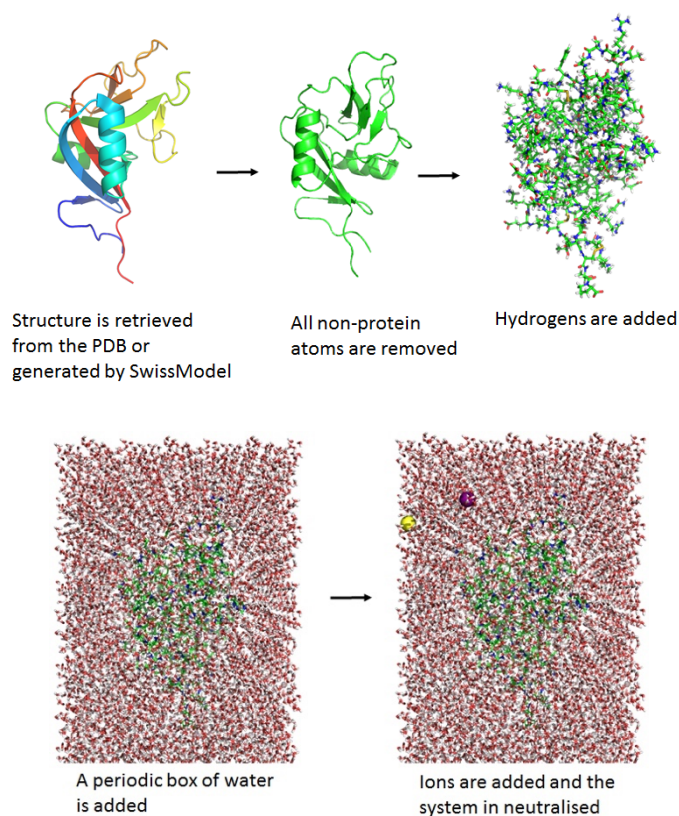
Figure 2-7 **A schematic overview of the MD simulation set-up.**

2.5.2 Running Molecule Dynamic Simulations

Simulations were performed using pmemd.cuda in AMBER14 program (Case et al., 2014) with a time-step of 2fs. Firstly, the system is minimized with protein constrained to equilibrate the solvent, then the entire system is minimized followed by a slow heating step to bring the system to 300K. Next, the system was relaxed, initially with the protein heavy atoms restrained, followed by the protein released in a NTP ensemble. The systems were subjected to 1 µs MD simulations at a constant temperature of 300K, controlled by Langevin dynamics and Berendsen barostat for pressure control. Please refer to Appendix C for full software instructions.

2.5.3 Calcium Ion binding plots

The trajectories were analysed to produce calcium ion binding plots in the Ipython notebook Jupyter (Perez and Granger, 2007). Details of the python script are found in Appendix D.

2.5.4 Trajectory Analysis

Mdcrd files were converted to dcd files using the CPPTRAJ program (Roe and Cheatham III, 2013), which also removed non-alpha carbon atoms. Calcium ion distances were also calculated in CPPTRAJ. The data was analysed with the R-based program Bio3d (please refer to Appendix E for script)(Grant et al., 2006), where structures were aligned with the exclusion of the termini as they radically altered the PCA (residues 15-125).

Principal component analysis (PCA) was conducted in Gromacs (version 5.1.3) (Abraham et al., 2015). PCA is a statistical technique to find the principal components of a set of data (Kerrigan, 2013). The principal component is not a standard x- or y-axis but an axis that describes the most variance. This is useful to know, since it is able to reduce data dimensions, simplifying it as well as making data easier to visualise. For MD trajectories, the principal component reveals the most important motions in proteins.

Files were converted into a Gromacs compatible files by CCPTRAJ and PARMED (Jason Swails, University of Florida).

Chapter 3 Calcium studies of derCD23

3.1 Introduction

The type II transmembrane glycoprotein and C-type lectin CD23/FcεRII is responsible for a plethora of biological responses (Acharya et al., 2010). In the context of asthma, CD23's role in IgE homeostasis is considered essential, alongside its roles in enhanced antigen presentation, and gut lumen transcytosis that lead to epitope spreading (Gould and Sutton, 2008). When examining the structure of CD23, it becomes apparent that unlike most antibody receptors, which are members of the immunoglobulin gene super-family, CD23 is a C-type lectin (Fridman, 1991). C-type lectins usually bind to carbohydrate via calcium ion interactions, such as mannose-binding protein (Weis, 1992), a member of the CTL family, and they frequently bind to two or more calcium ions (Drickamer, 1999). Uncommonly for a C-type lectin, human CD23 binds to IgE via protein-protein-interactions and only binds to one calcium ion, whereas an alignment of human CD23 sequences from other species and homologous proteins predicts two calcium-ion binding sites (Zelensky and Gready, 2003)(Zelensky and Gready, 2005).

This divergence of the primate and rodent lineages regarding calcium ion binding is shown in Figure 3-1, which displays calcium-coordinating residues in the auxiliary calcium-binding site (as stated in chapter 1) of human CD23 and in Figure 3-2. In primates, these residues have been substituted by residues with an opposite charge so that they no longer co-ordinate a calcium ion. In mice, calcium ions are mandatory for CD23 to bind to the ligand IgE via protein-protein interactions (Richards and Katz, 1990) while in humans, calcium is not vital for non-carbohydrate CD23-IgE interactions (Richards and Katz, 1990; Hibbert et al., 2005) In the presence of calcium, two additional salt bridges are formed between the human CD23-IgE complex which results

in a temperature-dependent 30-fold increase in affinity (Yuan et al., 2013). Considering the spatial proximity of the principal calcium-binding site to the IgE-binding-residues within loop 4 (Arg 253 to Glu 257) and loop 1 (Leu 226 to Glu 231) of human CD23 in Figure 3-2, it is clear how disordering or ordering of the loops by calcium could affect loop conformation and affinity of CD23 to IgE.

```
              220        230        240        250        260        270
               .          .          .          .          .          .
Human     GSWIGLRNLDLKGEFIWVDGSHVDYSNWAPGEPTSRSQGEDCVMMRGSGRWNDAFCDR

Gorilla   GSWIGLRNLDLKGEFIWVDGSHVDYSNWAPGEPTSRSQGKDCVMMRGSGRWNDAFCHR

Pig       GSWIGLRDLDIEGEFIWMDEKPLDYSNWQPGEPNDAGQGEHCVMMQASGQWNDAFCGS

Dog       GTWIGLRDLDREGEFIWMDENPLNYSNWRPGEPNNGGQGEDCVMMQGSGQWNDAFCGS

Rat       ESWIGLQDLNMEGEFVWPDGSPVGYSNWNPGEPNNGGQGEDCVMMRGSGQWNDAFCRS

Mouse     DSWIGLQDLNMEGEFVWSDGSPVGYSNWNPGEPNNGGQGEDCVMMRGSGQWNDAFCRS

                                          EPN
```

Figure 3-1 **A sequence alignment of the principal and auxiliary calcium binding sites of CD23 in various species.** The principal binding site residues are highlighted in blue and the auxiliary calcium binding site resides are highlighted in magenta. The green residue E257 is involved in both calcium binding sites. The EPN carbohydrate binding motif is highlighted in yellow. The amino acid numbering scheme is shown above the sequences.

Human CD23 has an affinity for calcium of 1.5 mM, and this is close to physiological calcium concentrations in blood, ranging between 1 and 1.5 mM (Yuan et al., 2013). This signifies that the calcium-bound and calcium-free CD23 might be in equilibrium and thus confer sensitivity to calcium-facilitated ligand interactions. Calcium-bound membrane-CD23 binds to allergen-bound IgE, and undergoes receptor-mediated endocytosis, where the change in calcium concentrations in endosomes (to ~10 μM) (Andersen and Moestrup, 2014) results in the release of IgE and the allergen. The loss of the auxiliary calcium-binding site in human CD23 might result in decreased sensitivity to IgE-allergen binding, and downstream this may affect the release of the

ligand within the endosome and the subsequent processing and presentation of the allergen to the immune system. This reduction in calcium sensitivity might be enough to tip the scales such that the ligand (allergen-bound IgE) is released into a different vesicle (a late endosome rather than an earlier vesicle) and this might be enough for the ligand to embark on a different route of enzymatic digestion, and display on MHC class II molecules. The different peptides presented to naïve T-cells might deviate the immune response from Th1 cells to the Th2 cell response, which recruits IgE-producing B cells, mast cells and eosinophils, all involved in allergic inflammation and thus resulting in an increased chance of an asthmatic reaction (Mudde et al., 1995a). In this context, calcium is posing an additional layer of regulation over IgE.
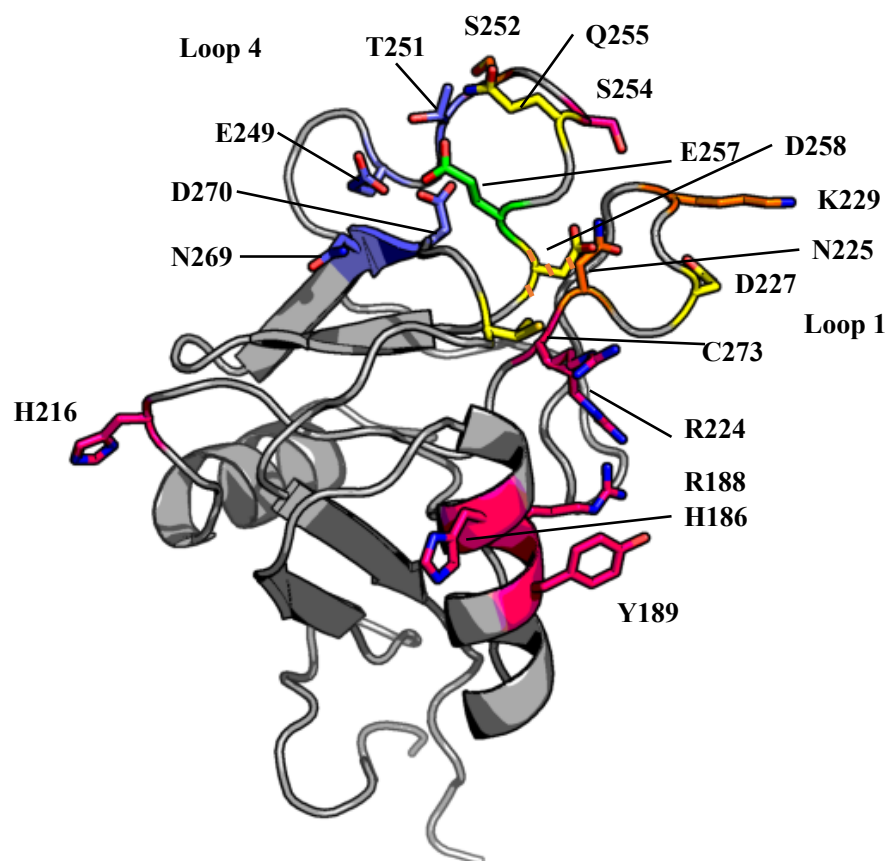
Figure 3-2 **Important residues in WT human derCD23**. Blue carbon atoms = the principal calcium-ion-binding site, orange = "auxiliary" non-functional calcium-binding site, green carbon atoms = E257, a residue involved in both principal and "auxiliary" sites, pink carbon atoms = residues involved in IgE binding, yellow carbon atoms = additional residues involved in IgE binding upon calcium binding, striped yellow and orange carbon atoms = D258 a residue part of the "auxiliary" calcium binding site and makes additional salt bridges to IgE on calcium binding. Residues and loops labelled, pdb file: 4G9A.

102

This has led to the hypothesis which suggests human CD23 may have lost its ability to bind a second calcium ion in the auxiliary calcium-binding site, in order to enhance its binding to human IgE. To test this hypothesis, a set of mutations was designed in human derCD23 that would enable the restoration of the auxiliary calcium binding site, and this focused on mutating the residues shown in Figure 3-3.
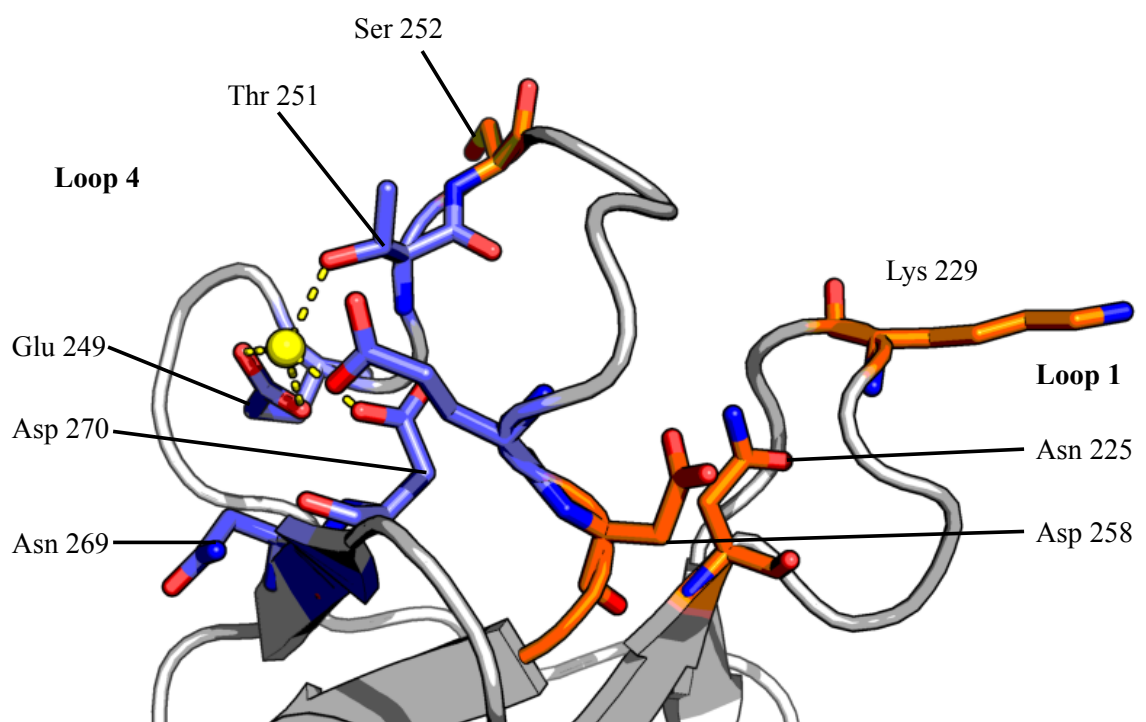
Figure 3-3 **The calcium binding site of human derCD23**. The principal calcium-ion-binding site (blue) and "auxiliary" site (orange) in wild-type human CD23. Calcium ion shown in yellow, residues and loops labelled, pdb file: 4G9A.

3.1.1 Mutation Design Rationale

Residues to be mutated were selected based on the alignment of human WT and mouse WT derCD23 as shown in Figure 3-1. The proteins have 70% sequence identity and comparisons to other C-type lectin structures imply that residue E257 is involved in both calcium-binding sites. Not wishing to destroy the primary site, a panel of three

mutant proteins were designed as illustrated in Figure 3-4. Each set of mutants were designed in a consecutive fashion so that it would only take a single set of additional primers to add the extra mutations into the DNA for the next mutant. In mutant A, (Figure 3-4A) the mutations introduced cause a charge reversal from mostly postively charged side chains in the WT protein, to amino acid residues containing carboxylate groups to aid calcium ligation, as most calcium ions in proteins are chelated by oxygen-containing groups (Yang et al., 2003). Mutant B has an additional mutation T251N (Figure 3-4B), because although it is involved in the principal calcium-binding site, the asparagine is conserved in C-type lectins and is part of the 'EPN' carboyhdrate binding motif. Residues R253 and S254 are not involved in calcium ion co-ordination, however they were mutated to glycines as found in the mouse CD23 sequence, which increased the flexibility of loop 4 depicted in mutant C Figure 3-4C. We speculated that these mutations could aid in either calcium or carboyhdrate binding.

A   **B**   **C**
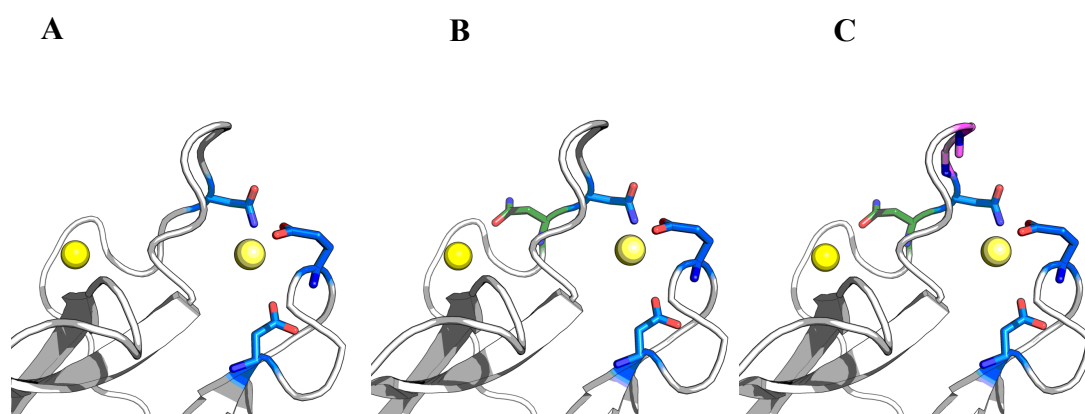
Figure 3-4 **Models of the mutant proteins, focusing on the calcium binding sites.** derCD23 in grey, in blue – mutations N225D, K229E and S252N, in green – additional mutation T251N, and in pink additional mutations R253G and S254G. Carbon atoms coloured either blue, green or magenta, red = oxygen, blue = nitrogen. Yellow spheres = calcium ions (predicted places).

3.2 Results

3.2.1 Production of WT human derCD23 and human derCD23 mutants

The pET5a vector containing the human WT derCD23 sequence was provided by Dr. Rebecca Beavil and was transformed into BL21 cells for protein expression. DerCD23 is a proteolytically cleaved fragment of CD23 which contains the C-type lectin domain and no glycosylated residues (Hibbert, 2005). For the auxiliary calcium-binding site mutants, the DNA was first mutated with the primers as described in the mutation design rationale and section 2.1.7.

3.2.2 Expression of human and mutant derCD23

Human WT and the mutant derCD23 proteins were grown in autoinduction medium, which negates the dependency to induce protein expression by ITPG as cells begin protein expression once all the glucose in the media has been consumed and the *E.coli* begin feeding on lactose (Studier, 2005). All the derCD23 recombinant protein is expressed as insoluble inclusion bodies, as shown in Figure 3-5. The insoluble inclusion bodies must first be washed to remove traces of lipid and carbohydrate contaminants from the *E.coli* cells, to improve protein purity before solubilising the protein for the refold step as explained in sections 2.2.5 and 2.2.6.
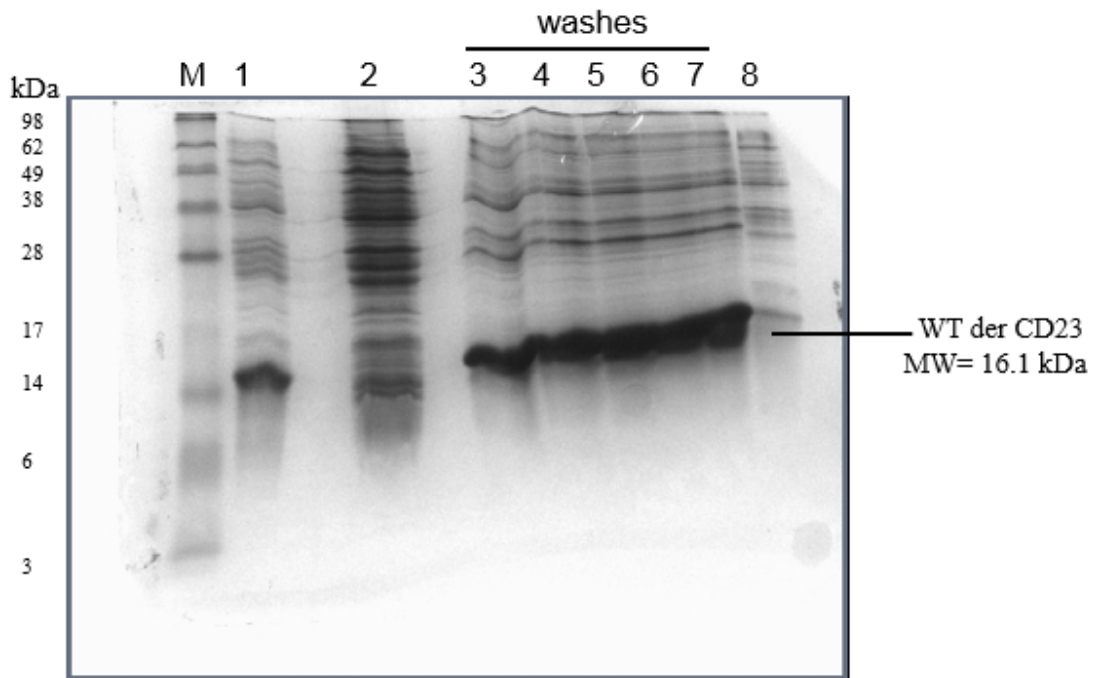
Figure 3-5 **15% SDS-PAGE of wild type derCD23 stained with Comassie Stain.** M = Molecular weight markers, lane 1 = insoluble fraction, lane 2 = soluble fraction, lane 3-7 = inclusion body washes, lane 8 = Solubilised protein in 6M guanidinium hydrochloride buffer.

3.2.3 Refolding

Protein expressed as inclusion bodies take longer to produce than soluble proteins. The recombinant protein is produced at a greater density within the inclusion body and thus its potential yield is higher. The refold process is slow as derCD23 must correctly form all four of its disulphide bonds at 4°C, in the dark, and the oxidised-glutathione and the light-sensitive cysteine in the buffer would promote denaturation and reduction of the protein. To achieve correct folding, a method modified from Taylor et al. (1992) was utilised in which the protein was slowly diluted into a large volume of refold buffer, to ensure monomeric folded protein.

35

3.2.4 Purification

The derCD23 construct does not have a purification tag at either terminus and thus hydrophobic interaction chromatography (HIC) was chosen to purify the protein as depicted by the chromatogram in Figure 3-6. The mutations introduced into derCD23 involved charge reversals that could have altered the isoelectric point, whilst the hydrophobicity would remain similar and allow for the same conditions of purification of all the derCD23 proteins. The eluted proteins were very pure, as indicated by the single bands in Figure 3-7, and HPLC SEC analysis revealed the derCD23 proteins to be homogeneously monomeric as shown by the monodisperse peaks in Figure 3-8.
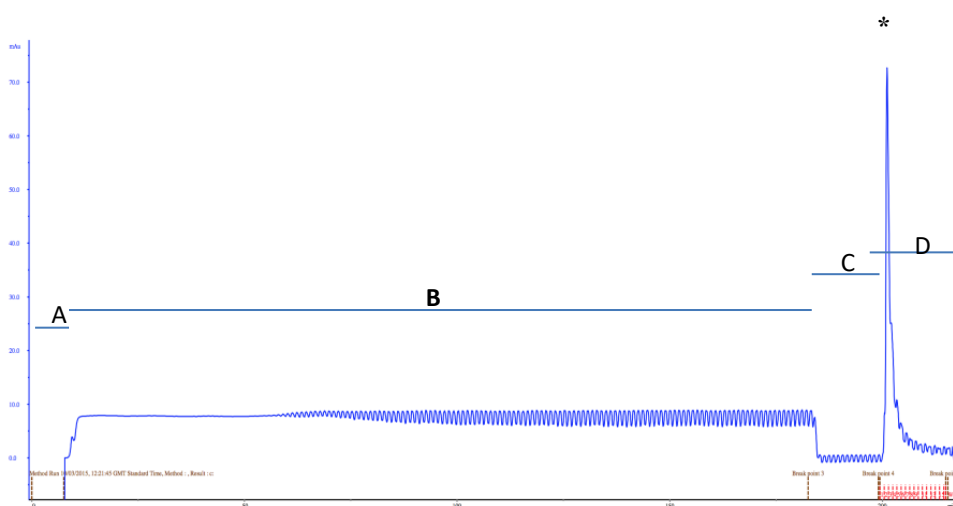
Figure 3-6 **Chromatogram of derCD23 purified by HIC.** The protein solution (blue line = UV absorbance at 280 nm) was loaded onto a 10ml phenyl sepharose 6 fast flow column (GE Healthcare), equilibrated in buffer A (1.5 M ammonium sulphate, 25 mM Tris, 2 mM $CaCl_2$ pH7.5) and eluted in a step-wise gradient of buffer B (25 mM, 2 mM $CaCl_2$ pH7.5), run at 10 ml/min. **A** = column equilibration, **B** = sample load, **C** = column wash, **D** = column elute with fractions shown on x-axis. * = A single peak in fraction 2 indicates pure protein eluted. The fluctuations of the trace are due to the age of the UV lamp.
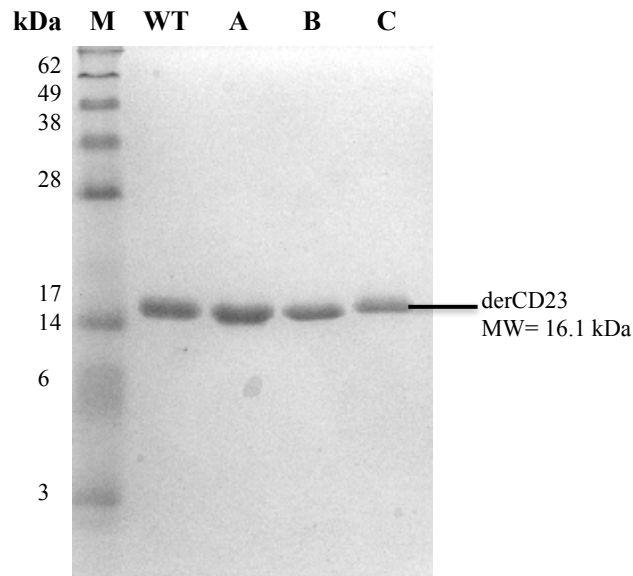
Figure 3-7 **16% SDS-PAGE of purified wild type and mutant derCD23 proteins stained with Comassie Stain.** M = Molecular weight markers, WT = WT derCD23 sample, A = mutant A, B = mutant B, C = mutant C. 2 μg of each protein was run on the gel. The single band indicates a pure protein sample with no degradation (no bands beneath the intense band) and no aggregation (no bands above the intense band).
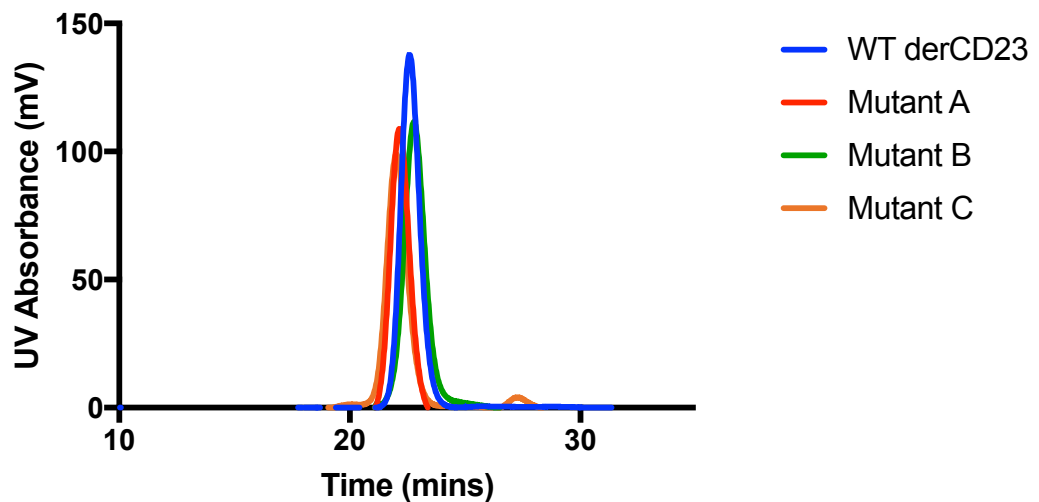
Figure 3-8 **Chromatographic profiles of derCD23 proteins determined by analytical gel filtration HPLC using a Superdex 200 column.** Between 100-140 μg of purified derCD23 was injected onto a Superdex 200 10/300 GL column equilibrated with TBS-Ca^{2+} buffer (pH 7.5), flow rate = 0.75ml/min. The well resolved monomeric peaks at 22 mins indicate a highly pure homogenous protein. If additional peaks were present to the left, this would indicate that aggregate protein was present and would require further purification.

The yields of folded protein from both the mutants and WT human derCD23 were similar, except for mutant A. As mutant A did not express as well as the other derCD23 mutant proteins at either 37°C or 18°C, it was considered to be an exception. As such, the majority of experiments conducted, utilised mutants B and C alongside the WT protein. Not all the solubilised inclusion bodies were used for the re-fold step as described in section 2.2.6, as an estimated 50 mg/L could be purified from 1 L of autoinduction medium. For one round of refolding and protein purification, that used part of the total isolated inclusion bodies, the yield of purified derCD23 protein was at about 12% of the total protein within the inclusion bodies.

3.2.5 Calcium affinity of the derCD23 proteins

To confirm that the proteins had folded properly, a ^{1}H NMR experiment was performed; the spectra are shown in Figure 3-9. The presence of the four characteristic leucine and isoleucine methyl peaks upfield of 0 ppm indicated that the derCD23 proteins were correctly folded, this finding was indicative of buried sidechain-protons packed next to aromatic residues in the hydrophobic core of the protein. The lack of peaks in this area would indicate that the protein has not folded into its tertiary structure. This finding was important, as prior research has documented that changes in the hydrophobic core are capable of decreasing calcium affinity, even when residues are distant from calcium-binding sites (Quesenberry and Drickamer, 1992). Chemical shift perturbations upon calcium binding can also be detected in Figure 3-10 with the small peak at -0.4 ppm not detected upon addition of 4 mM $CaCl_2$. The leucine and isoleucine methyl peaks experienced peak narrowing with the addition of calcium ions, which meant that the rate of exchange for the protons increased as the chemical microenvironment altered.

Larger chemical shift perturbations upon addition of 4 mM CaCl$_2$ could be detected in the mutant proteins, especially in mutant C as shown in Figure 3-11. The peaks at 0.25 and 0.30 ppm display larger chemical shifts than the WT derCD23 protein in Figure 3-10.
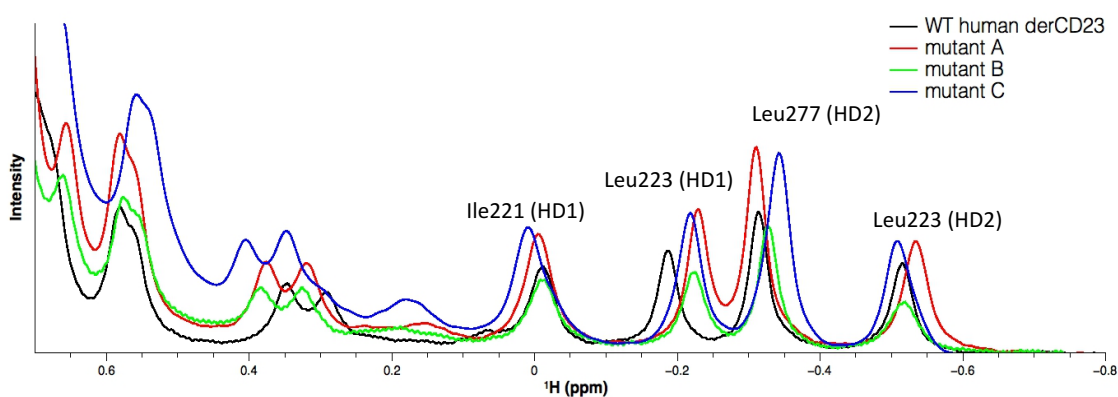
Figure 3-9 **1D-^{1}H NMR spectra of wild type and mutant derCD23 proteins in the presence of 4 mM CaCl$_2$.** Black = WT human derCD23, red = mutant A, green = mutant B, blue = mutant C. The leucine and isoleucine methyl group protons are labelled, labels taken from BMRB database entry 6732, (Hibbert, 2005).
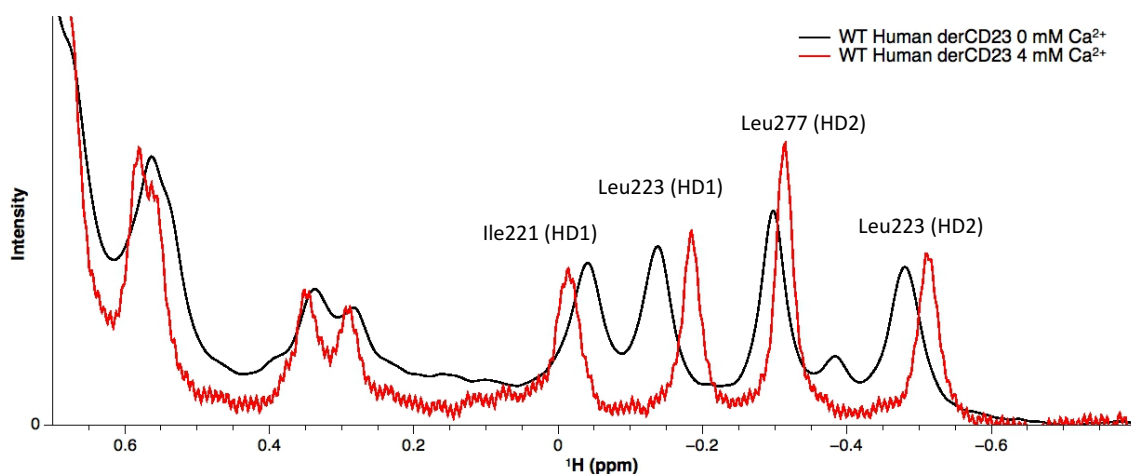
Figure 3-10 **1D-^{1}H NMR spectra of wild type derCD23 with and without calcium.** Black = zero CaCl2, red = with 4 mM CaCl2. The leucine and isoleucine methyl group protons are labelled, labels taken from BMRB database entry 6732, (Hibbert, 2005).
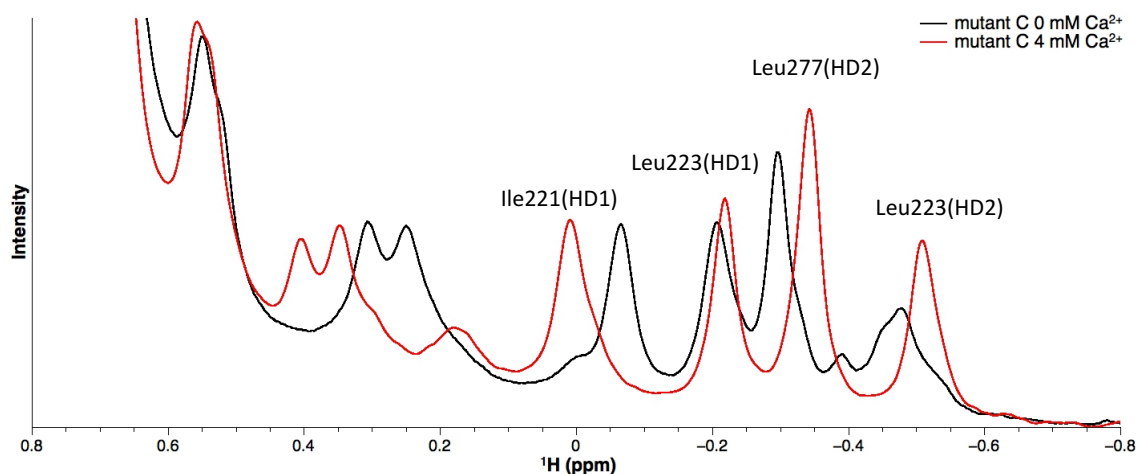
Figure 3-11 **1D-^{1}H NMR spectra of mutant C with and without calcium.** Black = zero CaCl$_2$, red = with 4 mM CaCl$_2$. The leucine and isoleucine methyl group protons are labelled, labels taken from BMRB database entry 6732, (Hibbert, 2005).

A similar experiment by SEC HPLC, in which calcium was titrated into the mobile phase, indicated a conformational change upon calcium binding for mutant C, which displayed a dramatic change in elution time as Figure 3-12. Upon addition of calcium, the hydrodynamic radius of the protein changed as the protein appeared to become smaller and more compact, since the protein eluted at a longer time point. The shape of the chromatographic peak also changed from a monodisperse single peak at 15.8 mins to two peaks spanning a wider base at 17 mins. This change was not seen in the WT or the other mutant derCD23 proteins (first three graphs of Figure 3-12) suggesting that the one-calcium bound state known for WT human CD23 could not be detected by SEC HPLC. The two peaks of mutant C might correspond to the two-calcium bound state and the apo-protein, alternatively, perhaps the second-calcium binding site in the mutant proteins is less sensitive and is unable to bind to second calcium at 2 mM Ca^{2+}. Changes such as these have also been observed in other CTLs as reported by Gingras et al., (2011), Sulmann et al. (2014) and Holla and Skerra (2011). The change in hydrodynamic radius could be explained by the calcium ion constraining the loops into a more compact conformation, limiting their movement and thus causing the protein to

become more spherical than oval-like in shape and passing through the column beads at a slower rate to elute later in time.
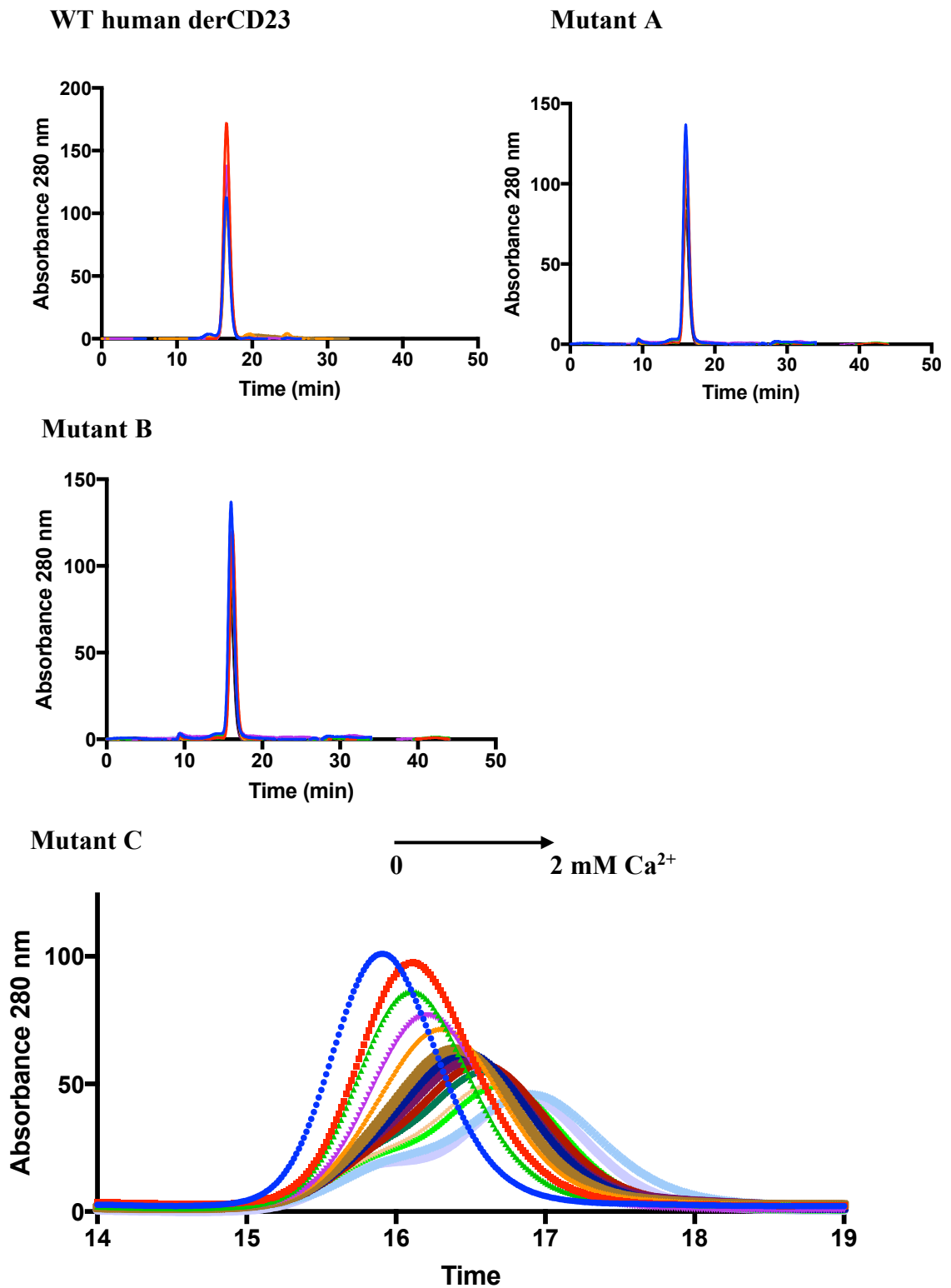
Figure 3-12 **Chromatographic profiles of WT and mutant derCD23 calcium titrations.** WT and mutant derCD23 proteins were run in calcium-free TBS on an analytical superdex 200 10/300 GL gel filtration column (GE Healthcare) at 4°C. All proteins were subjected to the same increasing step-wise concentrations of Ca^{2+} added to the sample and mobile phase from 0 (dark blue) to 2 mM $CaCl_2$ (light blue), flow rate = 0.75 ml/min.

3.2.6 Calcium binding assays

The size of a calcium ion (van der Waals radius of 0.231 nm) as a ligand, compared to a protein (~4.5 nm for derCD23) is 20 times smaller. The small size, and mass of the calcium ion (40 g/mol) means that this mass is below the detection limit for SPR. The mutant derCD23-calcium interaction could be investigated by the highly sensitive technique, Isothermal titration calorimetry (ITC), considered the gold standard for quantitative measurements of biomolecular interactions (Krainer and Keller, 2015).

The low affinity (K_D for calcium of 1.5 mM) and low protein concentration (50 μM, based on the WT human derCD23 experiments documented in Yuan et al. 2013) resulted in a low c value defined by equation 1 in section 2.3.1. 'C' is defined as the protein concentration divided by the dissociation constant, K_D. Under low c-value conditions, the isotherm binding curve is hyperbolic in shape, rather than sigmoidal as reported in section 2.3.1. In these circumstances, the stoichiometry or n value is fixed as a constant value during the fit. For the WT derCD23 protein titrated with calcium, n was fixed at 1 as was found by the X-ray crystallographic studies (Wurzburg et al., 2006; Yuan et al., 2013).

Figure 3-13A shows calcium binding to WT human CD23 with a K_D value of 0.63 mM, which is within the expected experimental range given standard errors, and using the same set-up as used by Yuan et al 2013. The ITC binding isoform curve results from plotting the analysed heat released per injection vs the molar ratio of titrant to the protein. In both figures Figure 3-13A and B, the molar ratio is quite high since the concentration of Ca^{2+} is increased so that the isotherm curve can reach some degree of saturation. Curve-fitting to the Wiseman isotherm can be difficult for systems with low

c values (please refer to Figure 2-2, Turnbull and Daranas, 2003) as such, for both the WT and mutant B derCD23 proteins, the stoichiometry value, n, was assumed to be 1 and kept constant during the curve-fitting process. Mutant B, in Figure 3-13B, bound to calcium about 3-fold tighter than WT human derCD23 with a K_D of 0.22 mM, as summarised in Table 3-1, although the affinity remained in the mM range which is typical for a CTL.
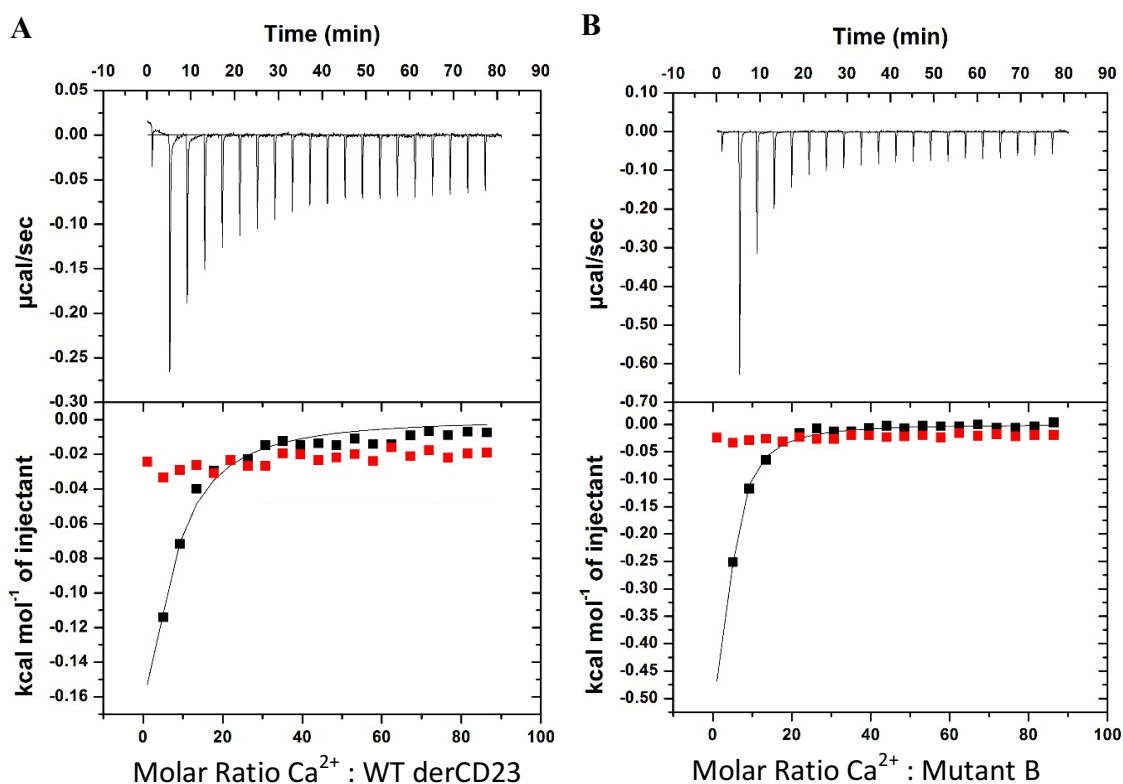
Figure 3-13 **Calorimetric titration of 20mM calcium to WT and mutant B derCD23. Top panel of each graph:** Heat generated by each injection of Ca^{2+} at each time interval. **Bottom panel of each graph:** Integration of each peak and the effective amount of heat produced. Red squares = 20 mM Ca^{2+} buffer into Ca^{2+} free buffer**.** Black curve and squares = derCD23 protein. **A)** 20 mM Ca^{2+} was titrated into 50μM apo WT derCD23. **B)** 20 mM Ca^{2+} titrated into 50μM mutant B derCD2. Both proteins measured at 25°C and fitted to a 1:1 binding event curve and plotted against the molar ratio of Ca^{2+} to protein (concentration range 0 – 20 mM).

The experiment with Mutant C was limited by the protein concentrations that could be achieved. At 500 μM of protein, or 10-fold more concentrated than in Figure 3-13, a

very different binding isoform could be seen in Figure 3-14 with at least two different binding events which were fitted to a 2:1 binding model. The biphasic shape of this curve is similar to those of proteins reported to bind to two metal ions as documented by Henzl et al. (2008), Zhang et al. (2009), and Brautigam (2015). Early in the titration at the low molar ratio, the higher affinity calcium-binding site was occupied first, followed by the lower affinity site, until saturation.

Thermodynamic parameters were measured by taking the measurements for mutant B and fixing them in a sequential binding site model. This meant that the higher affinity calcium-binding site had a similar K_D value to mutant B of 0.22 mM, as reported in Table 3-1. This binding event was enthalpically favourable with a $-T\Delta S$ of 1.30. This suggests that as the calcium ions bind to the mutant C protein, counterions and water molecules are being displaced. The second, lower affinity calcium-binding site has a K_D value of 1.10 mM, and is enthalpically unfavourable due to the positive enthalpy term. These values usually coincide with the formation of hydrogen bonds and van der Waals interactions.
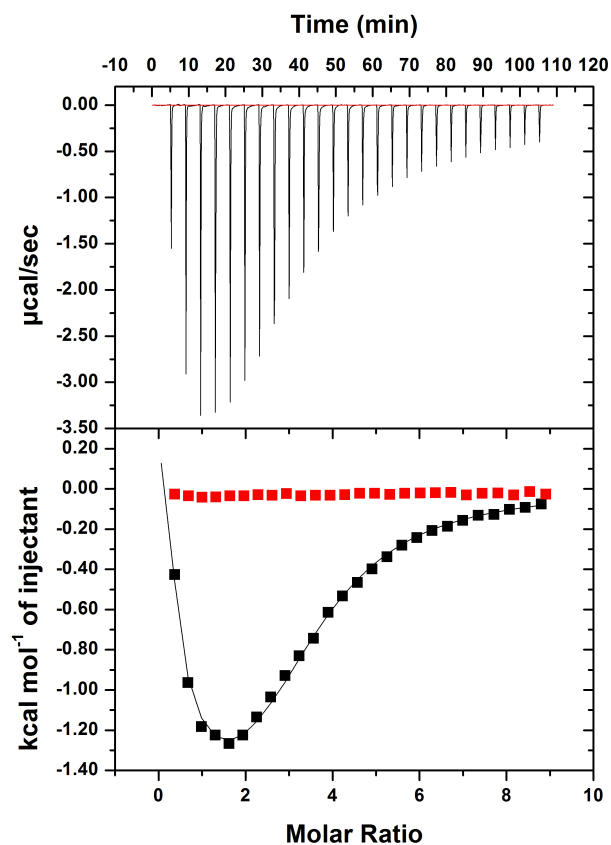
Figure 3-14 **Calorimetric titration of calcium mutant C measured by ITC. Top panel:** Heat generated by each injection of Ca^{2+} at each time interval. **Bottom panel:** Integration of each peak and the effective amount of heat produced. Red squares = 10 mM Ca^{2+} buffer into Ca^{2+} free buffer. Black curve and squares = mutant C. 10 mM Ca^{2+} titrated into 500 μM mutant C derCD23 measured at 25°C and fitted to a 2:1 binding model and plotted against the molar ratio of ligand to protein (concentration range 0 – 10mM).

Protein	WT derCD23	Mutant B	Mutant C
K_D (mM)	0.63	0.22	0.22 ± 0.04
ΔH (kcal/mol)	-2.23 ± 1.23	-2.96 ± 0.99	-6.3 ± 0.2
ΔG (kcal/mol)	-4.4	-5.00	-5.00
$-T\Delta S$ (kcal/mol)	-2.17	-2.04	1.30
K_{D2} (mM)			1.10 ± 0.20
ΔH_2 (kcal/mol)			0.95 ± 0.27
ΔG_2 (kcal/mol)			-4.10
$-T\Delta S_2$ (kcal/mol)			-5.05

Table 3-1 **The thermodynamic parameters for calcium binding to derCD23 mutants.** K_D = the dissociation equilibrium constant, ΔH = change in enthalpy (heat energy), ΔG = change in free energy, -$T\Delta S$ = temperature multiplied by the change of disorder of the system. The second half of the table refers to the 2nd phase of mutant C.

3.2.7 X-ray Crystallographic studies of the derCD23 proteins

An ideal method for confirming the presence, location and ligation chemistry of calcium to the mutant proteins would be by X-ray crystallography to determine high-resolution structures. Mutations to a protein can radically alter the crystallisation conditions for a protein, so initially commercial screens were set-up for mutant proteins B and C. The commercial screens established the initial 'hit' from which a matrix grid was performed to vary the growing conditions and further optimise the crystals. Figure 3-15 displays the optimisation process with the protein crystals growing larger each time. Potential hits were first scanned for salt detection with the PX scanner, which screens crystals with X-rays without the need to remove them from the drop. It collects the diffraction pattern of the crystal through the sealed plastic tray in which the crystals

grow, as shown in Figure 3-16.

A

B

C

500 px ~ 0.44 mm
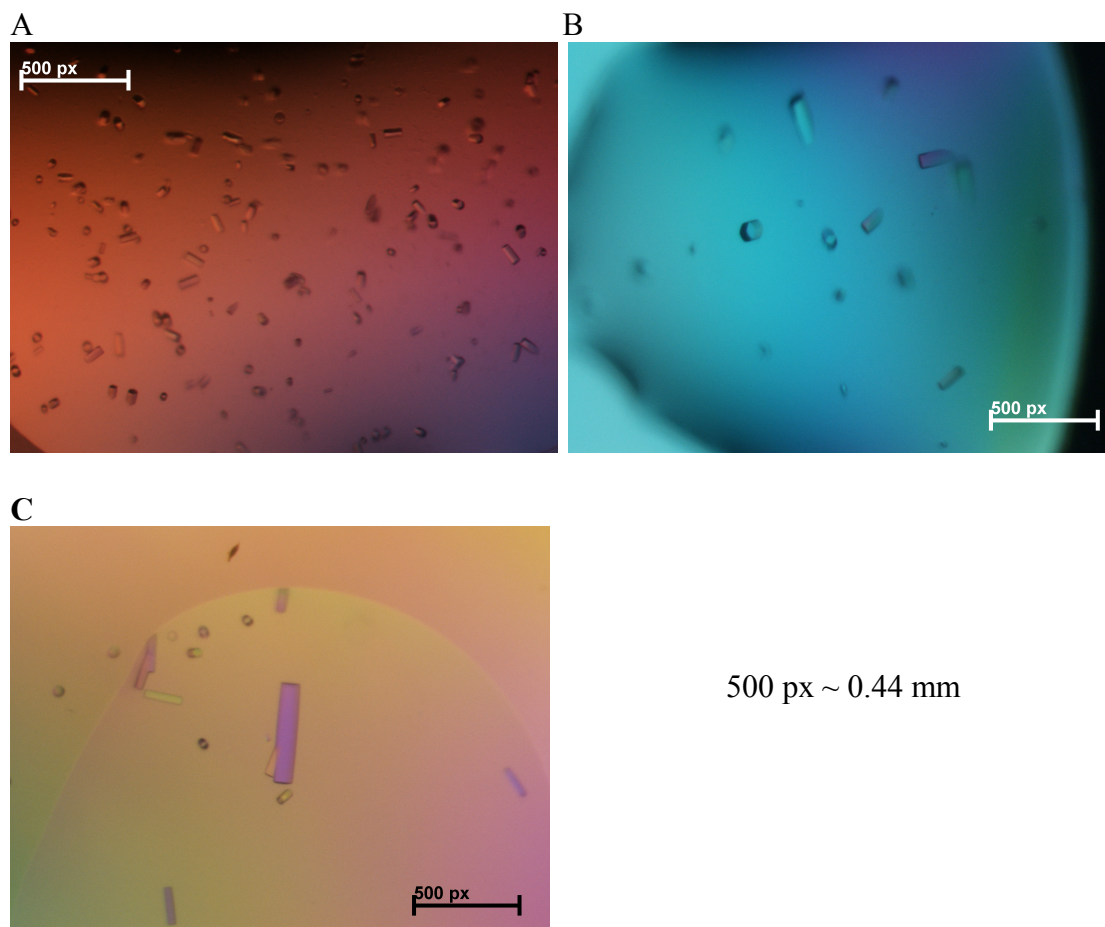
Figure 3-15 **Optimising mutant B crystallisation conditions. A –** The initial E1 hit in the commercial screen. The crystals took ~ 2 weeks to grow in conditions of 0.1 M Na-cacodylate pH 6.5 and 1 M tri-sodium citrate. **B –** These conditions formed the basis of an optimisation grid where the concentration of tri-sodium citrate ranged from 0.8 to 1.3 M and the pH of Na-cacodylate ranged from pH 6 to pH 6.75. **C** – the crystals were optimised further with a finer grid. The crystals are hexagonal rods. **A, B and C** are coloured due to a polarizer attached to the microscope; crystalline objects exhibit bifringence properties and change colour as the polarizer is rotated.

119
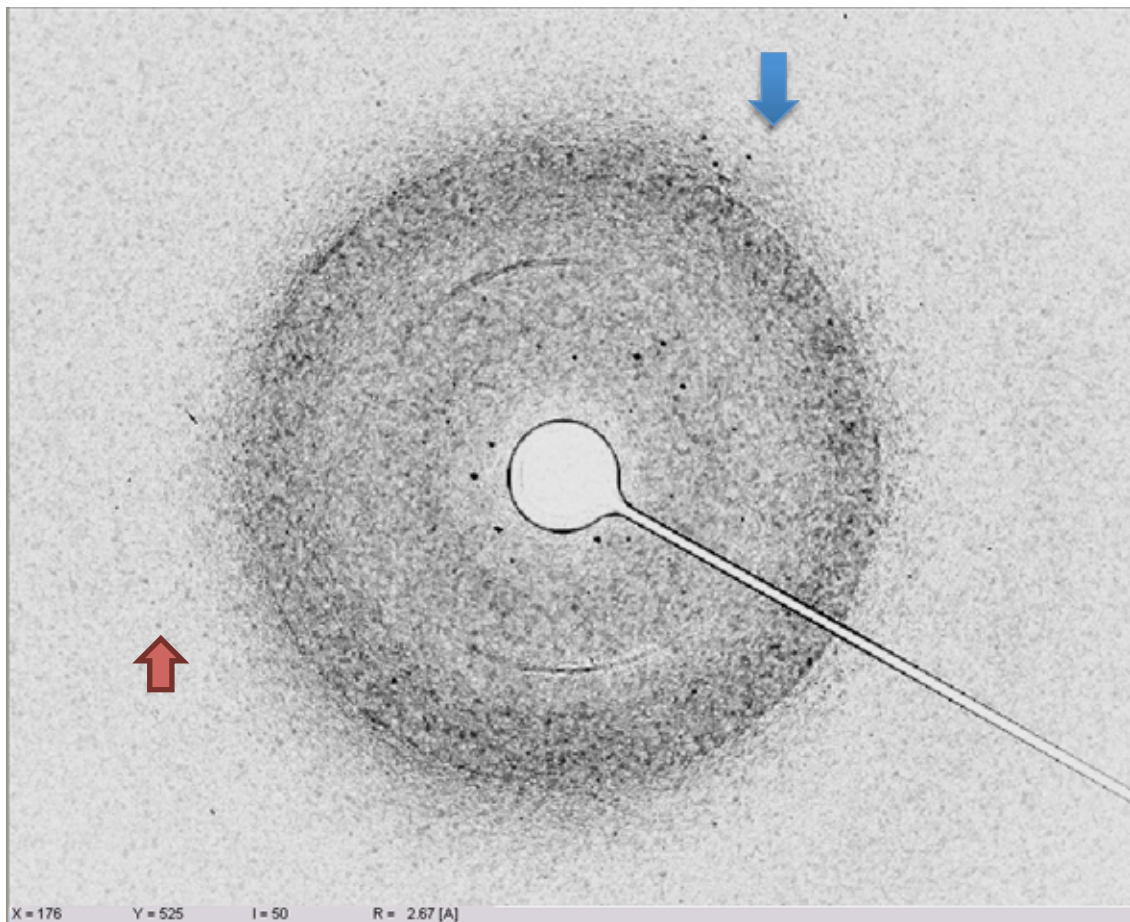
Figure 3-16 **Checking for salt crystals.** The large crystal shown in Figure 3-15**C** was tested on the PX scanner (section 2.4.3) for salt detection. A 40 min X-ray exposure has determined that it is a protein crystal causing the diffraction pattern seen. Many spots are observed at all distances from the centre, which indicates that this is a protein crystal. In contrast, a salt crystal would show relatively few, widely spaced spots. The lollipop-shaped space is the beam stop that prevents the X-rays damaging the detector. The spots furthest away from the centre give the highest resolution data; the spot indicated by the orange arrow corresponds to a resolution of ~ 2.67 Å and the blue arrow, a resolution of ~3.74 Å. It can be expected that diffraction will improve by 1-2 Å when the crystal is exposed to the more powerful X-rays at the synchrotron.

In order to obtain a calcium-bound structure, two methods were employed; one to set-up crystal trials in which the protein dispensed had been incubated with calcium, the other with calcium soaks. The protein was initially incubated with $CaCl_2$ to allow calcium to bind the C-type lectin at 4mM and then 10 mM, both concentrations which are above the K_D of the interaction. Calcium soaks were performed with three different solutions of glucose and other additives (tri-sodium-acetate and Na-cacodylate found in the

crystal plate reservoir) with calcium concentrations at 10 mM, and incubated for four days before cyro-cooling the crystals as explained in chapter 2, sections 2.4.5 and 2.4.6.

A   **B**
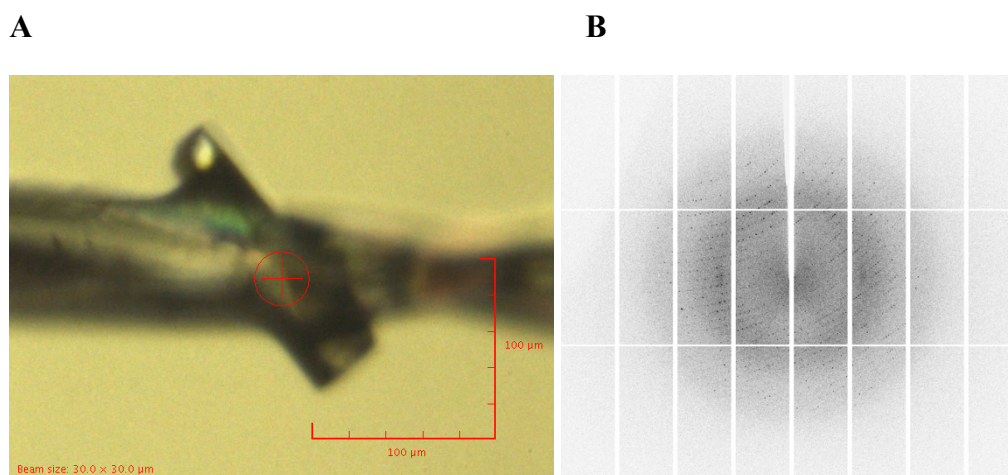
Figure 3-17 **derCD23 mutant B crystal and diffraction pattern**. **A** - The crystal shown in Figure 3-15C cryocooled and mounted on a loop. The red crosshair is the area where the beam will hit the crystal. The red labelled axis is in μm. **B** – Best diffraction pattern for mutant B, collected at Diamond Light Source, beam line I04-1; it has improved resolution to 1.37 Å from Figure 3-16.

The crystal structures of mutant B and C derCD23 were solved to a resolution of 1.37 Å and 1.89 Å respectively by molecular replacement using the WT human derCD23 protein (PDB file: 4G9A with the mutated residues removed, section 2.4.8). Differences between the WT human derCD23 structure and the mutant B structure can be seen in Figure 3-18A, which shows the mutant B structure superposed on to the WT derCD23 structure (PDB: 4G96). The overall Cα root mean square deviation (RMSD) is 0.62 Å (over 122 Cα pairs, excluding loop 1) indicating the backbone has changed little except for loop 1. Loop 1 is expected to have large movements, however, in this case, the loop has moved outward to make a crystal contact with a neighbouring molecule as seen in Figure 3-18B. This interaction is an artefact from the crystallisation process. Loop 4 is not well defined as the protein is very dynamic in this location and no electron density could be detected. Three of the four mutated residues had good electron density and could be identified for mutant B (Figure 3-18C-E). The mutation T251N had poor

electron density since it precedes loop 4, for which no electron density could be found. No calcium could be detected at either the primary or the auxiliary calcium-binding site.

Mutant C is superposed on to the WT derCD23 structure (PDB: 4G96) in Figure 3-19A, and shows a similar movement of loop 1 as the same crystal contact is formed, as depicted in Figure 3-19E. The electron densities for both proteins appear similar with good backbone and side-chain electron density identified for mutations N225D and K229E (Figure 3-19B and C). T251N had poor side-chain electron density (Figure 3-19D) and no backbone electron density was found for the S252N, R253G or S254G mutations. As with mutant B, the protein crystallised with no calcium bound.
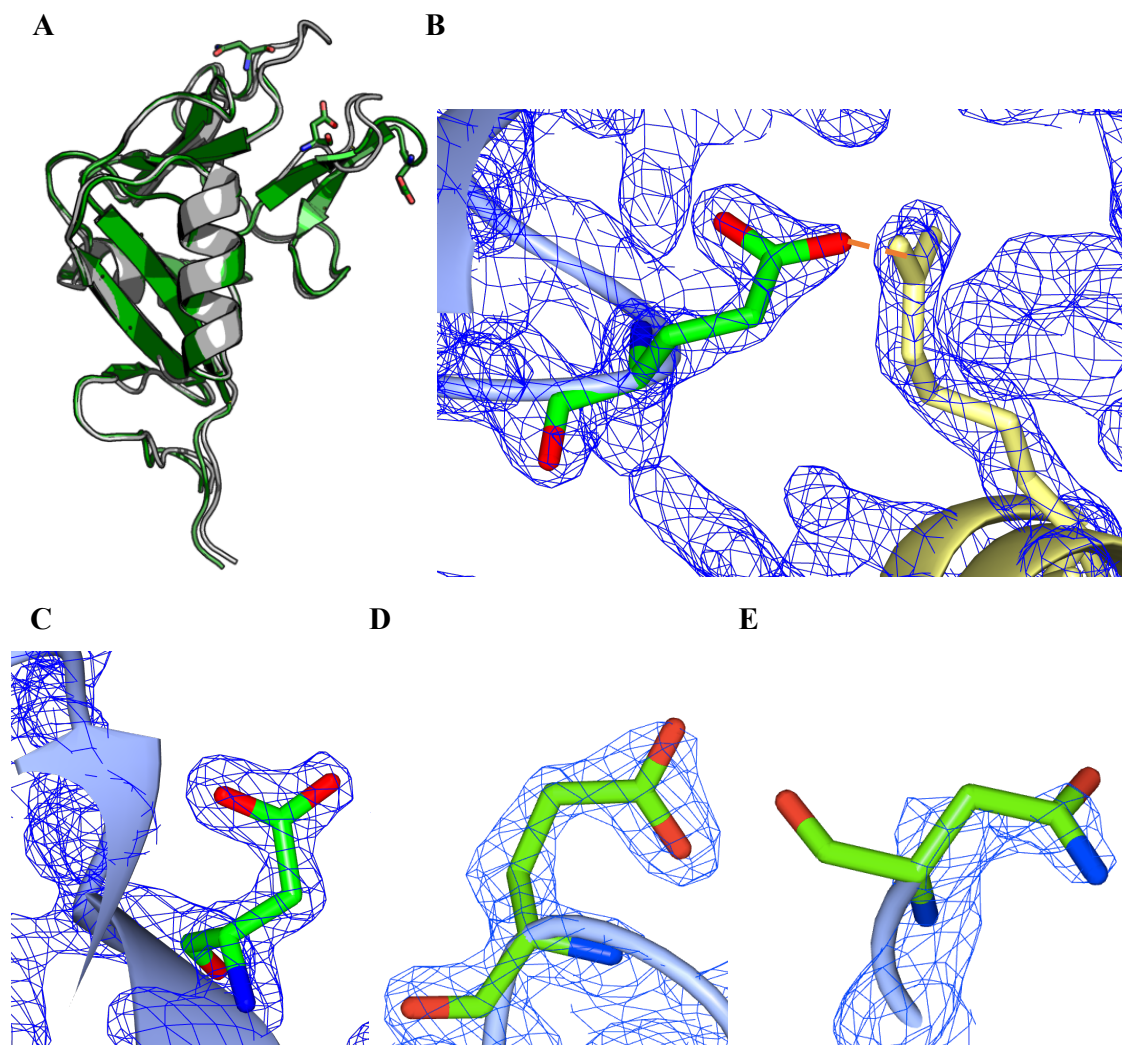
Figure 3-18 **Structure of mutant B derCD23 superposed onto WT human derCD23. A)** Green = mutant B derCD23, mutated residues are shown as sticks. Grey = WT human der CD23. **B)** New crystal contact between E229 (on newly positioned loop 1) and Arg188 of a symmetry equivalent (in yellow). **C-E)** Proof of mutations: blue chicken wire = 2Fo-Fc electron density map, light blue = protein backbone, green = carbon atoms, red = oxygens, blue = nitrogen. Contoured at 0.5 e/Å^3. **C)** Mutation N225D **D)** Mutation K229E **E)** Mutation T251N.
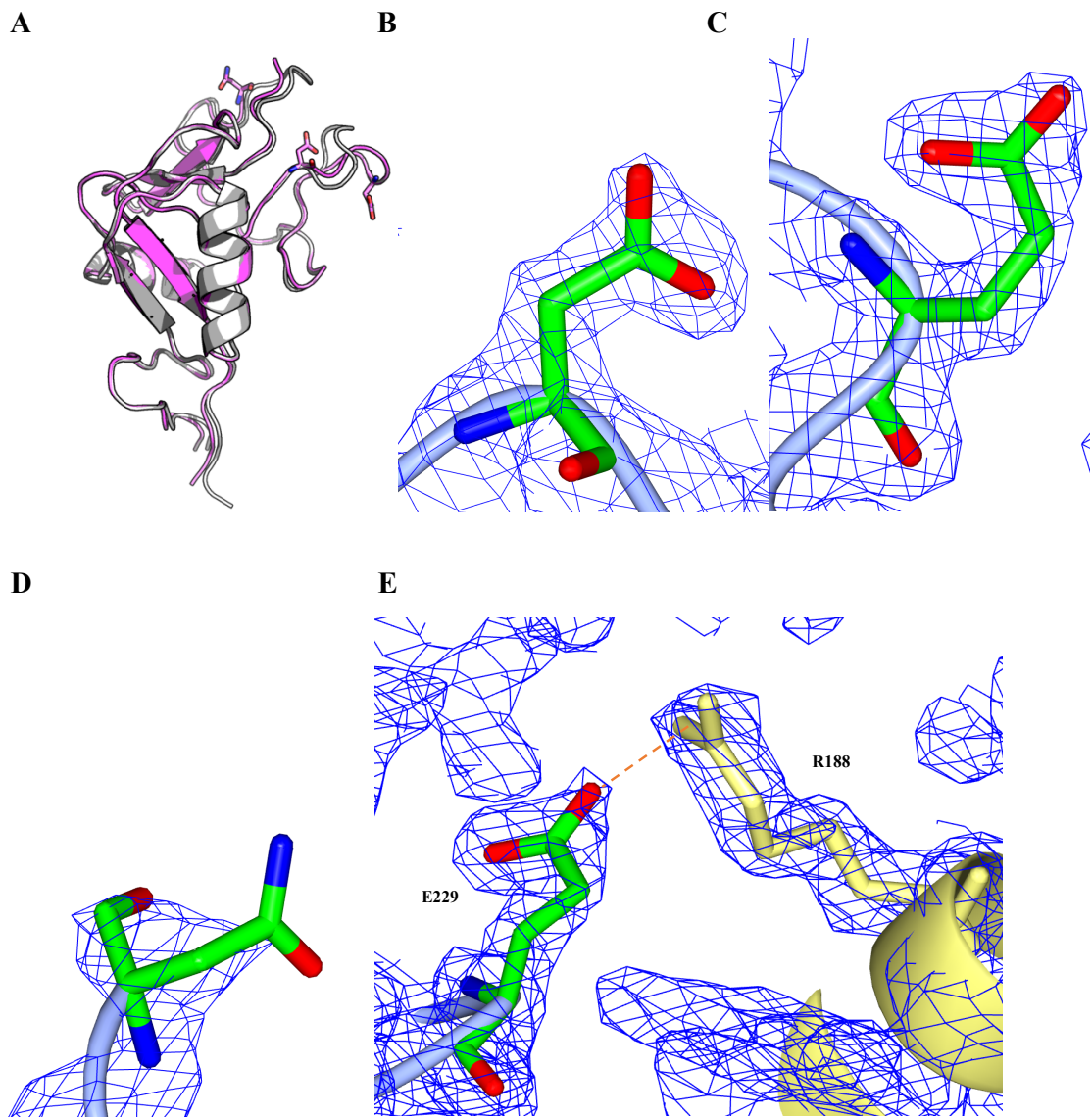
Figure 3-19 **Structure of mutant C derCD23 superposed onto WT human derCD23. A)** Pink = mutant C derCD23, mutated residues are shown as sticks. Grey = WT human der CD23. **B-D)** Proof of mutations: blue chicken wire = 2Fo-Fc electron density map, light blue = protein backbone, green = carbon atoms, red = oxygens, blue = nitrogen. Contoured at 0.5 e/Å^3. **B)** Mutation N225D **C)** Mutation K229E **D)** Mutation T251N. **E)** New crystal contact between E229 (on newly positioned loop 1) and Arg188 of a symmetry equivalent (in yellow).

Protein	Mutant B	Mutant C
Space Group	P 6 2 2	P 6 2 2
Unit Cell Parameters (Å, ˚)	a = 115.01 α = 90	a = 115.01 α = 90
	b = 115.01 β = 90	b = 115.01 β = 90
	c = 45.67 γ = 120	c = 45.67 γ = 120
No. of mol/a.u.	1	1
Resolution Range (Å)	28.75 -1.37 (1.42-1.37)	32.92–1.89 (1.92–1.89)
Total reflections	75362 (7406)	29336 (2878)
Unique reflections	37683 (3703)	14669 (1439)
Completeness (%)	0.99 (1.00)	0.98 (1.00)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Wilson B-factor (Å)	16.01	19.24
I/ σ (I)	13.08 (2.64)	19.34 (5.33)
$CC_{1/2}$	0.992 (0.562)	0.996 (0.714)
R_{means}	0.03 (0.39)	0.04 (0.46)
R_{merge}	0.02 (0.27)	0.03 (0.32)
No. of reflections used	37387 (3703)	14442 (1288)
No. of reflections in R_{free} set	1869 (167)	1445 (130)
R_{work}	0.278 (0.403)	0.184 (0.360)
R_{free}	0.273 (0.398)	0.214 (0.379)
No. of atoms	1035	1249
No. of protein residues	127	128
R.M.S bond length (Å)	0.007	0.007
R.M.S bond angle (˚)	0.98	0.93
Ramachandran Statistics (%):		
Favoured	97	96
Allowed	3	2.3
Outliers	0	0

Table 3-2 **Data Collection and refinement statistics for mutants B and C.** a.u. = asymmetric unit. R.M.S = Root mean square deviation between B-factors for bonded main chain atoms. Values in parentheses are for the outer resolution shell.

Despite various attempts to crystallise the mutant derCD23 proteins with calcium, no structures with calcium co-ordinated by the protein were obtained. This led to answers being sought using a different structural technique. The 1D ^{1}H-NMR spectra had already revealed the calcium sensitivity of the mutant proteins (Figure 3-11) and more advanced NMR techniques could disclose further information about calcium binding. For this, ^{15}N labelled protein had to be produced.

3.2.8 ^{15}N Labelled derCD23 protein expression

To perform more complex NMR experiments such as heteronuclear single quantum coherence spectroscopy (HSQCs), the protein was initially ^{15}N labelled, and then refolded and purified as mentioned previously in sections 2.2.2 and 2.2.10. To generate ^{15}N labelled recombinant protein, the *E.coli* was grown in minimal auto-induction media containing ^{15}N labelled ammonium sulphate, which provided the only nitrogen source, ensuring that ^{15}N was incorporated into the protein backbone and side-chains. An expression test of the ^{15}N protein in Figure 3-20 resembled the unlabelled protein in Figure 3-5, with the protein solubility unchanged since the protein only expressed as inclusion bodies. Figure 3-21 displayed the NMR spectrum of the WT human derCD23 ^{15}N labelled protein, which had folded correctly. The peaks were manually assigned based on the peak list in BMRB database entry 6732, (Hibbert, 2005) but re-numbered to fit the full protein sequence numbering scheme.
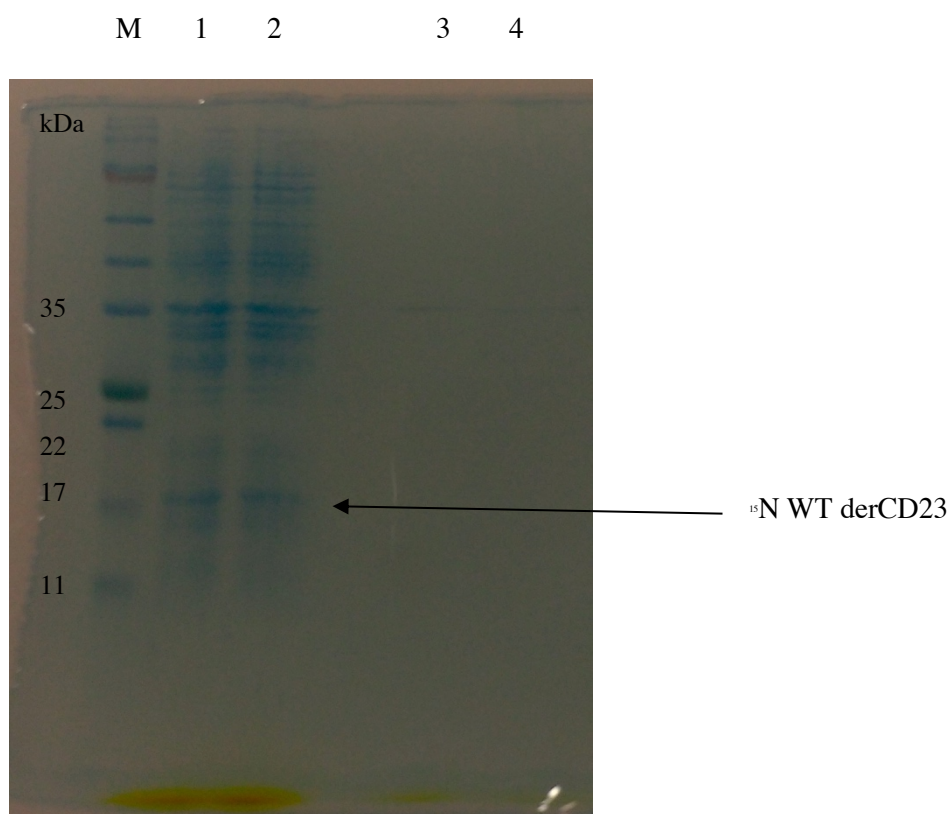
Figure 3-20 **15% SDS-PAGE of ^{15}N wild type derCD23 stained with Comassie Stain.** M = Molecular weight markers, lanes 1 and 2 = insoluble fraction, lanes 3 and 4 = soluble fraction. Expected size of ^{15}N labelled WT derCD23 = 16kDa.

A calcium titration performed with ^{1}H ^{15}N-HSQCs in Figure 3-22 revealed a few residues with significant chemical shift changes (>0.08 ppm, displaying a progression of peaks across the spectrum). The WT derCD23 spectra is in good agreement with previous experiments (Hibbert, 2005). Each set of concentric circles denotes one single amide peak, representing the hydrogen and nitrogen atoms in an amide group (NH) located mostly within the protein backbone amides, or less frequently in the amide groups of the side chains of asparagine, glutamine, tryptophan or histidine (HE).

Figure 3-21 **¹H-¹⁵N-HSQC Spectrum of WT human derCD23 with 0 mM calcium.** Peaks are labelled according to BMRB database entry 6732, (Hibbert, 2005). The assigned peaks are backbone amide proteins unless labelled with HE which corresponds to a side-chain amide.

Figure 3-22 **Calcium perturbations of WT human derCD23 captured by ¹H-¹⁵N-HSQCs. A)** Colours red through purple to grey show increasing calcium concentration where red = 0 mM, coral = 0.1 mM, gold = 0.3 mM, pale green = 0.6 mM, green = 1 mM, blue = 2 mM, purple = 4 mM, magenta = 10 mM, maroon = 25 mM and grey = 50 mM CaCl₂.

Most residues are unaffected by calcium in WT derCD23 since the peaks do not move upon addition of calcium, it is only the calcium-contact residues that display progressive peak shifts. These changes were shown more clearly by contrast with the spectrum for mutant B in Figure 3-23A, which showed more extensive changes. D283 in Figure 3-23B, is an example of a residue unaffected by calcium binding. In the mutant B protein spectrum, more peaks were affected by the binding of calcium, in particular, W184, I221, L223, W234, G237, N269 and D270. Figure 3-23C displayed the uniform movement of residue Trp234 as it experienced different microenvironments due to calcium binding. Residue 270 in Figure 3-23D is intriguing as it shows a very large proton and chemical shift change (> 0.1 ppm) and changes direction about half-way through the titration since in the first part of the chemical shift there is a greater change in the nitrogen axis while in the second half there is a greater change in the hydrogen axis. ^{15}N chemical shifts are sensitive to hydrogen bonding to the carbonyl of the preceding residue (the Ψ angle of the preceeding residue) which has less effect on the proton shift (Williamson, 2013). This effect of calcium addition on residue 270 suggests a two-step process in which the first calcium binding site of higher binding affinity binds calcium at lower concentrations causing a conformational change in the loop (a change in dihedral angle) and then the second calcium binding site is occupied at the higher calcium concentrations which affects the ^{15}N amide by hydrogen bonding interactions. The saturation binding curves for each of the identifiable residues (residues 184, 221, 223, 237, 269 and 270) with large measured chemical shifts, such as those depicted in Figure 3-23 could be grouped into 2 sets with each set roughly corresponding to the binding of a calcium ion, as shown in Figure 3-24A. The calcium binding affinities for mutant B were found to be 1.7 mM and 3.62 mM. From these results, it may be concluded that mutant B binds to two calcium ions.
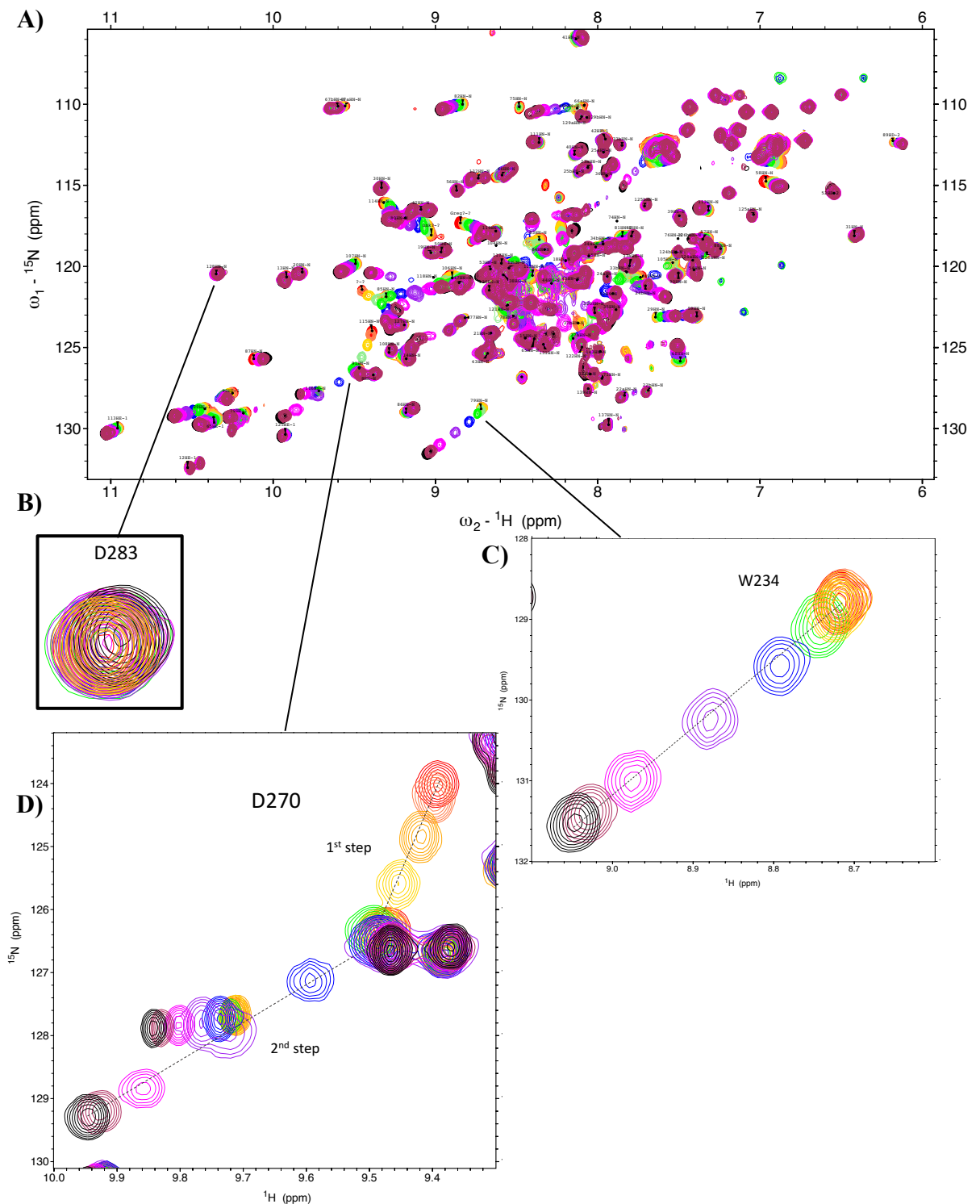
Figure 3-23 **Calcium perturbations of mutant B derCD23 captured by ^{1}H-^{15}N-HSQCs. A)** Colours red through purple to black show increasing calcium concentration where red = 0 mM, coral = 0.1 mM, orange = 0.3 mM, gold = 0.6 mM, green = 1 mM, blue = 2 mM, purple = 4 mM, magenta = 10 mM, maroon = 25 mM and black = 50 mM CaCl$_2$. **B – D) Specific residues B)** Residue 283 is unaffected by calcium binding **C)** Residue 234 is affected by higher concentrations of calcium and thus the peaks shift. **D)** Residue 270 exhibits a two-step course with the peaks initially moving downwards as the peak is sensitive to low concentrations of calcium along the nitrogen axis. At higher calcium concentrations, the peaks more equally affected in both the nitrogen and hydrogen axes and change the trajectory. This discontinuity indicates a two-step binding process.

The NMR spectra peaks that exhibit large chemical shifts ($\Delta\delta_{HN} \leq 0.1$ ppm) can be visualised on the structure as shown in Figure 3-24A. The majority of the residues with large chemical perturbations in mutant B used to determine the K_D of the auxiliary calcium binding site lie just below the proposed calcium binding site. The residues in loop 1 could not be identified in the spectra as they have not been assigned in the ^{1}H-^{15}N-HSQC, which may correspond to unlabelled peaks with chemical shifts. To test this, a 3D data acquisition leading to a ^{15}N filtered NOESY data set was obtained and the protein backbone traced to try to identify further residues, especially residues in loops 1 and 4. Unfortunately, this proved to be too difficult because the disorder of loop 1 could not be traced (data not shown).

A) **B)**
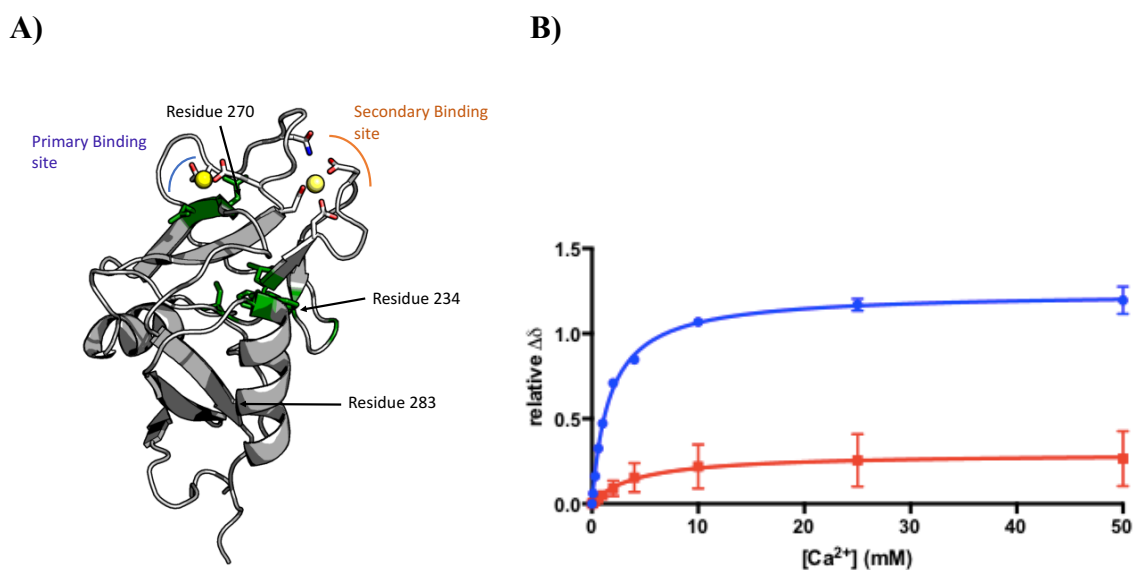
Figure 3-24 **Chemical shifts mapped onto the structure of derCD23 are used to determine calcium binding affinities. A)** Residues in green = residues in the NMR spectra of Figure 3-23 that displayed the largest chemical shifts upon titrating with calcium. Stick residues in grey = residues involved in calcium binding. **B)** Binding isotherm for mutant B based on change in chemical shift position, blue = auxiliary site, red = primary site. Results and standard deviations are calculated from residues 184 (backbone and sidechain), 221 (backbone and sidechain), 223, 234, 237, 269 and 270. The mutant B K_D values for calcium are 1.7 mM ± 0.07 and 3.6 mM ± 0.34.

Figure 3-25 displays the ^{1}H-^{15}N-HSQC spectra for mutant C which is again different from Figure 3-22 and Figure 3-23. The two-step shift of residue 270 of mutant B could not be identified in mutant C, but residue Trp234 (around 8.7, 131 ppm on the spectrum) showed a larger chemical perturbation with peak splitting occurring as well. Despite the structures of mutant B and C being identical (section 3.2.7), the six mutations in mutant C have changed local dynamics so that peaks in the spectra are harder to identify. Some peaks have a different location in Figure 3-25, and would not be identified without further analysis such as ^{15}N filtered NOESY data, which previously had not been forthcoming.
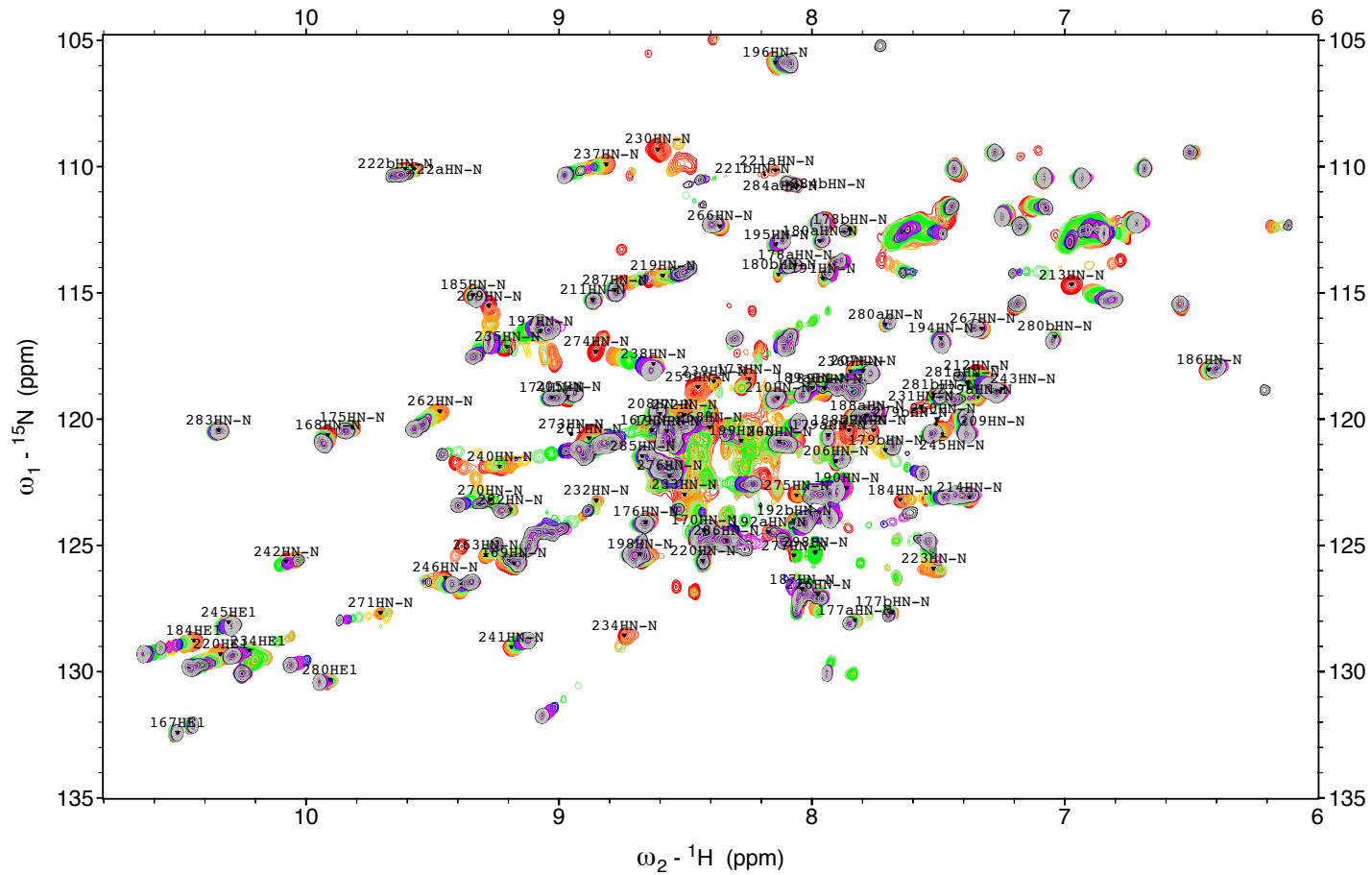
Figure 3-25 **Calcium perturbations of mutant C derCD23 captured by ^{1}H-^{15}N-HSQCs. A)** Colours red through purple to grey show increasing calcium concentration where red = 0 mM, coral = 0.1 mM, orange = 0.2 mM, gold = 0.3 mM, pale green = 0.4 mM, green = 0.6 mM, blue = 0.8 mM, purple = 1 mM, magenta = 2 mM, maroon = 4 mM, grey = 10 mM and black = 25 mM CaCl₂.

3.2.9 Carbohydrate Binding Assays

Human WT derCD23 contains an 'EPT' motif, a variation of the 'EPN' carbohydrate binding site, which usually preferentially binds to mannose in other CTLs (van den Berg et al., 2012). Both mutants B and C had the 'EPN' motif restored, and could potentially bind to carbohydrate moieties. Earlier studies have suggested that WT human CD23 can bind to both mannose and galactose (Kijimoto-Ochiai and Toshimitsu, 1995; Wurzburg et al., 2006) but other studies could not detect any binding to carbohydrates (Hibbert, 2005) as documented for WT human CD23 on the
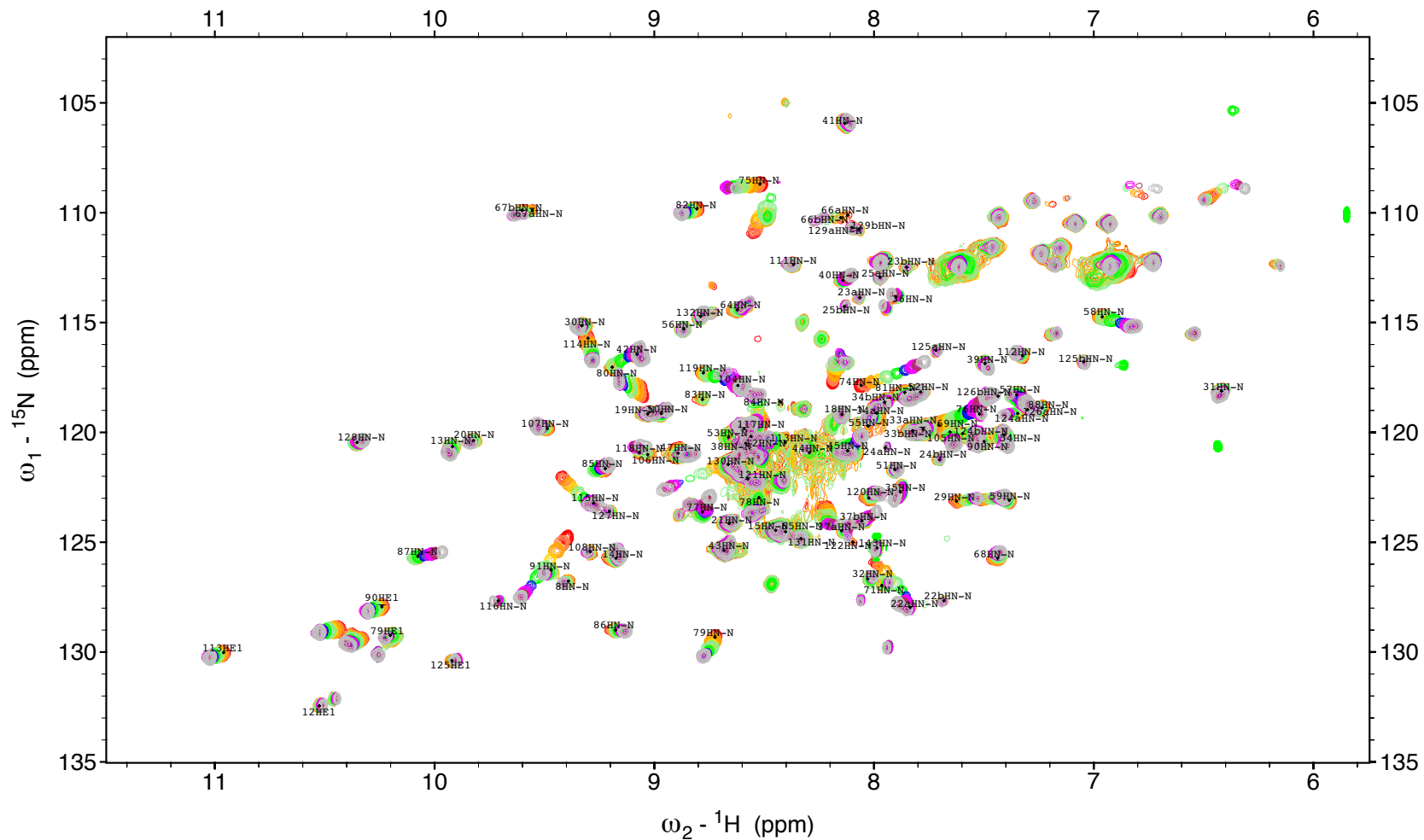
Consortium for Functional Glycomics website:

http://www.functionalglycomics.org/glycomics/publicdata/selectedScreens.jsp - Glycan Array: 634.

To test the sugar binding potential of the mutants, a mannose titration was completed in a similar way to the calcium titration. Figure 3-26 displays a ^{1}H-^{15}N-HSQC with mutant C, 25 mM $CaCl_2$ and mannose. Chemical shifts for specific residues were recognised, which differed from the chemical shifts identified upon calcium titration. This suggested specific binding as opposed to non-specific binding. Unfortunately, these peaks were unassigned residues, representing new peaks not found in WT human derCD23, and most likely belonging to loops that are better structured and less dynamic in mutant C than in the WT protein. When the chemical shifts of these unassigned residues were measured, as shown in Figure 3-27, they revealed a very weak binding affinity to mannose with a K_D of around 30 mM, which is outside the expected range for a CTL (100 μM – 10 mM) (Weis, 1997). The glycan-binding K_D is dependent on the presence of calcium in other CTLs, which is why the mannose titrations were performed in the presence of 25 mM $CaCl_2$, since calcium is vital for co-ordinating the protein side

chains for optimal glycan binding and forming direct co-ordination bonds with the glycan (Weis and Drickamer, 1996). The very weak binding of mutant C could be accounted for by the lack of an aromatic residue or a histidine in loop 4, which would facilitate hydrophobic packing against the non-polar face of a monosaccharide ring (Weis and Drickamer, 1996).
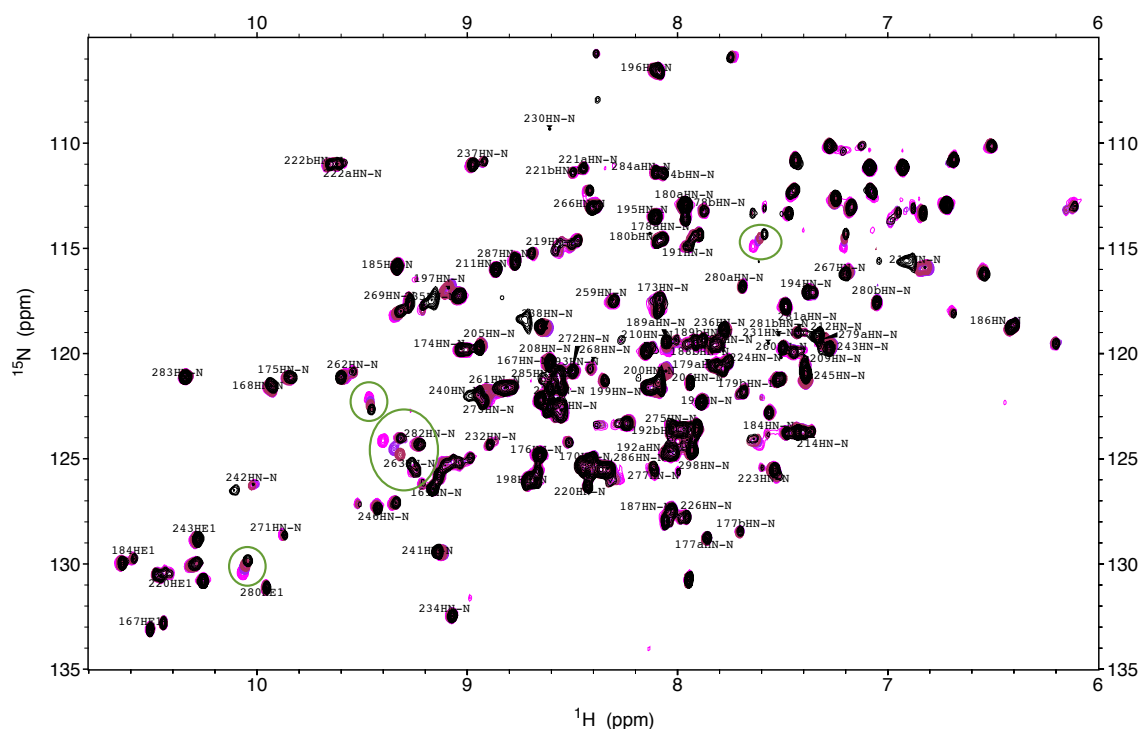
Figure 3-26 **Chemical shifts of mutant C on binding to mannose in the presence of 25 mM CaCl₂, captured by ¹H-¹⁵N-HSQC.** Colours magenta to black show increasing mannose concentrations where magenta is 0 mM mannose (and 0 mM CaCl₂), purple is 10 mM mannose and 25 mM Ca^{2+}, maroon is 20 mM mannose and 25 mM Ca^{2+} and black is 40 mM mannose and 25 mM Ca^{2+}. Green circles = residues that show chemical shift perturbations upon addition of mannose.
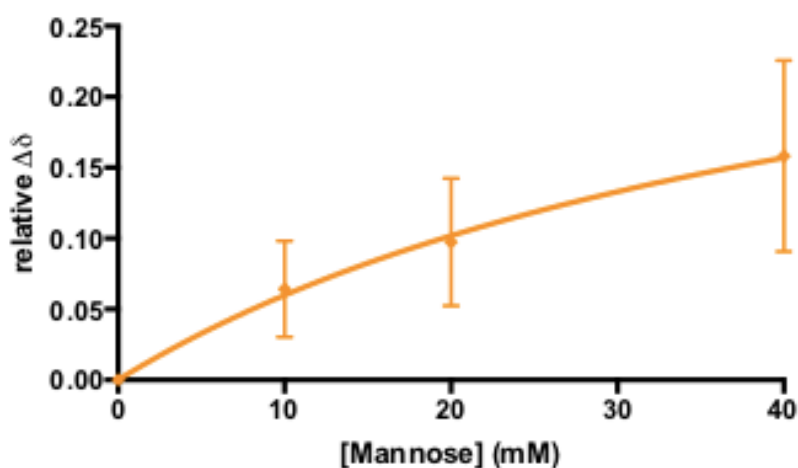
Figure 3-27 **Binding isotherm for mutant C based on changes in chemical shift position.** Orange = average of the glycan binding site. Results and standard deviations are calculated from residues which are not involved in oligomerisation, residue identities cannot be definitely identified. The proposed mutant C K_D value for mannose is 32.4 mM $\pm$ 10.2.

Testing for an individually selected monosaccharide was initially preferred to establish carbohydrate binding, but longer oligosaccharides as well as a greater variety of other monosaccharides needed to be tested with a glycan array. The RayBiotech Human Glycan Array 100 contains a range of glycans including N-linked glycans found on proteins such as antibodies, milk glycans, Lewis antigens and smaller units of various monosaccharides which could be found on pathogens, for testing in parallel. The glycans are printed onto a glass slide in quadruplicate via chemical linkers. The derCD23 proteins were first prepared by biotinylating them, which allowed for glycan binding to be detected via a Cy3 labelled streptavidin. Figure 3-28A displays the results of the glycan array without calcium. The strongest binding was to the positive control, which consists of a standardized concentration of biotinylated Immunoglobulin G printed directly onto the array (in quadruplicate) and is represented by the highest bar on the furthest right of Figure 3-28A and Figure 3-29B as expected. The negative control (to the right of the positive control) is protein-containing buffer printed onto the array (protein undisclosed by the manufacturer).

In Figure 3-28A, the overall pattern of derCD23 proteins binding to glycans can be observed. The largest increase in binding was seen in the aminoglycosides (glycans 8, 96, 98 and 99). The first peak, corresponding to peak 5, is the monosaccharide Rhamnose (glycan 5), which only mutant B could bind to. Peak 10 is discounted due to high fluorescence intensity from the water control. Figure 3-28B zooms in on the lower fluorescence intensities, which reveal other glycan binders above the threshold of significance, which is above background level and has an increase in fluorescence signal intensity of $\geq$ 1.5-fold. These include lower peaks such as 56 and 97, which are also aminoglycosides. Figure 3-28B also shows how only the mutant derCD23 proteins were capable of binding to oligosaccharides 29 and 69. The disaccharides 4-P-GlcNAc-β-1,4-Man-β-Sp (glycan 28) and Glc-α-1,2-Gal-α-Sp (glycan 48) also resulted in minor binding with the mutant derCD23 proteins only, as listed in Table 3-3. Table 3-3 has grouped the identified glycans into strong and weak binders dependent on their fluorescence intensity signals. For a full list of the glycans in this array, please refer to Appendix B.

Figure 3-29A shows the same glycan array repeated in the presence of 4 mM calcium. Mutant B appears to bind to many more glycans, although little difference from Figure 3-28 to Figure 3-29 could be discerned with the WT human derCD23 and mutant C proteins. The largest increase in binding for mutant B was seen in the aminoglycosides (glycans 8, 56, 96, 97, 98 and 99, which represented binding with more than 50 normalised fluorescence units). It must be noted that the manufacturer could not confirm whether the buffer was PBS free – this was a cause for concern since calcium in the presence of PBS forms calcium phosphate and causes proteins to precipitate (private correspondence with manufacturer).

Aminoglycosides are a class of antibiotics derived from the *Streptomyces* genus used in treating Gram negative infections. Aminoglycosides are polycations at physiological pH, and might have bound to a positive patch found on derCD23 where lysines (non-conserved with respect to other CTLs), and arginines clustered. The specificity of the binding is not clear, however, aminoglycosides have bactericidal activity against mycobacteria, and CD23 is known to mediate antimicrobial activity in human macrophages (Mossalayi et al., 2009).
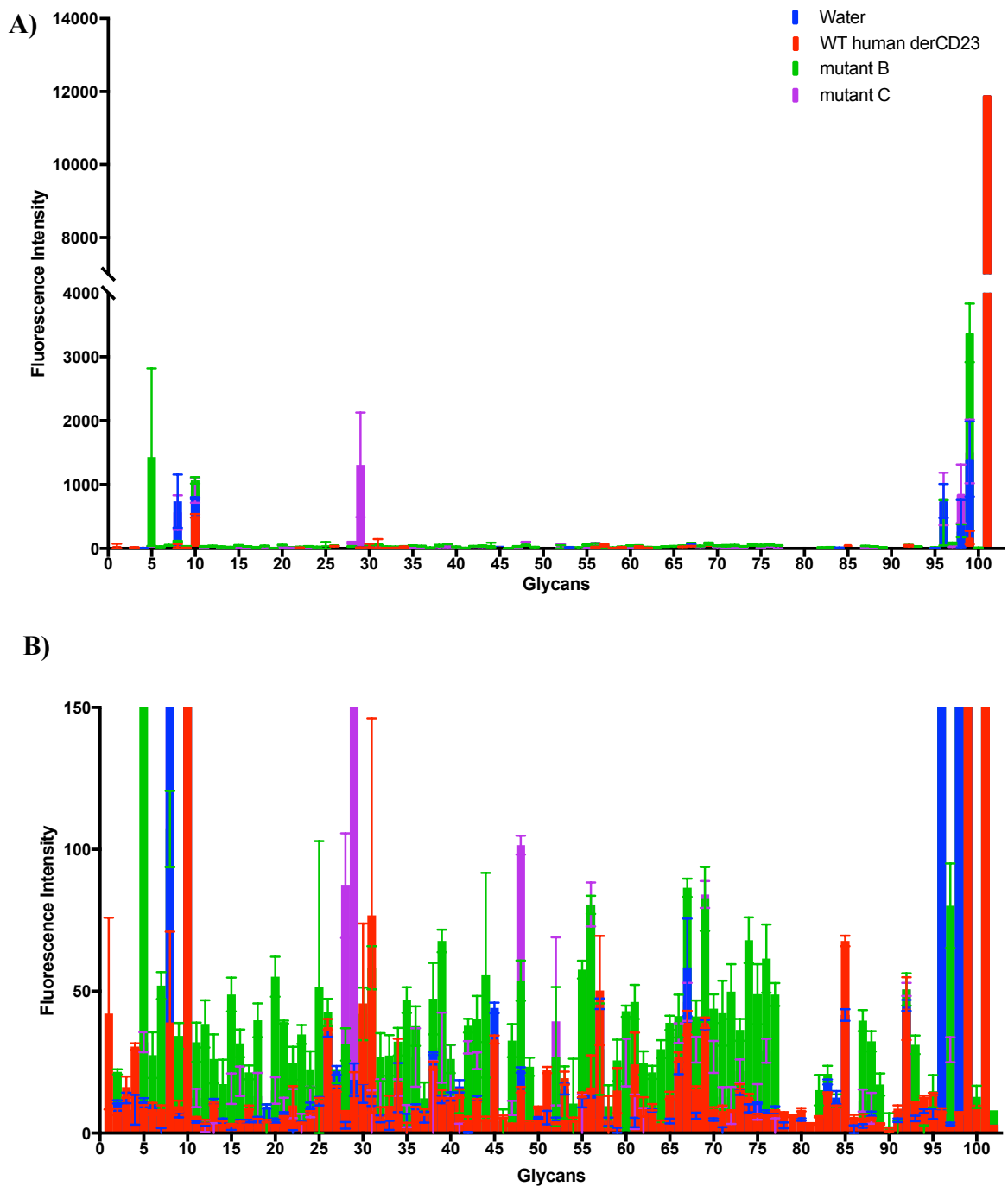
Figure 3-28 **Glycoarray analysis of binding specificities of WT human derCD23 and mutants B and C without calcium.** Blue = water (control), red = WT human derCD23, green = mutant B, purple = mutant C. Each of the 100 glycans is represented four times on the printed array. The positive and negative controls are shown at the end of the graph (glycan 101 and 102). The data is sorted and averaged before subtracting the background signal and normalising the data with respect to the positive control. The average binding to each glycan is shown as fluorescence intensity ± standard error of the mean. **A)** The complete graph **B)** Focusing on low fluorescence intensity of 0 – 150.
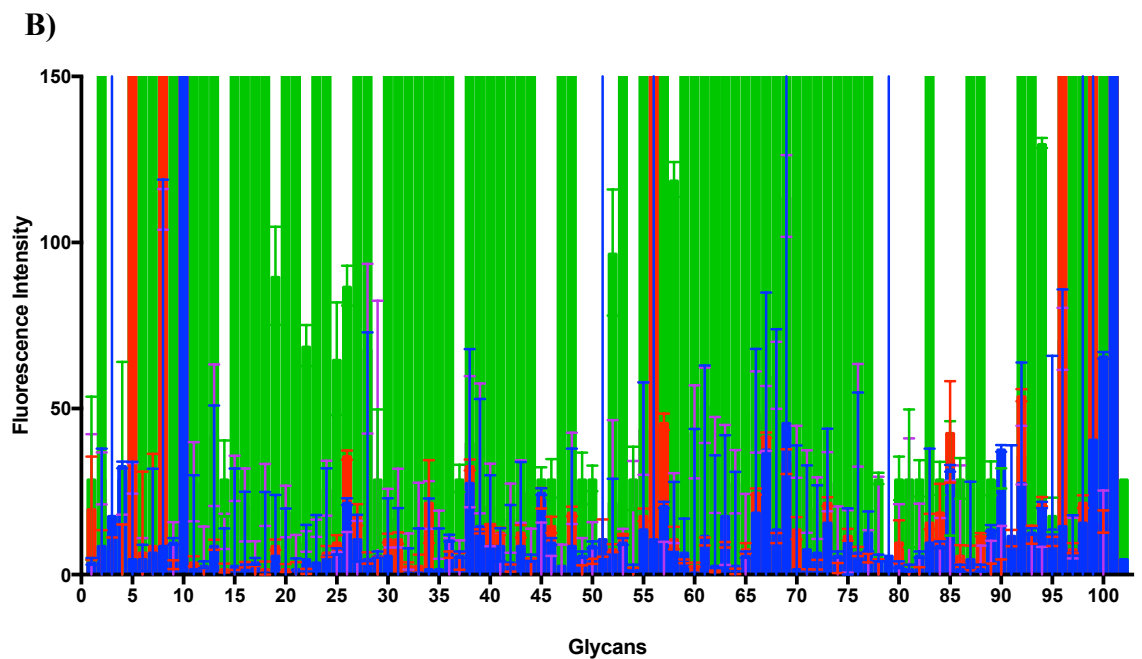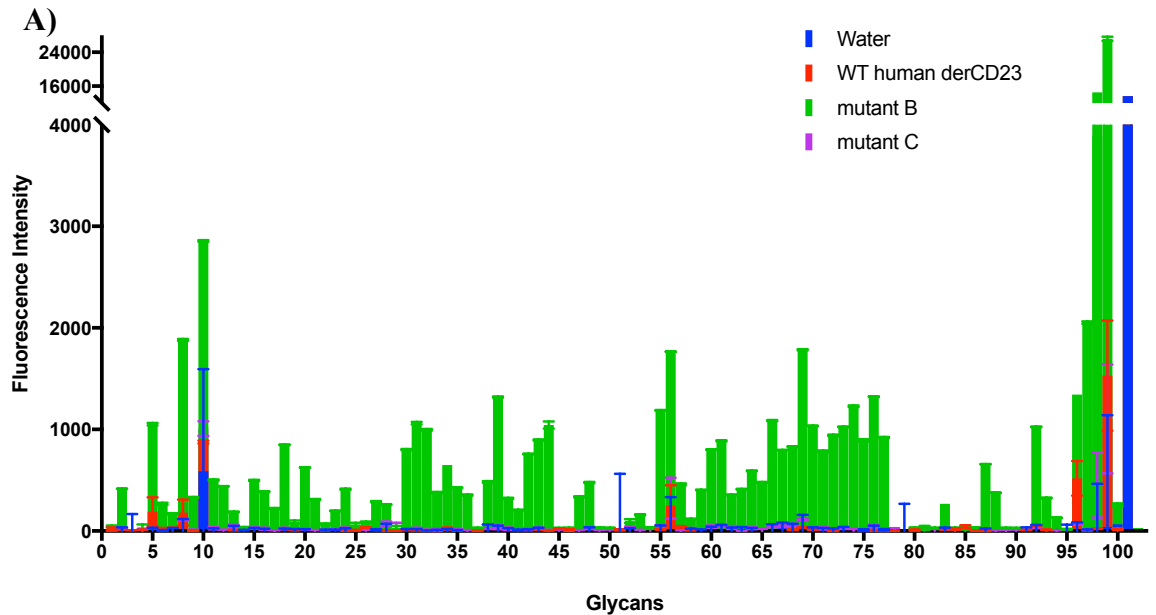
Figure 3-29 **Glycoarray analysis of binding specificities of WT human derCD23 and mutants B and C in the presence of 4mM CaCl₂.** Blue = water (control), red = WT human derCD23, green = mutant B, purple = mutant C. Each of the 100 glycans is represented four times on the printed array. The positive and negative controls are shown at the end of the graph (glycan 101 and 102). The data is sorted and averaged before subtracting the background signal and normalising the data with respect to the positive control. The average binding to each glycan is shown as fluorescence intensity ± standard error of the mean. **A)** The complete graph **B)** Focusing on low fluorescence intensity of 0 – 150.

Glycan no.	Name	Structure	WT	B	C
99	Neomycin trisulfate		++	+++	++
98	Geneticin Disulfate Salt (G418)		+	+	++
97	Kanamycin sulfate			+	
96	Gentamicin Sulfate		++	+	++
69	Neu5Ac-α-2,6-Gal-β-1,3-(Neu5Ac-α- 2,6)-GalNAc-β-Sp			+	+
56	Sisomicin Sulfate			+	+
48	Glc-α-1,2-Gal-α-Sp			+	+
29	Glc-α-1,2-Gal-α-1,3-Glc-α-Sp				++
28	4-P-GlcNAc-β-1,4-Man-β-Sp				+

Continued on next page

142
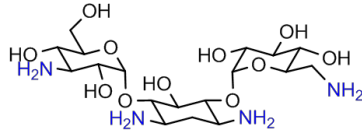
Glycan No.	Name	Structure	WT	B	C
08	Tobramycin		++	+	+
05	Rhamnose				++

Table 3-3 **The major glycan binding partners in the carbohydrate array.** Glycans bound by mutant derCD23 proteins with more than 100 normalised fluorescence units. Sp = linker. Glycan Symbol chart displayed below. +++ =< 1500 normalised fluorescence units, ++ =< 600, + =< 80.

Glycan Symbol

3.2.10 The effect of the derCD23 mutations on IgE-Fc3-4 binding

Surface plasmon resonance (SPR) was utilised to understand how the mutations altered the binding affinity of the CD23 mutant proteins to IgE in the presence, and absence of calcium by measuring the equilibrium dissociation constants (K_D) as illustrated in Figure 3-30A. To ensure a 1:1 binding ratio of derCD23 to Fcε3-4, the derCD23 proteins were immobilised to a CM5 chip by amine coupling at a low density of between 70 – 100 resonance units (RU). At high ligand densities, the shorter distance between derCD23 molecules meant that the homodimer Fcε3-4 could bind to two molecules of derCD23, which could ultimately complicate the analysis. K_D and B_{max}

143

were two main values extracted from the binding experiments. B_{max} is defined as the maximum signal when all the ligand binding sites are occupied by the analyte. When the derCD23 mutants were immobilised on the chip, the Bmax changed from 74.96 RU for the WT to 38.17 and 40.23 RU for the mutant proteins as summarised in Table 3-4, (sensorgrams displayed in Figure 3-30B-D), suggesting that maximum binding capacity had changed. This could be due to the coupling reaction, since amine coupling is considered to be random and would immobilise the protein in different orientations or even immobilise the protein in such an orientation in which the binding site is no longer exposed.

Protein	K_D (µM) with 4 mM Ca^{2+}	B_{max} (RU)
WT derCD23	0.40	75
Mutant A	0.82	40
Mutant B	0.83	40

Table 3-4 **Summary of the SPR fitting results of Fcε3-4 binding to immobilised derCD23 proteins in the presence of 4 mM Ca^{2+}.** Summary of results displayed in Figure 3-30. B_{max} = maximum binding capacity.
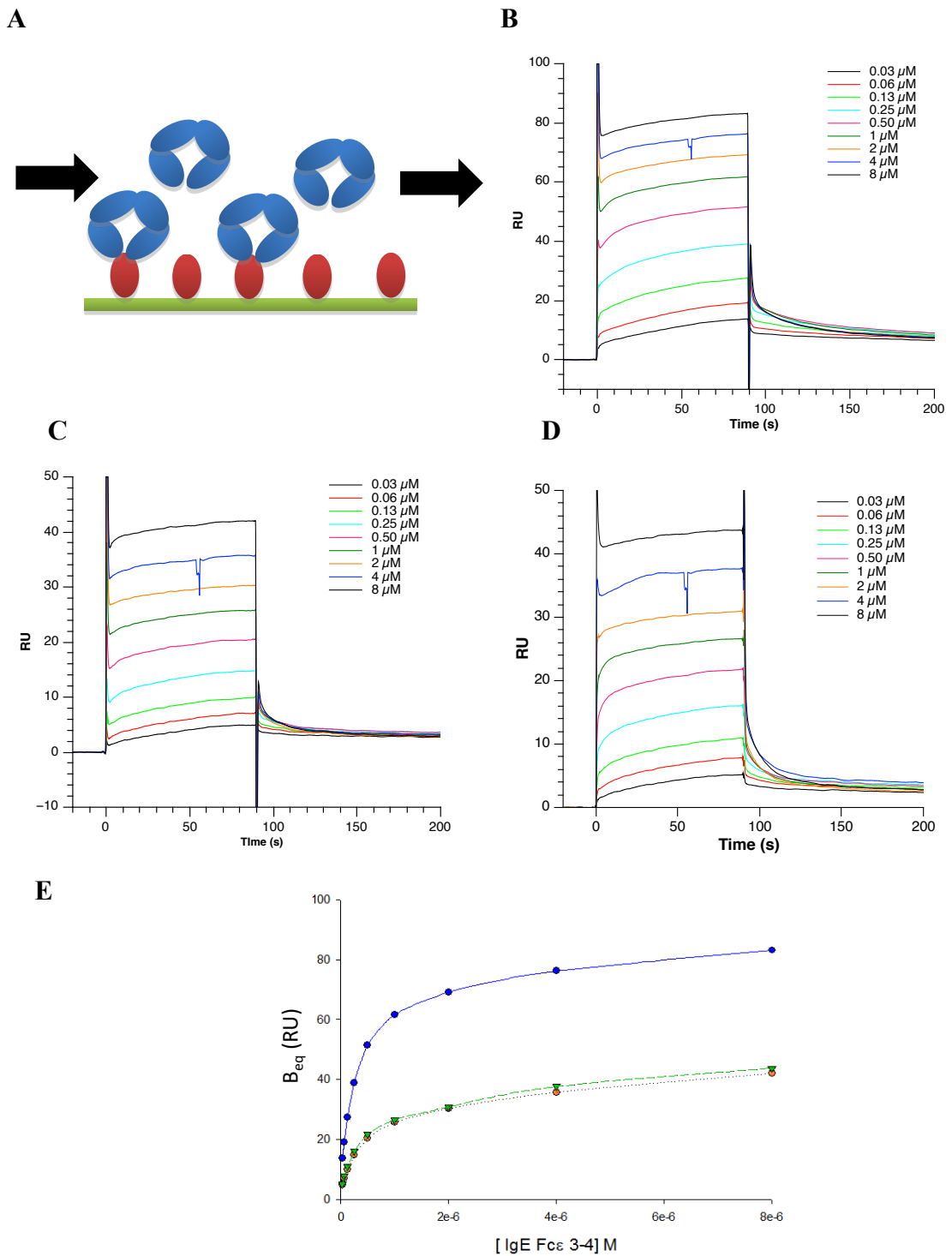
Figure 3-30 **Steady State SPR analysis of Fcε3-4 binding to immobilised derCD23 proteins in the presence of 4mM calcium. A)** A diagram of the experimental set up where derCD23 (red) is covalently captured on the sensor chip (green) and the analyte, Fcε3-4 (blue) is flowed over the surface of the sensor chip, the arrows indicate the direction of the analyte flow. **B-D)** SPR sensorgrams showing the binding of Fcε3-4 to immobilised derCD23 proteins determined over a range of ligand concentrations from 30 nM to 8 μM. **B)** WT derCD23 **C)** mutant A **D)** mutant B. **E)** Steady state binding curve generated from the sensorgrams. Blue = WT derCD23, green = mutant A, orange = mutant B. The K_D is the concentration of analyte at which 50 % of the ligand binding sites are occupied.

Protein	K_D (μM) with Ca^{2+}	Rmax (RU)	Chi^2 (RU^2)
WT derCD23	0.97	94.96	1.90
Mutant 3a	0.823	39.17	2.20
Mutant 3b	0.833	40.23	3.28

To alleviate the problem, the experiment was performed with an alternative set-up in which biotinylated IgE-Fc was immobilised to a streptavidin chip at a high and a low density and the prepared derCD23 proteins were the analytes to be flowed over the surface. Figure 3-30 displays the equilibrium binding graph and Table 3-5 displays a summary of the values obtained from the experiment. Briefly, in the presence of 4 mM $CaCl_2$, the K_D value altered in a peculiar way; initially the least murine-like protein, mutant A, bound 3-fold less strongly to IgE-Fc, however, a trend then emerges in that the more mouse-like the protein becomes, the stronger the binding to IgE-Fc so that mutant C has a similar K_D value as the WT human CD23. This accompanied a change in the B_{max} value for the mutants indicating that most of the mutant protein molecules were unable to bind to IgE-Fc, but a small subset was in a conformation that was able to bind.

Protein	K_D (μM)	B_{max} (%WT)
With 4 mM Calcium		
WT human CD23	0.71	100
Mutant A	2.80	72.4
Mutant B	1.77	40.3
Mutant C	0.33	5.37
Without Calcium		
WT human CD23	22.5	100
Mutant A	36.0	100
Mutant C	24.0	34.5

Table 3-5 **Summary of the SPR fitting results.** B_{max} = maximum binding capacity.
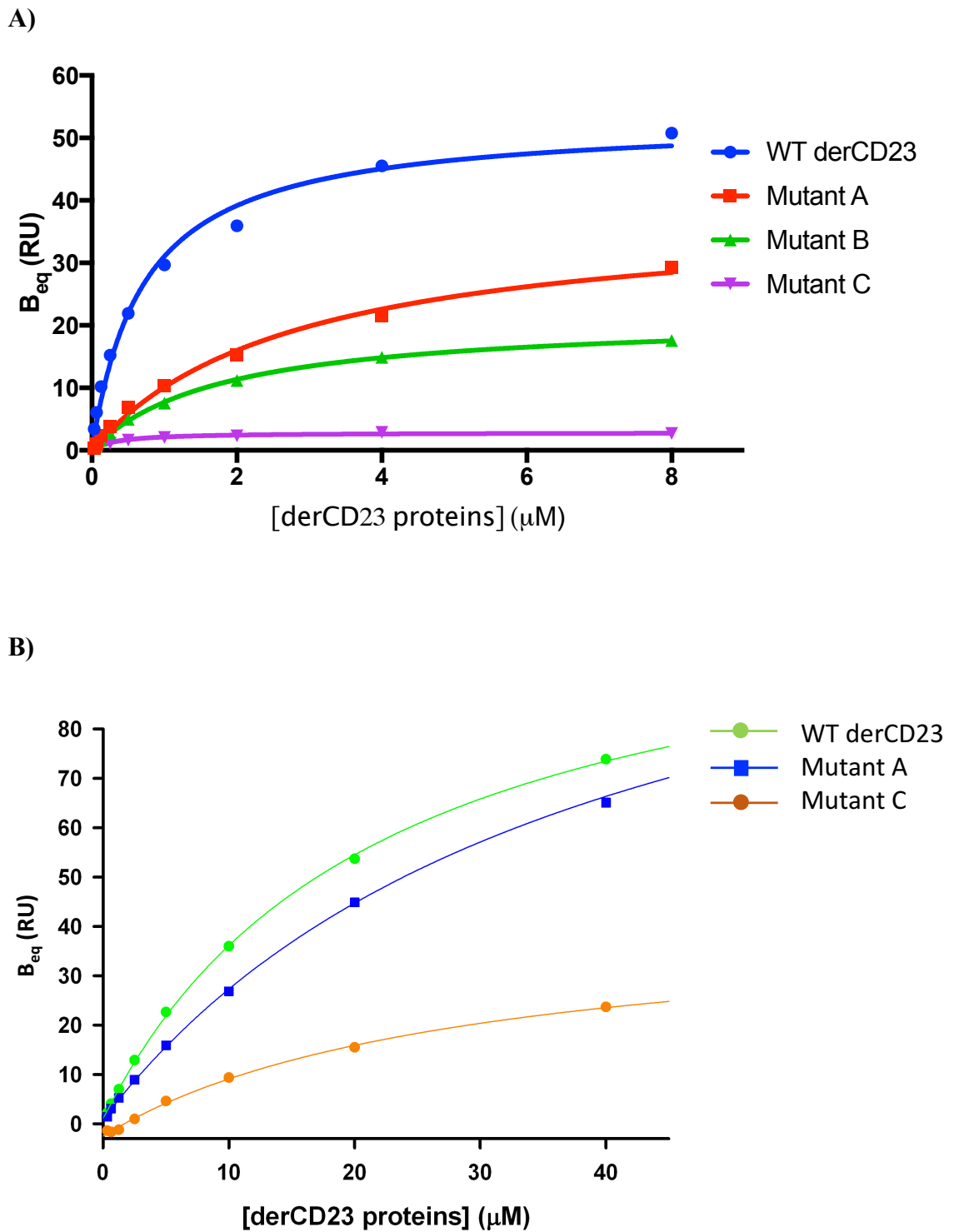
Figure 3-31 **Determination of the equilibrium constants of Fcε3-4 binding to derCD23 proteins in the presence and absence of calcium.** Points taken from equilibrium phase of the association for each concentration are plotted against the range of Fcε3-4 concentrations. **A)** In the presence of 4 mM Ca^{2+} in HBS. **B)** In the absence of calcium, 10 mM EDTA in HBS running buffer.

3.3 Discussion

To understand why human CD23 has lost its calcium-binding abilities in its "auxiliary" calcium-binding site, and how this might affect binding to its most important ligand, IgE, the auxiliary calcium-binding site was restored by designing a panel of mutant derCD23 proteins. Residues mutated were chosen based on an alignment with mouse CD23, an animal crucial to *in vivo* experiments. This protein panel had binding affinities for both calcium and IgE measured by a range of biophysical techniques. Greater understanding of why human CD23 distinctively binds to only one calcium ion may shed further knowledge into the mechanisms that induce allergic sensitisation in humans.

Mutant B clearly demonstrated two calcium-binding sites as recorded by NMR experiments (Figure 3-23). One calcium-binding site in mutant B produced a similar binding affinity to the WT derCD23 primary calcium-binding site of 1.7 mM, whilst the other binding site exhibits 2-fold weaker binding at 3.6 mM. This is in a similar range found for CTLs (Weis, 1997). Zelensky and Gready state that from an evolutionary point of view, for CTLs, the calcium-binding sites do not act cooperatively despite being spatially close to each other (Zelensky and Gready, 2005). It is unclear if mutant A has two calcium binding sites, since biophysical studies on the protein were left incomplete due to it expressing poorly, and focus was shifted to the other two mutant proteins.

Measuring mM binding affinities posed a challenge as most techniques are not sensitive enough, however, it is important to keep in mind that weak binding of calcium plays a vital role for the functioning of CTLs. Mutant C, despite not having displayed the clear two-step progression of residue 270 mutant B in the NMR titrations, also indicates two-

calcium binding sites as suggested by the unusual HPLC trace as shown in Figure 3-12, and the two events in its thermodynamic signature in ITC displayed in Figure 3-14. This ITC experiment is unfortunately limited by derCD23 solubility; if the concentration of the protein exceeds 500 μM the viscosity of the sample makes it too difficult to load into the apparatus, and the poor stability of the protein would cause it to precipitate.

The slight differences in calcium affinities measured by NMR and ITC arise from the techniques measuring different properties (Table 3-1 and Figure 3-24). Despite both being solution techniques that require no chemical modification or conjugation, ITC focuses on the thermodynamic parameters (the macroscopic binding details) whereas in NMR, only certain protein residues that display chemical perturbations are focused on (the microscopic affinities).

Ongoing crystallisation efforts are trying to reveal greater insight into the calcium co-ordination of the mutant proteins. Co-crystallising the protein with various concentrations of $CaCl_2$ (4 mM – 10 mM $CaCl_2$) added to the protein buffer 24 hr prior to distribution in the crystallisation plates did not result in a calcium-bound form of the mutant proteins. Although the calcium soaks did not work, they did not destroy the crystal as the solution passed through the solvent channels of the crystal. This suggests that lattice packing interactions were stronger and would not accommodate calcium-binding. The crystal of derCD23 mutant B can be seen in Figure 3-17 with the collection data and refinement statistics described in

Table 3-2. In the body, human WT derCD23 would usually bind to calcium in the neutral extracellular milieu, and calcium would only be expected to dissociate in an acidic endocytic vesicle. The mutant proteins crystallised in slightly acidic pH conditions of pH 6 – 6.75, which is equivalent of the environment found in early

endosomes, whereas late endosomes have a pH of 4.5-5.5 and this is where ligand dissociation occurs. Despite larger crystals growing at pH 6.5, finding conditions closer to physiological neutral pH may have resulted in calcium-bound structures. An alternative to calcium, would be to substitute lanthanide ions which still leave the CTL functional (Weis et al., 1991). If crystallisation conditions could be optimised, then structures binding to glycans would be equally interesting.

Restoring the auxiliary calcium binding site also re-instated the 'EPN' carbohydrate binding motif and this enabled the mutant proteins to bind to several glycans. Except for the aminoglycosides found in the glycobiology consortium arrays, WT human derCD23 was not found to bind to any glycans, in accordance with the findings of Hibbert and Jardetzky (Hibbert, 2005; Wurzburg et al., 2006). Perhaps, as Drickamer (Drickamer, 1999) suggests, although CD23 may be a CTL, its functions diverged earlier than other carbohydrate binding proteins, and it never developed the extra structural features on the exterior of the protein to facilitate glycan binding. (Drickamer and Taylor, 2015; Wurzburg et al., 2006).

Mutant B, however, binds to several glycans containing glucose, rhamnose and galactose residues on the array; potentially mutant B might be recognising the orientation of the hydroxyl groups on the pyranose ring structure, since D-galactose has the same configuration of hydroxyl groups at C2 and C4 as L-rhamnose (Quesenberry and Drickamer, 1992; Weis et al., 1998). It is interesting to note that mutant C, which contains all the mutations of mutant B, does not recognise the glycans equally well, and this could be because the glycines in mutant C cause the loops to become too flexible and are no longer able form hydrogen bonds with the glycan.

IgE is the most glycosylated immunoglobulin, exhibiting significant interindividual variation (Plomp et al., 2014). The several N-linked glycosylation sites contain branched glycans composed of mainly mannose, galactose, fucose, N-acetylglucosamine and N-acetylneuraminic acid. Although the derCD23 mutants did bind to some mannose and galactose containing glycans, the lack of binding data in the glycan array to fucose, N-acetylglucosamine and N-acetylneuraminic acid suggests that the mutant proteins are not capable of binding to human IgE via carbohydrate interactions. This finding implies that mouse CD23 is also incapable of binding to mouse IgE via carbohydrate interactions, which agrees with previous experiments (Richards and Katz, 1990).

The method of glycan presentation is important, and differing results can be generated from glycans presented on a protein backbone versus a glass slide with linkers. Most of the glycans assessed in the array were considered not to be of human biological significance. Although it is interesting to note that the alpha-3 link in glycan 10 of the array, which illicits a highly immunogenic response is not found in humans, yet would cause a significant immune reaction. The specificity to this linker is called into question as number 17 although highly unusual and not found in humans also contains this link but no proteins bound to it at all (Personal communication with Prof. Joy Burchell). It appears that no true specificity was found with human WT derCD23 or the mutant proteins B and C.

The mutant derCD23 proteins are more mouse-like in sequence but not in function since murine CD23 can only bind to murine IgE in the presence of calcium, rendering a binding affinity of 10^8 M^{-1} (Richards and Katz, 1990; Lee et al., 1986). In contrast, human WT derCD23 binds more weakly to human IgE with a $K_D \sim 10^{-6} - 10^{-7}$ M but is

not abosolutely dependent on calcium (McCloskey et al., 2007; Hibbert et al., 2005). The mutant proteins all displayed weaker binding affinities for IgE than WT derCD23 in the presence of calcium, except for mutant C as documented in Table 3-5. Despite the mutant derCD23 proteins exhibiting a trend of stronger binding to IgE-Fc the more mouse-like the protein is, the changes in B_{max} suggest a degree of allosteric modulation occurring in IgE-Fc (Davies et al., 2017). This allosteric modulation could explain the change in the derCD23 proteins' affinities for IgE-Fc based on the conformation and available binding sites of IgE-Fc.

In the absence of calcium, the effect of the allosteric modulation is less apparent. Mutant A appears unaffected since it has the same B_{max} value as the WT derCD23 while mutant C is affected, but the B_{max} does not change as much as when calcium is present in the experiment. Furthermore, WT human derCD23 displayed the highest calcium sensitivity for IgE binding (Table 3-5) with a 32-fold difference in the K_D value in the presence of, as well as the in the absence of, calcium.

SPR analysis revealed that all the mutants had weaker binding to IgE than WT derCD23 in the absence of calcium (Table 3-5), and decreased calcium sensitivity thus supporting the hypothesis that human derCD23 lost its ability to bind to two calcium ions, which led to an increase in binding activity toward IgE. In the context of the whole protein, when membrane-bound, this could affect receptor-mediated endocytosis. The calcium-based gain of function in human CD23 might be enough to tip the scale in favour of an allergic response, especially in IgE-mediated antigen presentation. More allergens might be processed and displayed to the immune system as pathogens, reducing tolerance and this might contribute to the phenomenon that humans experience allergies and asthmatic attacks.

Chapter 4 Molecular Dynamics Simulations Studies of derCD23

4.1 Introduction

Proteins are very dynamic and most structural studies only permit us to see 'snapshots' or certain possible protein conformations. In order to escape the constraints of experiments and to sample as many protein conformations as possible, Molecular Dynamics (MD) simulations were used to understand if it would be possible to predict the effects of the mutations in the mutant derCD23 proteins, described previously in Chapter 3, *in silico*. A typical constraint of structural studies, with CD23 in mind, is that not all residues of a protein are resolved. The residues of loop 4 in derCD23 couldn't be assigned with experimental techniques such as NMR or X-ray crystallography (Dhaliwal et al., 2013; Hibbert et al., 2005), this was overcome with MD simulations. The simulations described in this chapter allow for comparison with experimental studies in Chapter 3. MD simulations also allow for experiments to be conducted with models when the structure of a protein does exist. In this chapter, studies on mouse derCD23 (for which no structure currently exists) are also presented to understand better the differences between human and mouse CD23.

4.1.1 Introduction to Molecular Dynamics Simulations

MD simulations can explore shorter timescales than that of most laboratory based experimental methods, to investigate fast events in protein structural dynamics; such as side chain flips as shown in Figure 4-1. In classical molecular mechanics, Newton's second law of motion (F = m.a) is used, which is not as accurate as quantum mechanics,

but is computationally faster and allows for much longer timescales to be observed (1 ns -1 μs compared to ps for quantum mechanics) (Dror et al., 2010; Ode et al., 2012).

The starting point for any molecular dynamics simulation is a model of the protein, which could have been obtained experimentally by NMR, X-ray crystallography or cryo-electron microscopy, or alternatively the model could have been built on a homologous structure or by theoretical programs. The model provides a set of Cartesian coordinates for each atom in the system (usually in pdb file format) and describes the order of the atoms connectivity, therefore considered as the starting conformation for a simulation (usually saved as an inpcrd file in Amber notation). The second requirement for a MD simulation is topological information that describes how the atoms are connected, and their property types such as charge etc. This information is stored on a database as part of the Amber software (Case et al., 2014), and in Amber notation this information is stored as a prmtop file for the simulation.
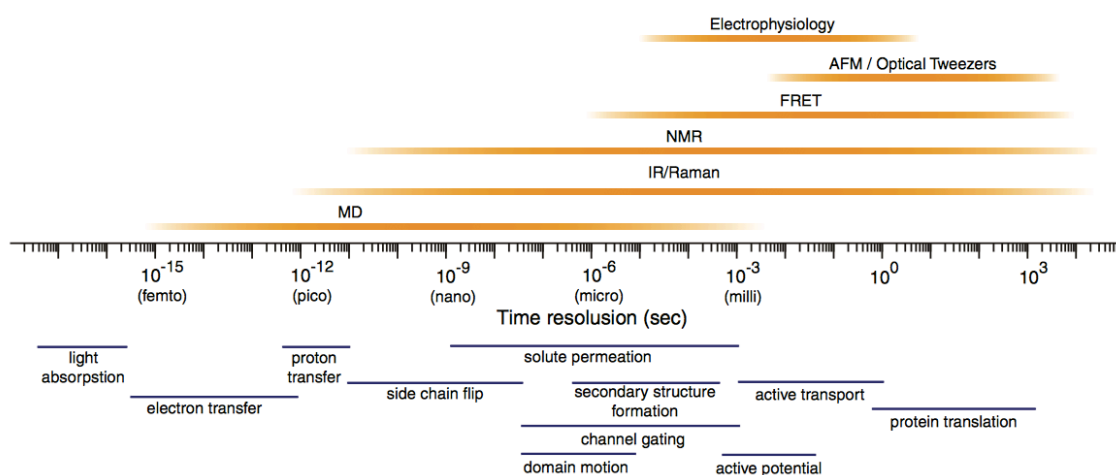
Figure 4-1 **The various timescales covered by biophysical techniques.** AFM = atomic force microscopy, FRET = fluorescence resonance energy transfer, IR = infrared resonance, NMR = Nuclear Magnetic resonance, MD = Molecular Dynamics simulations. The timescale covered by each biophysical technique is shown in orange above the timeline. The timescales of some fundamental atomic and molecular scale motions are shown below the timescale with purple lines. Image reproduced from (Ode et al., 2012).

The third condition for setting-up a MD simulation is choosing the force field which contains all the parameters to describe how the atoms move and interact in a MD simulation, which is a mathematical equation describing the energy in a system. A typical force field (FF) is described in Equation 5 and the parameters they control are illustrated in Figure 4-2. Additional FFs, for RNA, DNA, lipids and glycans are available for simulations containing these non-protein components.

$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)]$$
$$+ \sum_{\text{improper}} V_{imp} + \sum_{\text{LJ}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}},$$

Equation 5 **A typical equation for a force field.** The first four terms refer to bond stretching, angle bending, dihedral and improper torsions, which are intramolecular contributions to the total energy (U) while the last two terms describe van der Waals interactions by means of the Lennard-Jones potential and electrostatic interactions by Coulombic interactions.

A set of instructions on how to run the MD simulation and its boundaries such as the system temperature, pressure and time are saved as a mdin file in Amber notation. All this information is fed into a simulation program to run the MD simulation and the position, velocity and acceleration of each atom is calculated according to Newton's equations of motion. The initial velocities are chosen randomly but within the limitations of temperature stated in the instructions. The atomic coordinates are propagated through time using the Verlet algorithm (Verlet, 1967) which are based on integrations of Newton's equations of motion. The Verlet algorithm requires the position of atoms at two previous time points in order to calculate the position of atoms at the next time step.

Larger time steps within the simulation are allowed with the use of the SHAKE algorithm (Ryckaert et al., 1977). SHAKE acts to constrain the Verlet algorithm and

improves its efficiency by removing bond stretching of bound hydrogen atoms. By using fixed bond lengths, it freezes the fastest motions in the simulation with negligible effect on the accuracy and quality of the simulation while increasing the processing speed (Adcock and McCammon, 2006). The atomic coordinates from each time step is then connected to become the trajectory of the object which is analysed by a separate set of analysis programmes while the trajectory can be visualised by molecular structure software e.g. UCSF Chimera or VMD (Case et al., 2014).
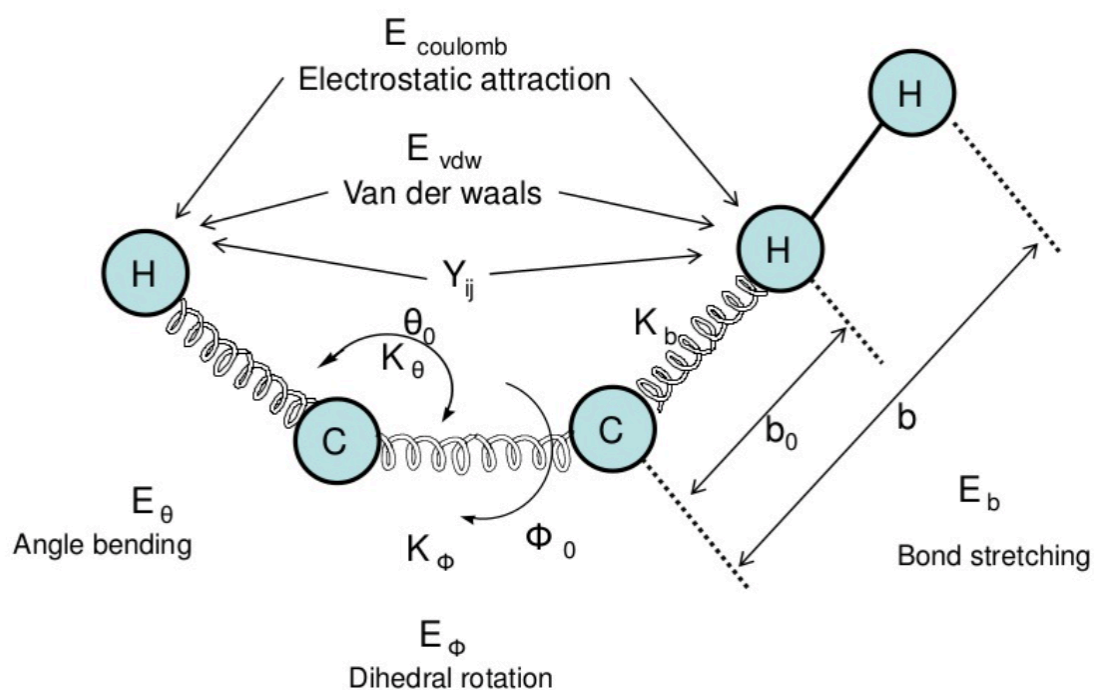
Figure 4-2 **An illustration of the parameters measured.** In classical molecular mechanics, atoms are treated as spheres (blue) joined by springs. Bonded interactions include K_b for bond length, bond angle K_θ and dihedral angles K_φ. Nonbonded interactions include van der Waals forces(E_{vdw}), and electrostatic potential ($E_{coulomb}$). Van der Waals forces are very important in a protein since they affect the secondary structure and how the protein organises in a complex or an assembly. Image reproduced from lecture slide, (Levitt, 2014).

4.1.2 derCD23 residues involved in calcium and IgE binding

Chapter 3 describes the difficulty in determining the stoichiometry of calcium binding experimentally, and the effect the mutant derCD23 proteins have on IgE-binding. To re-iterate residues of interest in the simulation, Figure 4-3 illustrates the IgE-Fc binding residues and calcium binding residues.
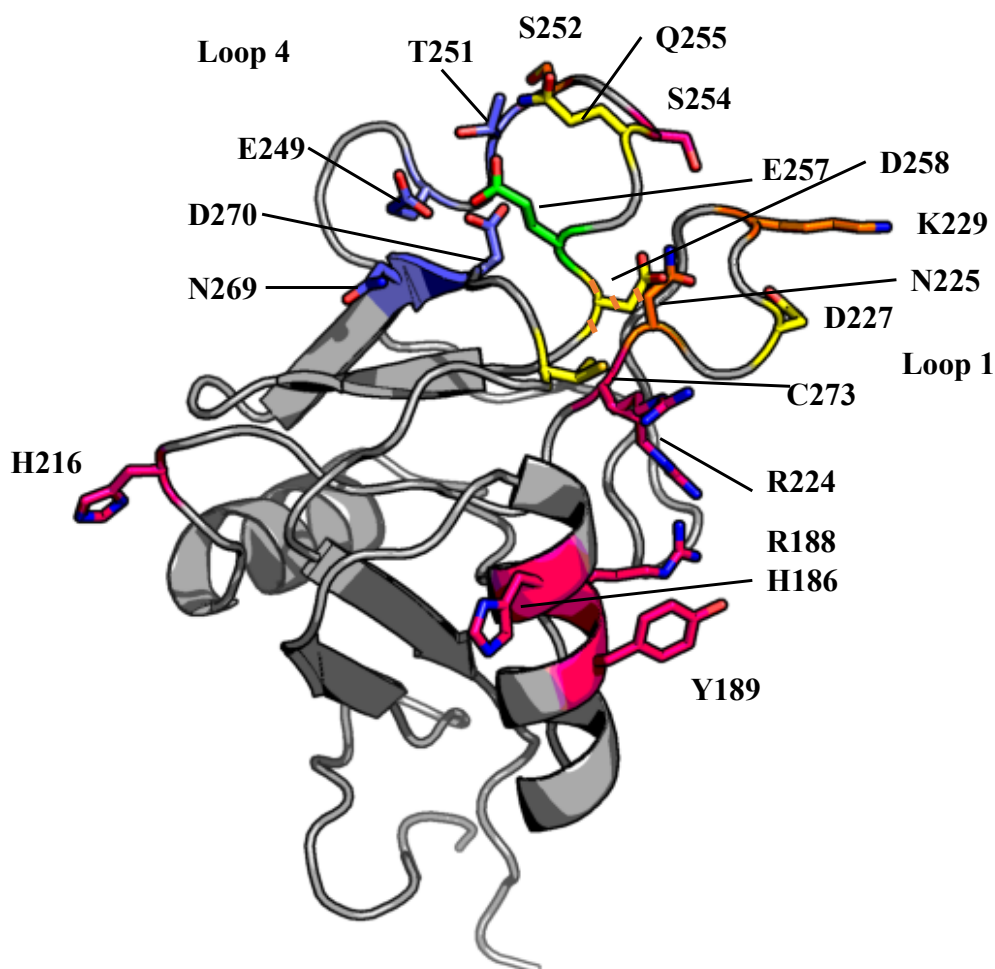
Figure 4-3 **The structure of WT human derCD23.** Blue carbon atoms = The principal calcium-ion-binding site, orange = "auxiliary" site, green carbon atoms = E257, a residue involved in both principal and "auxiliary" sites, pink carbon atoms = residues involved in IgE binding, yellow carbon atoms = additional residues involved in IgE binding upon calcium binding, striped yellow and orange carbon atoms = D258 a residue part of the "auxiliary" site and makes additional salt bridges to IgE on calcium binding. Residues and loops labelled, pdb file: 4G9A.

4.2 Results

4.2.1 A Note on Nomenclature

Most molecular dynamics simulation software designates CA to represent a Cα carbon. A calcium ion is also represented as either CA, or ca or Ca, which mostly registers as a Cα carbon, this unfortunately means that it has been particularly difficult to build and analyse a calcium-binding protein system. The systems were all run in Amber14 with the Amber ff14SB forcefield (Case et al., 2014).

4.2.2 Summary of Molecular Dynamics Simulations Datasets

Table 4-1 provides a summary of all the protein models run for molecular dynamics simulations, using the same nomenclature for the derCD23 mutants as chapter 3. All simulations were run with, and without calcium. For the calcium simulations, one bound calcium ion was modelled for the WT human derCD23 protein, and two ligated calcium ions were modelled for the other derCD23 proteins. The calcium-containing simulations began with calcium bound to the protein, since WT human derCD23 has such a weak binding affinity for calcium at $K_D \sim 1.5$ mM (Yuan et al., 2013). It would not be expected to see solution-free calcium bind to the protein within the maximal simulation timeframe of 1 μs, due to slow kinetics associated with conformational changes that occur upon calcium binding (Ng and Weis, 1998).

Simulation	Description	Time (µs)	No. of Runs
WT_0CA	Apo WT derCD23 without calcium	1	3
WT_1Ca	WT derCD23 with 2Ca^{2+}	1	3
A_0Ca	Triple mutant without calcium µs	1	3
A_2CA	Triple mutant with 2Ca^{2+}	1	3
B_0Ca	Quadruple mutant without calcium	1	3
B_2Ca	Quadruple mutant with 2Ca^{2+}	1	3
C_0Ca	Six residue mutant without calcium	1	3
C_2Ca	Six residue mutant with 2Ca^{2+}	1	3
Mouse_0Ca	Modelled mouse derCD23 without calcium	1	3
Mouse_2Ca	Modelled mouse derCD23 with 2Ca^{2+}	1	3

Table 4-1 **Description of Molecular Dynamics simulations performed in this study.** WT = WT human derCD23, A = human derCD23 with the following point mutations: N225D, K229E, S252N B= human derCD23, with the following mutations: N225D, K229E, T251N, S252N C= WT human derCD23 with the following mutations: N225D, K229E, T251N, S252N, R253G, S254G. Mouse = WT mouse derCD23 model.

4.2.3 Trajectory Statistics

All the MD simulations remained stable throughout their trajectories as shown by the average Cα Root Mean Square Deviation (RMSD), and the small standard deviations as shown in Table 4-2. The RMSD measures the structural distance between co-ordinate sets in a trajectory, and is defined by Equation 6. Interestingly, the RMSD is always smaller in the calcium-bound states, hence in agreement with the observations made for mutant C in section 3.2.5, which shows that the calcium adheres to the loops that form part of the calcium-binding sites, stabilizing the protein and constraining the loops. The stabilizing and protein-shape compacting influence of calcium is also reflected in the

total surface area, which is reduced for the calcium-bound proteins; apart from mutant C and murine derCD23. This exception may be due to the additional glycines found in the loop regions of these most 'mouse-like' proteins that may cause the loops to fall into conformations not seen in the other proteins, and thus extending the overall surface area. Conversely, the difference of compactness between the calcium-bound and apo protein simulations do not show any difference in column three of Table 4-2, which documents the total radius of gyration across the entire simulation time and is described by Equation 7. The change in radius translates to a change of less than 1 % of the protein volume. Both the RMSD and Rg confirm that the protein's tertiary structure remained stable because in an unfolding protein these values would increase over the course of the simulation.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(u_i - v_i)^2}$$

Equation 6 **The Root Mean Square Deviation (RMSD).** N = number of atoms, u and v are the vector coordinates of two different points in the structure.

$$Rg = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - r_{cm})^2}$$

Equation 7 **The radius of gyration.** n = number of visits to location i, r_i is the vector of the ith atom's coordinates in space and r_{cm} = position of the centre of mass of the protein.

The exposed surface area (defined by the LCPO algorithm (Weiser et al., 1999)) of both calcium-binding sites reported in Table 4-2 also decreased upon calcium-binding as the loops bent over the calcium ion. The deeper auxiliary binding site showed greater reduction in the surface area compared to the relatively shallow principal site. This was

true for all the simulated proteins, except mouse derCD23 which could be due to model inaccuracies, since an experimental structure has yet been to be confirmed. The individual surface areas of residues involved in IgE binding were expected to increase with calcium binding, so that more contacts could be made between derCD23 and IgE-Fc but with no clear trend for all the simulations, this might be considered a too simplistic view as binding is more dependent on atoms being at correct distances and orientations for hydrogen bonds and salt bridges to form.

Protein	RMSD (Å)	Rg (Å)	Total SA (Å^2)	IgE-SA (Å^2)	Ca1-SA (Å^2)	Ca2-SA (Å^2)
WT	0.99	13.37	6824.1	770.0	218.8	147.1
	(0.31)	(0.036)	(123.3)	(32.4)	(24.5)	(17.7)
WT_1Ca	0.94	13.37	6640.3	701.8	149.9	76.8
	(0.28)	(0.046)	(96.4)	(27.0)	(26.4)	(8.6)
A	1.11	13.40	6820.5	715.1	281.5	135.0
	(0.14)	(0.006)	(95.8)	(26.5)	(16.1)	(14.7)
A_2Ca	0.89	13.33	5113.2	720.8	229.1	79.4
	(0.11)	(0.007)	(67.5)	(27.6)	(14.6)	(9.1)
B	1.12	13.32	6800.5	715.7	279.9	136.2
	(0.51)	(0.082)	(100.9)	(32.2)	(19.7)	(22.1)
B_2Ca	0.95	13.37	6792.7	689.3	232.6	88.0
	(0.22)	(0.017)	(89.3)	(25.3)	(14.6)	(8.4)
C	1.22	13.40	6741.4	694.3	300.0	148.1
	(0.24)	(0.032)	(115.7)	(38.4)	(21.4)	(26.2)
C_2Ca	1.01	13.45	6784.2	743.4	223.2	83.0
	(0.21)	(0.013)	(110.5)	(34.4)	(15.4)	(10.1)
WT Mouse	1.40	13.00	5056.4	218.1	155.0	123.7
	(0.55)	(0.010)	(70.9)	(16.0)	(15.0)	(11.0)
Mouse_2Ca	1.06	13.00	6657.3	326.5	174.3	122.3
	(0.34)	(0.064)	(97.9)	(18.2)	(10.0)	(5.2)

Table 4-2 **Comparison between average geometric properties of all the MD simulations.** Average values of three runs displayed taken onwards from frame 30 (1.5 ns) to frame 20000 (1 us). Values in parentheses are the standard deviation. Rg = radius of gyration. RMSD and Rg values for Cα atoms only. Total SA = Total surface area of protein calculated with the LCPO algorithm. IgE-SA = Surface area of IgE-binding residues on CD23 (residues His186, Arg188, Tyr189, Arg224, Asp227, Ser254, Gln255, Asp258 and Cys273). Ca1-SA = Surface Area of the principal calcium binding site (residues Asp 225, Glu229, Thr251 and Asp258). Ca2-SA = Surface Area of the auxiliary calcium binding site (residues Glu249, Thr251 and Asp270). Values in brackets are the standard deviation.

4.2.4 Root Mean Square Fluctuations (RMSF)

Another measurement used to analyse movements of the structure during the simulation is root mean square fluctuations (RMSF), defined by Equation 8. It calculates the average atomic mobility of each of the Cα atoms during the course of the simulation so that portions of the structure that are fluctuating the most can be identified easily. In derCD23, the non-secondary structure components of the protein such as the loops would be expected to fluctuate the most since α-helices and β-sheets would be held in position by a network of hydrogen bonds. Indeed, in Figure 4-4 the regions that fluctuate the most are loops 1 and 4, while fluctuations at residues 261-263 and 266-274 correspond to turns connecting β strands and atomic movement at residues 212-217 correspond to a turn between β strand 2 and α helix 2 which contains the IgE-binding residue His 216.

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(v_t - \overline{v})^2}$$

Equation 8 **The Root Mean Square Fluctuation (RMSF).** T = the number of trajectory frames, $\overline{v}$ = the mean structure and v_t = the position of atoms in residue i after superposition on the mean structure.

If calcium is acting to constrain the movements of the loops, it would be expected that less mobility is seen in the calcium-bound structures than within the calcium free simulations. Accordingly, in the calcium-bound structures displayed in Figure 4-4, there is overall less motion (largest fluctuation ~ 2.5 Å) than in the apo-protein in Figure 4-5 (largest fluctuation ~ 4 Å). In loop 1, Figure 4-4C there is little variance between the derCD23 proteins, however, in loop 4, especially for the mouse protein and mutant C, a much larger fluctuation can be observed this might be due to the glycines, only found at

163

the base of loop 4 in these two proteins, which allow for greater spatial degrees of freedom.

Without calcium, the effect of the altered residues leads to a slightly more pronounced movement in loop 1, with all the proteins moving on average 1 Å more than in the calcium-bound structures. Mutant C shows the largest increase in RMSF in loop 1 as shown in Figure 4-5C, although it is the closest to murine derCD23 by sequence, two residues still differ, the aspartic acid and the adjacent leucine are substituted with an asparagine and an isoleucine in the mouse protein. The isoleucine may explain why the murine derCD23 is less flexible than mutant C in loop 1 due to the branched nature of the Cβ of the residue, which would restrict the conformations that the main chain could adopt. Asp227, a residue found in loop 1 of the WT human and mutant derCD23 proteins, forms a hydrogen bond with IgE-Fc in the presence of calcium. The greater movement of loop 1 could influence the kinetics of the CD23-IgE interaction. In general, loop 4 became slightly more flexible, most noticeably in mutant 4, and in the mouse protein. Both these proteins contain glycine substitutions, enabling the protein backbone to explore more conformational space as the constraints were released.

Another area of difference, is the region containing the calcium co-ordinating residues Asn269 and Asp270 of β strand 5 at the principal calcium binding site. All the mutant and mouse proteins are less mobile than the WT human protein at this location. The reason for this is not clear, since no additional stabilising hydrogen bond could be found involving those residues during the MD simulations. Examining solvent accessibility (Zhang and Lazim, 2017) or using techniques such as elastic network models (Fuglebakk et al., 2013, 2015) could further explore the protein motions.
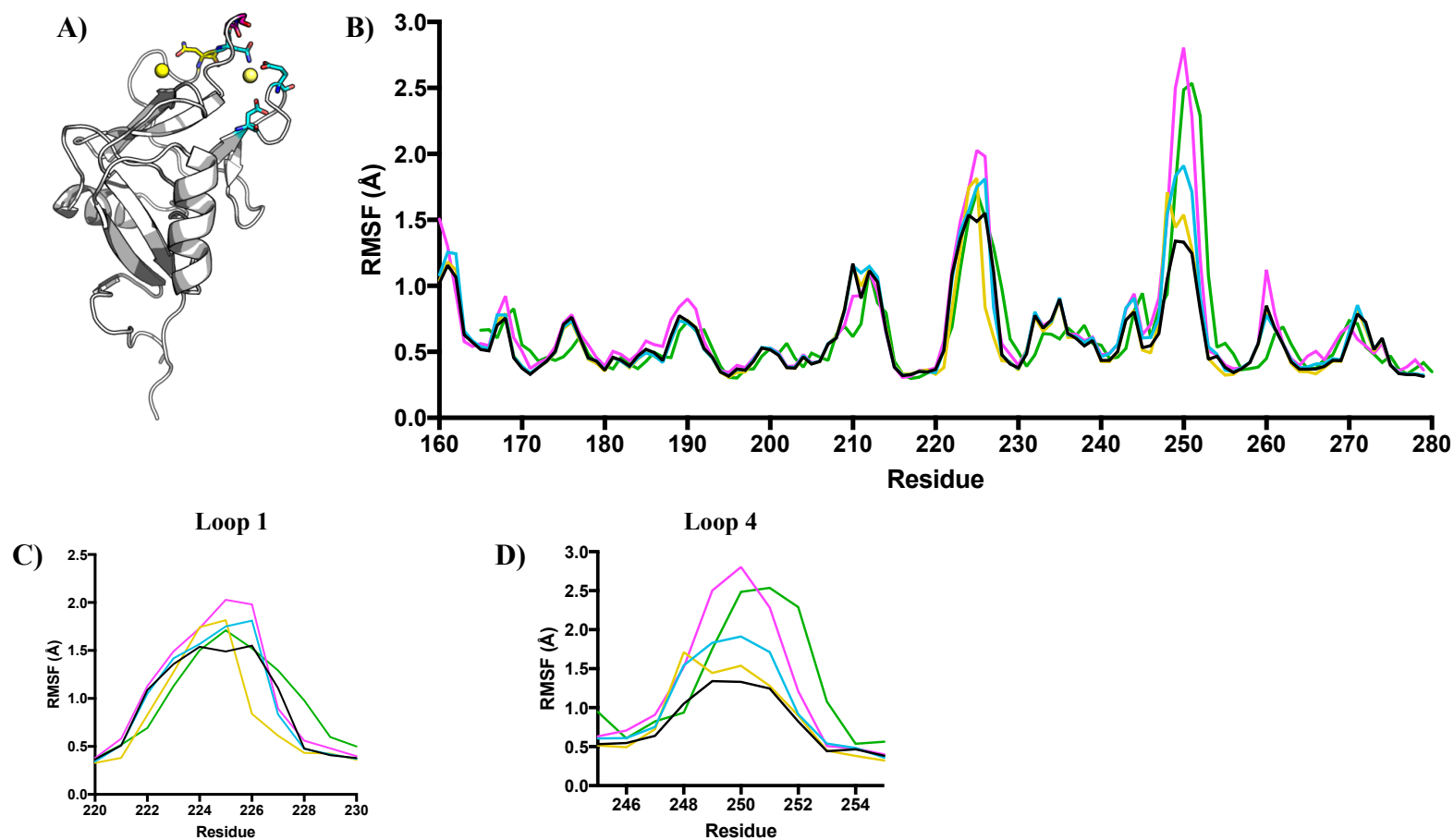
Figure 4-4 **Structure and RMSF values of the calcium-bound simulations.** Black = WT human derCD23, light-blue = mutant A, yellow = B, magenta = C, green = WT mouse derCD23, these colours also match those of the graphs. **A)** displays a cartoon figure of derCD23 with the different residues of each mutant protein highlighted; yellow spheres give the position of the two calcium ions. **B)** shows the RMSF of the Cα of each residue across the whole of the proteins, averaged over triplicate simulations. The RMSF is only shown for residues 159-280 which excludes the unstructured N and C terminal domains, residue numbers correspond to the numbering scheme of the protein in the pdb file. **C)** is a zoom-in of panel B focusing on loop 1. **D)** is a zoom-in of panel B focusing on loop 4.
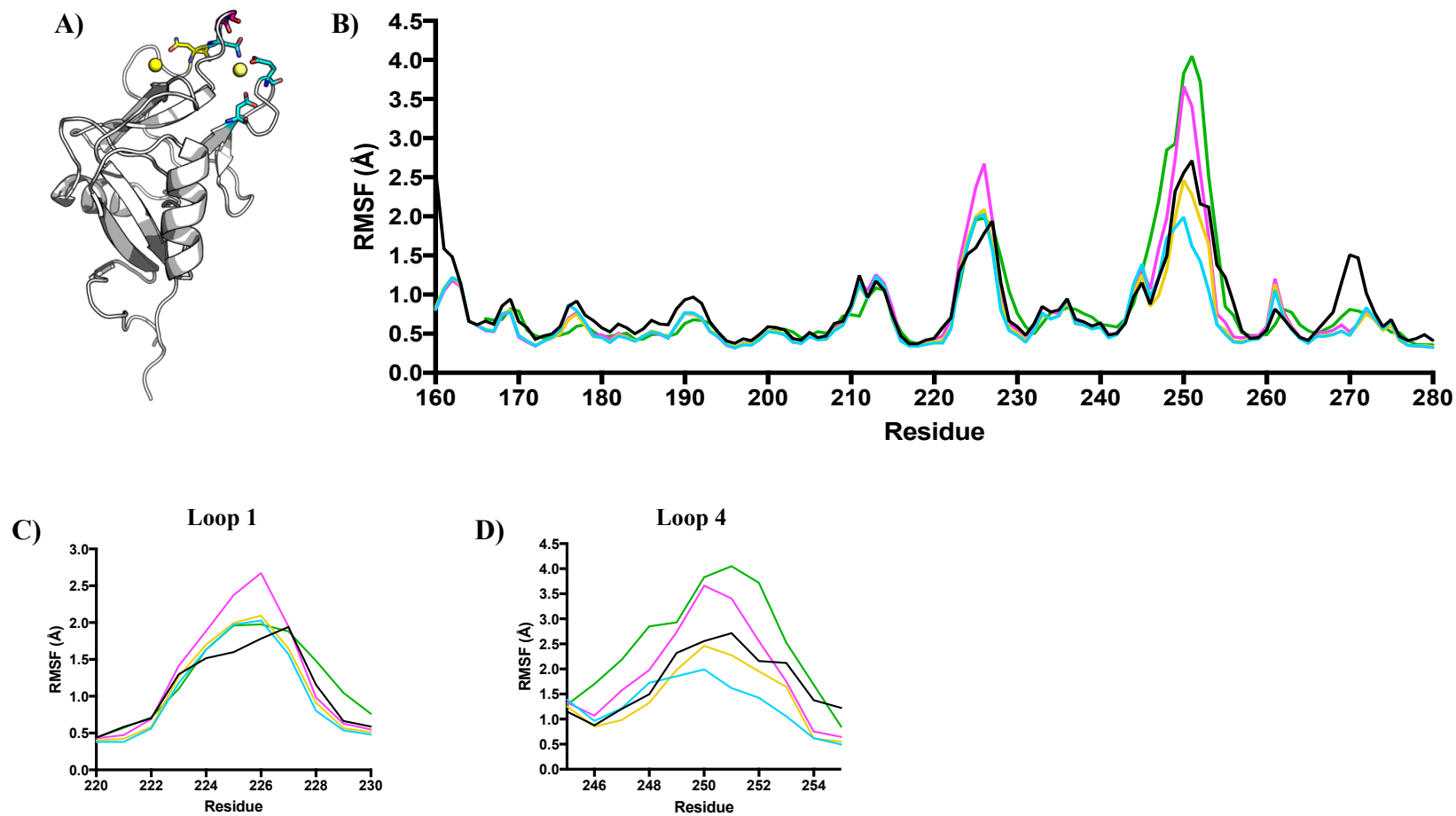
Figure 4-5 **Structure and RMSF values of the proteins without calcium. A)** A cartoon figure of derCD23 with the different residues of each mutant protein highlighted; yellow spheres give the position of the two calcium ions. Light-blue = mutant A, yellow = B and magenta = mutant C, these colours also match those of the graphs. **B)** black = WT human derCD23, green = WT mouse derCD23. The RMSF profile across the whole of the proteins with residue number corresponding to the numbering scheme of the protein. **C)** is a zoom-in of panel B focusing on loop 1. **D)** is a zoom-in of panel B focusing on loop 4.

4.2.5 Comparison of molecular motions seen by Molecular Dynamics Simulations and X-ray Crystallography

The MD simulation on average enables a greater freedom of movement for the protein to explore more conformations than might be captured experimentally. On comparing the crystallographic B-factors with the RMSF from the MD simulations (Kuzmanic and Zagrovic, 2010), the RMSF showed more peaks and greater flexibility than the B-factors as shown in Figure 4-6. This verifies that both techniques can locate the flexible regions of the protein and that the protein during simulation does not unfold. There is reasonable agreement with peaks at residues 225-230 and 252-258 corresponding to loop 1 and loop 4 respectively but there are also local areas of differing mobility. One area of discrepancy, in which the RMSF showed an increase and the crystallographic B-factors did not, is region 210 to 218 matching the end of helix a2 with residues connecting to β strand 2 and spatially-close residues 263-266. Within the asymmetric unit, these residues are involved in a tight network of hydrogen bonds that have enabled the protein to pack tightly (4 separate chains in one asymmetric unit), whereas in the MD simulation, these residues are less constrained and show greater mobility as they are free to interact with the solvent. This region also contains derCD23 residue His216 involved in binding to Cε2, but this derCD23- Cε2 interface is not as significant as the derCD23/Cε3-4 binding site. The derCD23- Cε2 interface has a 12-fold weaker binding affinity than the larger derCD23- Cε3-4 binding site (Dhaliwal et al., 2017). Changes in His216 orientation would have minimal impact on IgE binding. The residues of 274-278, linking β strands 4 and 5 also show a greater amplitude of movement in MD simulations than in the crystal structure, because they are involved in a hydrogen-binding network between two different chains within the asymmetric unit.

Another factor to consider that differs between the two sets of experiments, is the temperature the experiment is performed at. The protein is crystalline for cryo-X-ray crystallography and data is collected at cryo-temperatures ~ 100 K (Pflugrath, 2004) while in MD simulations the protein is in simulated solvent for the MD simulation, with simulations run at 300 K and subjected to greater thermal motions. It should be noted that the thermal settings for the MD simulations could be changed with ease, but are set at biologically relevant temperatures.
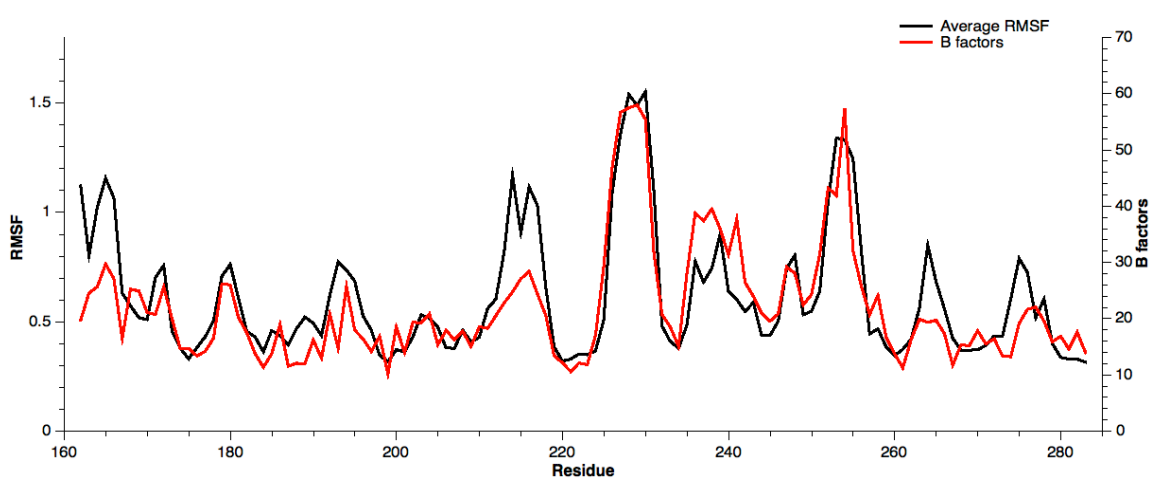
Figure 4-6 **Comparison of the average Cα RMSF of the WT-1Ca²⁺ Cα MD simulation to the B-factors of chain A pdb:4G9A (derCD23 – calcium bound).** RMSF displayed in black with axis on left, B-factors from the PDB file shown in red with axis on the right. Residue numbering matches that of the pdb file.

4.2.6 Effect of Calcium on the Protein

Overall, from observations made in Figure 4-7, there is a marked reduction in the motion of two regions of the protein backbone on the addition of calcium. It appears that calcium is acting to fasten its coordinating residues, and the bulk of the flexible loops. In WT human derCD23, only one calcium was modelled in the principal site whereas in the other proteins two calcium ions were modelled, one in the principal site and one in the auxiliary site. The calcium coordinating residues, Asn269 and Asp270,

which are not part of the loops, also show a difference in peak height as they sample conformational space. A restriction in the motion of loop4 is observed in all the proteins, and most pronounced in the mouse protein, although this may be a restriction of the model since a crystal structure does not yet exist. Compared to the single-calcium binding protein, no additional trends could be observed with the two-calcium binding proteins. Residue His216 appears to be unaffected by calcium which matches previously made observations of the derCD23/Fcε2-4 crystal structure. His216 makes one hydrogen bond to Thr260 in Cε2 but it is not close to the calcium binding site of derCD23 (Dhaliwal et al., 2017).

WT
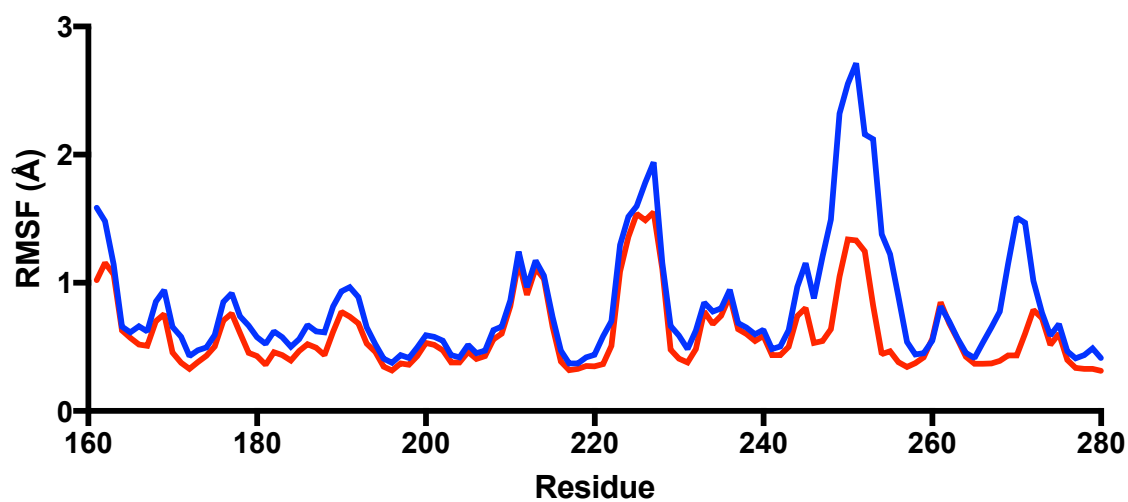
Mutant A
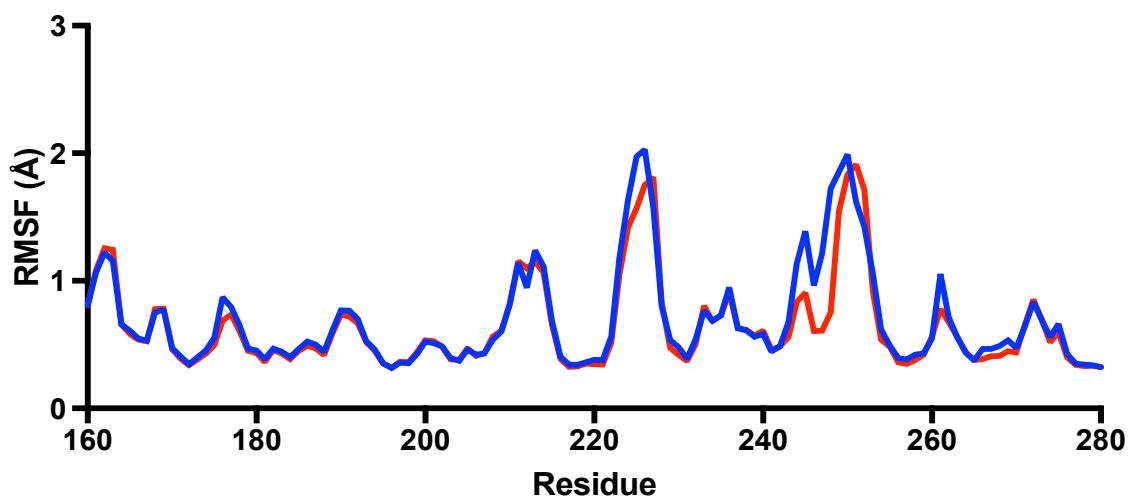
Mutant B
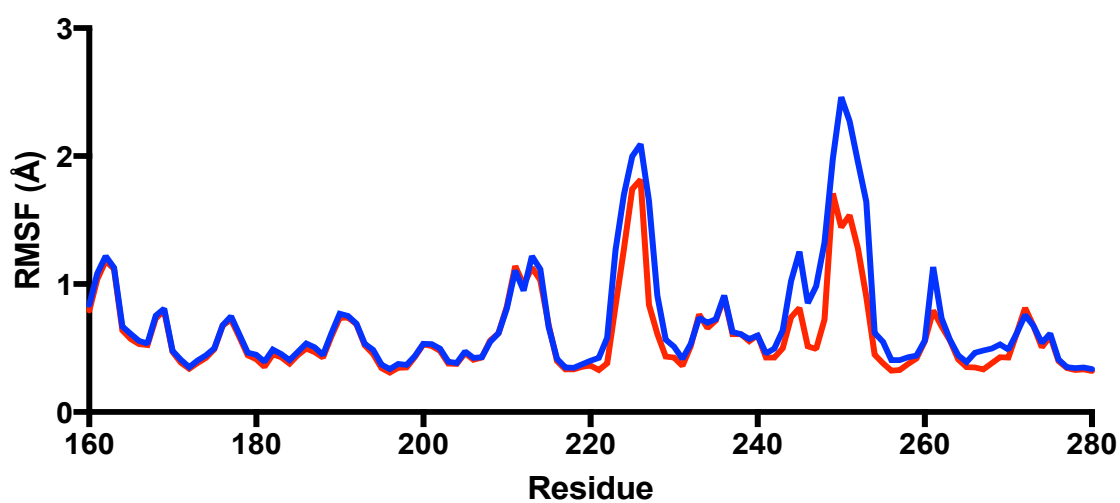
Mutant C
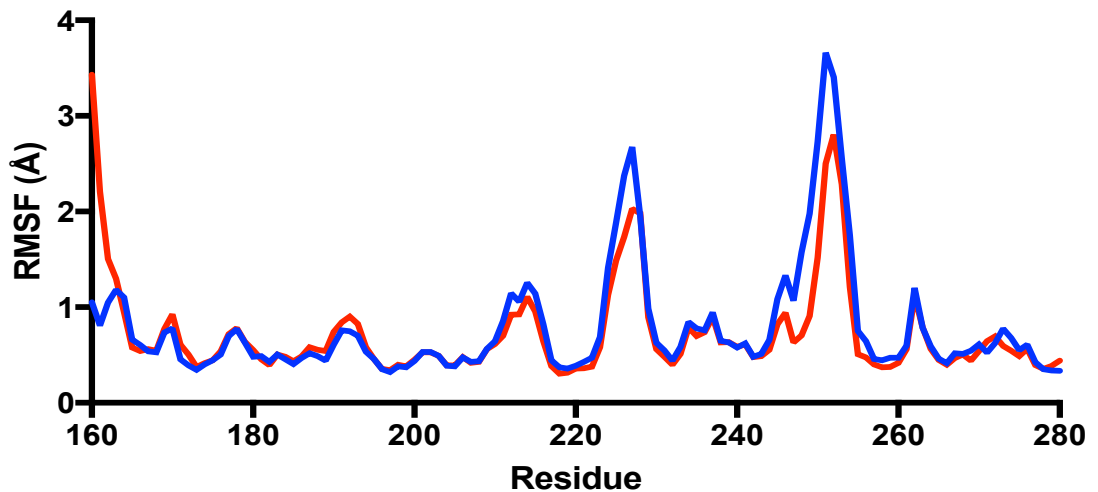
Mouse WT derCD23
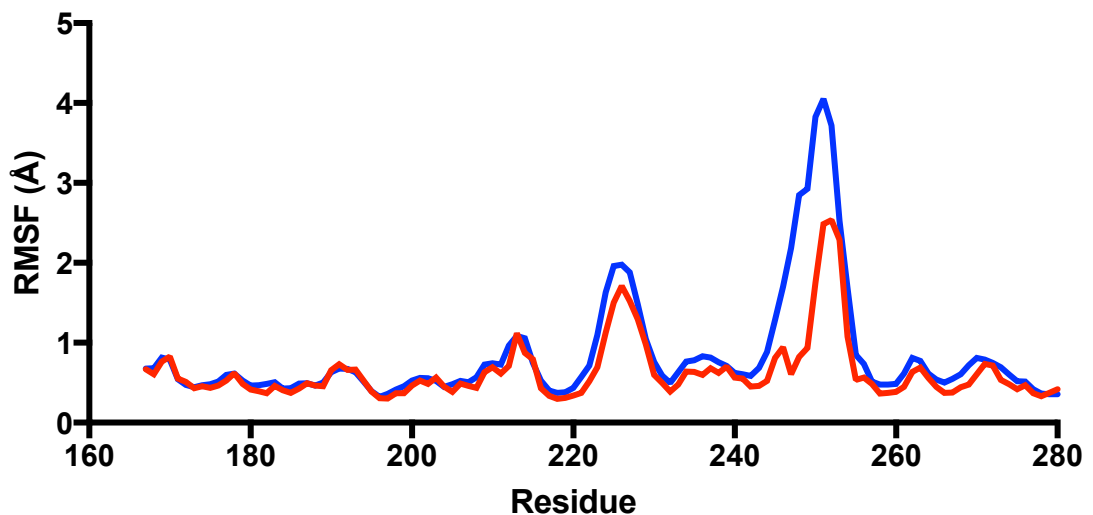
Figure 4-7 **The effect of calcium binding on the proteins.** The RMSFs were measured using the Cα of each residue from the entire trajectory (1μs) for both apo-protein in blue and protein + calcium in red, and are averaged over triplicate simulations for all proteins. The RMSF is only shown for residues 160 to 280 which excludes the unstructured and flexible N and C terminal domains.

4.2.7 Calcium Ion Distance Tracking Plots

These Calcium-binding graphs of Figure 4-8 plot the distance between calcium and the protein residues throughout the trajectory of the simulation. For the majority of cases, the close proximity of calcium to a residue results in a dark colour indicating a calcium co-ordinating residue, lighter colours indicate residues involved adjacent to the metal

ion binding residues that move in correlation with it. The smaller inset graph focuses on one residue within the calcium-binding site, and plots the distance of calcium over the entire 1 µs of the trajectory. A combination of both graphs indicate that the calcium remained bound and the darkest marks correspond to the calcium binding residues. In the WT graph in panel A, Figure 4-8. The direct calcium-binding residues of the principal binding sites D270, E249 and T251 are always displayed in black, indicating that they are always co-ordinating the calcium throughout the trajectory. N269, adjacent to D270 which in the crystal structure co-ordinates the calcium ion by a water molecule is also within a <0.3 nm distance of the calcium ion.

Panels B-E, display an additional graph for the same protein featuring the auxiliary calcium site, which shows more variation. There are two clear blocks of colour, the first block relating to loop1, whilst block two corresponds with loop 4 and the end of the beta strand preceding loop 4. The movement of the loops differs between the proteins; with the auxiliary calcium-binding site situated between the principal, and the auxiliary site; this calcium would have greater effect on loop movement. T251N in mutant B, C and mouse located at the tip of loop 4, does not appear to co-ordinate the calcium ion directly, only being within 0.5 nm of the calcium ion (expected calcium-ligation residue distance is ~ 0.27 nm). This residue is involved in the principal binding site, and in trying to make the proteins more 'mouse-like', the residue was changed to an asparagine. Initial model bias in the X-ray structures may account for why this residue still appears to be involved in the principal binding site.

WT derCD23 1Ca²⁺

Mutant A 2Ca²⁺

Mutant A 2Ca²⁺ 2ⁿᵈ Calcium

Mutant B 2Ca²⁺

Mutant B 2Ca²⁺ 2ⁿᵈ Calcium

Mutant C 2Ca²⁺

Mutant C 2Ca²⁺ 2ⁿᵈ Calcium
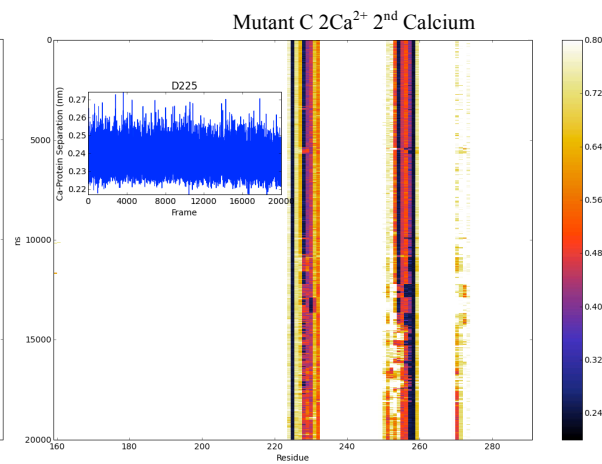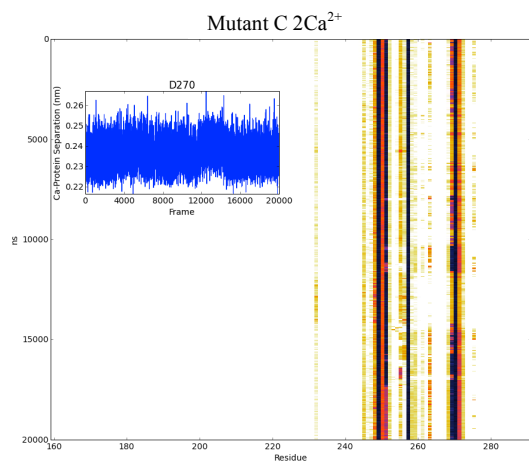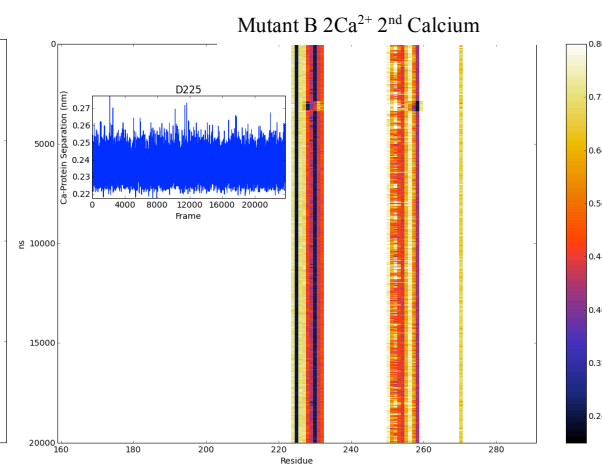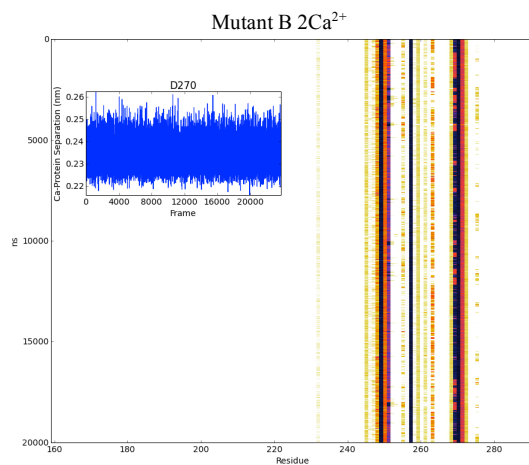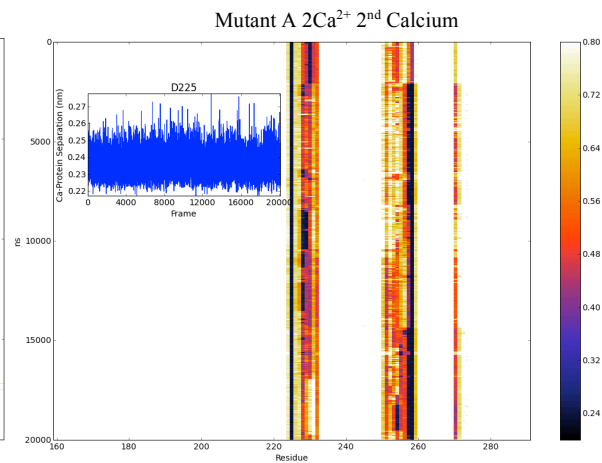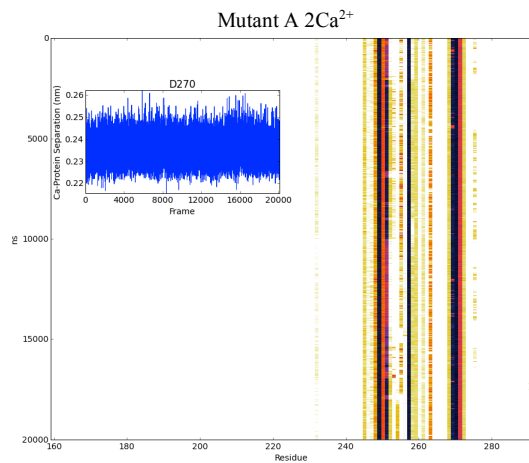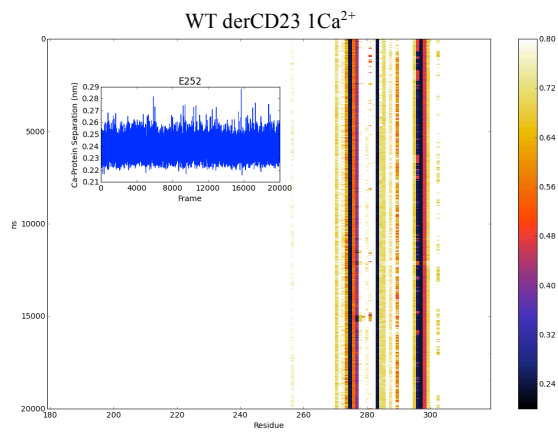
Figure continues on next page.
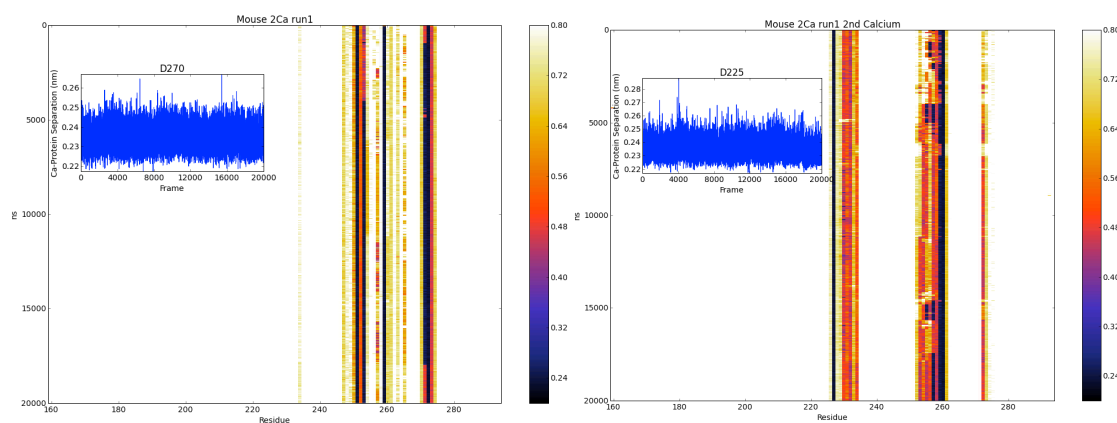
Figure 4-8 **Calcium ion binding graphs.** Representative graphs from the molecular dynamics simulations, y-axis corresponds to frame. Colour-scheme on right-hand side of the graph, white = calcium not within proximity of the calcium, black = calcium within 0.24 nm of the calcium ion. Residues corresponds to the numbering scheme of WT human derCD23. Ns = frames throughout trajectory.

4.2.8 Principal Component Analysis (PCA)

The RMSF graphs of Figure 4-4, Figure 4-5 and Figure 4-7, highlighted the sections of the structure that fluctuated the most (and the least) from the mean structure. To look at the most relevant motions of the trajectory, PCA (David and Jacobs, 2014) was applied to the backbone atoms in the concatenated simulation models (all repeats for one simulation model were concatenated). Eigenvector one describes the largest motions in the models and is also known as the principal component.

The first eigenvector was visualised in porcupine plots shown in Figure 4-9 and Figure 4-10, in which the arrows describe the direction and magnitude of the dominant motion of eigenvector 1. Several of the models express a 'breathing' motion in which loops 1 and 4 are moving to and fro from each other. The WT human derCD23 model without any calcium is the only model which displays movement around the principal calcium binding site with a lifting of the β4 strand and loop 4 only as shown in Figure 4-9. Calcium bound-mutant A displays the opposite behaviour to apo-WT model in which

174

the principal calcium binding site is occupied, and the area does not move, however loop 1 still expresses some flexing motion despite the auxiliary calcium site being occupied as well. The predicted auxiliary binding site would sit deep in the pocket compared the shallower principal binding site and depending on the backbone, could still exhibit some flexibility in the residues surrounding the calcium.

The other simulation models display this 'breathing' motion, as well as a slight twisting motion. The largest movement in loop 4 is exhibited by apo-mutant C, which might be due the point mutations since the glycines at the base of loop 4 would allow for greater conformational flexibility. Interestingly, the WT mouse model without calcium also displays movements like that of the calcium-free WT human protein. It too displays movements in loop 4 but of a larger magnitude. This finding might be due to the extra loop mobility provided by the glycine residues found at the base of loop 4 in the mouse, but not human sequence. With calcium added to the mouse derCD23 protein displayed in Figure 4-10, the movement in loop 4 is reduced, however, there is also pronounced flexibility in loop 1, not seen in the calcium-free simulation. Perhaps, binding of a calcium ion to the principal site stabilises loop 4 but the second calcium ion binding site might be less defined with the calcium ion associating and dissociating to cause the dynamics observed in loop 1.

0 CA   + CA

WT-0Ca   WT-1Ca

A_0Ca   A_2Ca

B_0Ca   B_2Ca

C – 0Ca   C_2Ca
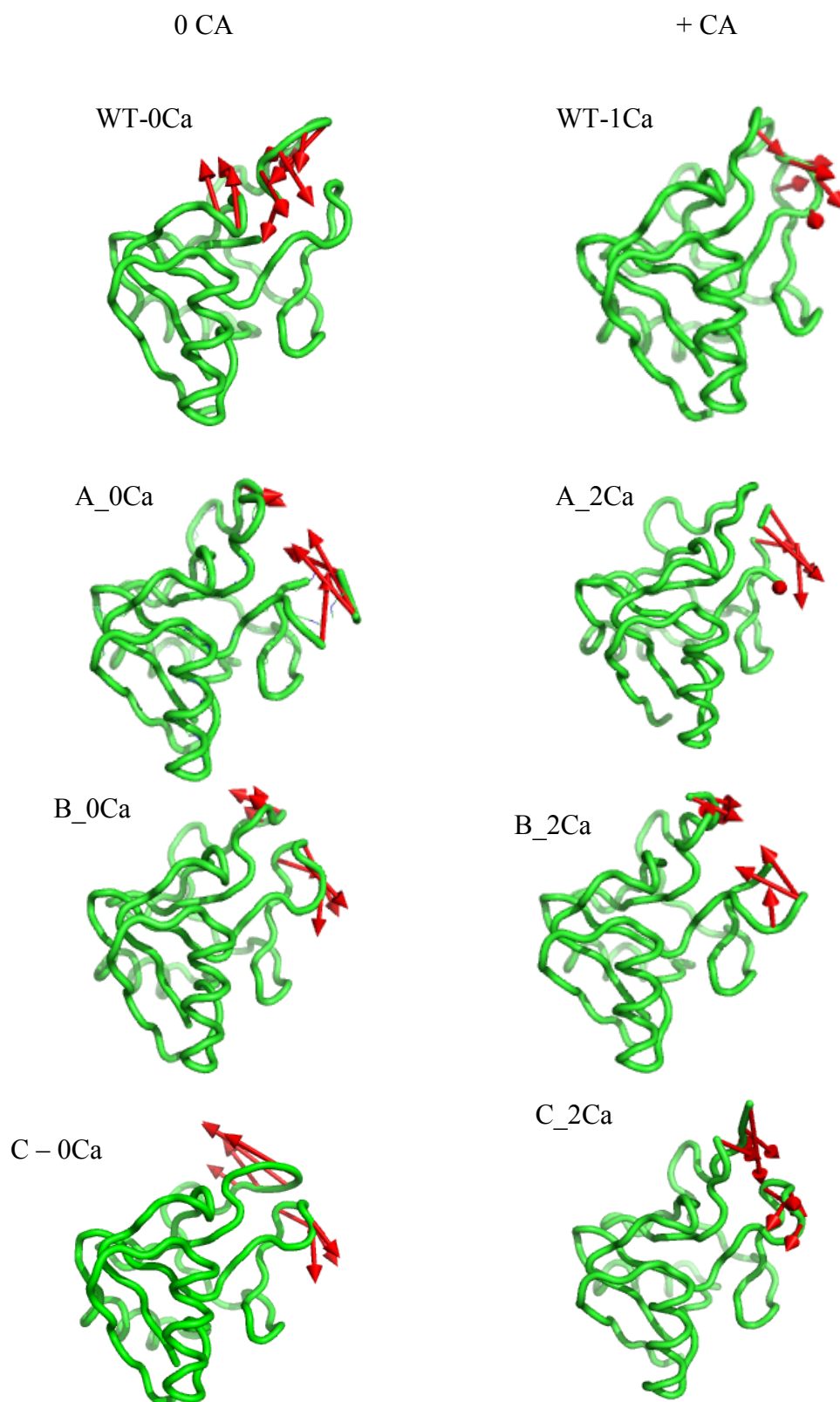
Figure 4-9 **Porcupine plot of the largest motions described by the first eigenvector of the human and mutant derCD23 proteins.** Cα backbone displayed as a green backbone trace, red arrows depict the direction and magnitude of motion. The direction of the arrows is an indication of the range of motions experienced by those atoms, it does not signify a final conformation. Left hand side shows simulations without calcium. Right hand side shows simulations with calcium.

Mouse_0Ca   Mouse + Ca
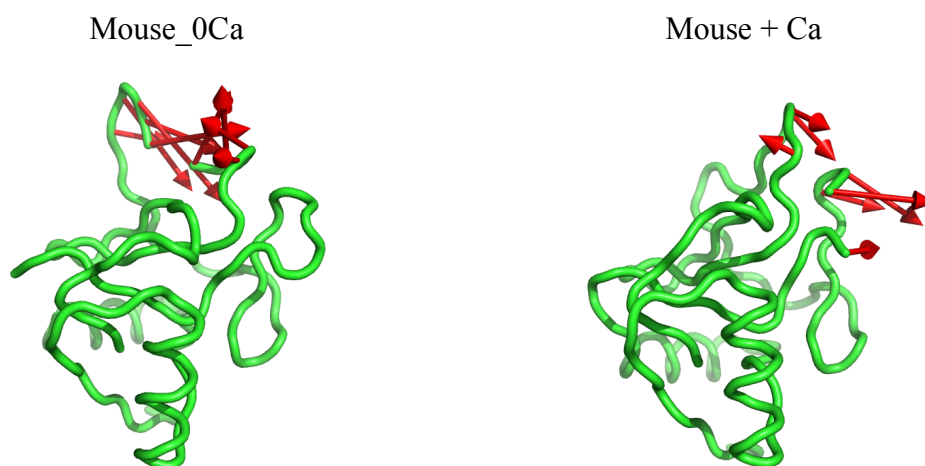
Figure 4-10 **Porcupine plot of the largest motions described by the first eigenvector of mouse derCD23 proteins.** Cα backbone displayed as a green backbone trace, red arrows depict the direction and magnitude of motion. The direction of the arrows is an indication of the range of motions experienced by those atoms, it does not signify a final conformation. Left hand side shows simulations without calcium. Right hand side shows simulations with calcium.

4.2.9 The Effect on IgE-binding residues

New conformations of derCD23 proteins were found during the trajectories as shown in Figure 4-11. Previously, molecule D of PDB file 4J6P contained a conformation of WT human derCD23 with the loops closest together (Dhaliwal et al., 2013), whereas the MD simulation has provided at least one new conformation, shown in Figure 4-11A. This conformation has been designated the 'closed' conformation due to the proximity of the calcium-binding loops. A new conformation in which the loops 1 and 4 are furthest apart is displayed in panel B of Figure 4-11, designated the 'open' conformation. The advantage with the MD simulations is that the loops are always well defined and the atom co-ordinates can always be distinguished since there is no absent electron density for the loops, frequently found in X-ray crystallography and NMR.
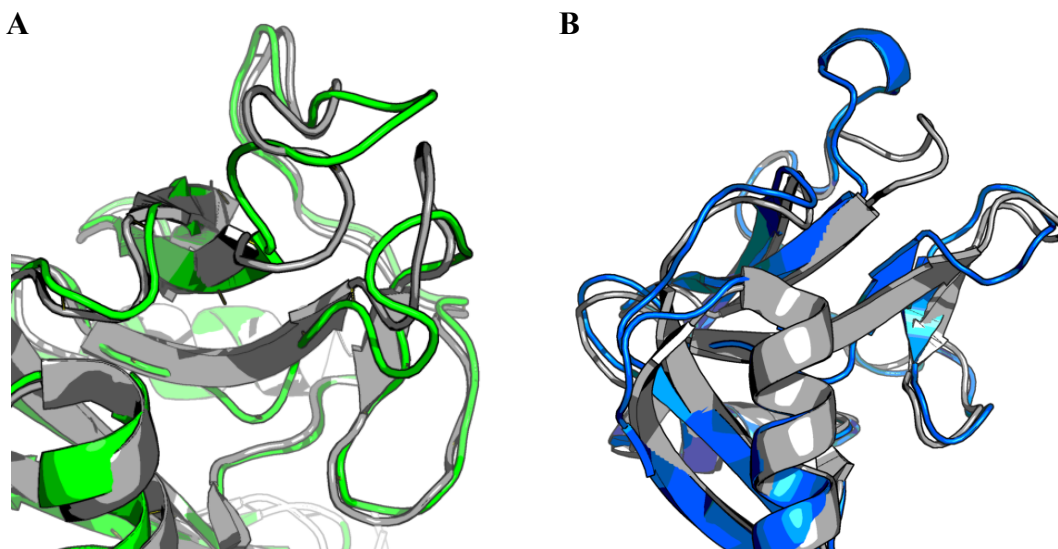
Figure 4-11 **Demonstrations of new conformations discovered through MD simulations compared to the most extreme conformations found by X-ray crystallography. A)** The most 'closed' conformation observed in the MD simulations. An alignment of a single MD trajectory frame (green) to PDB file: 4J6P (grey). **B)** The most 'open' conformation observed in the MD simulations. An alignment of a single MD trajectory frame (blue) to PDB file: 4J6Q (grey).

The binding site between human derCD23 and IgE-Fc$_{3-4}$ involves the base of loop4, and helix α1 of derCD23 and the loops AB, CD and EF of IgE-Fc. The conformation from Figure 4-11 B has been superposed onto the crystal structure of apo and calcium bound derCD23/IgE-Fc$_{3-4}$ complex, and several steric clashes are seen, indicating that this conformation would not be amenable for binding. For comparison, the crystallographically solved complex is shown in Figure 4-13, which displays minimal clashes. It is important to note, however, that the loops of both proteins are highly mobile and could easily complement each other in other ways than described in Figure 4-12.
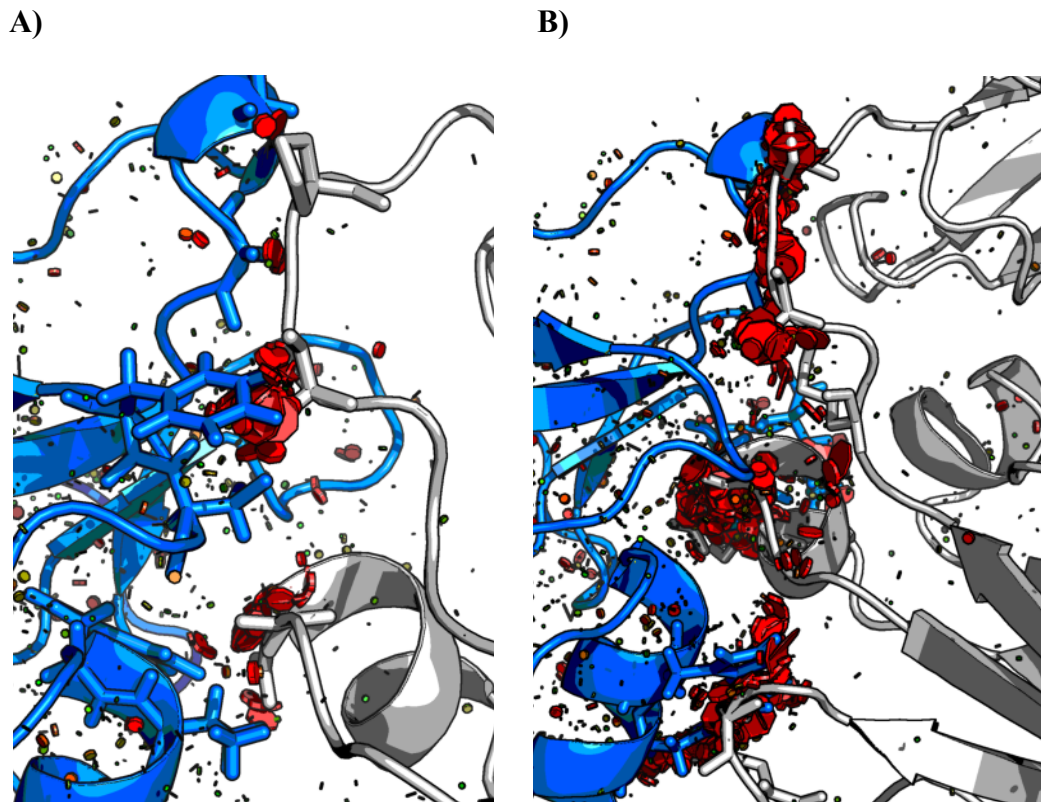
Figure 4-12 **Clashes arising from superimposing the most 'open' WT human CD23 MD simulation structure on IgE-Fc$_{3-4}$. A)** An alignment of a single MD trajectory frame (blue) to IgE-Fc$_{3-4}$ PDB file: 4EZM (grey). **B)** An alignment of a single MD trajectory frame (blue) to calcium-bound-IgE-Fc$_{3-4}$ PDB file: 4GKO (grey). The red disks display van der Waals overlaps or steric clashing.
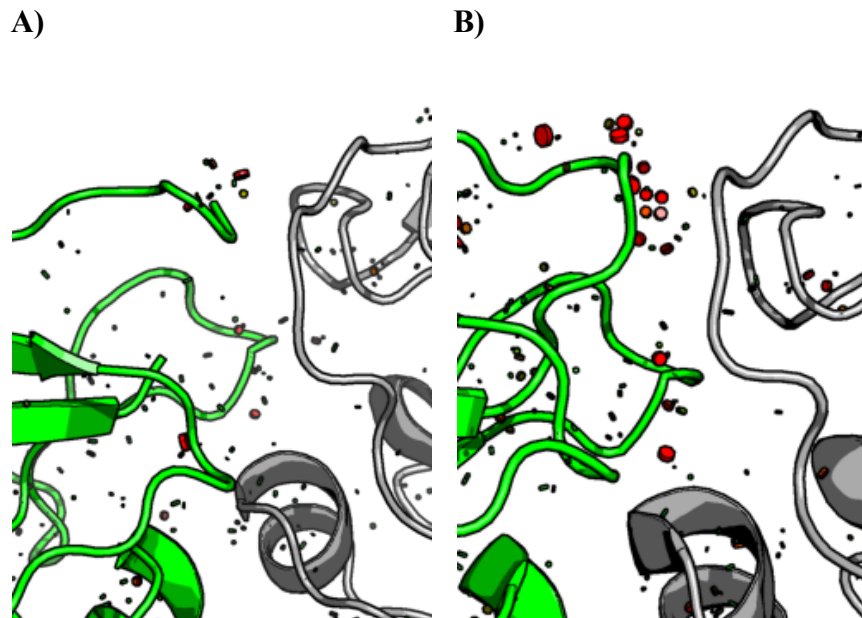
Figure 4-13 **The calcium-free and calcium-bound crystallographic structures of the CD23/IgE-Fc$_{3-4}$ complex.** Green = derCD23, grey = IgE-Fc$_{3-4}$ **A)** The calcium-free complex, PDB file: 4EZM. **B)** The calcium-bound complex, PDB file: 4GKO. The red disks display Van der Waals overlaps or steric clashing.

These clashes are also seen for the most 'closed' conformation (Figure 4-11A) depicted in the MD trajectories in Figure 4-14. It is interesting to note that the hydrogen bond that bridges loop 4 and loop 1 between Asn 225 and Asp 258 in WT human derCD23 remains in all the mutations as this involves the backbone and not side chain atoms.

A)   B)
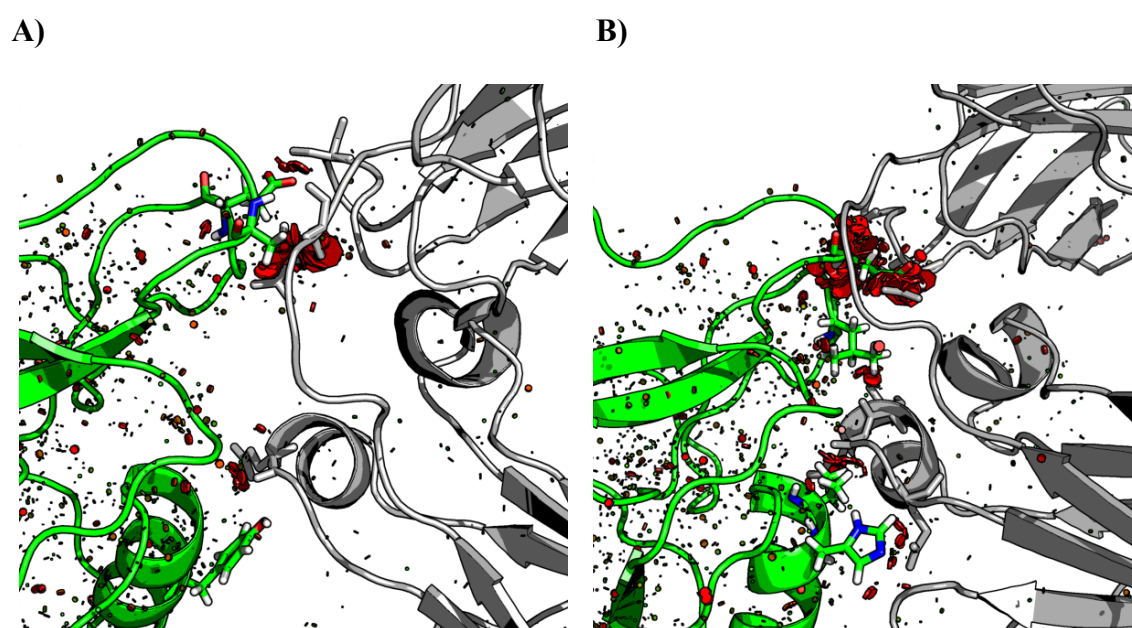
Figure 4-14 **Clashes arising from superimposing the most 'closed' WT human CD23 MD simulation structure on IgE-Fc$_{3-4}$. A)** An alignment of a single MD trajectory frame (blue) to IgE-Fc$_{3-4}$ PDB file: 4EZM (grey). **B)** An alignment of a single MD trajectory frame (blue) to calcium-bound-IgE-Fc$_{3-4}$ PDB file: 4GKO (grey). The red disks display van der Waals overlaps or steric clashing.

4.2.10 Modelling Three Calcium Ions in derCD23

Mutant C contains the same calcium ligating residues as DC-SIGNR, and it is possible for CTLs to bind to more than 2 calcium ions in general. In the case of derCD23, sequence analysis, molecular dynamics simulations and ITC results point to two calcium binding events in mutant C. Superpositioning mutant C onto the IgE-Fcε3-4 complex reveals another reason why a third calcium ion would be unlikely to bind as shown in Figure 4-15. A conserved arginine residue (Arg 440) in IgE occupies the site

of the third calcium ion and the positively charged side chain would repel the third calcium ion.
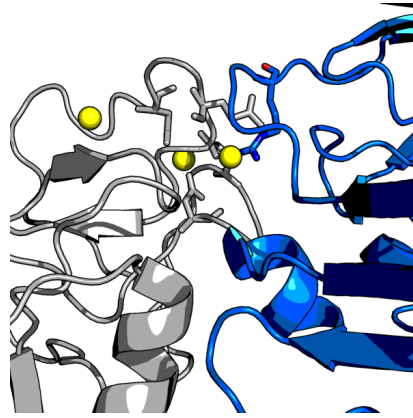
Figure 4-15 **Mutant C modelled with 3 calciums and superpositioned onto the derCD23/IgE-Fcε3-4 complex.** Yellow spheres = calcium ions modelled on DC-SIGNR, pdb file: 1K9J. Grey = mutant C, blue = IgE-Fcε3-4 pdb file: 4GKO.

4.2.11 Predictions on Mouse derCD23 binding to Mouse IgE-Fc$_{3-4}$

Currently, there is no structure of the mouse proteins, thus the impact of loop movement on binding to murine IgE-Fc is only speculative. For both mouse CD23 and mouse IgE, the overall fold of the proteins is similar, but there are some disparities between the sequences of the human and mouse protein. Small changes include H186 mutated to Q186 in humans, and H216 in human CD23 mutated to lysine in mouse, this residue is involved in binding to Thr260 in Cε2 (Dhaliwal et al., 2017). These histidine mutations have a similar backbone and the H216K mutation would not weaken the His216-Thr260 salt bridge. The change from Y189 in humans to F189 still maintains the aromatic nature of the residue, whereas the R224Q mutation found in mice would disrupt the salt bridge between R224 (in human CD23) and E412 (in human and mouse IgE-Fc).

Further analysis of the modelled complex, shown in Figure 4-16 illustrates the mouse structure from the MD simulations and a SWISS Model of mouse-IgE aligned to pdb file 4EZM (human X-ray structure). The most striking difference is the lack of hydrogen bonds (only 4 hydrogen bonds are predicted for the mouse CD23 to mouse IgE interaction), especially those involving residue E412. This finding might significantly reduce the affinity of mouse CD23 to mouse IgE. Mouse derCD23 contains similar electrostatics as human WT derCD23, so it is to be expected that interactions of mouse derCD23 and its ligands, and self-association will be similar to that seen in human WT derCD23. It is likely that the flexible regions and side-chains of both proteins alter in ways that the computer model cannot predict, to still allow for similar affinity interaction.
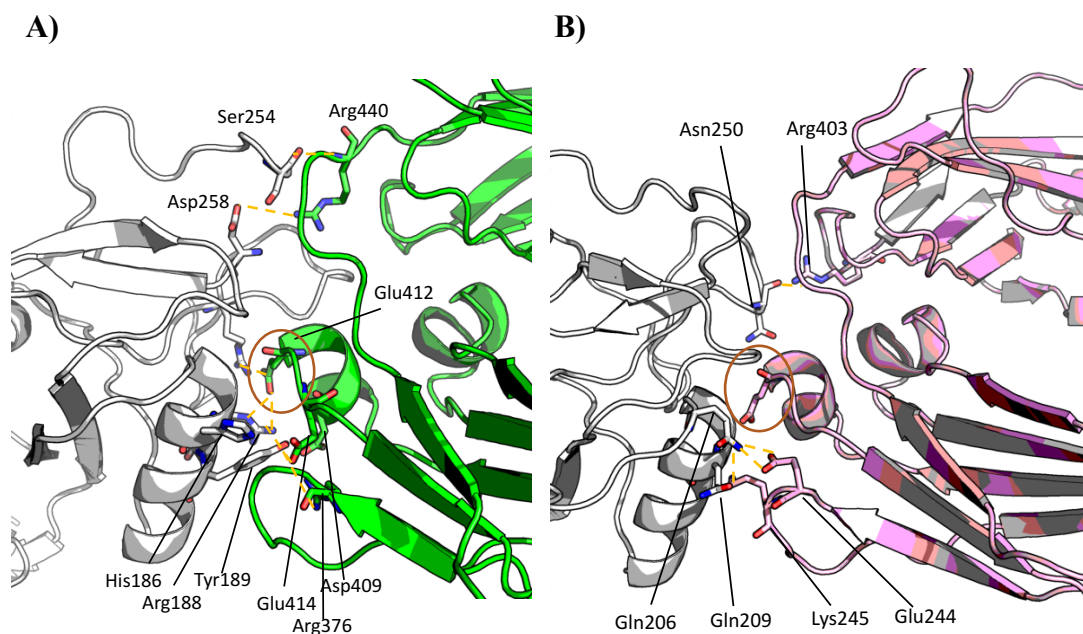
Figure 4-16 **Model of the human derCD23/IgE-Fc3-4 complex and a model of mouse-derCD23 binding to mouse-IgE.** Predicted hydrogen bonds shown with yellow dashes with residues involved in hydrogen bonds labelled. Glu412 is circled in orange. **A)** Human derCD23 (grey) bound to human Fcε3-4 (green). Pdb file: 4EZM. **B)** A model of m-derCD23 (grey) modelled on PDB file: 1S14 and a model of mouse-IgE (pink) modelled on PDB file: 4GKO and aligned to human derCD23-IgE PDB file: 4EZM.

4.3 Discussion

Half of the calcium-facilitated IgE-binding residues in derCD23 are found in loops 1 and 4, with Asn269 and Asp270 of β strand 4 also co-ordinating calcium ions. The proximity of ligand-binding residues and the allostery caused in IgE by calcium-bound derCD23 might be involved in the mechanism that releases IgE-allergen complexes in endosomes and recycles the receptors back to the plasma membrane. This mechanism could be involved in the processes that exacerbate the symptoms of allergies such as facilitated-antigen presentation.

These loop-dwelling residues are difficult to examine experimentally due to a lack of electron density in X-ray structures and poor definition in NMR; thus in this chapter, with the drawbacks experienced in experiments such as poor protein expression, precipitating protein and crystal-packing constraints, these challenges are overcome in MD simulations, which explore the role calcium plays in the highly mobile loops and IgE-binding regions of derCD23. These simulations were run with the WT human derCD23 proteins, as well as the mutations documented in Chapter 3 to see if an *in silico* prediction tool could be built.

The MD simulation systems built were stable, as summarised in Table 4-1 and Table 4-2 and allowed for greater space and protein conformations to be explored as recorded by the extra motions caught by the RMSF in Figure 4-4 and Figure 4-5. Figure 4-6 shows good correlation between the crystallographic B-factors that reflect the mobility of the crystalline protein, and the RMSD of the simulation protein with the simulation showing greater flexibility than the X-ray structure, this is in good agreement with previous findings (Beck et al., 2008).

183

The addition of calcium to the structures does not affect the global conformational flexibility of the protein. Surprisingly, this addition only has a minor effect on the local flexibility of the calcium binding loops as shown in Figure 4-9. Perhaps the averaging of the motions for the analysis has distorted any subtle insights into the protein's dynamics.

Flexibility of the IgE-binding site of derCD23 has been documented with a range of Cε3-Cε4 interdomain angles compatible with CD23 binding (Dhaliwal et al., 2014, 2012; Yuan et al., 2013; Dhaliwal et al., 2013). It is vital to remember than in solution or *in vivo*, IgE would be bending and sampling several of these conformations. Unlike the crystallographically solved structures of derCD23 (Dhaliwal et al., 2013), superimposing the most extreme structures found in the MD simulations on the derCD23/IgE-Fcε3-4 complex, does result in clashes with the antibody as shown in Figure 4-11, Figure 4-12 and Figure 4-14. This suggests that there are some conformations of derCD23 that are not capable of binding directly to IgE and expands on the work reported by Dhaliwal et al.

The effects of the mutations posed a challenge to elucidate in this system, as an equal amount of flexibility is found within the calcium binding loops, suggesting that other compensatory mechanisms are at work. Other *in silico* systems as documented by Yang et al. (Sang et al., 2014)(Yang et al., 2014b), suggest that the effect of individual mutations can be interpreted, but unfortunately these computational studies do not explain the experimental mutagenesis studies documented in Chapter 3.

Despite the MD system not being able to act as an *in silico* prediction model for mutations in derCD23, force fields are constantly improving (Beauchamp et al., 2012; Maier et al., 2015) and using more advanced approaches to MD could result in a greater depth of understanding. For example, a system for derCD23 in which calcium binds, unbinds and binds again could yield binding affinities that could be used to predict ITC results. Greater integration between the MD simulations and experimental findings could lead to virtual tool to screen protein mutant before they are generated *in vitro*.

Despite NMR being an in-solution technique, the loops of derCD23 could not be observed in NMR due to the intermediate exchange timescale while in X-ray crystallography, crystals contacts may occur as documented in Figure 3-18B or poor or no electron density can be found for mobile loop regions. MD simulations was able to provide insight to the loop movements with new conformations being observed in addition to those found by Dhaliwal et al. Understanding this plasticity is an important factor in small molecule drug design and new binding pockets could be exposed etc. Alternatively stabilising an unfavourable IgE-binding conformation of derCD23 could also be a beneficial therapeutic.

Chapter 5 Generation of the surface membrane IgE-Fc display library

5.1 Introduction

IgE in solution is usually found in a bent state (the acute angle between the two-fold axes of symmetry of the $(C\varepsilon 2)_2$ and $(C\varepsilon 4)_2$ domain pairs is at 62°), and is in an even more acutely bent state when bound to its receptor FcεRI (54°) (Holdom et al., 2011). IgE is very flexible, and can 'flip', passing through a fully extended state with the Cε2-4 domains aligned (as shown in Figure 1-5, Drinkwater et al., 2014). This dynamic movement has been proposed to allow the Cε2 domains of the immunoglobulin to flip over from one side of the Cε3-4 domains to the other, allowing the Fabs to occupy the optimal space for allergen binding (Drinkwater et al., 2014). Although, this conformational flexibility may hamper the conventional structure-activity relationship of drug development, it allows possibilities for allosteric intervention (Wurzburg and Jardetzky, 2009; Dhaliwal et al., 2013; Davies et al., 2017).

Although we know the protein interfaces between IgE and its receptors, shown in Figure 5-1 and Figure 5-2, the large range of conformational states and allosteric binding sites makes the rational process of protein engineering challenging. Previous mutational studies, in which individual interface residues were mutated to alanine, had their energetic contributions measured by an SPR assay and resulted in less than a 10 fold effect on the K_D value of the IgE-CD23 interaction (Borthakur et al., 2012). Combining several of these mutations has resulted in an unfolded protein (data not shown). This finding prompted a new mutagenesis approach, which proved less laborious and rendered a higher through-put with a greater chance of success. A library display approach, in which IgE-Fc molecules could be selected, was proposed as a

method to overcome this problem. Therefore, the aim was to design and construct a surface membrane IgE-Fc (memIgE-Fc) mutant library that could be screened to select mutants that could only bind to one receptor, and not the other. This library could then be used to yield research reagents to explore processes such as epitope spreading, or yield potential theranostics.
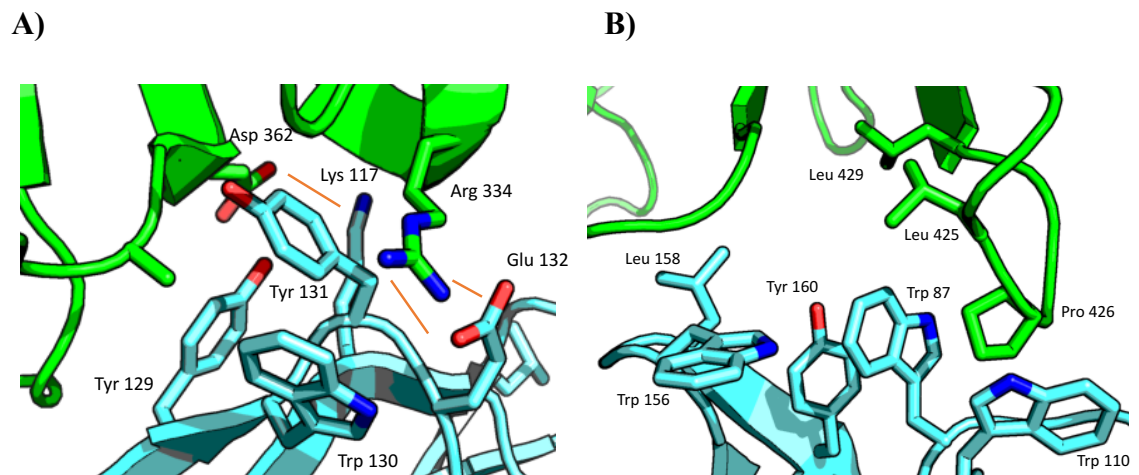
Figure 5-1 **IgE binding to FcεRI; interactions at both subsites.** Green = IgE-Fc, blue = FcεRI, orange = hydrogen bonds. **A)** Cε3 domain binds via two salt bridges (Arg334-Glu132 and Asp362-Lys117) to a chain of FcεRI. **B**) The proline sandwich at site 2, with Pro426 of Cε3 packed between Trp87 and Trp110 of the receptor (blue). PDB file: 1F6A.
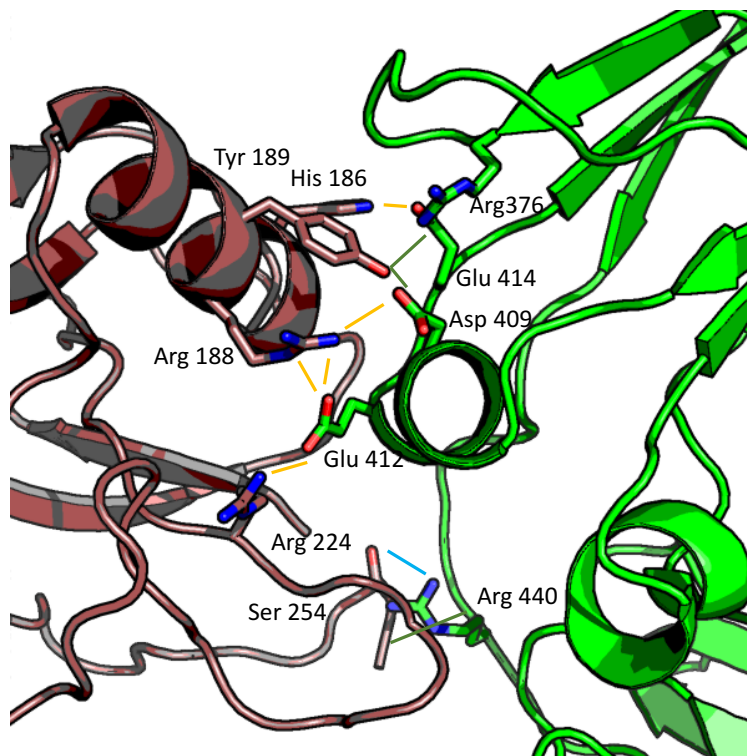
Figure 5-2 **IgE binding to derCD23.** Red = derCD23, green = IgE, yellow = salt bridges, green = hydrogen bonds, blue = a hydrogen bond found in 5 out of 6 molecules in the crystallographic unit cells. PDB file: 4EZM.

5.1.1 Introduction to Display Technologies

Display technologies such as phage display allow for the efficient presentation, screening and selection of protein or antibody fragment libraries on the surface of a range of scaffolds or hosts (e.g. mRNA, bacteriophage, a prokaryotic cell or a eukaryotic cell) against a specific target in a high-throughput manner. Phenotype (i.e. specificity/affinity) is linked to the genotype (i.e. DNA/RNA sequences) to enable complete identification of lead hits. The diversity of display technologies reflects their potency in a range of applications.

mRNA display libraries have the largest potential library size since they can present any protein secretable to the periplasmic space of *E.coli*. Yeast display libraries, although slightly smaller, can display a wider range of proteins including some with yeast-

specific glycosylation. For complete expression of mammalian-specific post-translational modifications and properly folded multi-domain protein, mammalian cells are required. The size of a display library is limited by the transformational efficiency of the cells. Even within mammalian cell display libraries, there is a range of presenting proteins that can be utilised, with the protein of interest (P.O.I) being expressed as a GPI-anchored-fusion protein for example on CHO cells (Armour et al., 2010; Soga et al., 2015), or as a full-length antibody (Akamatsu et al., 2007). Tethering the P.O.I to the cell membrane means that it is less constrained by orientation and the binding of accessory proteins. IgE is already expressed as an existing membrane-bound protein when part of the B-cell Receptor (BCR), in the later stages of B-cell maturation, and thus this construct can be used for the display library. Naïve B-cells express membrane-bound IgM and IgD as part of the BCR and express other isotypes upon activation.

5.1.2 Differences between membrane and soluble IgE

Membrane IgE differs from secreted IgE in that it contains three additional structural features encoded by exons M1 and M2, as shown in Figure 5-3. M1 encodes the extramembrane proximal domain (EMPD), which in humans can be either of the short (IgE_s) or long isoforms (IgE_L). Mice only express IgE with a short isoform, whilst human IgE_s is reported to be 52 amino acids shorter than IgE_L, and has more efficient transport to the plasma membrane (Batista et al., 1996). Findings from some studies suggest that the EMPD is a control element of apoptotic signalling when the B cell receptor (BCR) of an IgE expressing cell is engaged by antigen (Poggianella et al., 2006). The transmembrane and cytoplasmic domain are the other two structural features encoded by exons M1 and M2, while the cytoplasmic tail of IgE has no specific role on its own without its accessory proteins.
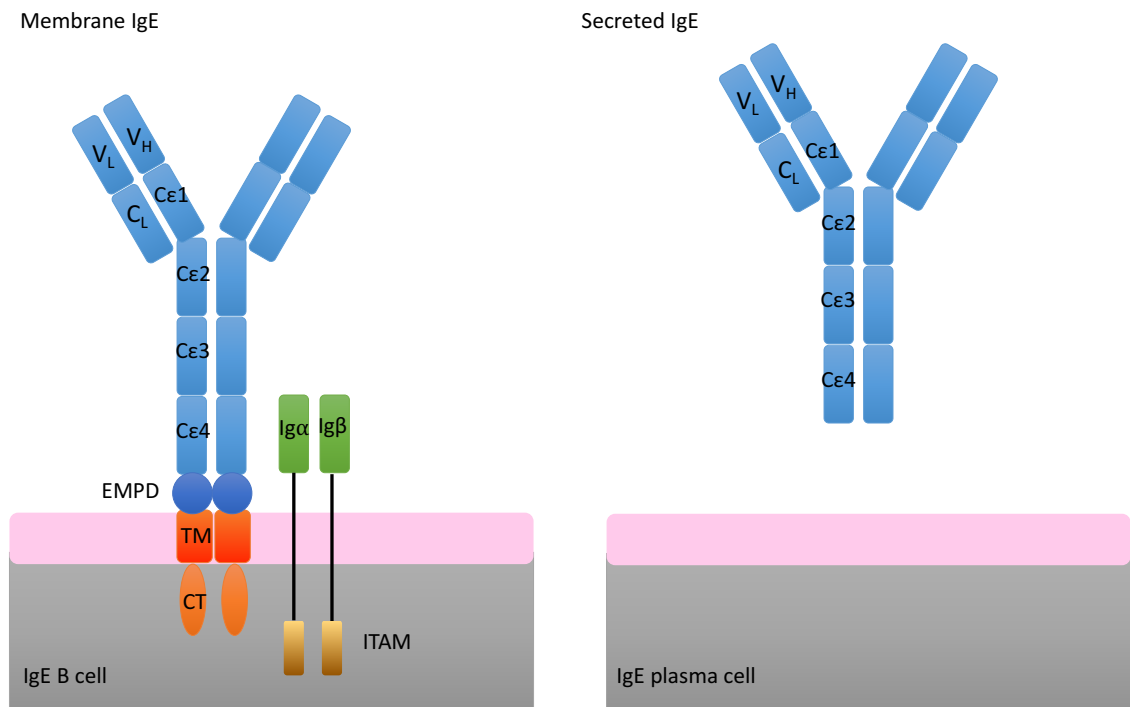
Figure 5-3 **Membrane and soluble Immunoglobulin E.** TM = transmembrane domain, CT = cytoplasmic tail. The two forms of IgE are produced by different cells: class-switched B-cells express membrane IgE while plasma cells secrete IgE.

5.1.3 Chaperone proteins

Membrane-bound IgE forms part of the BCR in which the antibody is obligatorily associated with other membrane proteins: CD79-a and CD79-b (Ig-α/Ig-β) as shown in Figure 5-3. These proteins are not only vital for signalling, but also for transporting and assembling the BCR within the plasma membrane. Experimental evidence suggest that without the extracellular and transmembrane domains of CD79, the BCR will not assemble in a B cell (Dylke et al., 2007); thus it is not clear whether membrane IgE-Fc will express in the plasma membrane of transfected HEK 293 cells alone.

5.1.4 The strategy for the generation of the surface memIgE-Fc display library

The first step in designing a library was to decide on the format and framework that would express well and accommodate mutations. My strategy was to generate a library consisting of the Fc region (Cε2 to Cε4) of the heavy chains solely, as it is this region that binds to the two receptors. This took into consideration the fact that the light chains or the Fab regions of the antibody are not required for binding to the IgE receptors and might obstruct the binding of the selection assay reagents to the membrane-bound protein. In addition, library selection does not depend on antigen capture or signalling, so these parts were omitted in order to simplify the construct; since the light and heavy chains no longer need to pair up, only the heavy chains need to homodimerise. This still presents a level of complexity since the two chains need to dimerise and form disulphide bonds between the Cε2 domains of each chain.

A single mutation in IgE-Fc (R334S) is known to significantly reduce binding of soluble IgE to FcεRI. This mutation is involved in two hydrogen bonds in subsite 1 (Figure 5-1A). R334S reduces IgE affinity for FcεRI ~50-fold and *in vivo* reduces the passive cutaneous anaphylaxis response by 75% (Henry et al., 1997; Hunt et al., 2008). Additional mutations in IgE-Fc acting co-operatively with R334S could further decrease the affinity for FcεRI such that cross-linking of IgE molecules on a mast cell is no longer able to occur, and the allergic cascade is not triggered. Therefore, a stable cell line with the R334S mutation was generated to act as a benchmark against which additional mutant IgE-Fcs could be assessed.

A small-scale surface memIgE-Fc display library was created as a proof-of-principle study, in which two IgE - FcεRI interface residues were mutated to all the other amino

acids. Residue, P426 (Figure 5-1B) at the second IgE subsite was chosen because it is involved in a 'proline sandwich', a dominant feature of FcεRI binding subsite 2. P462 makes several van der Waals contacts with Trp87 and Trp110 of the FcεRI α2 domain (Holdom et al., 2011) and mutating it would disrupt the buried hydrophobic surface area of the binding site. In addition, residue R334 was selected to determine if an alternative amino acid at position 334 (compared to the R334S mutation) may further decrease binding to FcεRI. The two residues were mutated and transfected into FlpIn™ HEK293 cells for membrane surface expression. FlpIn™ cells contain one Flp Recombination Target (FRT) site in a specific genomic location, so that when the cells are co-transfected with the expression vector containing the gene of interest and the vector encoding the Flp recombinase, the expression vector is integrated at the FRT site. It is vital that only one mutant IgE-Fc is displayed per cell otherwise it will not be possible to maintain the phenotype-genotype link. The hygromycin gene in the expression vector contains no promoter nor start codon and therefore it can only be expressed when integrated successfully into the host cell's genome at the FRT site. The FlpIn™ cells enable the generation of a stable cell line in only a month, compared to six months to a year for conventional stable cell line production (Lai et al., 2013).

The pipeline of this mammalian cell surface IgE-Fc display platform is displayed in Figure 5-4.

IgE

IgG

N-linked glycosylation

V_L V_H $C\varepsilon1$ C_L V_H V_L $C\varepsilon1$ C_L
$C\varepsilon2$ $C\varepsilon2$
$C\varepsilon3$ $C\varepsilon3$
$C\varepsilon4$ $C\varepsilon4$

V_L V_H $C\gamma1$ C_L V_H V_L $C\gamma1$ C_L
$C\gamma2$ $C\gamma2$
$C\gamma3$ $C\gamma3$

Preparation of rationally designed library of mutations

Addition of Flp-recombinase and transfection of mammalian Flp-In HEK293 cells

Integration and expression of membrane IgE-Fc under growth selection

Selection of proteins by FACS sorting

Recovery of Sequences

Further characterisation of proteins

Flp-In host cell

FRT Zeo

Hygromycin GOI FRT

pc DNA5/FRT with GOI

FLP

pOG44

+

FRT Hygromycin GOI FRT Zeo

10^0 10^1 10^2 10^3

BL1-A :: –Fl

5-4 **The mammalian cell surface IgE-Fc display platform pipeline.** Mutations are introduced into the expression vector by degenerate primers. The library of expression vectors is then co-transfect with pOG44, which expresses the recombinase necessary for creation of the stable cell line, into Flp-In HEK293 cells. This bipartite vector system generates a stable high expressing cell-line within one mon Stably transfected cells are positively selected by hygromycin. Single cell clones displaying one type mutant per clone with binding specific to one IgE-receptor are then isolated by flow cytometry employi a fluorescently labelled receptor as bait. Hits are then further characterised.

5.2 Results

5.2.1 Design and Construction of the membrane IgE vector

Proteins destined for the secretory pathway such as IgE usually contain a signal peptide at their N-terminus, which directs them to the correct pathway for secretion. The native signal sequence of IgE was chosen (GenBank: AAB59424.1, Figure 5-5A), as IgE is naturally secreted by plasma cells, which should be sufficient to shuttle the protein to the cell surface, rather than being directed to the proteasome for degradation.

In previous studies carried out by our laboratory, the soluble form of IgE-Fc had two of its surface glycosylation sites removed to ease crystallisation (the N265Q, N371Q construct)(Young et al., 1995). These glycosylation sites were re-introduced to the new construct to enable proper folding of the ectodomain of the protein. This is especially important to IgE since it is the most glycosylated antibody (Plomp et al., 2014). Another factor taken into consideration was the expression of the two human memIgE isoforms, which differ in the length of their EMPD. The short form of memIgE was chosen as it has been shown to be more stable and express better than the memIgE long form (Batista et al., 1996).

No changes were made to the native transmembrane domain sequence or the cytoplasmic tail, which acted as a linker for GFP and enabled space for the β-barrel protein to fold. Enhanced GFP (eGFP), which is 35 times brighter than WT GFP (Zhang et al., 1996), folds fast and matures quickly, was chosen as the transfection marker and to also provide semi-quantitative analysis for the FACS selection assay (see section 200). The expected protein construct and its vector is displayed in Figure 5-5B while the anticipated expressed protein is shown as a cartoon in Figure 5-5C.
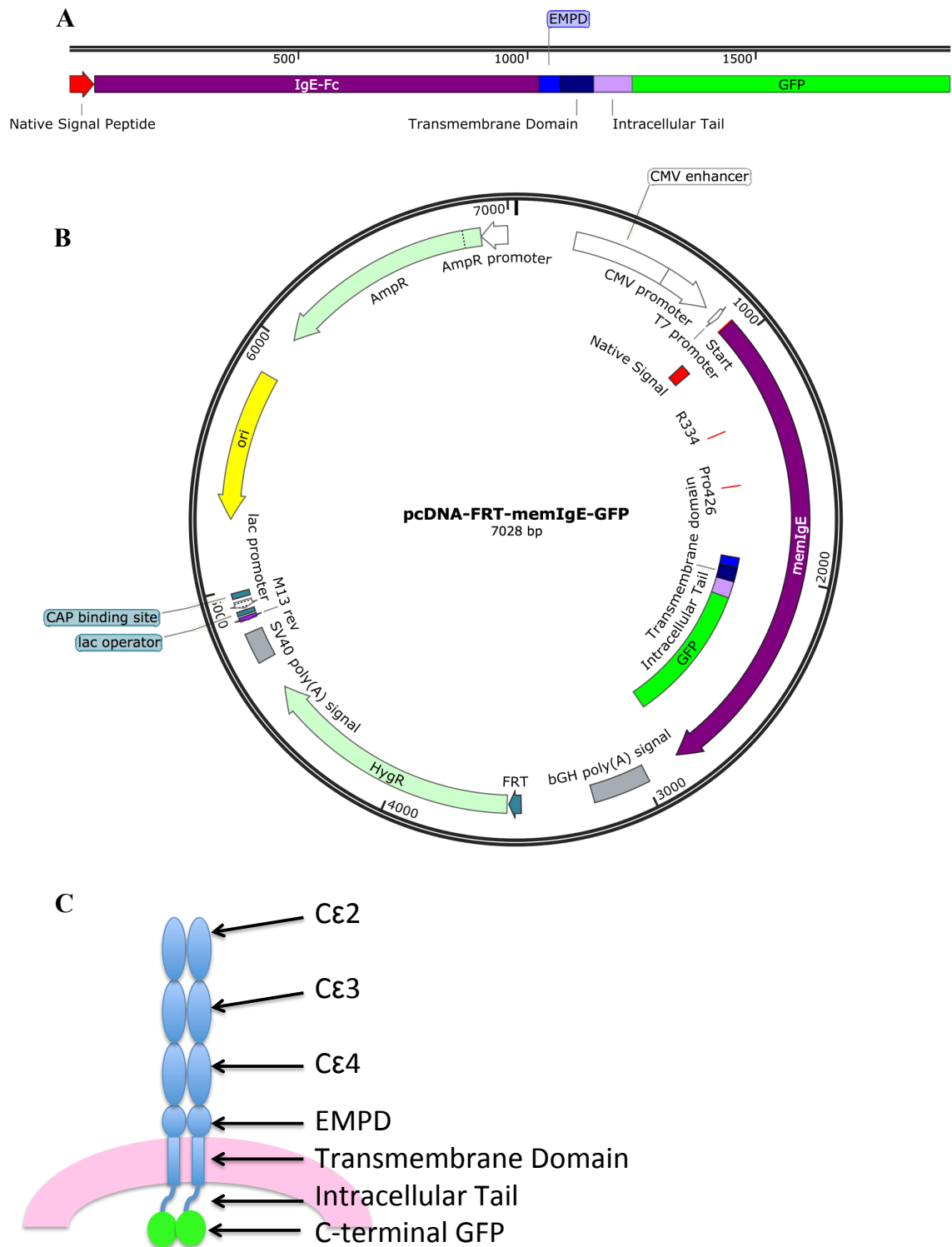
Figure 5-5 **A map of the memIgE-Fc-GFP construct and protein**. **A)** An overview of the designed protein. 1923 bp in length, numbers under the black line correspond to bp. Red = Native signal peptide, purple = IgE-Fc, blue = extramembrane-proximal domain (EMPD), navy = transmembrane domain, lilac = intracellular tail and green = GFP. **B)** The new Flp-In vector containing the designed protein. Features labelled and the two residues to be mutated are highlighted in red as R334 and Pro426. Images created in Snapgene. **C)** A cartoon representation of the proposed protein drawn linearly for ease. Blue = IgE protein, pink = cell membrane, green = GFP.

5.2.2 Proof-of-principle library generation – mutating the memIgE-Fc protein using the trick-22 method

This proof-of-principle library focused on mutating two residues, P426 and R334, within the Fcε region of the heavy chains, involved in IgE-FcεRI binding. The residues at these positions were mutated to all the other amino acids to create a library size of potentially 400 different memIgE-Fc mutant proteins for screening against selective binding to either CD23 or FcεRI. To create a 'smarter' library, the '22-c trick' was employed to reduce codon redundancy. The 22-c trick is so-called because it consists of a 22:20 codon to amino acids ratio as illustrated in Figure 5-6, and uses a mix of three degenerate-base containing primers (detailed in section 2.1.14) to create intelligent libraries and reduce subsequent screening efforts (Kille et al., 2013; Sanchis and Fernández, 2008). A mammalian display library would be smaller than a phage display library, hence it is important to maximise the sampling of the library. If 15 amino acid positions were chosen and mutated to all the canonical amino acids, this would result in $20^{15} = 3.27 \times 10^{19}$ possible combinations, too many to sample or transfect efficiently.
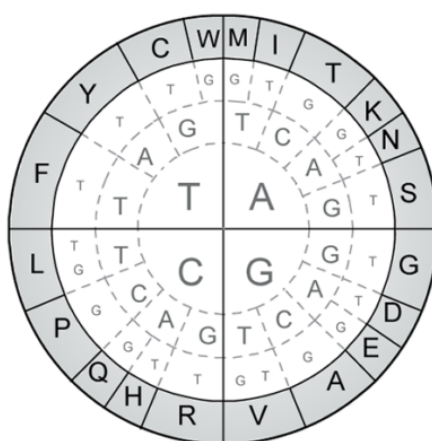
Figure 5-6 **An example of the redundancy of the genetic code using in the 22-c trick.** Codons are read from the most inner circle to the outside. The 20 amino acids for which the 22 codons code for are presented in one letter code in the outer grey shell. Image adapted from (Kille et al., 2013).

5.2.3 Validation of the surface memIgE-Fc display libraries at the DNA level

The mutational rate of the library was assessed using the semi-quantitative quick quality control (QQC) method, for which twenty colonies were randomly chosen and sequenced (Acevedo-Rocha et al., 2015). It was vital to check that the WT clones were not over-represented as this would reduce library quality. Figure 5-7 displays the results of the QQC. For the first nucleotide chart a distribution of 27%(A), 19%(T), 27%(G) and 27%(C) is expected, as detailed in section 2.1.15. R334S clones corresponded closely to the expected distribution with a slight bias for the wild type codon. A similar pattern was observed for the third position (expected 55%(T) and 45%(G)). The second nucleotide is expected to have a distribution of 32%(A), 32%(T), 18%(G) and 18% (C) which was not observed by either set of clones. Overall, the primers appeared more efficient for the R334 residue, which showed greater diversity than the P426 residue, as displayed in Figure 5-7.
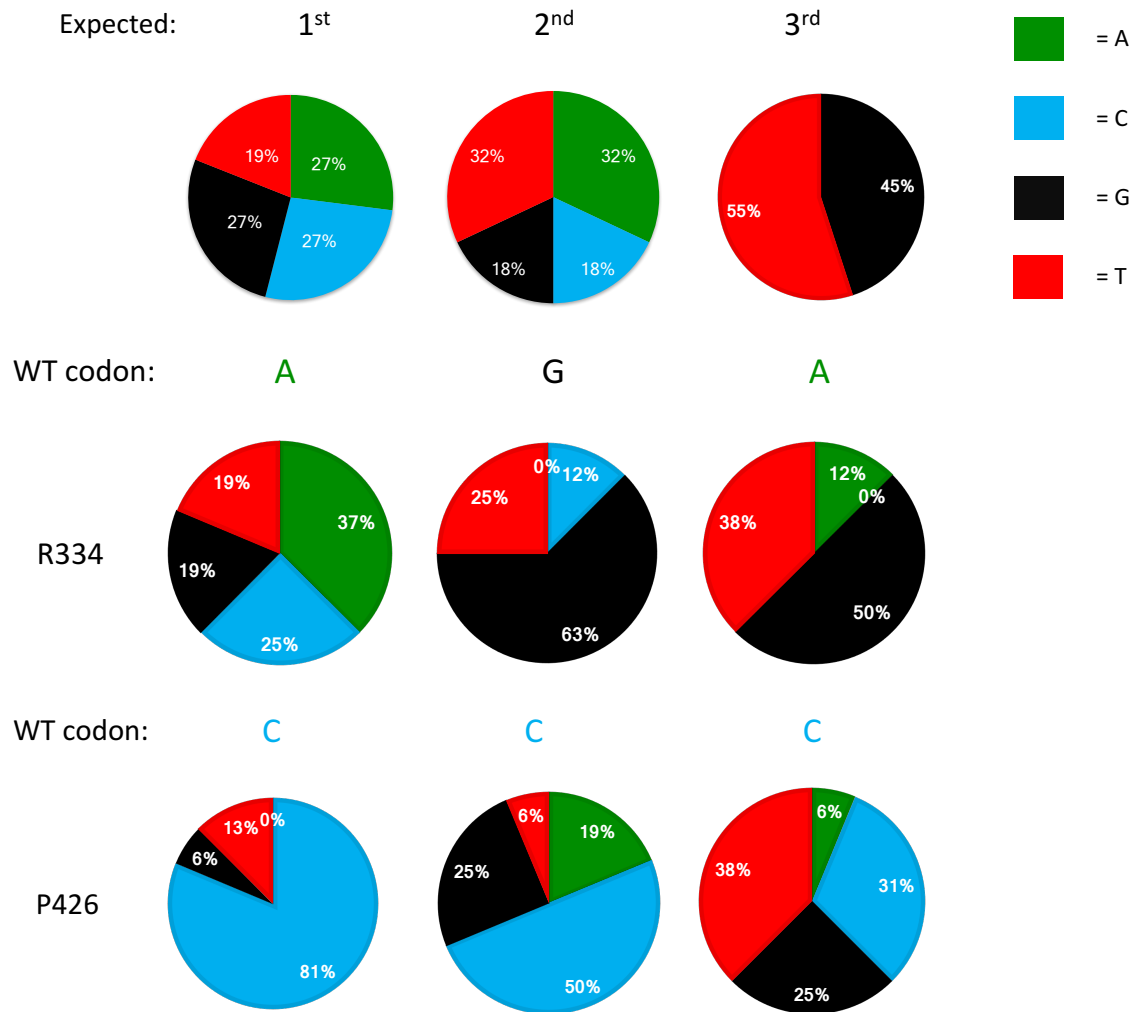
Figure 5-7 **Quick quality control of the library generated with the 22c-trick.** Obtained distribution of nucleotides for residues R334 and P426. The wild-type codon is presented above the pie-charts.

5.2.4 Generation of the membrane IgE cell line

FlpIn™HEK293 cells were co-transfected with the constructed plasmid library as described in Figure 5-5, and pOG44, which codes for the Flp recombinase. After selection with hygromycin, successfully transfected cells expressed GFP, whereas the non-transfected cells either turned necrotic or did not express the GFP of the protein construct. The green cells in Figure 5-8 show that the designed protein is being expressed and that at least the intracellular part folded correctly. To test whether the

extracellular portion of the membrane protein has folded, the cells were stained with anti-human-IgE polyclonal antibodies and analysed by two-colour flow cytometry.
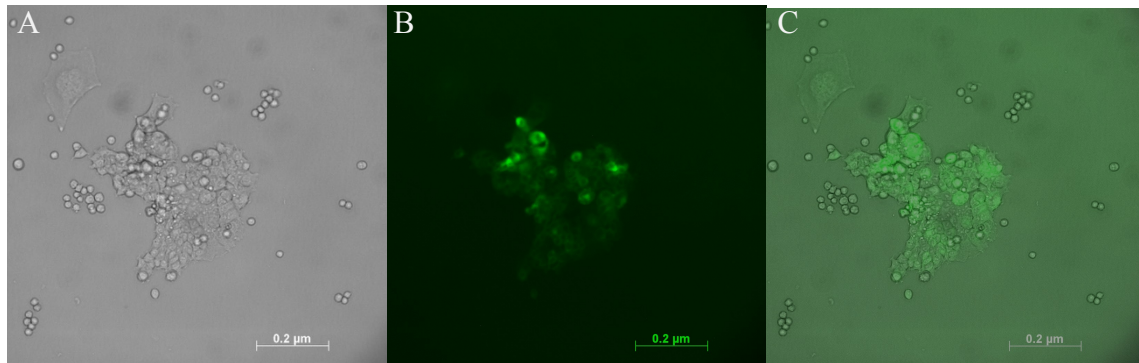
Figure 5-8 **Successful Flp-In transfections. A)** Brightfield **B)** GFP and **C)** merged epifluorescence images of successfully transfected cells in 6-well plates growing in phenol-red containing media, imaged at 20x magnification on the Axiovert 200 M inverted microscope. The necrotic, untransfected cells do not express GFP. Live cells express GFP-tagged IgE-Fc.

Figure 5-9 shows double positive cells for GFP and anti-human IgE PerCP Cy5.5, indicating that the polyclonal antibodies recognise the IgE-Fc and thus it can be concluded that the designed protein has been accurately targeted to the cell membrane and is fully folded. Non-transfected FlpIn™HEK293 cells did not express GFP, and the polyclonal anti-IgE antibodies did not stain for IgE either.
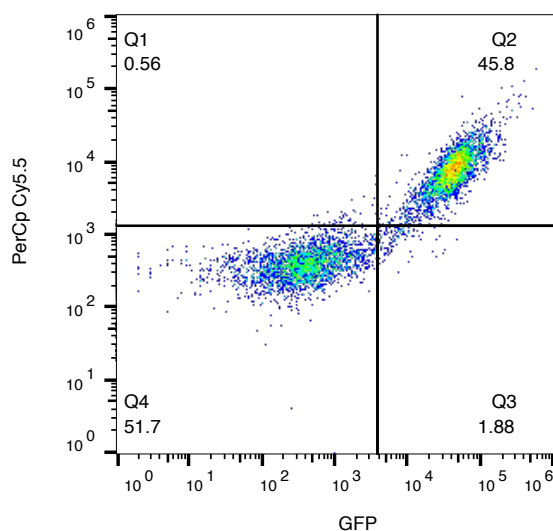
Figure 5-9 **Validation of the WT memIgE-Fc-GFP cell line.** Cells were gated for live cells first, based on forward and side scatter. Transfected FlpIn ™ cells in Q2 express GFP (cytoplasmic part of the protein construct is folded correctly and functioning) and bind to anti-human IgE polyclonal antibody (cells express the fully folded ectodomain). In Q4, untransfected FlpIn™ cells do not express GFP and do not stain for PerCP 5.5.

5.2.5 Production of selection assay reagents: derCD23-Alexa488

The first reagent to test was the fluorescently labelled WT human derCD23 protein, the production of which is detailed in section 3.2.1, and the fluorescent labelling detailed in section 2.2.26. Monomeric WT human derCD23 has a relatively low affinity for IgE with a K_a of 10^6 -10^7 M^{-1}. It was not known whether the protein would have to be an oligomer to work as a reliable flow cytometry reagent since the avidity effect increases the affinity of trimeric CD23 to a K_a of 10^8 -10^9 M^{-1}(Gould and Sutton, 2008). Figure 5-10 displays the amine-coupled WT derCD23-Alexa488 protein used in the pilot study.
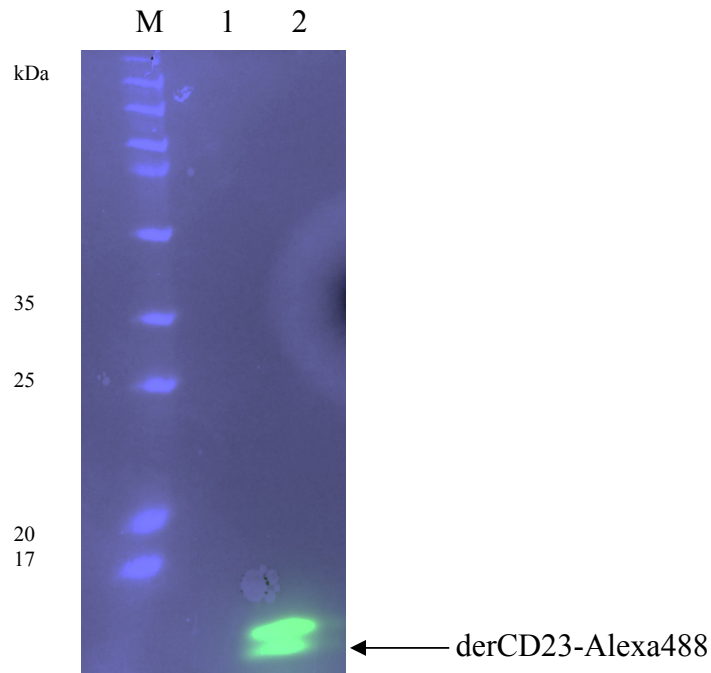
Figure 5-10 **Labelling of derCD3 with Alexa488.** 15% SDS-PAGE captured with multichannel fluorescence; pale blue = brightfield channel to capture markers, green= Alexa488 channel. Lane 1 = WT protein, lane 2 = derCD3-Alexa488, expected size = 17 KDa.

While the transfections were in progress, a pilot study was performed with WEHI cells (Batista et al., 1996), which are mouse B-cells able to express full-length short isoform of membrane-bound human IgE, to test binding of the derCD23-Alexa488 to memIgE. Figure 5-11A demonstrates how monomeric derCD23-Alexa488 did not have sufficient affinity to bind to the membrane-IgE-Fc protein.

To address this problem, a trimeric CD23 protein (triCD23) was produced, and conjugated to Alexa647, which was then bound to WEHI cells and these were tested by flow cytometry. The next section describes the production of triCD23, whilst Figure 5-11B demonstrates that triCD23 conjugated to Alexa647 was able to bind to WEHI cells. The larger shift to the right gives a larger window to observe any decreases in binding as a result of mutation, which could be further optimised.
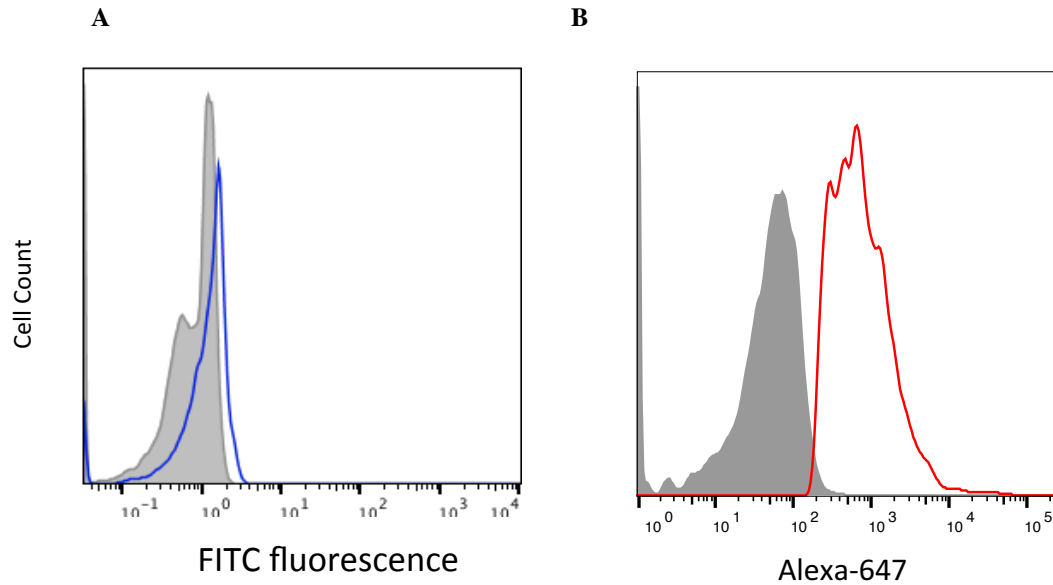
Figure 5-11 **Binding of Alexa647-conjugated monomeric derCD23 and trimeric triCD23 to WEHI cells.** Grey = unstained cells, coloured lines = stained cells. **A)** Cells stained with monomeric derCD23-FITC. **B)** Cells stained with triCD23-Alexa647.

5.2.6 Production of selection assay reagents: Expression, refolding and purification of recombinant triCD23

Soluble triCD23 contains the C-type lectin domain and an engineered stalk with a synthetic trimerisation sequence (Cooper et al., 2012). TriCD23 has a greater affinity for IgE-Fc than derCD23 due to its avidity (Shi et al., 1997; Cooper et al., 2012; Borthakur et al., 2012). The pET151-GA-tri vector was kindly provided by Rebecca Beavil and BL21 star DE3 cells were transformed with the vector for protein expression. The expression test in Figure 5-12 shows an induced band at 30 kDa, which is not seen in the supernatant lane, indicating that the protein is expressed in inclusion bodies.
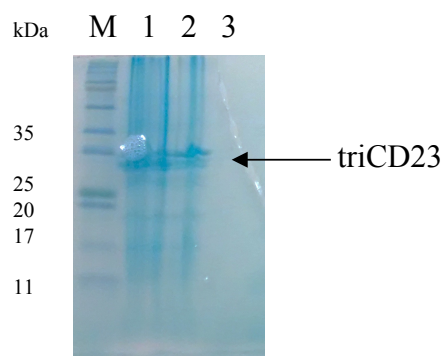
Figure 5-12 **15 % SDS-PAGE expression test of triCD23 stained with Comassie stain.** M = Marker, 1 and 2 = insoluble fraction, lane 3 = supernatant. Samples are reduced, and a band representing triCD23 can be observed in 1 and 2 indicating that triCD23 is expressed in inclusion bodies.

Next, triCD23 was refolded in a buffer similar to the refold buffer of WT human derCD23, except that imidazole was added to the refold buffer since the protein is His-tagged. During the refold, the protein was diluted much less than the monomeric derCD23 so that it could trimerise. The protein was purified by passing it over a Histrap column and concentrated prior to further purification by size exclusion chromatography (SEC). Fractions for peak PN4 were pooled and concentrated as shown in Figure 5-13. This resulted in highly pure protein as displayed in Figure 5-14.
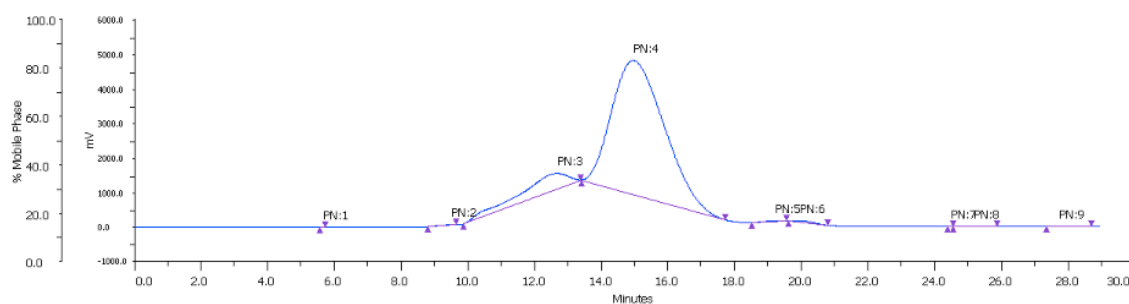
Figure 5-13 **Size exclusion chromatographic profile of triCD23 using a Superdex 200 column.** Fractions for peak PN4 were collected on a Superdex 200 column equilibrated with 0.5M Tris, 100 mM NaCl pH 7.5, flow rate = 0.75ml/min.
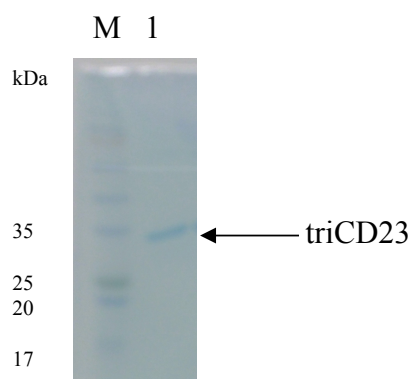
Figure 5-14 **15 % SDS-PAGE of purified triCD23 stained with Comassie stain.** M = marker, lane 1 = purified reduced triCD23, expected size = 28 kDa.

5.2.7 Conjugating fluorophores to triCD23 and the αγ-fusion protein

TriCD23 was conjugated to Alexa-647, a photostable red dye with a high quantum yield. This fluorophore was chosen in order to minimise the possibility of spectral overlap with the transfected GFP. Figure 5-15 displays the spectra of Alexa647 and GFP (expressed intracellularly by the cells) with the optical configurations of the flow cytometer.
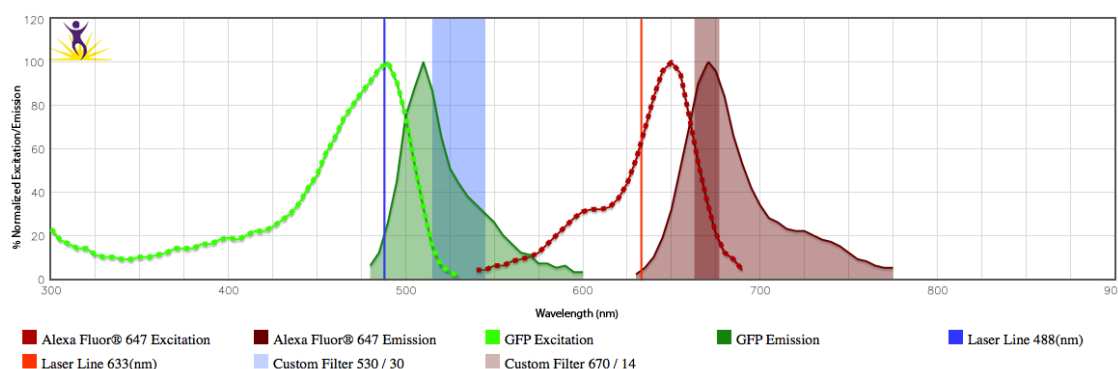
Figure 5-15 **Fluorophore spectrum.** Created by Biolegend website. Dotted and solid lines represent the excitation and emission spectra, respectively. Solid vertical lines represent the laser line while the thicker shaded vertical lines represent the flow cytometer's filters. The emission spectrum of GFP and Alexa647 do not overlap.

Alexa 647 was conjugated to triCD23 by amine coupling (Figure 5-16). In addition, IgG-sFcεRIα, the extracellular region of FcεRIα fused to an IgG4-Fc (αγ-fusion protein; kindly provided by Rebecca Beavil) (Shi et al., 1997) was also conjugated to Alexa647 by amine coupling. To optimise the degree of labelling with the αγ-fusion protein, varying molar excesses of dye were tried as documented in Figure 5-16.
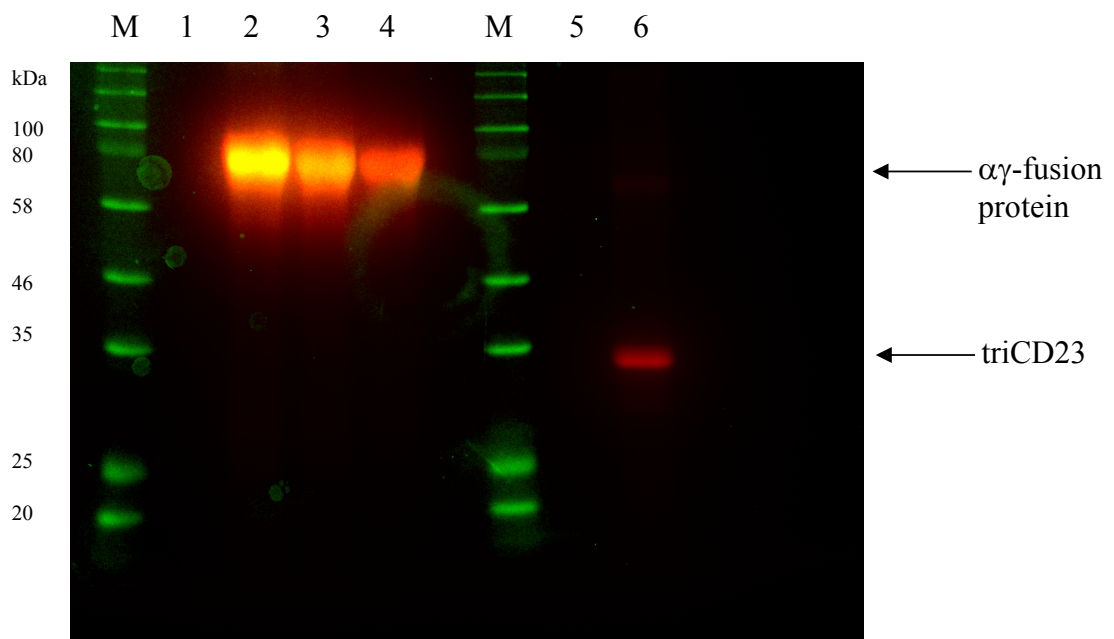
Figure 5-16 **Labelling of αγ-fusion protein and triCD3 with Alexa647.** 12% SDS-PAGE gel captured with multichannel fluorescence; green = brightfield channel to capture markers, red= Alexa647 channel. 1 = WT αγ-fusion protein, 2 = 20x molar dye excess, 3 = 10 x molar dye excess, 4 = 5x molar dye excess, 5= WT triCD23, 6 = labelled triCD23.

The αγ-fusion protein-Alexa647 was stable (no degradation bands in Figure 5-16) and had expressed at high yield. In contrast triCD23 had a low yield and a longer production time. Subsequent experiments therefore employed the αγ-fusion protein to assess the IgE-FcεRI binding, which involves the R334 binding site.

5.2.8 Validation of selection assay reagents: αγ-fusion protein-Alexa647 and FcεRI-Alexa647

The IgE-FcεRI binding site had two potential assay reagents, the αγ-fusion protein-Alexa647 and the labelled high affinity IgE binding receptor, FcεRI. Both of these reagents were tested on WEHI B cells expressing the short isoform (Batista et al., 1996) by flow cytometry as shown in Figure 5-17. Both Alexa-647 labelled reagents could bind to IgE expressed on the WEHI B-cells suggesting that membrane IgE spends more time bent, rather than fully extended on the cell surface for it to accommodate binding of these reagents. The αγ-fusion protein-Alexa647 showed greater binding to the cells in panel B of Figure 5-17 and thus was chosen for further experiments.
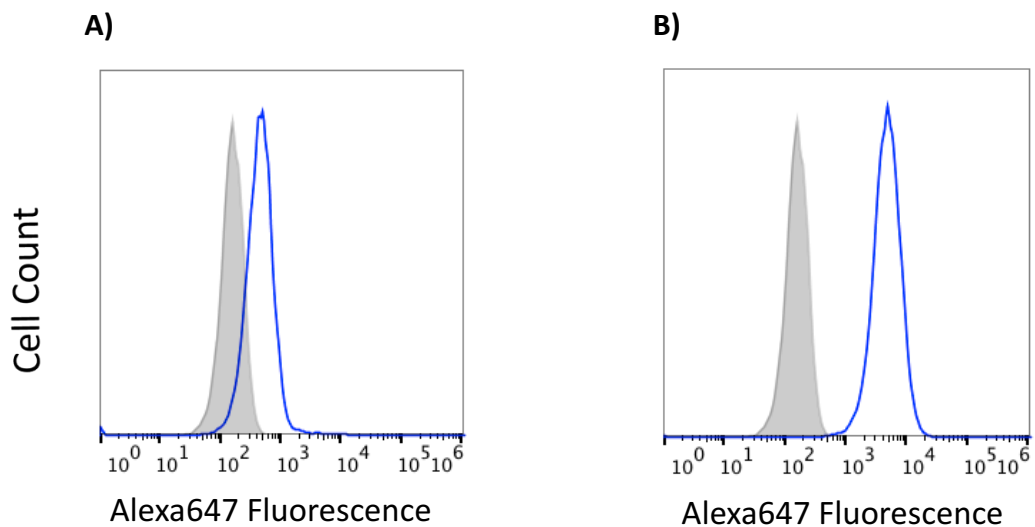
Figure 5-17 **Binding of Alexa64-conjugated FcεRI and αγ-fusion protein to WEHI cells.** Grey = unstained cells, blue line = stained cells. **A)** Cells stained with FcεRI-Alexa647 **B)** Cells stained with αγ-fusion protein-Alexa647.

5.2.9 Validation of the WT memIgE-Fc-GFP and R334S memIgE-Fc-GFP cell lines

To analyse the memIgE-Fc-GFP library, two main control cell lines were required. These were the WT memIgE-Fc-GFP cell lines, and a known effective mutation: R334S (the R334S memIgE-Fc-GFP cell line), which served as a benchmark for the library. All the cells were positive for GFP. Figure 5-18 shows the addition of the Alexa-647 conjugated αγ-fusion protein to the cell lines and a clear difference can be seen in the histograms with the R334S cell line mean fluorescence intensity (MFI) binding 4x less well than the WT cell line. In previous studies, it is known that the R334S mutation is responsible for a 50-fold decrease in affinity for FcεRI (Hunt et al., 2008) so any cells from the library that express a better than a 4-fold decrease in MFI would be expected to have a greater than 50-fold decrease in affinity for FcεRI.
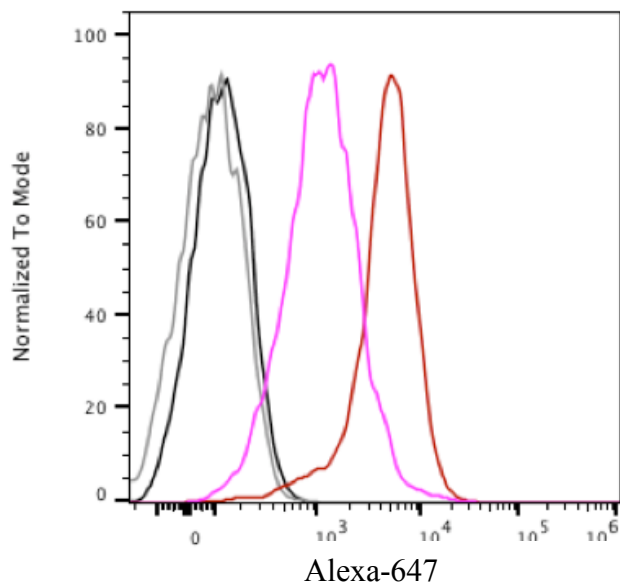
Sample	MFI - Alexa647
WT cells only	121
R334S cells only	84.8
R334S cells and αγ-fusion protein-Alexa647	1092
WT cells and αγ-fusion protein-Alexa647	4695

Figure 5-18 **Histogram of αγ-fusion protein-Alexa647 binding to the WT and R334S cell lines.** Red = WT memIgE-Fc-GFP cell line with αγ-fusion protein-Alexa647 at 1 ug/ml, magenta = R334S memIgE-Fc-GFP cell line with αγ-fusion protein-Alexa647 at 1 ug/ml, grey = R334S memIgE-Fc-GFP cells only, black = WT memIgE-Fc-GFP cells only. **Table below:** The MFI for each sample.

5.2.10 Selection assay of memIgE-Fc-GFP library and its validation

The selection assay for the memIgE-Fc-GFP library is a flow-cytometry based assay in which cells are first gated on GFP and then binding to Alexa-647 labelled IgE-receptors is measured. A typical analysis method is shown in Figure 5-19. The memIgE-Fc-GFP mutant library was pooled and treated as one sample to determine the scope of the

library with results shown in Figure 5-20. The FlpIn™ untransfected cells (red) do not express the memIgE-Fc-GFP and thus no GFP or Alexa647 MFI was measured. The WT memIgE-Fc-GFP cells (grey) bind to αγ-fusion protein-Alexa647 with the highest MFI of 40963. The library (green) shows distinct populations which bind to αγ-fusion protein-Alexa647 protein with apparent lower affinity than R334S memIgE-Fc-GFP cell line (Figure 5-20). The library has one negative peak that does not appear to bind to αγ-fusion protein-Alexa647 protein. Another broader peak binds with reduced MFI compared to the R334S memIgE-Fc-GFP cell line, and overall the library had a 6-fold reduction in binding compared to the WT cell line.
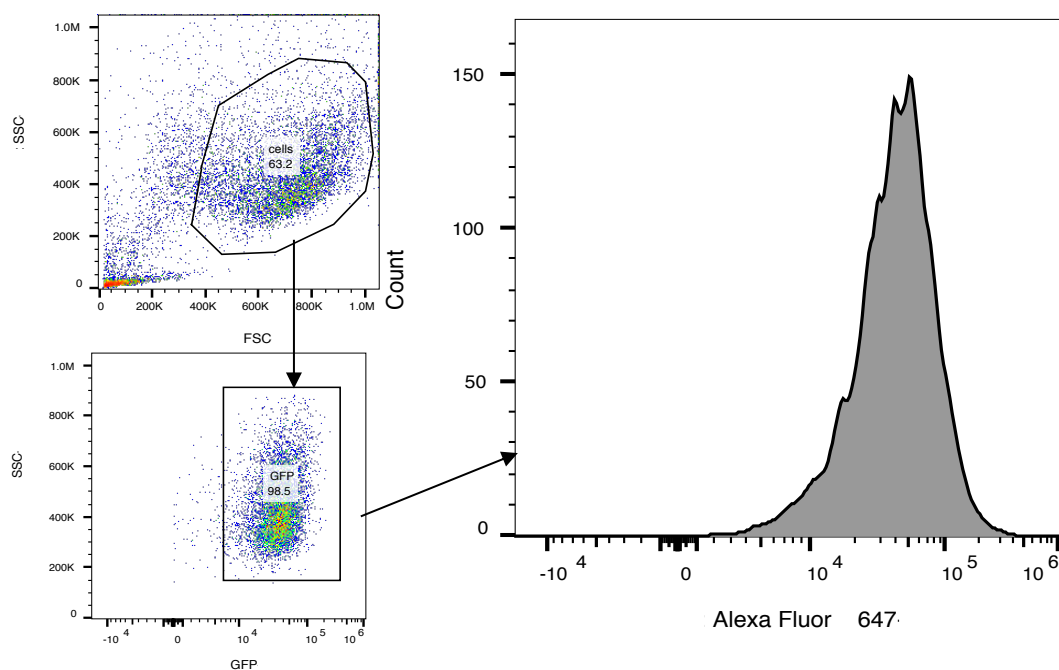
Figure 5-19 **Representative Flow Cytometry analysis with gating and selection strategies displayed.** Live WT memIgE-Fc-GFP cells are gated of which only GFP positive cells are chosen for further analysis. The main graph displays a histogram of the Alexa647 positive cells indicating that they are expressing the memIgE-Fc protein and can bind to αγ-fusion protein-Alexa647.
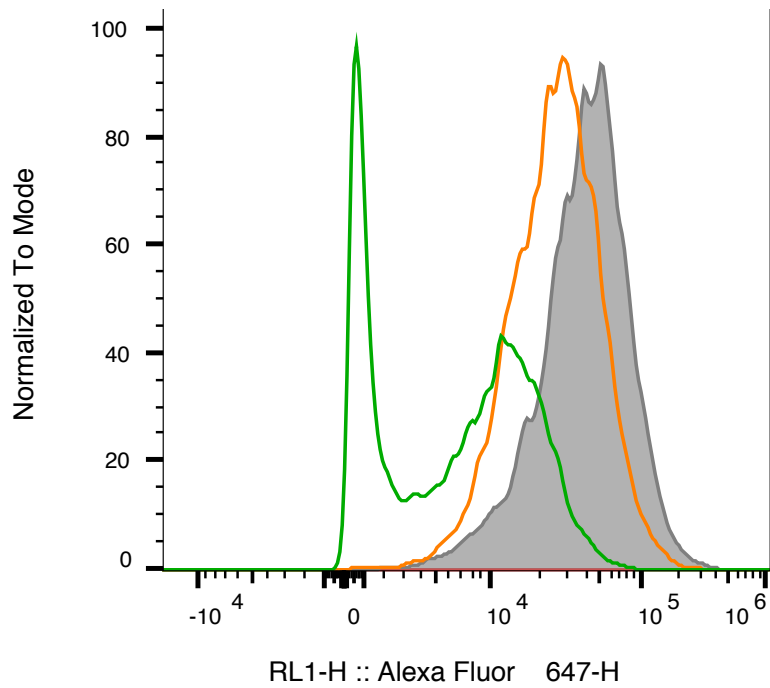
Colour	Sample Name	Subset Name	Count	MFI – Alexa647
	memIgE-Fc-GFP mutant library	GFP	5861	7474
	R334SmemIgE-Fc-GFP cells	GFP	6469	25680
	WTmemIgE-Fc-GFP cells	GFP	6222	40963
	Untransfected FlpIn™HEK cells	GFP	0	n/a
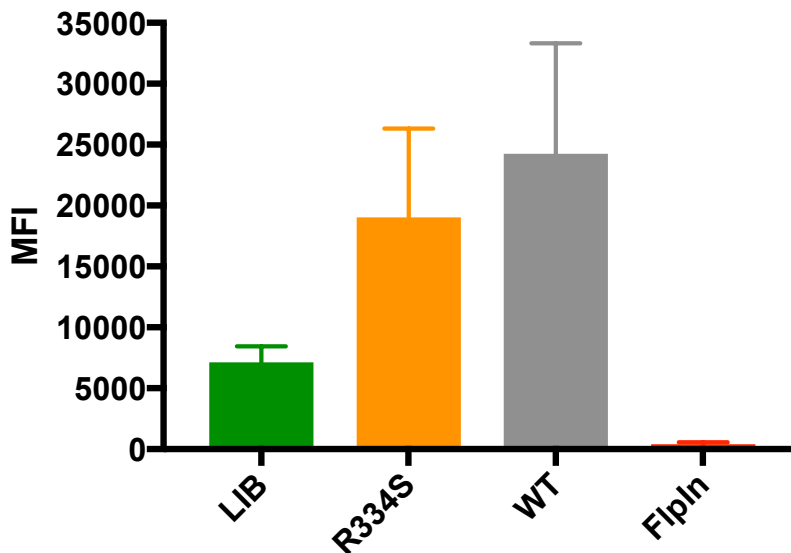
Figure 5-20 **Histogram values of αγ-fusion-protein-Alexa647 binding to the WT, R334S and library memIgE-Fc-GFP proteins and the untransfected FlpIn™HEK cells.** Grey = WT memIgE-Fc-GFP cells, orange = R334SmemIgE-Fc-GFP cells, green =mutant library memIgE-Fc-GFP cells, red = untransfected FlpIn™HEK cells. MFI counts listed below histogram. Table displays MFI values, error bars = standard error of mean (SEM), n =3.

5.3 Discussion

Despite knowing the residues involved in binding at the IgE-FcεRI and IgE-CD23 interface, the exact roles of both IgE receptors in the normal physiology of IgE is not complete. The mechanisms of antigen presentation and epitope spreading that involve the IgE receptors could also be improved upon. The problem is further compounded by the highly flexible nature of IgE since it can twist from an acutely bent conformation (in which the constant domains are bent over one another) to an extended one (where the constant domains are fully extended) (Drinkwater et al., 2014; Holdom et al., 2011). These conformational changes result in two spatially separate receptor binding sites which are allosterically linked (Dhaliwal et al., 2012). These structural features have complicated the development of IgE inhibitors, since biologics are intrinsically complex and allosteric proteins are deemed a challenging target for small molecules. Targeting individual residues at the IgE receptor-binding interface has resulted in either little change in binding affinity or unfolded protein (Borthakur et al., 2012).

To overcome these challenges, the Fc portion of IgE itself was engineered to create a library of IgE-Fc mutants, and a selection assay was developed and validated to identify single receptor-binding mutants. This mammalian cell-surface display library is the first to employ engineering of antibody constant domains with respect to receptor binding. Previous mammalian cell-surface antibody display libraries, despite expressing full-length antibodies, have focused on expanding the diversity of variable domains (Waldmeier et al., 2016; Zhou et al., 2010; King et al., 2014). Those studies have also focused on IgG, which has a very different conformation in solution compared to the acutely bent IgE (Zheng et al., 1992) and has fewer glycosylation sites than IgE. This mem-IgE-Fc-GFP library also maintains the linked phenotype to genotype feature found in phage display since each cell only expresses one type of mutant Fc, and the

conservation of this feature in this display system is vital for the enrichment of proteins or the directed evolution of lead proteins.

The size of mammalian cells (~100x larger than phage) impose a limit on the size of the library, and thus when creating a small library, it is vital to maximise diversity. The approach to make a 'smarter', less empirically-driven library was to utilise the '22c-trick' to limit codon degeneracy so that each amino acid was represented more evenly (Kille et al., 2013). Figure 5-7 indicates the diversity of the memIgE-Fc-GFP library, which was more successful for the R334 site than the P426 site, although this could just reflect the small sample size. The degenerate primers could have had a bias while being manufactured and the mutagenesis protocol could be optimised further by altering annealing temperatures of the primers or by changing the ratio of the degenerate primers in the PCR mix. This small proof-of-concept study has however produced a maximal protein diversity size of 400.

For these studies, several stable cell lines were generated: WT memIgE-Fc-GFP and R334S memIgE-Fc-GFP as controls and the memIgE-Fc-GFP DNA library was successfully generated by a batch transfection. Figure 5-20 demonstrates the library contains mutant proteins with a range of apparent affinities for αγ-fusion-protein-Alexa647, however, a non-uniform distribution was observed.

The two-colour FACS selection assay designed here to analyse the stable cell lines is adaptable and alternative fluorophores could be used. If the emission and excitation spectra do overlap then it would be best to create control cell lines, similar to that of flow cytometry compensation beads. In reference to Figure 5-5C, this would entail expressing only the transmembrane domain and the GFP portion of the protein and none

of the extracellular protein to create an internal standard. Figure 5-16 shows how different molar dye ratios were tested; this is important to obtain a greater sample range in the flow cytometry (Figure 5-18) so that reductions in binding can be observed more clearly, but it is also important to note that a brighter labelled protein does not necessarily mean a better reporter, since the dyes could conjugate to residues important for binding.

When generating the selection assay reagents, it became clear (Figure 5-11) that the best method of increasing the affinity of CD23 would be to increase the avidity of the protein. There are many approaches to oligomerising proteins, such as biotinylating the protein and joining it to streptavidin, however, repeated need to purify the protein and dialyse it after each step would significantly reduce protein yield. The best solution was to produce a soluble trimeric CD23 protein and conjugate it directly to a fluorophore, reducing the number of processing steps and maximising yield of the reagent. A further possibility to optimise the reagent assay would be to redesign the protein construct and incorporate a fluorescent protein tagged to the terminus of the protein to avoid the need for conjugation.

The potential assay reagents for the Fcε-FcεRI receptor binding site demonstrated a clear preference for αγ-fusion protein-Alexa647 over FcεRI-Alexa647 (Figure 5-17). This might be due to the heterogeneous glycosylation patterns of FcεRI-Alexa647, which is heavily glycosylated (Kanellopouloss et al., 1980) although not a requirement for binding to IgE, which is based on protein-protein interactions (Garman et al., 2000). The carbohydrates on FcεRI are required for efficient folding and solubility of the receptor (Robertson, 1993; Letourneur et al., 1995). In these results the FcεRI-Alexa647 may have caused greater steric hindrance than that of the αγ-fusion protein. Another

213

reason for the better binding observed by the αγ-fusion protein is that it is a dimer, and a possible avidity effect might allow for 2:1 binding not seen with FcεRI-Alexa647. The greater availability of the αγ-fusion protein also led to the decision to choose αγ-fusion protein-Alexa647 for future experiments.

This pipeline project could be converted to high-throughput in many ways. First, instead of a batch co-transfection for the library, individual transfections in 24-well plates could be conducted (transfections in 96-well plates were found to be inefficient, data not shown). The transfected cells could then be sorted on their ability to bind to the fluorescently labelled reagents. If the order is maintained, then the identity of the cells is known post-sorting if the plate layout is known.

Alternatively, with novel emerging technology that could soon be applied, the sorted repertoire of mutant IgE-Fc cells could be analysed by next-generation sequencing and protein mass spectrometry instead of single-cell PCR analysis (Lavinder et al., 2015). This method builds on single-cell PCR analysis, to provide a low-cost technology. It uses a flow-focusing device like flow cytometry to capture individual cells in a single drop of fluid containing mRNA captures beads and lysis solution. The mRNA from the lysed cells anneals to the beads which are then purified and emulsified with reverse-transcription solution. After reverse-transcription, the product is purified and sequenced by Ilumina sequencing technology (DeKosky et al., 2016; McDaniel et al., 2016). Thereafter, the DNA sequences could be recovered for analysis leading to further rounds of mutagenesis, or further protein production in order to better characterise the mutant IgE-Fc protein by SPR.

The selected mutant IgE-Fcs generated in the library can act as tools to examine the role of IgE receptor in antigen presentation and epitope spreading as well as discovering new ways of inhibiting the allergic response. This potentially could lead to the development of new therapeutics for allergen immunotherapy (to induce allergen tolerance) or for tumour immunotherapy.

Chapter 6 Summary and Future Work

The pathophysiological role of IgE in allergy had been the focus of immunoglobulin E studies for several decades. The exact roles of this elusive antibody in normal physiology and homeostasis are still not fully known. Historically, IgE was known to be involved in mediating a type 2 immune response against macroparasites such as helminths, ticks and certain parasites, although IgE is not critical to clear infections naturally (Harris and Gause, 2011). Nowadays, IgE is more commonly associated with allergies, a misdirected immune response against common environmental stimuli. Recent studies have also associated IgE with tumour surveillance (Jensen-Jarolim et al., 2010; Nigro et al., 2016). To better understand the true function of IgE and its role within a network of IgE receptors and proteins, new tools are required to investigate IgE.

In chapter 5, I present a toolbox of new human IgE-Fc mutant proteins that can used to further investigate the functional role of IgE. The toolbox is presented as a surface display library, in which mutant membrane IgE-Fc proteins are displayed on mammalian HEK293 cells. The IgE-Fc domains have been mutated at receptor binding sites to modulate the Fc effector functions to disable binding to one or other of the two IgE receptors. This library has a proven diversity as indicated by the QQC controls described in section 5.2.3, which can be further expanded upon in the future. The surface display library and designed FACS screen have demonstrated a range of binding affinities for FcεRI as shown in Figure 5-20.

The known R334S mutant (Henry et al., 1997), which acted as benchmark for these studies, has shown a 50-fold reduction in IgE affinity for FcεRI and a proportionate reduction of degranulation *in vitro* when compared to WT IgE (Hunt et al., 2008). The

developed FACS assay demonstrated IgE-Fc mutants with a lower affinity for FcεRI than the known R334S mutant. These IgE-Fc mutants have the possibility of being even more effective than the R334S mutant at reducing mast cell degranulation *in vitro*. The R334S mutant resulted in a 75% reduction in passive cutaneous anaphylaxis response *in vivo* (Hunt et al., 2008); the new mutant-Fc proteins might be able to surpass this reduction and might indeed be below the affinity limit required to trigger mast cell degranulation *in vivo*.

The designed pipeline for the surface display library uses laboratory equipment commonly found in molecular biology and cell culture laboratories, making it an accessible protocol. This library also makes it significantly easier to study IgE, studies of which are usually hampered by the low serum concentration (Dullaers et al., 2012) and the rarity of IgE-producing B-cells (Zuidscherwoude and van Spriel, 2012), as well as the dynamics of the protein itself (Drinkwater et al., 2014). The memIgE-Fc library produced a range of human IgE-Fc proteins *in vitro* with modulated Fc effector function. This small proof-of-principle study can be used to build a more complex library allowing for greater screening power, which can then be applied to defining IgE's parameters in FAP and epitope spreading. This demonstration of Fc engineering, which is less common then Fab engineering, could also be used in an industrial setting since Fc optimisation is critical for antibody pharmacokinetics and pharmacodynamics (Tiller and Tessier, 2015).

The membrane IgE-Fc protein construct I designed and created, has made it possible for the group to study membrane-bound IgE and its signalling. The memIgE-Fc construct has been adapted to express light chains and the remainder of the heavy chains, so that an entire membrane-IgE, capable of binding to allergens, is presented on the membrane

surface of B cells. The same protein construct contains the EMPD, a region for which researchers have a very limited knowledge. With the advent of monoclonal antibodies targeting this region (Feichtner et al., 2008; Vigl et al., 2017), this could be a chance to further explore its role. The crystal structure of IgE-Fc and FcεRI provided great insight into IgE biology and the triggering of the allergic response, and perhaps a crystal structure of the EMPD will also reveal more about its role, and the epitope being targeted by the anti-IgE-EMPD monoclonal antibodies.

With the exact role of IgE in the immune system being debated, the role of its receptors is also under examination. The low affinity receptor for IgE, CD23, can control IgE synthesis, dependent on whether it is membrane-bound or soluble, and on its oligomerisation status (Gould and Sutton, 2008). This regulation might be further modulated by the metal ion, calcium (Yuan et al., 2013). In chapters 3 and 4, I present *in vitro* experimental results and theoretical *in silico* simulations of CD23 to investigate further the role of calcium. To explore if human CD23 lost its ability to bind a second calcium ion and whether this uniquely affects its binding to human IgE, I created human CD23 mutants that ought to bind to two calcium ions.

C-type lectins do not have high affinities for calcium like antibodies do for their antigens; the affinity of a CTL for calcium is within the mM range. The difficulty of measuring mM interactions with biophysical techniques, which are usually employed for tighter binding properties, made the task of elucidating calcium's role challenging. Calcium titrations with NMR confirmed that a second calcium binding site had been introduced into mutant C, as shown in figure 3-22. Many previous studies of calcium in CTLs have been conducted by X-ray crystallography (Drickamer, 1999). Unfortunately, X-ray crystallographic studies with the mutant derCD23 proteins could not reveal the

co-ordination of the calcium ions since only the metal-free form of the protein crystallised. Despite various attempts to crystallise the calcium-bound protein, a task already found difficult with the WT human derCD23 protein (Yuan, 2012), the new crystal contact formed between molecules of the mutant proteins was more energetically favourable than binding of calcium (Figures 3-17 and 3-18).

The mutant derCD23 proteins demonstrated lower calcium sensitivity and reduced IgE-Fc binding affinities as shown by the SPR experiments in Figure 3-31. WT human derCD23 was the most sensitive to calcium with a 20-fold difference in K_D in the presence of calcium, while mutant C was unable to bind to IgE with calcium present. This data supports the notion that the loss of binding to a second calcium ion in human CD23, resulted in a gain-of-function, in which human CD23 binds to human IgE with greater affinity and sensitivity to calcium.

In C-type lectins, calcium plays a vital role in co-ordinating carbohydrate binding, although CD23 is unusual in its carbohydrate-independent interaction with IgE. A carbohydrate array reported in section 3.2.9, corroborates previous findings that WT human derCD23 is unable to bind to glycans (Hibbert 2005 and array 634 conducted by the Consortium for Functional Glycomics). Mutant B and C which contain the 'EPN' carbohydrate binding motif, displayed limited binding to some glycans (Figures 3-26 and 3-27), however, this appeared to be non-specific. A NMR mannose titration with mutant C (Figure 3-26) revealed very weak binding to mannose ~ 30 mM, outside the expected range for a CTL (with dissociation constants usually in the 0.1 – 10 mM range) (Weis, 1997), and perhaps not within a functionally relevant range. This low affinity might be due to lack of an aromatic residue in loop 4, which would facilitate hydrophobic packing.

It was thought that the MD simulations might be able overcome challenges faced by experimental approaches, and predict the effect of the mutations on calcium binding to CD23. Only predictions for mutant B binding to two calcium ions matched the experimental data. Although MD simulations could provide insight into the calcium-coordinating loop activities, these could not be compared to the experimental data, as the loops could not be identified in structural studies such as NMR or X-ray crystallography (Hibbert et al., 2005; Dhaliwal et al., 2012). MD simulations did however identify effects on loop 4 (which contain residues involved in both calcium-binding sites, figure 5-2), that lead to reduced loop mobility when calcium bound. Unfortunately, this MD system was not well characterised enough to be used as an *in silico* prediction tool.

There are various other types of MD simulations that could be used to estimate binding affinities, which could complement SPR studies. Alternatively, more complex types of MD simulations such as the MM-PBSA approach (Case et al., 2014) and the Free Energy Workflow Tool (Case et al., 2014), could be used to calculate binding energies.

The mutant proteins described in both chapters 3 and 4, are an attempt to further our knowledge of mouse CD23 and the differences between mouse and human CD23. Studies on murine derCD23 could advance our understanding of the different calcium binding and dissociation properties between the species. For example, could the calcium sensitivity have implications for the efficiency of FAP and epitope spreading? In humans, is IgE homeostasis more sensitive to calcium levels and thus more prone to dysregulation and allergy? Hopefully the tools created and described in this thesis can help answer these questions.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindah, E. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25.

Acevedo-Rocha, C. G., Reetz, M. T. & Nov, Y. (2015) Economical analysis of saturation mutagenesis experiments. *Sci. Reports*. **5**, 1–12.

Acharya, M., Borland, G., Edkins, L., Maclellan, L. M., Matheson, J., Ozanne, B. W. & Cushley, W. (2010) CD23/FcεRII: molecular multi-tasking. *Clin. Exp. Immunol.* **162**, 12–23.

Achatz, G., Lamers, M. & Crameri, R. (2008) Membrane Bound IgE: The Key Receptor to Restrict High IgE Levels. *Open Immunol. J.* **1**, 25–32.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221.

Adcock, S. A. & McCammon, J. A. (2006) Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615.

Akamatsu, Y., Pakabunto, K., Xu, Z., Zhang, Y. & Tsurushita, N. (2007) Whole IgG surface display on mammalian cells: Application to isolation of neutralizing chicken monoclonal anti-IL-12 antibodies. *J. Immunol. Methods*. **327**, 40–52.

Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. (2012) H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, 1–5.

Anandan, C., Nurmatov, U., Van Schayck, O. C. P. & Sheikh, A. (2010) Is the prevalence of asthma declining? Systematic review of epidemiological studies.

Allergy Eur. J. Allergy Clin. Immunol. **65**, 152–167.

Andersen, C. B. F. & Moestrup, S. K. (2014) How calcium makes endocytic receptors attractive. *Trends Biochem. Sci.* **39**, 82–90.

Armitage, R. J., Goff, L. K. & Beverley, P. C. (1989) Expression and functional role of CD23 on T cells. *Eur. J. Immunol.* **19**, 31–35.

Armour, K. L., Smith, C. S. & Clark, M. R. (2010) Expression of human FcγRIIIa as a GPI-linked molecule on CHO cells to enable measurement of human IgG binding. *J. Immunol. Methods.* **354**, 20–33.

Arnold, J. N., Radcliffe, C. M., Wormald, M. R., Royle, L., Harvey, D. J., Crispin, M., Dwek, R. A., Sim, R. B. & Rudd, P. M. (2004) The Glycosylation of Human Serum IgD and IgE and the Accessibility of Identified Oligomannose Structures for Interaction with Mannan-Binding Lectin. *J. Immunol.* **173**, 6831–6840.

Aubry, J.-P. & Pochon, S. (1992) CD21 is a ligand for CD23 and regulates IgE production. *Nature.* **358**, 505–507.

Aubry, J. P., Pochon, S., Gauchat, J. F., Nueda-Marin, A., Holers, V. M., Graber, P., Siegfried, C. & Bonnefoy, J. Y. (1994) CD23 interacts with a new functional extracytoplasmic domain involving N-linked oligosaccharides on CD21. *J. Immunol.* **152**, 5806–5813.

Baker, K., Lencer, W. I. & Blumberg, R. S. (2010) Beyond IgA: the mucosal immunoglobulin alphabet. *Mucosal Immunol.* **3**, 324–325.

Bartlett, W. C., Kelly, A. E., Johnson, C. M. & Conrad, D. H. (1995) Analysis of Murine Soluble FcεRII Sites of Cleavage and Requirements. *J. Immunol.* **154**, 4240–4246.

Batista, F., Anand, S., Presani, G., Efremov, D. G. & Burrone, O. R. (1996) Functionally Distinct B Cell Antigen Receptors. *J. Exp. Med.* **184**, 2197–2205.

Beauchamp, K., Lin, Y.-S., Das, R. & Pande, V. (2012) Are Protein Force Fields

Getting Better? A Systemic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **8**, 1409–1414.

Beck, D. a C., Jonsson, A. L., Schaeffer, R. D., Scott, K. a., Day, R., Toofanny, R. D., Alonso, D. O. V & Daggett, V. (2008) Dynameomics: Mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Eng. Des. Sel.* **21**, 353–368.

van den Berg, L. M., Gringhuis, S. I. & Geijtenbeek, T. B. H. (2012) An evolutionary perspective on C-type lectins in infection and immunity. *Ann. N. Y. Acad. Sci.* **1253**, 149–158.

Bestagno, M., Vangelista, L., Mandiola, P. A., Mukherjee, S., Sepúlveda, J. & Burrone, O. R. (2001) Membrane immunoglobulins are stabilized by interchain disulfide bonds occurring within the extracellular membrane-proximal domain. *Biochemistry (Mosc.)*. **40**, 10686–10692.

Bettler, B., Hofstetter, H., Raot, M., Yokoyamat, W. M., Kilchherr, F. & Conrad, D. H. (1989) Molecular structure and expression of the murine lymphocyte low-affinity receptor for IgE (FcεRII). *Immunology*. **86**, 7566–7570.

Borthakur, S., Hibbert, R. G., Pang, M. O. Y., Yahya, N., Bax, H. J., Kao, M. W., Cooper, A. M., Beavil, A. J., Sutton, B. J., Gould, H. J., et al. (2012) Mapping of the CD23 binding site on immunoglobulin E (IgE) and allosteric control of the IgE-FcεRI interaction. *J. Biol. Chem.* **287**, 31457–31461.

Bothwell, J. H. F. & Griffin, J. L. (2011) An introduction to biological nuclear magnetic resonance spectroscopy. *Biol. Rev.* **86**, 493–510.

Brautigam, C. A. (2015) Fitting two- and three-site binding models to isothermal titration calorimetric data. *Methods*. **76**, 124–136.

British Thoracic Society (2016) British guideline on the management of asthma. Scottish Intercollegiate Guidelines Network (September).

Cambi, A., Koopman, M. & Figdor, C. G. (2005) How C-type lectins detect pathogens. *Cell. Microbiol.* **7**, 481–488.

Carlsson, F., Hjelm, F., Conrad, D. H. & Heyman, B. (2007) IgE enhances specific antibody and T-cell responses in mice overexpressing CD23. *Scand. J. Immunol.* **66**, 261–270.

Carroll, W. D., Lenney, W., Child, F., Strange, R. C., Jones, P. W., Whyte, M. K., Primhak, R. a & Fryer, a a (2006) Asthma severity and atopy: how clear is the relationship? *Arch. Dis. Child.* **91**, 405–409.

Case, D. A., Darden, T., Iii, T. E. C., Simmerling, C., Brook, S., Roitberg, A., Wang, J., Southwestern, U. T., Duke, R. E., Hill, U., et al. (2014) Amber 14. University of California, San Francisco

Chengcheng, S., Zhang, S. & Huang, H. (2015) Choosing a suitable method for the identification of replication origins in microbial genomes. *Front. Microbiol.* **6**, 1–18.

Conrad, D. H., Ford, J. W., Sturgill, J. L. & Gibb, D. R. (2007) CD23 : An Overlooked Regulator of Allergic Disease. *Curr. Allergy Asthma Rep.* **7**, 331–337.

Cooper, A. M., Hobson, P. S., Jutton, M. R., Kao, M. W., Drung, B., Schmidt, B., Fear, D. J., Beavil, A. J., McDonnell, J. M., Sutton, B. J., et al. (2012) Soluble CD23 controls IgE synthesis and homeostasis in human B cells. *J. Immunol.* **188**, 3199–3207.

David, C. C. & Jacobs, D. J. (2014) Principal Component Analysis: A Method for Determing the Essential Dynamics of Proteins. *Methods Mol. Biol. (Clifton, N.J.)*. **1084**, 193–226.

Davies, A. M., Allan, E. G., Keeble, A. H., Delgado, J., Cossins, B. P., Mitropoulou, A. N., Pang, M. O. Y., Ceska, T., Beavil, A. J., Craggs, G., et al. (2017) Allosteric mechanism of action of the therapeutic anti-IgE antibody omalizumab. *J. Biol.*

Chem. **292**, 9975–9987.

DeKosky, B. J., Lungu, O. I., Park, D., Johnson, E. L., Charab, W., Chrysostomou, C., Kuroda, D., Ellington, A. D., Ippolito, G. C., Gray, J. J., et al. (2016) Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc. Natl. Acad. Sci.* **113**, 2636–2645.

Dhaliwal, B., Pang, M. O. Y., Yuan, D., Beavil, A. J. & Sutton, B. J. (2014) A range of Cε3–Cε4 interdomain angles in IgE Fc accommodate binding to its receptor CD23. *Struct. Biol. Commun.* **70**, 305–309.

Dhaliwal, B., Pang, M. O. Y., Yuan, D., Yahya, N., Fabiane, S. M., McDonnell, J. M., Gould, H. J., Beavil, A. J. & Sutton, B. J. (2013) Conformational plasticity at the IgE-binding site of the B-cell receptor CD23. *Mol. Immunol.* **56**, 693–697.

Dhaliwal, B., Yuan, D., Pang, M. O. Y., Henry, A. J., Cain, K., Oxbrow, A., Fabiane, S., Beavil, A. J., McDonnell, J. M., Gould, H. J., et al. (2012) Crystal structure of IgE bound to its B-cell receptor CD23 reveals a mechanism of reciprocal allosteric inhibition with high affinity receptor FcεRI. *PNAS.* **109**, 12686–12691.

Dhaliwal, B., Pang, M. O. Y., Keeble, A. H., James, L. K., Gould, H. J., McDonnell, J. M., Sutton, B. J. & Beavil, A. J. (2017) IgE binds asymmetrically to its B cell receptor CD23. *Sci. Reports.* **7**, 1–6.

Dombrowicz, D., Flamand, V., Brigman, K. K., Koller, B. H. & Kinet, J. P. (1993) Abolition of anaphylaxis by targeted disruption of the high affinity immunoglobulin E receptor alpha chain gene. *Cell.* **75**, 969–976.

Dorrington, K. J. & Bennich, H. H. (1978) Structure-function relationships in human immunoglobulin E. *Immunol. Rev.* **41**, 3–25.

Drickamer, K. (1999) C-type lectin-like domains. *Curr. Opin. Struct. Biol.* **9**, 585–590.

Drickamer, K. & Taylor, M. E. (2015) Recent insights into structures and functions of C-type lectins in the immune system. *Curr. Opin. Struct. Biol.* **34**, 26–34.

Drinkwater, N., Cossins, B. P., Keeble, A. H., Wright, M., Cain, K., Hailu, H., Oxbrow, A., Delgado, J., Shuttleworth, L. K., Kao, M. W.-P., et al. (2014) Human immunoglobulin E flexes between acutely bent and extended conformations. *Nat. Struct. Mol. Biol.* **4**, 1–10.

Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. (2010) Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.* **135**, 555–562.

Dullaers, M., De Bruyne, R., Ramadani, F., Gould, H. J., Gevaert, P. & Lambrecht, B. N. (2012) The who, where, and when of IgE in allergic airway disease. *J. Allergy Clin. Immunol.* **129**, 635–645.

Durham, S. R., Walker, S. M., Varga, E. M., Jacobson, M. R., O'Brien, F., Noble, W., Till, S. J., Hamid, Q. a & Nouri-Aria, K. T. (1999) Long-term clinical efficacy of grass-pollen immunotherapy. *N. Engl. J. Med.* **341**, 468–475.

Dutta, A. K., Rösgen, J. & Krishna, R. (2015) Using Isothermal Titration Calorimetry to Determine Thermodynamic Parameters of Protein-Glycosaminoglycan Interactions. *Methods Mol. Biol. (Clifton, N.J.)*. **1**, 315–324.

Dylke, J., Lopes, J., Dang-lawson, M., Machtaler, S. & Matsuuchi, L. (2007) Role of the extracellular and transmembrane domain of Ig-α/β in assembly of the B cell antigen receptor (BCR). *Immunol. Lett.* **112**, 47–57.

Edkins, A. L., Borland, G., Acharya, M., Cogdell, R. J., Ozanne, B. W. & Cushley, W. (2012) Differential regulation of monocyte cytokine release by αV and β2 integrins that bind CD23. *Immunology*. **136**, 241–251.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501.

Engeroff, P., Fellmann, M., Yerly, D., Bachmann, M. F. & Vogel, M. (2017) A novel recycling mechanism of native IgE-antigen complexes in human B cells facilitates

transfer of antigen to dendritic cells for antigen presentation. *J. Allergy Clin. Immunol.* **1**, 1–18.

Feichtner, S., Infuhr, D., Achatz-Straussberger, G., Schmid, D., Karnowski, A., Lamers, M., Rhyner, C., Crameri, R. & Achatz, G. (2008) Targeting the Extracellular Membrane-Proximal Domain of Membrane-Bound IgE by Passive Immunization Blocks IgE Synthesis In Vivo. *J. Immunol.* **180**, 5499–5505.

Fitzhugh, D. J. & Lockey, R. F. (2011) Allergen immunotherapy: a history of the first 100 years. *Curr. Opin. Allergy Clin. Immunol.* **11**, 554–559.

Fridman, H. (1991) Fc receptors and immunoglobulin binding factors. *FASEB J.* **5**, 2684–2690.

Fuglebakk, E., Tiwari, S. P. & Reuter, N. (2015) Comparing the intrinsic dynamics of multiple protein structures using elastic network models. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 911–922.

Fuglebakk, E., Reuter, N. & Hinsen, K. (2013) Evaluation of protein elastic network models based on an analysis of collective motions. *J. Chem. Theory Comput.* **9**, 5618–5628.

Fujita, H., Soyka, M. B., Akdis, M. & Akdis, C. a (2012) Mechanisms of allergen-specific immunotherapy. *Clin. Transl. allergy*. **2**, 2–8.

Galli, S. J. & Tsai, M. (2012) IgE and mast cells in allergic disease. *Nat. Med.* **18**, 693–704.

Garman, S. C., Kinet, J. P. & Jardetzky, T. S. (1998) Crystal structure of the human high-affinity IgE receptor. *Cell.* **95**, 951–961.

Garman, S. C., Wurzburg, B. A., Tarchevskaya, S. S., Klnet, J. P. & Jardetzky, T. S. (2000) Structure of the Fc fragment of human IgE bound to its high-affinity receptor FcεRIα. *Nature*. **406**, 259–266.

Gerasimenko, J. V, Tepikin, a V, Petersen, O. H. & Gerasimenko, O. V (1998)

Calcium uptake via endocytosis with rapid release from acidifying endosomes. *Curr. Biol.* **8**, 1335–1338.

Gilbert, H. E., Asokan, R., Holers, V. M. & Perkins, S. J. (2006) The 15 SCR Flexible Extracellular Domains of Human Complement Receptor Type 2 can Mediate Multiple Ligand and Antigen Interactions. *J. Mol. Biol.* **362**, 1132–1147.

Gingras, A. R., Girija, U. V., Keeble, A. H., Panchal, R., Mitchell, D., Moody, P. C. E. & Wallis, R. (2011) Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: implications for complement activation. *Structure.* **19**, 1635–1643.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. (2005) H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, 368–371.

Gould, H. J. & Sutton, B. J. (2008) IgE in allergy and asthma today. *Nat. Rev. Immunol.* **8**, 205–217.

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. (2006) Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics.* **22**, 2695–2696.

Haniuda, K., Fukao, S., Kodama, T., Hasegawa, H. & Kitamura, D. (2016) Autonomous membrane IgE signaling prevents IgE-memory formation. *Nat. Immunol.* **17**, 1–11.

Harris, N. & Gause, W. C. (2011) To B or not to B: B cells and the Th2-type immune response to helminths. *Trends Immunol.* **32**, 80–88.

van der Heijden, F. L., van Neerven, R. J. & Kapsenberg, M. L. (1995) Relationship between facilitated allergen presentation and the presence of allergen-specific IgE in serum of atopic patients. *Clin. Exp. Immunol.* **99**, 289–293.

Henry, a J., Cook, J. P., McDonnell, J. M., Mackay, G. a, Shi, J., Sutton, B. J. & Gould, H. J. (1997) Participation of the N-terminal region of Cε3 in the binding of human

IgE to its high-affinity receptor FcεRI. *Biochemistry (Mosc.)*. **36**, 15568–15578.

Henzl, M. T., Davis, M. E. & Tan, A. (2008) Divalent ion binding properties of the timothy grass allergen, Phl p 7. *Biochemistry (Mosc.)*. **47**, 7846–7856.

Hermann, P., Armant, M., Brown, E., Rubio, M., Ishihara, H., Ulrich, D., Caspary, R. G., Lindberg, F. P., Armitage, R., Maliszewski, C., et al. (1999) The vitronectin receptor and its associated CD47 molecule mediates proinflammatory cytokine synthesis in human monocytes by interaction with soluble CD23. *J. Cell Biol.* **144**, 767–775.

Heyman, B. (2002) IgE-mediated enhancement of antibody responses: the beneficial function of IgE? *Allergy*. **57**, 577–585.

Heyman, B., Tianmin, L. & Gustavsson, S. (1993) In vivo enhancement of the specific antibody response via the low-affinity receptor for IgE. *Eur. J. Immunol.* **23**, 1739–1742.

Hibbert, R. (2005) *Characterising the Interactions of CD23 with Multiple Ligands*. University of Oxford. Thesis.

Hibbert, R. G., Teriete, P., Grundy, G. J., Beavil, R. L., Reljic, R., Holers, V. M., Hannan, J. P., Sutton, B. J., Gould, H. J. & McDonnell, J. M. (2005) The structure of human CD23 and its interactions with IgE and CD21. *J. Exp. Med.* **202**, 751–760.

Holdom, M. D., Davies, A. M., Nettleship, J. E., Bagby, S. C., Dhaliwal, B., Girardi, E., Hunt, J., Gould, H. J., Beavil, A. J., McDonnell, J. M., et al. (2011) Conformational changes in IgE contribute to its uniquely slow dissociation rate from receptor FcεRI. *Nat. Struct. Mol. Biol.* **18**, 571–576.

Holla, A. & Skerra, A. (2011) Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. *Protein Eng. Des. Sel. PEDS*. **24**, 659–669.

Hunt, J., Bracher, M. G., Shi, J., Fleury, S., Dombrowicz, D., Gould, H. J., Sutton, B. J. & Beavil, A. J. (2008) Attenuation of IgE affinity for FcεRI radically reduces the allergic response in vitro and in vivo. *J. Biol. Chem.* **283**, 29882–29887.

Incorvaia, C. & Mauro, M. (2015) Evolution of anti-IgE treatment. *Cell. Mol. Immunol.* **13**, 1–2.

Jensen-Jarolim, E., Achatz, G., Turner, M. C., Karagiannis, S. N., Legrand, F. & Capron, M. (2010) AllergoOncology : the role of IgE-mediated allergy in cancer. *Allergy*. **63**, 1255–1266.

Josephs, D. H., Spicer, F., Karagiannis, P., Gould, H. J. & Karagiannis, S. N. (2014) IgE immunotherapy A novel concept with promise for the treatment of cancer. *mAbs*. **6**, 54–72.

Kanellopouloss, J. M., Liuj, T., Poyg, G. & Metzgerg, H. (1980) Composition and Subunit Structure of the Cell Receptor for Immunoglobulin E. *J. Biol. Chem.* **255**, 9060–9066.

Karagiannis, S. N., Wang, Q., East, N., Burke, F., Riffard, S., Bracher, M. G., Thompson, R. G., Durham, S. R., B, L., Balkwill, F. R., et al. (2003) Activity of human monocytes in IgE antibody- dependent surveillance and killing of ovarian tumor cells. *Heart Lung*. **33**, 1030–1040.

Karagiannis, S. N., Warrack, J. K., Jennings, K. H., Murdock, P. R., Christie, G., Moulder, K., Sutton, B. J. & Gould, H. J. (2001) Endocytosis and recycling of the complex between CD23 and HLA-DR in human B cells. *Immunology*. **103**, 319–331.

Karagiannis, S. N., Bracher, M. G., Hunt, J., McCloskey, N., Beavil, R. L., Beavil, A. J., Fear, D. J., Thompson, R. G., East, N., Burke, F., et al. (2007) IgE-antibody-dependent immunotherapy of solid tumors: cytotoxic and phagocytic mechanisms of eradication of ovarian cancer cells. *J. Immunol.* **179**, 2832–2843.

Karagiannis, S. N., Josephs, D. H., Karagiannis, P., Gilbert, A. E., Saul, L., Rudman, S. M., Dodev, T., Koers, A., Blower, P. J., Corrigan, C., et al. (2012) Recombinant IgE antibodies for passive immunotherapy of solid tumours: From concept towards clinical application. *Cancer Immunol. Immunother.* **61**, 1547–1564.

Keller, S., Sanderson, M. P., Stoeck, A. & Altevogt, P. (2006) Exosomes: from biogenesis and secretion to biological function. *Immunol. Lett.* **107**, 102–108.

Kerrigan, J. (2013) *GROMACS Introductory Tutorial*.

Kijimoto-Ochiai, S. (2002) CD23 (the low-affinity IgE receptor) as a C-type lectin : a multidomain and multifunctional molecule. *Cell. Mol. life Sci. C.* **59**, 648–664.

Kijimoto-Ochiai, S. & Toshimitsu, U. (1995) CD23 molecule acts as a galactose-binding lectin in the cell aggregation of EBV-transformed human B-cell lines. *Glycobiology.* **5**, 443–448.

Kikutani, H., Yokota, A., Uchibayashi, N., Yukawa, K., Tanaka, T., Sugiyama, K., L. Barsumian, E., Suemura, M. & Kishimoto, T. (1989) 'Structure and Function of Fcε Receptor II (FcεRII/CD23): A Point of Contact Between the Effector Phase of Allergy and B Cell Differentiation', in *Ciba Foundation symposium*. p. 23–31; discussion 31.

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z. G., Opperman, D. J., Reetz, M. T. & Acevedo, J. P. (2013) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92.

Kindt, T. J., Goldsby, R. A., Osborne, B. A. & Kuby, J. (2007) *Kuby immunology.* Sixth Edit. New York: W.H. Freeman.

Kinet, J.-P. (1999) THE HIGH-AFFINITY IgE RECEPTOR (FcεRI): From Physiology to Pathology. *Annu. Rev. Immunol.* **17**, 931–972.

King, D. J., Bowers, P. M., Kehry, M. R. & Horlick, R. a (2014) Mammalian cell

display and somatic hypermutation in vitro for human antibody discovery. *Curr. Drug Discov. Technol.* **11**, 56–64.

Klunker, S., Saggar, L. R., Seyfert-Margolis, V., Asare, A. L., Casale, T. B., Durham, S. R. & Francis, J. N. (2007) Combination treatment with omalizumab and rush immunotherapy for ragweed-induced allergic rhinitis: Inhibition of IgE-facilitated allergen binding. *J. Allergy Clin. Immunol.* **120**, 688–695.

Krainer, G. & Keller, S. (2015) Single-experiment displacement assay for quantifying high-affinity binding by isothermal titration calorimetry. *Methods.* **76**, 116–123.

Kulus, M., Hébert, J., Garcia, E., Fowler Taylor, A., Fernandez Vidaurre, C. & Blogg, M. (2010) Omalizumab in children with inadequately controlled severe allergic (IgE-mediated) asthma. *Curr. Med. Res. Opin.* **26**, 1285–1293.

Kuzmanic, A. & Zagrovic, B. (2010) Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* **98**, 861–871.

Laemmli, U. . (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature.* **227**, 680–685.

Lai, T., Yang, Y. & Ng, S. K. (2013) Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production. *Pharmaceuticals.* **6**, 579–603.

Lavinder, J. J., Horton, A. P., Georgiou, G. & Ippolito, G. C. (2015) Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Curr. Opin. Chem. Biol.* **24**, 112–120.

Lee, R. T., Hsu, T.-L., Huang, S. K., Hsieh, S.-L., Wong, C.-H. & Lee, Y. C. (2011) Survey of immune-related, mannose/fucose-binding C-type lectin receptors reveals widely divergent sugar-binding specificities. *Glycobiology.* **21**, 512–520.

Lee, W. T., Conrad, D. H., Lee, W. T. & Conrad, D. H. (1986) Murine B cell hybridomas bearing ligand-inducible Fc receptors for IgE . *J. Immunol.* **136**, 4573–

4580.

Lemieux, G., Blumenkron, F., Yeung, N., Zhou, P., Williams, J., Grammer, A. C., Petrovich, R., Lipsky, P. E., Moss, M. L. & Werb, Z. (2007) The low affinity IgE receptor (CD23) is cleaved by the metalloproteinase ADAM10. *J. Biol. Chem.* **282**, 14836–14844.

Leslie, A. G. W., Johnson, O. & Powell, H. R. (2011) Recent developments in the mosflm package. *MRC Lab. Mol. Biol. Artic.* **9**, 1–8.

Letourneur, O., Sechi, S., Willette-Brown, J., Robertson, M. W. & Kinet, J. P. (1995) Glycosylation of human truncated FcεRI α chain is necessary for efficient folding in the endoplasmic reticulum. Journal of Biological Chemistry 270 (14) p.8249–8256.

Levitt, M. (2014) 'Birth and future of multi-scale modeling of macromolecules', in *9th Randall Lecture*. p. Lecture slide.

Levy, M. L. (2015) The national review of asthma deaths: What did we learn and what needs to change? *Breathe*. **11**, 15–24.

Li, Zhang, Z.-R., Schluesener, H. J. & Xu, S.-Q. (2006) Role of exosomes in immune regulation. *J. Cell. Mol. Med.* **10**, 364–375.

Li, H., Nowak-Wegrzyn, A., Charlop-Powers, Z., Shreffler, W., Chehade, M., Thomas, S., Roda, G., Dahan, S., Sperber, K. & Berin, M. C. (2006) Transcytosis of IgE-antigen complexes by CD23a in human intestinal epithelial cells and its role in food allergy. *Gastroenterology*. **131**, 47–58.

Lowe, P. J., Tannenbaum, S., Gautier, A. & Jimenez, P. (2009) Relationship between omalizumab pharmacokinetics, IgE pharmacodynamics and symptoms in patients with severe persistent allergic (IgE-mediated) asthma. *Br. J. Clin. Pharmacol.* **68**, 61–76.

Lynch, N. R., Hagel, I. A., Palenque, M. E., Di Prisco, M. C., Escudero, J. E., Alejandra

Corao, L., Sandia, J. A., Ferreira, L. J., Botto, C., Perez, M., et al. (1998) Relationship between helminthic infection and IgE response in atopic and nonatopic children in a tropical environment. *J. Allergy Clin. Immunol.* **101**, 217–221.

Maenaka, K., Van der Merwe, P. A., Stuart, D. I., Jones, E. Y. & Sondermann, P. (2001) The Human Low Affinity Fcγ Receptors IIa, IIb, and III Bind IgG with Fast Kinetics and Distinct Thermodynamic Properties. *J. Biol. Chem.* **276**, 44898–44904.

Maier, J., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. & Simmerling, C. (2015) ff 14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff 99SB. *J. Chem. Theory Comput.* **11**, 3696–3713.

Martin, R. K., Brooks, K. B., Henningsson, F., Heyman, B. & Conrad, D. H. (2014) Antigen transfer from exosomes to dendritic cells as an explanation for the immune enhancement seen by IgE immune complexes. *PLoS ONE.* **9**, 1–6.

Mathews, J. a, Gibb, D. R., Chen, B.-H., Scherle, P. & Conrad, D. H. (2010) CD23 Sheddase A disintegrin and metalloproteinase 10 (ADAM10) is also required for CD23 sorting into B cell-derived exosomes. *J. Biol. Chem.* **285**, 37531–37541.

Mattila, P., Joenväärä, S., Renkonen, J., Toppila-Salmi, S. & Renkonen, R. (2011) Allergy as an epithelial barrier disease. *Clin. Transl. allergy.* **1**, 1–8.

McCloskey, N., Hunt, J., Beavil, R. L., Jutton, M. R., Grundy, G. J., Girardi, E., Fabiane, S. M., Fear, D. J., Conrad, D. H., Sutton, B. J., et al. (2007) Soluble CD23 monomers inhibit and oligomers stimulate IgE synthesis in human B cells. *J. Biol. Chem.* **282**, 24083–24091.

McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. (2016) Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat. Protoc.* **11**, 429–442.

McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011) Presenting your structures: The CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394.

McSharry, C., Xia, Y., Holland, C. V. & Kennedy, M. W. (1999) Natural immunity to Ascaris lumbricoides associated with immunoglobulin E antibody to ABA-1 allergen and inflammation indicators in children. *Infect. Immun.* **67**, 484–489.

Mossalayi, M. D., Vouldoukis, I., Mamani-Matsuda, M., Kauss, T., Guillon, J., Maugein, J., Moynet, D., Rambert, J., Desplat, V., Mazier, D., et al. (2009) CD23 mediates antimycobacterial activity of human macrophages. *Infect. Immun.* **77**, 5537–5542.

Mudde, G. C., Bheekha, R. & Bruijnzeel-Koomen, C. a (1995a) Consequences of IgE/CD23-mediated antigen presentation in allergy. *Immunol. Today.* **16**, 380–383.

Mudde, G. C., Bheekha, R. & Bruijnzeel-Koomen, C. a (1995b) IgE-mediated antigen presentation. *Allergy.* **50**, 193–199.

Mukherjee, M., Stoddart, A., Gupta, R. P., Nwaru, B. I., Farr, A., Heaven, M., Fitzsimmons, D., Bandyopadhyay, A., Aftab, C., Simpson, C. R., et al. (2016) The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Med.* **14**, 113.

Nadler, L. M., Stashenko, P., Hardy, R., van Agthoven, A., Terhorst, C. & Schlossman, S. F. (1981) Characterization of a human B cell-specific antigen (B2) distinct from B1. *J. Immunol.* **126**, 1941 LP-1947.

van Neerven, R. J. J., Knol, E. F., Ejrnaes, A. & Würtzen, P. (2006) IgE-mediated allergen presentation and blocking antibodies: regulation of T-cell activation in allergy. *Int. Arch. Allergy Immunol.* **141**, 119–129.

van Neerven, R. J., Wikborg, T., Lund, G., Jacobsen, B., Brinch-Nielsen, A., Arnved, J.

& Ipsen, H. (1999) Blocking antibodies induced by specific allergy vaccination prevent the activation of CD4+ T cells by inhibiting serum-IgE-facilitated allergen presentation. *J. Immunol.* **163**, 2944–2952.

Ng, K. K. & Weis, W. I. (1998) Coupling of Prolyl Peptide Bond Isomerization and Ca2+ Binding in a C-type Mannose-Binding Protein. *Biochemistry (Mosc.).* **37**, 17977–17989.

Nigro, E. A., Brini, A. T., Yenagi, V. A., Ferreira, L. M., Achatz-Straussberger, G., Ambrosi, A., Sanvito, F., Soprana, E., van Anken, E., Achatz, G., et al. (2016) Cutting Edge: IgE Plays an Active Role in Tumor Immunosurveillance in Mice. *J. Immunol.* **197**, 2583–2588.

Nunes, C., Pereira, A. M. & Morais-Almeida, M. (2017) Asthma costs and social impact. *Asthma Res. Pract.* **3**, 1–11.

Ode, H., Nakashima, M., Kitamura, S., Sugiura, W. & Sato, H. (2012) Molecular dynamics simulation in virus research. *Front. Immunol.* **3**, 1–9.

Palaniyandi, S., Tomei, E., Li, Z., Conrad, D. H. & Zhu, X. (2011) CD23-dependent transcytosis of IgE and immune complex across the polarized human respiratory epithelial cells. *J. Immunol.* **186**, 3484–3496.

Palaniyandi, S., Liu, X., Periasamy, S., Ma, A., Tang, J., Jenkins, M., Tuo, W., Song, W., Keegan, a D., Conrad, D. H., et al. (2015) Inhibition of CD23-mediated IgE transcytosis suppresses the initiation and development of allergic airway inflammation. *Mucosal Immunol.* **8**, 1262–1274.

Palomares, O., Sanchez-Ramon, S., Davila, I., Prieto, L., Perez de Llano, L., Lleonart, M., Domingo, C. & Nieto, A. (2017) dIvergEnt: How IgE Axis Contributes to the Continuum of Allergic Asthma and Anti-IgE Therapies. *Int. J. Mol. Sci.* **18**, 1328–1342.

Pawankar, R., Canonica, G. W., Holgate, S. & Lockey, R. F. (2011) *White Book on*

Allergy 2011-2012 Executive Summary.

Perez, F. & Granger, B. (2007) IPython : A System for. *Comput. Sci. Eng.* **9**, 21–29.

Pflugrath, J. W. (2004) Macromolecular cryocrystallography - Methods for cooling and mounting protein crystals at cryogenic temperatures. *Methods*. **34**, 415–423.

Pierce, M. M., Raman, C. S. & Nall, B. T. (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods (San Diego, Calif.)*. **19**, 213–221.

Plomp, R., Hensbergen, P. J., Rombouts, Y., Zauner, G., Dragan, I., Koeleman, C. A. M., Deelder, A. M. & Wuhrer, M. (2014) Site-Specific N-Glycosylation Analysis of Human Immunoglobulin E. *J. Proteome Res.* **13**, 536–546.

Pochon, S., Graber, P., Yeager, M., Jansen, K., Bernard, a R., Aubry, J. P. & Bonnefoy, J. Y. (1992) Demonstration of a second ligand for the low affinity receptor for immunoglobulin E (CD23) using recombinant CD23 reconstituted into fluorescent liposomes. *J. Exp. Med.* **176**, 389–397.

Poggianella, M., Bestagno, M. & Burrone, O. R. (2006) The Extracellular Membrane-Proximal Domain of Human Membrane IgE Controls Apoptotic Signaling of the B Cell Receptor in the Mature B Cell Line A20. *J. Immunol.* **177**, 3597–3605.

Poulsen, L. K. (2008) Facilitated antigen presentation: circumstantial evidence? *Clin. Exp. allergy J. Br. Soc. Allergy Clin. Immunol.* **38**, 1246–1248.

Quesenberry, M. S. & Drickamer, K. (1992) Role of Conserved and Nonconserved Residues in the Ca2+-dependent Carbohydrate-recognition Domain of a Rat Mannose-binding Protein. *J. Bol. Chem.* **267**, 10831–10841.

Redhu, N. S., Shan, L., Al-Subait, D., Ashdown, H. L., Movassagh, H., Lamkhioued, B. & Gounni, A. S. (2013) IgE induces proliferation in human airway smooth muscle cells: role of MAPK and STAT3 pathways. *Allergy, asthma, Clin. Immunol. Off. J. Can. Soc. Allergy Clin. Immunol.* **9**, 41–51.

Richards, M. L. & Katz, D. H. (1990) The binding of IgE to Murine FceRII is calcium-

dependent but not inhibited by carbohydrate. *J. Immunol.* **4**, 2638–2646.

Rihet, P., Demeure, C. E., Bourgois, A., Prata, A. & Dessein, A. (1991) Evidence for an association between human resistance to Schistosoma mansoni and high anti-larval IgE levels. *Eur. J. Immunol.* **21**, 2679–2686.

Ritzefeld, M. & Sewald, N. (2012) Real-Time Analysis of Specific Protein-DNA Interactions with Surface Plasmon Resonance. *J. Amin. Acids*. **2012**, 1–19.

Robertson, M. W. (1993) Phage and Escherichia coli Expression of the Human High Affinity Immunoglobulin E Receptor alpha-Subunit Ectodomain. *J. Biol. Chem.* **268**, 12736–12743.

Roe, D. R. & Cheatham III, T. E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular synamics trajectory data. *J Chem Theory Com*. **9**, 3084–3095.

Roth, M., Zhong, J., Zumkeller, C., S, C. T., Goulet, S. & Tamm, M. (2013) The Role of IgE-Receptors in IgE-Dependent Airway Smooth Muscle Cell Remodelling. *PLoS ONE*. **8**, 1–9.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.

Sanchis, J. & Fernández, L. (2008) Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution : application to difficult-to-amplify templates. *Appl. Microbiol. Biotechnol.* **81**, 387–397.

Sang, P., Yang, L.-Q., Ji, X.-L., Fu, Y.-X. & Liu, S.-Q. (2014) Insight derived from molecular dynamics simulations into molecular motions, thermodynamics and kinetics of HIV-1 gp120. *PLoS ONE*. **9**, 104714–104730.

Schulz, O., Sutton, B. J., Beavil, R. L., Shi, J., Sewell, H. F., Gould, H. J., Laing, P. & Shakib, F. (1997) Cleavage of the low-affinity receptor for human IgE (CD23) by

a mite cysteine protease: nature of the cleaved fragment in relation to the structure and function of CD23. *Eur. J. Immunol.* **27**, 584–588.

Shakib, F., Schulz, O. & Sewell, H. (1998) A mite subversive: cleavage of CD23 and CD25 by Der p 1 enhances allergenicity. *Immunol. Today*. **19**, 313–316.

Shi, J., Ghirlando, R., Beavil, R. L., Beavil, A. J., Keown, M. B., Young, R. J., Owens, R. J., Sutton, B. J. & Gould, H. J. (1997) Interaction of the Low-Affinity Receptor CD23 / FcεRII Lectin Domain with the Fcε3 - 4 Fragment of Human Immunoglobulin E. *Biochemistry (Mosc.)*. **2960**, 2112–2122.

Singer, J. & Jensen-Jarolim, E. (2014) IgE-based immunotherapy of cancer: Challenges and chances. *Allergy Eur. J. Allergy Clin. Immunol.* **69**, 137–149.

Soga, K., Abo, H., Qin, S., Kyoutou, T., Hiemori, K., Tateno, H., Matsumoto, N., Hirabayashi, J. & Yamamoto, K. (2015) Mammalian Cell Surface Display as a Novel Method for Developing Engineered Lectins with Novel Characteristics. *Biomolecules*. **5**, 1540–1562.

Stanworth, D. R. (1993) The discovery of IgE. *Allergy*. **48**, 67–71.

Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234.

Sulmann, S., Dell'Orco, D., Marino, V., Behnen, P. & Koch, K.-W. (2014) Conformational Changes in Calcium-Sensor Proteins under Molecular Crowding Conditions. *Chemistry*. **20**, 6756–6762.

Szczeklik, A., Dropihski, J. & Göra, P. F. (1993) Serum immunoglobulin E and sudden cardiac arrest during myocardial infarction. *Coron. Artery Dis.* **4**, 1–3.

Szczeklik, A. & Dawson, J. (2000) Neuraminidase inhibitor and amantadine Atopy and sudden cardiac For personal use only . Not to be reproduced without permission of The Lancet . *Lancet*. **355**, 2254.

Tanaka, A., Jinno, M., Hirai, K., Miyata, Y., Mizuma, H., Yamaguchi, M., Ohta, S.,

Watanabe, Y., Yamamoto, M., Suzuki, S., et al. (2014) Longitudinal increase in total IgE levels in patients with adult asthma: an association with poor asthma control. *Respir. Res.* **15**, 144–162.

Taylor, M., Pratt, K., Revell, D. F., Baker, K. C., Sumner, I. G. & Goodenough, P. W. (1992) Active papain renatured and processed from insoluble recombinant propapain expressed in Escherichia coli. *Protein Eng.* **5**, 455–459.

Taylor, M. E. & Drickamer, K. (2003) Structure-function analysis of C-type animal lectins. *Methods Enzymol.* **363**, 3–16.

Teng, T. Y. (1990) Mounting of Crystals for Macromolecular Crystallography in a Free-standing Thin Film. *J. Appl. Crystallogr.* **23**, 387–391.

Tiller, K. E. & Tessier, P. M. (2015) Advances in Antibody Design. *Annu. Rev. Biomed. Eng.* **17**, 1–26.

Tu, Y., Salim, S., Bourgeois, J., Di Leo, V., Irvine, E. J., Marshall, J. K. & Perdue, M. H. (2005) CD23-mediated IgE transport across human intestinal epithelium: Inhibition by blocking sites of translation or binding. *Gastroenterology*. **129**, 928–940.

Turnbull, W. B. & Daranas, A. H. (2003) On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *JACS*. **125**, 14859–14866.

Vercelli, D., Jabara, H., Lee, B.-W., Woodland, N., Geha, R. & Leung, D. (1988) Huam recombinant interleukin 4 induces FcεR2 / CD23 on normal human monocytes. *J. Exp. Med.* **167**, 1406–1416.

Verlet, L. (1967) Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103.

Vigl, B., Salhat, N., Parth, M., Pankevych, H., Mairhofer, A., Bartl, S. & Smrzka, O. W. (2017) Quantitative in vitro and in vivo models to assess human IgE B cell receptor crosslinking by IgE and EMPD IgE targeting antibodies. *J. Immunol.*

Methods. **449**, 28–36.

Wachholz, P. A., Soni, N. K., Till, S. J. & Durham, S. R. (2003) Inhibition of allergen-IgE binding to B cells by IgG antibodies after grass pollen immunotherapy. *J. Allergy Clin. Immunol.* **112**, 915–922.

Waldmeier, L., Hellmann, I., Gutknecht, C. K., Wolter, F. I., Cook, S. C., Reddy, S. T., Grawunder, U. & Beerli, R. R. (2016) Transpo-mAb Display: transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. *mAbs.* **0862**, 726–740.

Wan, T., Beavil, R. L., Fabiane, S. M., Beavil, A. J., Sohi, M. K., Keown, M., Young, R. J., Henry, A. J., Owens, R. J., Gould, H. J., et al. (2002) The crystal structure of IgE Fc reveals an asymmetrically bent conformation. *Nat. Immunol.* **3**, 681–686.

Weis, J. J., Toothaker, L. E., Smith, J. A., Weis, J. H. & Fearon, D. T. (1988) Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus and relatedness to other members of the family of C3/C4 binding proteins. *J. Exp. Med.* **167**, 1047–1066.

Weis, W. I. (1997) Cell-surface carbohydrate recognition by animal and viral lectins. *Curr. Opin. Struct. Biol.* **7**, 624–630.

Weis, W. I., Crichlow, G. V, Murthysq, H. M. K. & Hendricksons, W. A. (1991) Physical Characterization and Crystallization of the Carbohydrate- recognition Domain of a Mannose-binding Protein from Rat. *J. Biol. Chem.* **266**, 20678–20686.

Weis, W. I. (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature.* **360**, 127–134.

Weis, W. I., Taylor, M. E. & Drickamer, K. (1998) The C-type lectin superfamily in the immune system. *Immunol. Rev.* **163**, 19–34.

Weis, W. I. & Drickamer, K. (1996) Structural Basis of Lectin-Carbohydrate

Recognition. *Annu. Rev. Biochem.* **65**, 441–473.

Weiser, J., Shenkin, P. S., Still, W. C., Lcpo, O., Shenkin, P. S. & Still, W. C. (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*. **20**, 217–230.

Wilcock, L. K., Francis, J. N. & Durham, S. R. (2006) IgE-Facilitated Antigen Presentation: Role in Allergy and the Influence of Allergen Immunotherapy. *Immunol. Allergy Clin. North Am.* **26**, 333–347.

Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16.

Wright, J. D., Chu, H.-M., Huang, C.-H., Ma, C., Chang, T. W. & Lim, C. (2015) Structural and Physical Basis for Anti-IgE Therapy. *Sci. reports*. **5**, 11581–11595.

Wurzburg, B., Tarchevskaya, S. & Jardetzky, T. (2006) Structural changes in the lectin domain of CD23, the low-affinity IgE receptor, upon calcium binding. *Structure*. **14**, 1049–1058.

Wurzburg, B. A. & Jardetzky, T. S. (2009) Conformational Flexibility in Immunoglobulin E-Fc 3 − 4 Revealed in Multiple Crystal Forms. *J. Mol. Biol.* **393**, 176–190.

Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H. & Liu, S.-Q. (2014a) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **32**, 372–393.

Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H. & Liu, S.-Q. (2014b) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J. Biomol. Struct. Dyn.* **32**, 372–393.

Yang, W., Jones, L. M., Isley, L., Ye, Y., Lee, H.-W., Wilkins, A., Liu, Z., Hellinga, H. W., Malchow, R., Ghazi, M., et al. (2003) Rational design of a calcium-binding protein. *J. Am. Chem. Soc.* **125**, 6165–6171.

Yang, Z., Robinson, M. J., Chen, X., Smith, G. A., Taunton, J., Liu, W. & Allen, C. D. (2016) Regulation of B cell fate by chronic activity of the IgE B cell receptor. *eLife*. **5**, 1–31.

Young, R. J., Owens, R. J., Mackay, G. A., Chan, C. M. W., Shi, J., Hide, M., Francis, D. M., Henry, A. J., Sutton, B. J. & Gould, H. J. (1995) Secretion of recombinant human IgE-Fc by mammalian cells and biological activity of glycosylation site mutants. *Protein Eng. Des. Sel.* **8**, 193–199.

Yu, L. C. H., Montagnac, G., Yang, P.-C., Conrad, D. H., Benmerah, A. & Perdue, M. H. (2003) Intestinal epithelial CD23 mediates enhanced antigen transport in allergy: evidence for novel splice forms. *Am. J. Physiol. Gastrointest. Liver Physiol.* **285**, G223-34.

Yu, P., Kosco-Vilbois, M. H., Richards, M., Kohler, G. & Lamers, M. (1994) Negative Feedback regulation of IgE synthesis by murine CD23. *Nature.* **369**, 753–756.

Yuan, D., Keeble, A. H., Hibbert, R. G., Fabiane, S., Gould, H. J., McDonnell, J. M., Beavil, A. J., Sutton, B. J. & Dhaliwal, B. (2013) Ca2+-dependent Structural Changes in the B-cell Receptor CD23 Increase Its Affinity for Human Immunoglobulin E. *J. Biol. Chem.* **288**, 21667–21677.

Yuan, D. (2012) *Structural studies of human CD23 and its complexes*. King's College London. Thesis.

Zelensky, A. N. & Gready, J. E. (2003) Comparative Analysis of Structural Properties of the C-Type-Lectin-like Domain (CTLD). *Proteins Struct. Funct. Genet.* **477**, 466–477.

Zelensky, A. N. & Gready, J. E. (2005) The C-type lectin-like domain superfamily. *Fed. Eur. Biochem. Soc.* **272**, 6179–6217.

Zhang, D. & Lazim, R. (2017) Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Sci.*

Reports. **7**, 1–12.

Zhang, G., Gurtu, V. & Kain, S. R. (1996) An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227**, 707–711.

Zhang, L., Zheng, Y., Xi, Z., Luo, Z., Xu, X., Wang, C. & Liu, Y. (2009) Metal ions binding to recA inteins from Mycobacterium tuberculosis. *Mol. Biosyst.* **5**, 644–650.

Zheng, Y., Holowka, D., Baird, B. & Shopes, B. (1992) Dynamic Conformations Compared for IgE and IgGl in Solution and Bound to Receptors. *Biochemistry (Mosc.)*. **31**, 7446–7456.

Zhou, C., Jacobsen, F. W., Cai, L., Chen, Q. & Shen, W. D. (2010) Development of a novel mammalian cell surface antibody display platform. *mAbs*. **2**, 508–518.

Zuidscherwoude, M. & van Spriel, A. B. (2012) The origin of IgE memory and plasma cells. *Cell. Mol. Immunol.* **9**, 373–374.

Appendix A. The DNA sequence of the membrane IgE-Fc-GFP construct

ATGGACTGGACCTGGATCCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACT

CCTCCGTCGCCTCCAGGGACTTCACCCCGCCCACCGTGAAGATCTTACAGTC

GTCCTGCGACGGCGGCGGGCACTTCCCCCCGACCATCCAGCTCCTGTGCCTC

GTCTCTGGGTACACCCCAGGGACTATCAACATCACCTGGCTGGAGGACGGG

CAGGTCATGGACGTGGACTTGTCCACCGCCTCTACCACGCAGGAGGGTGAG

CTGGCCTCCACACAAAGCGAGCTCACCCTCAGCCAGAAGCACTGGCTGTCA

GACCGCACCTACACCTGCCAGGTCACCTATCAAGGTCACACCTTTGAGGAC

AGCACCAAGAAGTGTGCAGATTCCAACCCGAGAGGGGTGAGCGCCTACCTA

AGCCGGCCCAGCCCGTTCGACCTGTTCATCCGCAAGTCGCCCACGATCACCT

GTCTGGTGGTGGACCTGGCACCCAGCAAGGGGACCGTGAACCTGACCTGGT

CCCGGGCCAGTGGGAAGCCTGTGAACCACTCCACCAGAAAGGAGGAGAAG

CAGCGCAATGGCACGTTAACCGTCACGTCCACCCTGCCGGTGGGCACCCGA

GACTGGATCGAGGGGGAGACCTACCAGTGCAGGGTGACCCACCCCCACCTG

CCCAGGGCCCTCATGCGGTCCACGACCAAGACCAGCGGCCCGCGTGCTGCC

CCGGAAGTCTATGCGTTTGCGACGCCGGAGTGGCCGGGGAGCCGGGACAAG

CGCACCCTCGCCTGCCTGATCCAGAACTTCATGCCTGAGGACATCTCGGTGC

AGTGGCTGCACAACGAGGTGCAGCTCCCGGACGCCCGGCACAGCACGACGC

AGCCCCGCAAGACCAAGGGCTCCGGCTTCTTCGTCTTCAGCCGCCTGGAGG

TGACCAGGGCCGAATGGGAGCAGAAAGATGAGTTCATCTGCCGTGCAGTCC

ATGAGGCAGCGAGCCCCTCACAGACCGTCCAGCGAGCGGTGTCCGTAAATC

CCGAGCTGGACGTGTGCGTGGAGGAGGCCGAGGGCGAGGCGCCGTGGACG

TGGACCGGCCTCTGCATCTTCGCCGCACTCTTCCTGCTCAGCGTGAGCTACA

GCGCCGCCCTCACGCTCCTCATGGTGCAGCGGTTCCTCTCAGCCACGCGGCA

246

GGGGAGGCCCCAGACCTCCCTCGACTACACCAACGTCCTCCAGCCCCACGC

CATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGT

CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG

GCGAGGGCGATGCCACCTACGGCAAACTCACATTGAAATTTATCTGTACGA

CGGGCAAGCTCCCTGTTCCGTGGCCCACTCTCGTGACAACACTGACTTATGG

AGTCCAATGCTTTTCTCGATACCCCGACCACATGAAGCAGCATGACTTCTTC

AAATCAGCGATGCCCGAAGGATACGTTCAAGAGCGCACCATCTTCTTCAAG

GATGACGGGAACTATAAGACAAGGGCCGAAGTAAAATTTGAGGGTGACAC

GCTCGTAAATCGGATTGAACTTAAAGGCATCGATTTCAAGGAAGACGGTAA

TATCCTGGGGCACAAGCTCGAATACAATTACAACTCTCACAATGTCTACATC

ATGGCTGATAAACAGAAGAATGGGATTAAGGTCAATTTTAAAATTCGACAT

AACATAGAAGATGGCTCTGTGCAGTTGGCGGATCACTATCAACAGAATACG

CCCATCGGAGACGGGCCCGTACTTTGCCCGACAATCATTACCTTTCTACAC

AGTCAGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGC

TGGAGTTCGTGACCGCCGCCGGGATCTAG

Red = Native Signal Peptide, Purple = IgE-Fc$_{2-4}$, Blue = EMPD, Navy = Transmembrane Domain, Pink = Intracellular Tail, Green = eGFP

Appendix B. List of glycans on the RayBio Glycan Array

1 β-Glc-Sp

2 β-Gal-Sp

3 α-Man-Sp

4 α-Fuc-Sp

5 α-Rha-Sp

6 β-GlcNAc-Sp

7 β-GalNAc-Sp

8 Tobramycin

9 Gal-β-1,3-GlcNAc-β-Sp

10 Gal-α-1,3-Gal-β-1,3-GlcNAc-β-Sp

11 Neu5Ac-α-2,3-Gal-β-1,3-GlcNAc-β-Sp

12 Neu5Ac-α-2,6-Gal-β-1,3-GlcNAc-β-Sp

13 Neu5Gc-α-2,3-Gal-β-1,3-GlcNAc-β-Sp

14 Neu5Gc-α-2,6-Gal-β-1,3-GlcNAc-β-Sp

15 Gal-β-1,3-(Fuc-α-1,4)-GlcNAc-β-[Lewis A]-Sp

16 Gal-β-1,4-Glc-β-Sp

17 Gal-α-1,3-Gal-β-1,4-Glc-β-Sp

18 Gal-α-1,4-Gal-β-1,4-Glc-β-Sp

19 GlcNAc-β-1,3-Gal-β-1,4-Glc-β-Sp

20 GalNAc-β-1,3-Gal-β-1,4-Glc-β-Sp

51 GlcNAc-β-1,4-GlcNAc-β-Sp1

52 β-D-GlcA-Sp

53 Gal-β-1,4-(6S)GlcNAc-β-Sp

54 GlcNAc-α-1,3-(Glc-α-1,2-Glc-α-1,2)-Gal- α-1,3-Glc-α-Sp

55 Gal-β-1,3-GalNAc-β-1,4-(Neu5Gc-α-2,3)- Gal-β-1,4-Glc-β-Sp1

56 Sisomicin Sulfate

57 GalNAc-α-1,3-(Fuc-α-1,2)-Gal-β- [Blood A antigen trisaccharide]-Sp1

58 Fuc-α-1,2-Gal-β-1,4-GlcNAc-β- [Blood H antigen trisaccharide]-Sp1

59 Gal-α-1,3-(Fuc-α-1,2)-Gal-β- [Blood B antigen trisaccharide]-Sp1

60 Fuc-α-1,2-Gal-β-1,3-GlcNAc-β-1,3-Gal-β- 1,4-Glc-β- [LNFP I]-Sp1

61 Fuc-α-1,2-Gal-β-1,4-Glc-β- [Blood H antigen trisaccharide]-Sp1

62 Gal-α-1,3-(Fuc-α-1,2)-Gal-β-1,4-Glc-β- [Blood B antigen tetrasaccharide]-Sp1

63 (Fuc-α-1,2)-Gal-β-1,4-(Fuc-α-1,3)- GlcNAc-β- [Lewis Y]-Sp1

64 (Fuc-α-1,2)-Gal-β-1,3-(Fuc-α-1,4)- GlcNAc-β- [Lewis B]-Sp1

65 Gal-β-1,3-(Fuc-α-1,4)-GlcNAc-β-1,3-Gal- β-1,4-(Fuc-α-1,4)-Glc-β- [Lewis A]-Sp1

66 Gal-β-1,3-GalNAc-β-Sp1

67 Gal-β-1,3-(Neu5Ac-α-2,6)-GalNAc-β-Sp

68 Neu5Ac-α-2,6-Gal-β-1,3-GalNAc-β-Sp

69 Neu5Ac-α-2,6-Gal-β-1,3-(Neu5Ac-α- 2,6)-GalNAc-β-Sp

70 Neu5Ac-α-2,3-Gal-β-1,3-(Neu5Ac-α- 2,6)-GalNAc-β-Sp

21 Neu5Ac-α-2,3-Gal-β-1,4-Glc-β-Sp

22 Neu5Ac-α-2,6-Gal-β-1,4-Glc-β-Sp

23 Neu5Gc-α-2,3-Gal-β-1,4-Glc-β-Sp

24 Neu5Ac-α-2,6-Gal-β-1,4-Glc-β-Sp

25 Gal-β-1,4-(Fuc-α-1,3)-Glc-β-Sp

26 GalNAc-β-1,3-Gal-α-1,4-Gal-β-1,4-Glc-β-Sp

27 GlcNAc-β-1,6-GlcNAc-β-Sp

28 4-P-GlcNAc-β-1,4-Man-β-Sp

29 Glc-α-1,2-Gal-α-1,3-Glc-α-Sp

30 Gal-β-1,3-GalNAc-α-Sp

27 GlcNAc-β-1,6-GlcNAc-β-Sp

28 4-P-GlcNAc-β-1,4-Man-β-Sp

29 Glc-α-1,2-Gal-α-1,3-Glc-α-Sp

30 Gal-β-1,3-GalNAc-α-Sp

31 Gal-β-1,4-GlcNAc-β-Sp

32 Gal-β-1,4-(Fuc-α-1,3)-GlcNAc-β-[Lewis X]-Sp

33 Neu5Ac-α-2,3-Gal-β-1,4-(Fuc-α-1,3)-GlcNAc-β- [Sialyl Lewis X]-Sp

34 Neu5Ac-α-2,3-Gal-β-1,3 -(Fuc-α-1,4)-GlcNAc-β- [Sialyl Lewis A]-Sp

35 Neu5Gc-α-2,3-Gal-β-1,3-(Fuc-α-1,4)-GlcNAc-β- [Sialyl Lewis A]-Sp

36 Gal-α-1,4-Gal-β-1,3-GlcNAc-β-Sp

37 Gal-β-1,4-GlcNAc-β-1,3-Gal-β-1,4-Glc-β-[LNnT]-Sp

38 GlcA-β-1,4-GlcNAc-α-1,4-GlcA-β-Sp

39 GlcNAc-β-1,6-(Gal-β-1,3)-GalNAc-α-O-Ser- Sp4

40 Neu5Ac-α-2,3Gal-β-1,4-(6S)GlcNAc-β-Sp

71 Neu5Ac-α-2,6-(Neu5Ac-α-2,3)-Gal-β- 1,3-GalNAc-β-Sp

72 GalNAc-β-1,4-(Neu5Ac-α-2,3)-Gal-β-1,4- Glc-β- [GM2]-Sp

73 GalNAc-β-1,4-(Neu5Ac-α-2,8-Neu5Ac-α- 2,3)-Gal-β-1,4-Glc-β- [GD2]-Sp

74 Gal-α-1,4-Gal-β-1,4-GlcNAc-β-Sp1

75 β-D-Rha-Sp

76 Glc-α-1,4-Glc-β-Sp1

77 Glc-α-1,6-Glc-α-1,6-Glc-β-Sp1

78 Maltotriose-β-Sp1

79 Glc-α-1,6-Glc-α-1,6-Glc-β-Sp1

80 Maltotetraose-β-Sp1

77 Glc-α-1,6-Glc-α-1,6-Glc-β-Sp1

78 Maltotriose-β-Sp1

79 Glc-α-1,6-Glc-α-1,6-Glc-β-Sp1

80 Maltotetraose-β-Sp1

81 GlcNAc-α-1,4-GlcA-β-1,4-GlcNAc-α1,4- GlcA-β-Sp

82 Maltohexaose-β-Sp1

83 Maltoheptaose-β-Sp1

84 Acarbose-β-Sp1

85 D-pentamannuronic acid-β-Sp1

86 L-pentaguluronic acid-β-Sp1

87 D-cellose-β-Sp1

88 Gal-α-1,3-Gal-β-Sp1

89 β-1,4-Xylotetrose-Sp1

90 Chitin-trisaccharide-Sp1

41 GalNAc-β-1,4-GlcNAc-β-Sp2

42 Neu5Ac-α-2,8-Neu5Ac-α-2,3-Gal β-1,4-Glc-β-Sp

43   Neu5Gc-α-2,8-Neu5Ac-α-2,3-Gal-β-1,4-Glc-β-Sp

44 GalNAc-α-1,3-(Fuc-α-1,2)-Gal-β-1,4-Glc-β- [Blood A antigen tetrose]-Sp1

45 GlcNAc-β-1,2-Man-α-Sp

46 Neu5Ac-α-2,3-Gal-β-Sp1

47 Gal-β-1,3 -GalNAc-β-1,3-Gal-β-Sp1

48 Glc-α-1,2-Gal-α-Sp

49 Gal-β-1,4-(Fuc-α-1,3)-GlcNAc-β-1,3-Gal- β-Sp1

50 Neu5Ac-α-2,3-Gal-β-1,4-(Fuc-α-1,3)-Glc-β- [3-Sialyl-3-fucosyllactose/ F-SL]-Sp1

91 KDN-α-2,8-Neu5Ac-α-2,3-Gal-β-1,4-Glc- β-Sp

92 Neu5Ac-α-2,8-Neu5Gc-α-2,3-Gal-β-1,4- Glc-β-Sp

93 Neu5Ac-α-2,8-Neu5Ac-α-2,8-Neu5Ac-α- 2,3-Gal-β-1,4-Glc-β-Sp3

94 Neu5Ac-α-2,8-Neu5Ac-α-2,6-Gal-β-1,4- Glc-Sp5

95 Gal-β-1,3-GalNAc-β-1,4-(Neu5Ac-α-2,3)- Gal-β-1,4-Glc-β-Sp1

96 Gentamicin Sulfate

97 Kanamycin sulfate

98 Geneticin Disulfate Salt (G418)

99 Neomycin trisulfate SGP

100 SGP

Table 6-1 **A list of the glycans printed on the RayBio Glycan Array.** Sp = linkers used to bind glycan to the glass slide. Sp: $OCH_2CH_2CH_2NH_2$. Sp1: $NH(CH_3)OCH_2CH_2NH_2$. Sp2: $OCH_2CH_2NH_2$. Sp3: $O(CH_2)_3NHCOCH_2(OCH_2CH_2)_5CH_2CH_2NH_2$. Sp4: $OCH_2CH(COOH)NH_2$.

Sp5:
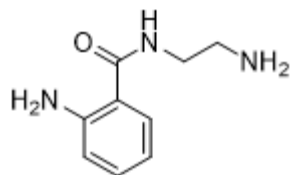
Appendix C. Input files for the Molecular Dynamics Simulations

System minimization, minimize only the water and restrain the protein:

```
System minimization:
  &cntrl                      #Namelist for basic controls
  imin=1,                     #Choose a minimization run
  ntmin=1,                    #Method  of  minimization  =  steepest
                              descent  for  ncyc  cycles  then  switch  to
                              conjugate gradient
  nmropt=0,                   #No NMR-specific analysis
  drms=0.1                    #Convergence   criterion   for   energy
                              gradient
  maxcyc=2000,                #Maximum number of minimization cycles
  ncyc=1500,                  #Switch to conjugate gradient algorithm
  ntx=1,                      #Input: read formatted inpcrd file
  irest=0,                    #Run new simulation
  ntpr=100,                   #Output: print energy information every
                              100 steps
  ntwr=100,                   #Output:  print  restart  file  every  100
                              steps
  iwrap=0,                    #Output:  do  not  wrap  coordinates  in
                              restart/trajectory files
  ntf=1,                      #Complete force evaluation
  ntb=1,                      #Equilibrate  at  constant  volume  with
                              periodic box boundaries
  cut=10.0,                   #Nonbonded cut-off (Angstroms)
  nsnb=20,                    #Nonbonded list update frequency
  igb=0,                      #No generalised Born solvation model
  ibelly=0,                   #No  groups  of  independently  restrained
                              atoms (belly type dynamics)
  ntr=1,                      #Use  restraints  (harmonic  potential)  on
                              atoms specified by restraintmask
restraintmask="!:WAT",        #Restraint information
restraint_wt=10.0,            #Force    constant    (kcal/mol-Å²)    for
                              restrained atoms
  &end                        #End of instructions
```

Let the water move while restraining the protein:

```
LET WATER MOVE:
  &cntrl                #Namelist for basic controls
  timlim=999999.,       #Time limit
  nmropt=0,             #No NMR-specific analysis
  imin=0,               #No minimization: molecular dynamics run
```

```
   ntx=1,              #Input: read formatted inpcrd file
   irest=0,            #Run new simulation
   ntrx=1,             #Formatted (ASCII) Cartesian coordinate
   ntxo=1,             #Formatted (ASCII) restart file
   ntpr=500,           #Output: print energy information every 500
                        steps
   ntwx=500,           #Output: print trajectory file every 500
                        steps
   ntwv=0,             #Output: no velocity trajectory file
   ntwe=0,             #Output: no energy and temperature file
   ntwr=5000,          #Output: print restart file every 5000 steps
   ntf=2,              #No force evaluation for bonds with H atoms
   ntb=2,              #Constant pressure with periodic box
                        boundaries
   cut=10.0,           #Nonbonded cut-off (Angstroms)
   nsnb=20,            #Nonbonded list update frequency
   nstlim=10000,       #Number of molecular dynamics steps
   nscm=2500,          #Remove translational center-of-mass motion
                        every 2500 steps
   iwrap=1,            #Output:     wrap     coordinates     in
                        restart/trajectory files
   t=0.0,              #Start time
   dt=0.002,           #Time step (ps)
   temp0=300.0,        #Reference temperature to be kept at
   tempi=200.0,        #Initial temperature
   tautp=0.5,          #Time constant (ps) for heat bath coupling
   ntt=1,              #Constant Temperature (Berendsen thermostat)
   ntp=1,              #Constant pressure with periodic boundary
                        conditions
   taup=1.0,           #Pressure relaxation time (ps)
   ntc=2,              #H atom bond lengths are SHAKE constrained
   tol=0.00001,        #Relative geometrical tolerance for
                        coordinate resetting in shake
   ibelly=0,           #No groups of independently restrained atoms
                        (belly type dynamics)
   ntr=1,              #Use  restraints  (harmonic  potential)  on
                        atoms specified by restraintmask
restraintmask="!:    #Restraint information
WAT",
restraint_wt=10.0    #Force constant (kcal/mol-Å$^2$) for restrained
,                     atoms
&end                  #End of instructions
```

Minimize the water and the protein:

```
System minimization:
  &cntrl              #Namelist for basic controls
```

```
   imin=1,              #Choose a minimization run
   ntmin=1,             #Method of minimization = steepest descent
                        for ncyc cycles then switch to conjugate
                        gradient
   nmropt=0,            #No NMR-specific analysis
   drms=0.1             #Convergence criterion for energy gradient
   maxcyc=2000,         #Maximum number of minimization cycles
   ncyc=1500,           #Switch to conjugate gradient algorithm
   ntx=1,               #Input: Read formatted inpcrd file
   irest=0,             #Run new simulation
   ntpr=100,            #Output: print energy information every
                        100 steps
   ntwr=100,            #Output: print restart file every 100
                        steps
   iwrap=0,             #Output: do not wrap coordinates in
                        restart/trajectory files
   ntf=1,               #Complete force evaluation
   ntb=1,               #Equilibrate at constant volume with
                        periodic box boundaries
   cut=10.0,            #Nonbonded cut-off (Angstroms)
   nsnb=20,             #Nonbonded list update frequency
   igb=0,               #No generalised Born solvation model
   ibelly=0,            #No groups of independently restrained
                        atoms (belly type dynamics)
   ntr=0,               #No positional restrains
&end                    #End of instructions
```

Heat the system while restraining the protein:

```
Heating System:
&cntrl                  #Namelist for basic controls
  imin=0,               #No minimization: molecular dynamics run
  nmropt=1,             #NMR restraints
  ntx=1,                #Input: read formatted inpcrd file
  irest=0,              #Input: run new simulation
  ntpr=500,             #Output: print energy information every 500
                        steps
  ntwr=500,             #Output: print restart file every 500 steps
  ntwx=500,             #Output: trajectory file every 500 steps
  iwrap=1,              #Wrap coordinates in restart/trajectory files
  ntf=2,                #No force evaluation for bonds with H atoms
  ntb=1,                #Equilibrate at constant volume with periodic
                        box boundaries
  cut=10.0,             #Nonbonded cut-off (Angstroms)
  nsnb=20,              #Nonbonded list update frequency
  igb=0,                #No generalised Born solvation model
```

```
   ibelly=0,              #No groups of independently restrained atoms
                          (belly type dynamics)
   ntr=1,                 #Use restraints (harmonic potential) on atoms
                          specified by restraintmask
   nstlim=250000,         #Number of molecular dynamics steps
   nscm=500,              #Remove translational center-of-mass motion
                          every 500 steps
   dt=0.002,              #Time step (ps)
   ntt=1,                 #Constant Temperature (Berendsen thermostat)
   temp0=0.0,             #Reference temperature to be kept at
   tempi=0.0,             #Initial temperature
   tautp=0.5              #Time constant (ps) for heat bath coupling
   ntc=2,                 #H atom bond lengths are SHAKE constrained
restraintmask=':1-        #Restraint information
135',
restraint_wt=10.0,       #Force constant (kcal/mol-Å$^2$) for restrained
                          atoms
&end
  &wt type='REST',        #relative NMR restraint energy term weights
                          are varied
  istep1=0,               #Apply change in conditions
  istep2=0,               #Conditions maintained until end of the run
  value1=1.0,             #Value of istep 1 change
  value2=1.0,&end         #Value of istep 2 change
  &wt                     #Varying target temperature
type='TEMP0',
  istep1=0,               #Apply change in conditions
  istep2=250000,          #End conditions at istep2
  value1=0.0,             #Value of istep 1 change
  value2=300,&end         #Value of istep 2 change
  &wt type='END'          #End varying conditions
&end                      #End of instructions
```

Relax the system, restraining the protein heavy atoms:

```
heat:
  &cntrl                  #Namelist for basic controls
  imin=0,                 #No minimization: molecular dynamics run
  irest=1,                #Input: restart previous MD
  ntx=5,                  #Input: read formatted coordinates and
                          velocities
  nstlim=250000,          #Number of molecular dynamics steps
  dt=0.002,               #Time step (ps)
  ntc=2,                  #H atom bond lengths are SHAKE constrained
  ntf=2,                  #No force evaluation for bonds with H atoms
```

```
    cut=10.0,              #Nonbonded cut-off (Angstroms)
    ntb=2,                 #Constant   pressure   with   periodic   box
                           boundaries

    ntp=1,                 #Constant   pressure   with   periodic   boundary
                           conditions

    taup=1.0,              #Pressure relaxation time (ps)
    ntpr=500,              #Output: print energy information every 500
                           steps

    ntwx=500,              #Output: trajectory file every 500 steps
    ntt=3,                 #Temperature control (Langevin dynamics)
    gamma_ln=2.0,          #Collision frequency (ps⁻¹)
    temp0=300.0,           #Reference temperature to be kept at
    iwrap=1,               #Output:      wrap      coordinates      in
                           restart/trajectory files

    ntr=1,                 #Use restraints (harmonic potential) on atoms
                           specified by restraintmask
restraintmask=':1-         #Restraint information
135',
restraint_wt=10.0,         #Force constant (kcal/mol-Å²) for restrained
                           atoms
 /
```

Relax the system:

```
Equilibrate:
  &cntrl                #Namelist for basic controls
  imin=0,               #No minimization: molecular dynamics run
  irest=1,              #Input: restart previous MD
  ntx=5,                #Input:   read   formatted   coordinates   and
                        velocities
  nstlim=2500000,       #Number of molecular dynamics steps
  dt=0.002,             #Time step (ps)
  ntc=2,                #H atom bond lengths are SHAKE constrained
  ntf=2,                #No force evaluation for bonds with H atoms
  ig=-1,                #Random seed for pseudo-random number
                         generator based on date and time
  cut=10.0,             #Nonbonded cut-off (Angstroms)
  ntb=2,                #Constant   pressure   with   periodic   box
                        boundaries
  ntp=1,                #Constant   pressure   with   periodic   boundary
                        conditions
  taup=2.0,             #Pressure relaxation time (ps)
  ntpr=1000,            #Output: print energy information every 1000
                        steps
  ntwx=1000,            #Output: trajectory file every 1000 steps
  ntt=3,                #Temperature control (Langevin dynamics)
```

```
  gamma_ln=2.0,      #Collision frequency (ps⁻¹)
  temp0=300.0,       #Reference temperature to be kept at
 /
```

Production Run:

```
ntpMD - Langevin Temp, Monte Carlo P,2fs steps, 50ps saves:
  &cntrl              #Namelist for basic controls
  imin=0,             #No minimization: molecular dynamics run
  irest=1,            #Input: restart previous MD
  ntx=5,              #Input:  read  formatted  coordinates  and
                      velocities
  nstlim=100000000,   #Number of molecular dynamics steps
  dt=0.002,           #Time step (ps)
  ntc=2,              #H atom bond lengths are SHAKE constrained
  ntf=2,              #No force evaluation for bonds with H atoms
  ig=-1,              #Random seed for pseudo-random number
                       generator based on date and time
  cut=10.0,           #Nonbonded cut-off (Angstroms)
  ntp=1,              #Constant  pressure  with  periodic  boundary
                      conditions
  barostat=2          #Monte  Carlo  barostat  used  for  pressure
                      control
  ntpr=25000,         #Output: print data every 25000 steps
  ntwx=25000,         #Output: trajectory file every 25000 steps
  ntwr=25000,         #Output:  print  restart  file  every  25000
                      steps
  ntt=3,              #Temperature control (Langevin dynamics)
  gamma_ln=5.0,       #Collision frequency (ps⁻¹)
  temp0=300.0,        #Reference temperature to be kept at
  ioutfm=1,           #Coordinates  and  velocities  written  in
                      NetCDF format
  iwrap=1,            #Output:     wrap     coordinates     in
                      restart/trajectory files
 /
```

Appendix D. Ipython Jupyter python script for calcium ion binding plots

```python
import pyemma
pyemma.__version__

%pylab inline
import numpy as np # import pyemma modules
import matplotlib.pylab as plt

import os
%pylab inline
matplotlib.rcParams.update({'font.size': 12})
import pyemma.coordinates as coor
import pyemma.msm as msm
import pyemma.plots as mplt
from pyemma import config
import matplotlib.pyplot as plt
import pyemma.coordinates as coor

trajfile = './WT/trj/WT_Ca1.nc' # renamed mdcrd trajectory file
topfile = './WT/WT_Ca3.pdb' # pdb file

feat = coor.featurizer(topfile)# just use all xyz-coordinates
feat.add_all()

feat.describe()[:5] # print output

inp = coor.source(trajfile, feat)
print ('trajectory length = ' , inp.trajectory_length(0))
print ('number of dimension = ',inp.dimension())

feat = coor.featurizer(topfile)

n_res = 135 # number of residues
ben_ind = 136 # Calcium index (with residues starting at 1)
ind_arr = np.zeros((n_res,2))
for i in range(n_res):
    ind_arr[i][0] = ben_ind-1
    ind_arr[i][1] = i
feat.add_residue_mindist(residue_pairs=ind_arr, scheme='closest-
heavy')

feat.describe()[:5] # print output

inp = coor.source(trajfile, feat)
print ('number of trajectories = ',inp.number_of_trajectories())
print ('trajectory length = ',inp.trajectory_length(0))
print ('trajectory time step = ',200.0 /
(inp.trajectory_length(0)),'ns')
print ('number of dimension = ',inp.dimension())
```

```python
inp.describe()[:5]

RES = inp.get_output() # get results
print ('number of trajectories = ', np.shape(RES)[0])
print ('number of frames = ', np.shape(RES)[1])
print ('number of dimensions = ',np.shape(RES)[2])

XY=RES[0]
print (np.shape(XY))
print (XY[1500,70:77])
# check Ca residue distance for res70 to res77 at frame 1500

# plot one residue to Ca separation
matplotlib.rcParams.update({'font.size': 18})
plt.figure(figsize(15,5))
plt.ylabel('Ca-Protein Separation (nm)')
plt.xlabel('Frame')
plt.xticks(np.arange(0,4001,500))
plt.title('E252')
plt.plot(XY[:,(72)-1 linewidth=2) # residue to plot

# plot Ca to all residues distance
figure(figsize(16,10))
plt.ylabel('ns')
plt.xlabel('residue')
imgplot =
plt.imshow(XY[:,1:135],aspect='auto',extent=(179,319,200,0),cmap
='CMRmap',interpolation='none')
# residues numbered according to mouseCD23 instead of 1 to 135
# colour of plot used is CMRmap
plt.colorbar(); imgplot.set_clim(0.2,1.0)
# plot from 0.2nm to 1nm only
plt.title('WT m-derCD23') # rename graph for mutants
# CA binds at res 72 aka E252 OR res 92 aka E272

# Combining both graphs
figure(figsize(18,12))
matplotlib.rcParams.update({'font.size': 18})
fig, ax1 = plt.subplots()
imgplot =
plt.imshow(XY[:,1:135],aspect='auto',extent=(179,319,200,0),cmap
='CMRmap',interpolation='none')
plt.colorbar(); imgplot.set_clim(0.2,1.0)
plt.title('WT m-derCD23') # rename graph for mutants
plt.ylabel('ns')
plt.xlabel('Residue')
left, bottom, width, height = [0.2, 0.6, 0.2, 0.2]
ax1 = fig.add_axes([left, bottom, width, height])
ax1.plot = plt.plot(XY[:,(72)-1],linewidth=1)
plt.xticks(np.arange(0,4001,1000))
plt.yticks(np.arange(0,4,1)) # adjust y axis as required
plt.title('E252')
plt.ylabel('Ca-Protein Separation (nm)')
plt.xlabel ('Frame')
# Repeat for CA137
```

Appendix E. Trajectory Analysis script in R

```
#Veronica's automated analysis for MD Simulations
for I in 1 2 3
do
echo "Analysis Beginning" $I

#Load and read the files and environment
R --save <<EOF
library(bio3d)
library(igraph)
dcd <- read.dcd("PROTEIN_NAME_run$I.dcd")
pdb <- read.pdb("CA.pdb")
print(pdb)
print(dcd)
ca.inds <- atom.select(pdb,"calpha", resno=4:125)
xyz <- fit.xyz(fixed=pdb\$xyz, mobile=dcd,
fixed.inds=ca.inds\$xyz,
mobile.inds=ca.inds\$xyz)

dim(xyz) == dim(dcd)

#Generate a new PDB with renumbered residues
nums <- as.numeric(pdb\$atom[,"resno"])
nums <- nums - (nums[1] - 159)
write.pdb(pdb=pdb, resno=nums, file="renum.pdb")

#RMSD
rd <- rmsd(xyz[1,ca.inds\$xyz], xyz[,ca.inds\$xyz])
plot(rd, typ="l", ylab="RMSD", xlab="Frame No.")
points(lowess(rd), typ="l", col="red", lty=2, lwd=2); dev.copy((png), file = "RMSD\$I.png")
write.table(rd, file="RMSD\$I.csv", col.names=FALSE, sep = ",")
dev.off()

hist(rd, breaks=40, freq=FALSE, main="RMSD_Histogram", xlab="RMSD")
lines(density(rd), col="gray", lwd=3); dev.copy((png), file = "Hist\$I.png")
dev.off()

#Radius of gyration
rg <- rgyr(xyz[,ca.inds\$xyz])
plot(rg, typ="l", ylab="Rgyr", xlab="Frame No.")
points(lowess(rg), typ="l", col="red", lty=2, lwd=2); dev.copy((png), file = "Rgyr\$I.png")
write.table(rg, file="Rgyr_run\$I.csv", col.names=FALSE, sep = ",")
rg[1:10]
dev.off()

#RMSF
```

```
rf <- rmsf(xyz[,ca.inds\$xyz])
plot(rf, ylab="RMSF", xlab="Residue Position", typ="l"); dev.copy((png), file =
"RMSF\$I.png")
write.table(rf, file="RMSF\$I.csv", col.names=FALSE, sep = ",")
dev.off()


#PCA
pc <- pca.xyz(xyz[,ca.inds\$xyz])
plot(pc, col=bwr.colors(nrow(xyz)) ); dev.copy((png), file = "PCA\$I.png")
dev.off()


hc <- hclust(dist(pc\$z[,1:2]))
grps <- cutree(hc, k=2)
plot(pc, col=grps); dev.copy((png), file = "Clusters\$I.png")
dev.off()


plot.bio3d(pc\$au[,1], ylab="PC1 (A)", xlab="Residue Position", typ="l")
points(pc\$au[,2], typ="l", col="blue"); dev.copy((png), file = "PCA1_2res\$I.png")
dev.off()
p1 <- mktrj.pca(pc, pc=1, b=pc\$au[,1], file="pc1.pdb")
p2 <- mktrj.pca(pc, pc=2,b=pc\$au[,2], file="pc2.pdb")
p3 <- mktrj.pca(pc, pc=3,b=pc\$au[,3], file="pc3.pdb")
write.ncdf(p1, "trj_pc1.nc")
write.ncdf(p2, "trj_pc2.nc")
write.ncdf(p3, "trj_pc3.nc")
plot.bio3d(pc\$au[,2], ylab="PC2 (A)", xlab="Residue Position", typ="l")
points(pc\$au[,3], typ="l", col="blue"); dev.copy((png), file = "PCA2_3res\$I.png")
dev.off()
plot.bio3d(pc\$au[,1], ylab="PC1 (A)", xlab="Residue Position", typ="l")
points(pc\$au[,3], typ="l", col="blue"); dev.copy((png), file = "PCA1_3res\$I.png")
dev.off()


#Residue Cross Correlation
cij <- dccm(xyz[,ca.inds\$xyz])
plot(cij); dev.copy((png), file = "CrossCorrel\$I.png"); dev.copy((png), file =
"CrossCorrel\$I.png")
write.table(cij, file="CrossCorrel\$I.csv", col.names=FALSE, sep = ",")
dev.off()



#Calculate suboptimal network paths
net <- cna(cij, cutoff.cij=0.3)
print(net)
x <- summary(net)
par(mfcol = c(1, 2), mar = c(0, 0, 0, 0))
plot(net, pdb, full = TRUE, vertex.label.cex=0.7)
plot(net, pdb)
pa <- cnapath(net, from=70, to=100, k=10)
print(pa)
```

```r
print(pa, label="HIVP", plot=TRUE); dev.copy((png), file = "HIVP\$I.png")
view.cnapath(pa, pdb=pdb, spline=TRUE, col='cyan', launch=FALSE)
write.table(net\$cij, quote=FALSE, row.names=FALSE, col.names=FALSE,
file="adj_forWISP.txt")

node.betweenness <- betweenness(net\$network)
plot(node.betweenness, xlab="Residue No", ylab="Centrality", type="h"); dev.copy((png), file
= "Centrality\$I.png")
dev.off()

#Write RMSF values into B column

pdbshort <-read.pdb("CA_short.pdb")
print(pdbshort)
view.dccm(cij, pdbshort, type="pdb", launch=FALSE)
write.pdb(pdb=NULL, xyz=pdbshort\$xyz, file="RMSF-Bfactors.pdb", b=rf)
write.pdb(pdbshort, b=normalize.vector(node.betweenness), file="betweenness.pdb")


q()
EOF

#Move files to Folder

mkdir Analysis_res4-125_Run$I
mv *.csv NaAnalysis_res4-125_Run$I
mv pc1.pdb Analysis_res4-125_Run$I
mv pc2.pdb Analysis_res4-125_Run$I
mv pc3.pdb Analysis_res4-125_Run$I
mv betweenness.pdb Analysis_res4-125_Run$I
mv renum.pdb Analysis_res4-125_Run$I
mv trj_pc1.nc Analysis_res4-125_Run$I
mv trj_pc2.nc Analysis_res4-125_Run$I
mv trj_pc3.nc Analysis_res4-125_Run$I
mv Rplots.pdf Analysis_res4-125_Run$I
mv corr.inpcrd.pdb Analysis_res4-125_Run$I
mv corr.pdb Analysis_res4-125_Run$I
mv RMSF-Bfactors.pdb Analysis_res4-125_Run$I
mv view.cnapath.pdb Analysis_res4-125_Run$I
mv view.cnapath.vmd Analysis_res4-125_Run$I
mv adj_forWISP.txt Analysis_res4-125_Run$I
mv CrossCorrel$I.png Analysis_res4-125_Run$I
mv HIVP\$I.png Analysis_res4-125_Run$I

done
```